http://icmic-conf.org/

ICMIC 2024

The 3rd International Conference on Mobile, Military, and Maritime IT Convergence

"Supporting the Harmonious Convergence of Mobile, Military, and Maritime"



PROCEEDINGS

Date August 7 (Wed) ~ 9 (Fri), 2024

Venue Concorde Hotel Kuala Lumpur, Malaysia

Organized by KICS The Korean Institute of Communications and Information Sciences; UTM Universiti Teknologi Malaysia

Technically Co-Sponsored by IEICE Communications Society

Patrons KIOST; IITP Institute of Information & Communications Technology Planning & Evaluation; Palantir Technologies; Hanwha Systems; LIG Nex1; PIOLINK; NS LAB Networked Systems Laboratory; JY System (주)제이와이시스템; MET-SOLUTIONS; ICT CRC ICT Convergence Research Center

General Information

The International Conference on Mobile·Military ·Maritime IT Convergence (ICMIC) aims to promote convergence activities of the three pillar fields of mobile, military, and maritime, encompassing fundamental and core wireless communication technologies. With a keen understanding that advanced wireless technologies leveraging artificial intelligence must confront escalating technological hurdles in shaping our future hyper-connected societies, the conference aims to spotlight these challenges, particularly within the realms of mobile radio, military operational technologies, and maritime communications. In view of these diverse requirements and demands, the ICMIC 2024 will serve as a timely and essential academic platform wherein researchers, academics, and industrial professionals convene to exchange perspectives, unveil cutting-edge technologies, and explore tailored solutions. The focal point of this conference lies in tackling the communication challenges across the three core pillars—mobile, military, and maritime—via ICT convergence or future advancements in communications. This encompasses a broad spectrum of industrial sectors, academia, and practical engineering applications. The conference will include keynote sessions, invited special sessions, and technical paper sessions. You are warmly encouraged to contribute your research and expertise to this dynamic platform.

Proceedings of International Conference on Mobile, Military, and Maritime IT Convergence (eISSN: 2983-3132)

Greetings

Messages from the General Co-Chairs of ICMIC 2024

It is with great pleasure and honor that I welcome you to the 2024 International Conference on M3 IT Convergence (ICMIC). As the General Chair, I am delighted to extend my warmest greetings to all participants, presenters, and distinguished guests who have joined us from around the globe.

This year marks a significant milestone in the history of our conference. As can be noted in the conference title (M3), we have introduced three new pillars as the main focus of our discussions and presentations: Mobile Technology, Maritime IT, and Military Technology. These areas have been selected for their growing importance and impact on our field, and we are thrilled to see the enthusiasm they have generated. I am certain that this expansion will serve as an excellent platform for even stronger growth. As such, we have received more than 130 papers from researchers and practitioners worldwide, reflecting the breadth and depth of interest and expertise in these critical domains.

Additionally, we are excited to announce that, for the first time, we are hosting the conference here in Malaysia, outside of Korea. This move represents our commitment to fostering a truly international dialogue and collaboration. Our chosen venue provides a unique setting that we hope will inspire fruitful discussions and new partnerships.

We have an exciting program planned, featuring keynote speeches from leading experts, insightful discussions, and opportunities for networking and knowledge exchange. We believe this year's conference will be a catalyst for innovative ideas and solutions that will shape the future of our field.

I would like to take this opportunity to express my deepest gratitude to our sponsors, Korean and Malaysian organizing committees, and volunteers for their hard work and dedication in making this event possible. Their unwavering support has been instrumental in bringing this conference to fruition.

Finally, I would like to reiterate that I deeply appreciate your participation and contribution to the success of this event. We look forward to engaging with you and witnessing the innovative research and developments that will be presented over the coming days.

I wish you a very fruitful and enjoyable time here in Kuala Lumpur.

Warm regards,

General Co-chair of ICMIC 2024
Prof. Yeon Ho Chung

First and foremost, I extend my sincere gratitude to all of you for participating in ICMIC 2024. This is the 3rd international conference held this year, following the successful events in previous years. It holds great significance as it focuses on mobile, military, and maritime IT convergence and aims to introduce and share the latest technologies in the related IT field.

We extend our heartfelt gratitude to the Korean Institute of Communications and Information Sciences (KICS). We are truly honored to host this prestigious event and are immensely grateful for the opportunity to collaborate with KICS, its esteemed committee, and dedicated members.

We are delighted to announce that this marks the 1st ICMIC conference outside Korea.

During the conference, a total of about 100 papers will be presented, accompanied by keynote speeches delivered by three distinguished foreign experts in both on-site and virtual sessions.

ICMIC 2024 is a dynamic international conference organized collaboratively by the The Korean Institute of Communications and Information Sciences (KICS), the largest academic institute in the field of ICT in Korea.

Furthermore, we are committed to elevating ICMIC to an even higher level, and we hope that all participants in this event will continue to collaborate and contribute to its development.

Over the course of three days, from 7-9 August, we gather in Concorde Hotel KL, the beautiful city of Malaysia. I wish you all fruitful discussions, achievements, and opportunities to rejuvenate.

Warm regards,

General Co-chair of ICMIC 2024
Prof. Sevia M. Idrus

Committees

International Advisory Committee

- Seong-Ho Jeong (President of KICS, Korea)
- Hyi Seung Lee (KIOST, Korea)
- Zabih Ghassemlooy (Northumbria University, UK)
- Murat Uysal (New York University Abu Dhabi, UAE)
- Dong Jin Kang (Korean Society of Oceanography, Korea)
- Chao Zhang (Tsinghua University, P.R.China)
- Murat Yuksel (The University of Central Florida, USA)

International Steering Committee

- Chang-jun Ahn (Chiba University, Japan)
- Trio Adiono (Institute of Technology at Bandung, Indonesia)
- Yeon Man Jeong (Gangneung-Wonju National University, Korea)
- Dong Wook Cho (Chungbuk Provincial University, Korea)
- Robert Bestak (Czech Technical University, Czech Republic)
- Dong Seong Kim (Kumoh National Institute of Technology, Korea)
- Manav R. Bhatnagar (IIT Delhi, India)
- Woo Yong Lee (ETRI, Korea)
- Francisco Jose Lopez Hernandez (Universidad Politecnica de Madrid, Spain)
- Jafry Din (Universiti Teknologi Malaysia, Malaysia)
- Mohd Naz'ri Mahrin (Universiti Teknologi Malaysia, Malaysia)
- Ali Selamat (Universiti Teknologi Malaysia, Malaysia)
- Himal A. Suraweera (University of Peradeniya, Sri Lanka)
- Prasant Kumar Sahu (IIT Bhubaneswar, India)
- Anand Srivastava (IIIT Delhi, India)
- Hong-Yeop Song (Yonsei University, Korea)

General Co-Chairs

- Yeon Ho Chung (Pukyong National University, Korea)
- Dong Myung Lee (Tongmyong University, Korea)
- Sevia M Idrus (Universiti Teknologi Malaysia, Malaysia)

General Vice Co-Chairs

- Seungjae Baek (KIOST, Korea)
- Siti Sophiayati Yuhaniz (Universiti Teknologi Malaysia, Malaysia)
- Jae Min Lee (Kumoh National Institute of Technology, Korea)

Technical Program Committee Co-Chairs

- Azizul Azizan (Universiti Teknologi Malaysia, Malaysia)
- Yonggang Kim (Kongju National University, Korea)

Finance Co-Chairs

- Teklu Merhawit Berhane (Pukyong National University, Korea)
- Siti Sophiayati Yuhaniz (Universiti Teknologi Malaysia, Malaysia)

Publicity Co-Chairs

- Sudhanshu Arya (VIT, Vellore, India)
- Muhammad Yusof Mohd Noor (Universiti Teknologi Malaysia, Malaysia)
- Tu Lam-Thanh (Ton Duc Thang University, Vietnam)

Registration Co-Chairs

- Taesoo Jun (Kumoh National institute of Technology, Korea)
- Sumiyaty Ambran (Universiti Teknologi Malaysia, Malaysia)

Publication Co-Chairs

- Dongwan Kim (Dong-a University, Korea)
- Muhammad Yusof Mohd Noor (Universiti Teknologi Malaysia, Malaysia)

Web Co-Chairs

- Byungju Lim (Pukyong National University, Korea)
- Siti Hasunah Mohammad (Universiti Teknologi Malaysia, Malaysia)

Local Arrangement Chair

- Nur Aqilah Yusri (Universiti Teknologi Malaysia, Malaysia)
- Puteri Nadiah Syamimi Said Jaafar (Universiti Teknologi Malaysia, Malaysia)

Technical Program Committee

- Sanjeev Sharma (IIT Varanasi, India)
- Insoo Sohn (Dongguk University, Korea)
- Chanjun Chun (Chosun University, Korea)
- Sudhanshu Arya (VIT, Vellore, India)
- Dongwan Kim (Dong-a University, Korea)
- Willy Anugrah Cahyadi (Telkom University, Indonesia)
- Gyung Bae Kim (Seowon University, Korea)
- Hyoyoung Jung (KOPTI, Korea)
- Indranath Chatterjee (Manchester Metropolitan University, UK)
- Hyuckjin Choi (Kyushu University, Japan)
- Hyuntae Cho (Tongmyong University, Korea)
- Inchul Shin (Pukyoung National University, Korea)
- Jin Hwan Koh (Gyeongsang National University, Korea)
- Ryangsoo Kim (ETRI, Korea)
- Suk Chan Kim (Pusan National University, Korea)
- Sung Hoon Lim (Hallym University, Korea)
- Sunyong Kim (Dongseo University, Korea)
- Taewoon Kim (Pusan National University, Korea)
- Wang Sang Lee (Gyeongsang National University, Korea)
- Wooyeol Choi (Chosun University, Korea)
- Yohan Kim (Dongseo University, Korea)
- Yonghun Kim (KATECH, Korea)
- Yujae Song (Yeungnam University, Korea)
- Minseok Han (Korea Naval Academy, Korea)
- Beomjun Kim (Yonam Institute of Technology, Korea)
- Vo Le Cuong (Hanoi University of Science and Technology, Vietnam)
- Kyungsook Kim (National Institute of Advanced Industrial Science and Technology, Japan)
- Minkyu Je (KAIST, Korea)
- Soon Ryang Kwon (Tongmyong University, Korea)
- Woocheol Kim (ETRI, Korea)
- Seunghyun Yoon (Korea Institute of Energy Technology, Korea)
- Norhana Arsad (Universiti Kebangsaan Malaysia, Malaysia)
- Widad Ismail (Universiti Sains Malaysia, Malaysia)
- Mohamad Sofian Abu Talip (Universiti Malaya, Malaysia)
- Fairuz Abdullah (Universiti Tenaga Nasional, Malaysia)
- Adam Wong (Universiti Teknikal Melaka, Malaysia)
- Wan Hafiza Wan Hassan (Universiti Terengganu Malaysia, Malaysia)
- Norliza Mohammad (Universiti Teknologi Malaysia, Malaysia)
- Aduwati Sali (Universiti Putra Malaysia)

Table of Contents

Oral Sessions

Oral Session 1A : Maritime/Underwater Communication Technology

Chair : Dr. Seungjae Baek (KIOST, Korea)

1A-1) Experimental Study of Light Wave Propagation in Underwater Optical Wireless Commmuncation Channel using Visible Light Communication Transceiver 1
Wan Hafiza Wan Hassan (Universiti Malaysia Terengganu), Kong Jia Wen (Intel Malaysia), Faezah Jasman (Princess Norah bint Abdulrahman University), Md Rabiul Awal and Muhd Shahir Mukhtar (Universiti Malaysia Terengganu), Sevia Mahdaliza Idrus (Universiti Teknologi Malaysia)

1A-2) Adjusting Repetitions for Error Reduction in Underwater Optical Wireless Communication 6
Huicheol Shin and Seungjae Baek (KIOST)

1A-3) Aerodynamically Considered Antenna Housing with Dielectric Lens for Extreme Environments 10
Jinkyu Bang, Man-Woong Heo and Seungjae Baek (KIOST), Jae Hee Kim (Korea University of Technology and Education)

1A-4) Reusable Fish Biologger 14
Juhyun Kim (KIOST), Chulsu Park (MET-Solutions), Seungjae Baek (KIOST)

1A-5) A Sonar Micro-Doppler Analysis Study for the Estimation of Diver Motion Parameters 17
Min Kim, Doyoung Kim, Jihyeong Lee, Jaeho Choi and Seungjae Baek (KIOST)

Oral Session 1B : 5G/6G, LTE, LTE-Advanced Communications I

Chair : Prof. Yonggang Kim (Kongju National University, Korea)

1B-1) Relative Precise Positioning Method for Two Moving Vehicles and UGV-UAV System based on Moving Baseline RTK 21
Gyeongmin Kim (Sejong University), Hyojung Yoon (Korea Institute of Aviation Safety Technology), Eunyeong Lee (LG Uplus), Wonwoo Park and Byungwoon Park (Sejong University)

1B-2) Performance Analysis of Multiple High-Altitude Platform Stations According to Architecture 25
jaeyeol lee, Won-Jae Lee, Jongtae Lee and Jae-Hyun Kim (Ajou University)

1B-3) An LSTM-based Enhanced Mobility Prediciton Mechanism 29
Kamrul Hasan and Seong Ho Jeong (Hankuk University of Foreign Studies)

1B-4) Deep Reinforcement Learning in Wireless Communications: Challenges and Solutions 31
Seungnyun Kim (Massachusetts Institute of Technology), Byonghyo Shim (Seoul National University)

1B-5) Spectum status and issues in Korea 35
Seung-Hoon Hwang (Dongguk University)

Oral Session 2A : Artificial Intelligence, Machine Learning, and Pattern Awareness I

Chair : Prof. Norhana Arsad (Universiti Kebangsaan Malaysia, Malaysia)

2A-1) Improved Human Echolocation Click Detection and Ranging with Gammatone Filter Banks 37
Nur Luqman Saleh, Alyani Ismail and Sali Aduwati (Universiti Putra Malaysia), Nur Emileen Abdul Rashid (Universiti Teknologi MARA)

2A-2) Exploring Machine Learning Models for Real-Time Maritime Vessel Monitoring on Interlinked Camera-Mounted Buoys 39
Michael Ralph Estrada, Margrette Yebes, Emilio Escobillio and Gabriel Avelino Sampedro (De La Salle University)

2A-3) Deep Learning Model for Anomaly Detection in Water Supply Pipes Based on Sensor Data 41
Jinhyung Cho, Hwansoo Kang and Heejin Lee (Dongyang Mirae University), Yoorim Cho (The University of Sheffield), Yudoo Kim (Dongyang Mirae University)

2A-4) Towards Time-Aware Decentralized Federated Learning for Bearing Fault Detection 43
Made Adi Paramartha Putra, Nengah Widya Utami and Gede Juliana Eka Putra (Primakara University), Nyoman Karna (Telkom University)

2A-5) Interactive Tablet Menu for Information-underprivileged Class 45
Yoon-Jong Lee, Yoon-Seo Koh, Min-Ju Jeon, Tae-Yun Rho, Han-Eum Lee, Su-Min Goo, Seo-il Han and Kwansang Na (Kyonggi University)

Table of Contents

Oral Session 2B : 5G/6G, LTE, LTE-Advanced Communications II

Chair : Dr. Siti Hasunah Mohammad (Pukyong National University, Korea)

2B-1) Impact of Beam Split on BER Performance of IRS-aided THz Communications 48
Sanjeev Sharma (Indian Institute of Technology (Banaras Hindu University) Varanasi), Yeon Ho Chung (Pukyong National University)

2B-2) Analysing of 90-100 GHz Radar using Triangular Trihedral Corner Reflector for Airport Runway Safety 52
Nur Aqilah Yusri, Sevia Mahdaliza Idrus, Sumiaty Ambran and Muhammad Al Farabi Muhamad Iqbal (Universiti Teknologi Malaysia), Nobuhiko Shibagaki (Hitachi Kokusai Electric Inc.), Shunichi Futatsumori (Electronic Navigation & Research Institute)

2B-3) Performance Comparison using various LDPC coded and CSK modulated Schemes 56
Daekyeong Kim amd Hong-Yeop Song (Yonsei University)

2B-4) Complexity analysis to obtain maximum performance of successive cancellation (SC) decoder for Reed-Muller codes with dynamic frozen bits 58
Woo Yong Lee and Keunyoung Kim (Electronics and Telecommunications Research Institute), Dongwook Cho (Korea Association of University, Research Institute and Industry)

2B-5) Differentiated Traffic Class Mean Delay of Degradation Attack in XG-PON 60
Fadila M Atan (Universiti Teknologi MARA Johor Branch), Nadiatulhuda Zulkifli, Sevia Mahdaliza Idrus, Nor Affida M.Zin, Nur Asfahani Ismail and Azizul Azizan (Universiti Teknologi Malaysia)

Oral Session 3A : Artificial Intelligence, Machine Learning, and Pattern Awareness II

Chair : Dr. Siti Hasunah Mohammad (Pukyong National University, Korea)

3A-1) Semantic Communication with Adversarial Defense for Malware Detection in Metaverse E-Commerce 64
Ebuka Chinaechetam Nkoro, Cosmas Ifeanyi Nwakanma, Jae-Min Lee and Dong-Seong Kim (Kumoh National Institute of Technology)

3A-2) A Survey on Cybersecurity Techniques for AI based Intelligent Systems 68
Insoo Sohn (Dongguk University)

3A-3) PhishNet: Exploring Machine Learning Techniques for Advanced URL Phishing Detection 69
Sean Andrei Olores, Renee Althea Khan, Shanley Valenzuela and Gabriel Avelino Sampedro (De La Salle University)

3A-4) Reinforcement Learning-Based Optimal Task Offloading for Multi-Tier Cloud-Edge Computing Systems 71
Avilia Kusumaputeri Nugroho (Pusan National University), Min Sang Yoon (Fujitsu Network Communications), Taewoon Kim (Pusan National University)

3A-5) Differentiated Traffic Class Delay Modeling for XG-PON Using ANN 75
Nur Asfahani Ismail, Farabi Iqbal, Sevia Mahdaliza Idrus, Nor Affida M.Zin and Fadila Mohd Atan (Universiti Teknologi Malaysia)

Oral Session 3B : 5G/6G, LTE, LTE-Advanced Communications III

Chair : Prof. Sumiaty Ambran (Universiti Teknologi Malaysia, Malaysia)

3B-1) Queuing Delay Analysis Based on Statistical Quality-of-Service Guarantee for Latency-Sensitive Mobile Edge Computing 79
Gusun Joung, Mintae Kim and Inkyu Lee (Korea University)

3B-2) WIP: Analysis of Distance-based Handover Parameters in LEO Satellite Communications 81
Jongtae Lee, Yujeung Ahn, Jaeyeol Lee, Wonjae Lee and Jae-Hyun Kim (Ajou University)

3B-3) A Review on Resource Allocation in FL-assisted IoT Networks: Overview, Research Challenges and Future Trends 83
Manh Cuong Ho, Dongwook Won, Thanh Phung Truong, Jaemin Kim and Sungrae Cho (Chung-Ang University)

3B-4) Advancing Green Network Communication through TRIZ: Analyzing Patented Technologies for Sustainable 6G Connectivity 86
Zulhasni Abdul Rahim, Muhammad Saqib Iqbal and Sevia Mahdaliza Idrus (Universiti Teknologi Malaysia)

Table of Contents

Oral Session 4A : Artificial Intelligence, Machine Learning, and Pattern Awareness III

Chair : Dr. Muhammad Yusof Mohd Noor (Universiti Teknologi Malaysia, Malaysia)

4A-1) Enhancing the Dynamic Range of Refractive Index Sensors Using Neural Network Classification with MMI Coreless Fiber.. 93
Athi Shankhara, Muhammad Yusof Mohd Noor, Fauzan Ahmad and Azizul Azizan (Universiti Teknologi Malaysia)

4A-2) Deep Neural Network Towards URL Classification Using Effective URL Tokens..... 97
Zafar Ali and Siti Sophiayati Yuhaniz (Universiti Teknologi Malaysia), Ghulam Mujtaba Shaikh (Sukkur Institute of Business Administration)

4A-3) On-Device Plastic Waste Classification for Waste Sorting Robots..... 101
Hyuntae Cho (Tongmyong University), Young Jin Lee (RobOne)

4A-4) SmartRSD: An Intelligent Multimodal Approach to Real-Time Road Surface Detection for Safe Driving..... 103
Adnan Md Tayeb, Mst Ayesha Khatun, Mohatsin Golam, Md Facklasur Rahaman, Ali Aouto, Oroceo Paul Angelo, Min-Seon Lee, Dong-Seong Kim, Jae-Min Lee and Jung-Hyeon Kim (Kumoh National Institute of Technology)

4A-5) Vision-Based Black Ice Detection using Convolutional Neural Networks for Road Safety..... 107
Ali Aouto, Jae-Min Lee and Jung-Hyeon Kim (Kumoh National Institute of Technology)

Oral Session 4B : Maritime Applications and Transporation Systems

Chair : Dr. Jinkyu Bang (KIOST, Korea)

4B-1) Input feature analysis for LSTM-based maximum wave height prediction..... 111
Geonwoo Hong and Yonggang Kim (Kongju National University)

4B-2) Derivation of controllable region of towfish according to change in towing point, towing speed, and center of gravity..... 115
Min-Kyu Kim, Jihyeong Lee and Sang-Ki Jeong (KIOST)

4B-3) A study on composite platform and field test..... 119
Jihyeong Lee, SangKi Jeong and Haeyong Park (KIOST)

4B-4) Enhancing Maritime Security via Blockchain Integration for Cross-Border Vehicle Authentication..... 123
Mohamed Abubakar Dini, Esmot Ara Tuli, Dong-Seong Kim and Taesoo Jun (Kumoh National Institute of Technology)

4B-5) Infrastructure-Linked CAV Overtaking Decision Algorithm..... 127
Jihun Kim, Youngjin Yoon, Seoksan Shin (Korea Automotive Technology Institute), Dong Seog Han (Kyungpook National University)

4B-6) Analysis of Safety Regulation Violations by Electric Kick Scooter Users..... 131
Chang-Ryul Park, Gangmin Lee, Yukang Kim and JaeYeong Lee (Kyonggi University), Dae-Won Jeong (401SPAG Co., Ltd.), Kyungyong Chung (Kyonggi University)

Oral Session 5A : Maritime Internet of Things

Chair : Dr. Azizul Azizan (Universiti Teknologi Malaysia, Malaysia)

5A-1) Beam Propagation Method Simulation of Uncoated Coreless Silica Fiber for Humidity Sensing..... 135
Athi Shankhara Velusamy, Muhammad Yusof Mohd Noor, Fauzan Ahmad and Siti Mahfuza Saimon (Universiti Teknologi Malaysia)

5A-2) Applicability of block pavement in a harbor based on deflection and settlement characteristics..... 139
Yong-Soo Ha (KIOST), Yun-Tae Kim and Minh-Vuong Pham (Pukyong National University), Haeyong Park, Sang Ki Jeong and Myounghak Oh (KIOST)

5A-3) Simulation of Bare Trench Silica Fiber Sensors for Humidity Sensing between Length of Depths of 50μm and 110μm..... 141
Athi Shankhara Velusamy, Muhammad Yusof Mohd Noor, Fauzan Ahmad and Siti Mahfuza Saimon (Universiti Teknologi Malaysia)

5A-5) Construction of a simulation environment for federated learning-based LEO satellite communications systems..... 145
Gwangun Yu and Yonggang Kim (Kongju National University)

Oral Session 5B : Advanced Communication Systems and Networks I

Chair : Prof. Widad Ismail (Universiti Sains Malaysia, Malaysia)

5B-1) Proof of Concept (POC) for Queue Management in Sustainable Smart Community Automated Convocation 147
Jing Huey Khor (University of Southampton (Malaysia Branch)), Widad Ismail (Universiti Sains Malaysia)

5B-2) Design of a Scheme to Enhance Real-Time Performance for DDS-Based Encrypted Communication in ROS2 151
Hee-Jae Shin, Jae-Min Lee, Tae-Soo Jun and Dong-Seong Kim (Kumoh National Institute of Technology)

5B-3) Hybrid Energy Harvesting Development for Internet of Things (IoT) 154
Shum Kar Yen and Widad Ismail (Universiti Sains Malaysia)

5B-4) Model Plant Mismatch in FCS-MPC Implementation: A Tuning Guideline for Robust Performance 158
Ahmad Takiyuddin Abdullah, Sevia Mahdaliza Idrus and Shahrin Md Ayob (Universiti Teknologi Malaysia)

5B-5) Recent Studies on Serverless Computing: Addressing the Cold Start Problem 162
Jaehwan Lee, Yeunwoong Kyung and Seungmin Oh (Kongju National University)

Oral Session 6A : Civil-Military ICT Convergence & System I

Chair : Prof. Taesoo Jun (Kumoh National Institute of Technology, Korea)

6A-1) Short-Packet Based Power Allocation Using Quantum Neural Network in Downlink NOMA Systems for Military Communication 164
Won Jae Ryu (ICT-CRC), Dong-Seong Kim (Kumoh National Institute of Technology)

6A-2) Federated Learning-Based Joint Radar-Communication mmWave Beamtracking for V2X-Communications 168
Sanjay Bhardwaj and Dong-Seong Kim (Kumoh National Institute of Technology)

6A-3) Node Election Scheme for Energy-Efficient Autonomous Underwater Vehicles 172
Young-June Kwon, Jae-Min Lee, Tae-Soo Jun and Dong-Seong Kim (Kumoh National Institute of Technology)

6A-4) Performance Analysis for Next Generation Integrated Naval Ship Network 176
Min-Hui Jang and Neul-Som Lee (Kumoh National Institute of Technology), Hyeong-Jin Kim (NS Lab Co., Ltd), Jae-Min Lee (Kumoh National Institute of Technology), Young-Keun Go and Nak-Jung Choi (Agency for Defense Development), Jae Ho Lee and Jaehak Yu (Hanwha Systems), Taesoo Jun and Dong-Seong Kim (Kumoh National Institute of Technology)

6A-5) Intent Inference Method for an Intent-based System in Military Communications 180
So-Yeon Lee (Soonchunhyang University), Jungwook Choi and Soobum Park (LIG Nex1), Dae-Young Kim (Soonchunhyang University)

Oral Session 6B : Advanced Communication Systems and Networks II

Chair : Dr. Wan Hafiza Wan Hassan (Universiti Malaysia Terengganu, Malaysia)

6B-1) Low Complexity MU-MIMO Scheduling for B5G Cellular Network 184
Satoshi Denno, Akihito Yamada and Yafei Hou (Okayama University)

6B-2) Channel Modeling Based on Ray Tracing for Wireless Transmission on Wood Materials 190
Nurul Fahmi Arief Hakim (Universitas Pendidikan Indonesia), Alyani Ismail, Nur Luqman Saleh and Aduwati Sali (Universiti Putra Malaysia), Nabil Taufiq Nur Izzat (Universitas Pendidikan Indonesia)

6B-3) A DDPG-based Framework to Optimize Hierarchical Rate Splitting Multiple Access 195
Anh-Tien Tran, Thanh Phung Truong, Manh Cuong Ho, Tung-Son Do, Quang Tuan Do and Dongwook Won (Chung-Ang University), Nhu-Ngoc Dao (Sejong University), Sungrae Cho (Chung-Ang University)

6B-4) Design of an Antenna for Electronic Article Surceillance System Capable of Receving all Polarizations 198
Jinkyu Bang (KIOST), Jae Hee Kim (Korea University of Technology and Education)

Oral Session 7A : Civil-Military ICT Convergence & System II

Chair : Prof. Jae Min Lee (Kumoh National Institute of Technology, Korea)

7A-1) Blockchain-Inspired GPS Spoofing Detection for Zero-Trust Mission-Critical Military IoT Networks 201
Ahmad Zainudin and Revin Naufal Alief (Kumoh National Institute of Technology), Made Adi Paramartha Putra (Primakara University), Dong-Seong Kim and Jae-Min Lee (Kumoh National Institute of Technology)

7A-2) Blockchain-enabled Token for Military Equipment Logistics 205
Ikechi Saviour Igboanusi, Cosmas Ifeanyi Nwakanma, Mirian Odigomma Igboanusi and Dong-Seong Kim (Kumoh National Institute of Technology)

7A-3) Detecting DDoS in Internet of Military Things (IoMT) Data Caching: Machine Learning Approach 209
Love Allen Chijioke Ahakonye, Cosmas Ifeanyi Nwakanma and Dong-Seong Kim (Kumoh National Institute of Technology)

7A-4) Revolutionizing Naval Logistics through NFT and Hyperledger Besu Integration 213
Gifar Arif Haryadi (Kumoh National Institute of Technology), Revin Naufal Alief (NS Lab Co., Ltd), Jae-Min Lee and Dong-Seong Kim (Kumoh National Institute of Technology)

Oral Session 7B : Optical Communication Technology I

Chair : Prof. Yonggang Kim (Kongju National University, Korea)

7B-1) Deterministic Bandwidth Allocation in Passive Optical Network with a Preemptive Approach 217
Nor Affida M.Zin, Sevia Mahdaliza Idrus, Arnidza Ramli, Fadila Mohd Atan and Nur Asfahani Ismail (Universiti Teknologi Malaysia)

7B-2) Performance Evaluation Transporting Multiservice RoF WDM System 221
R Mohamad (Telekom Research & Development), SM Idrus and NH Awang (Universiti Teknologi Malaysia), T Kanesan (Telekom Research & Development), S Yaakob (Universiti Putra Malaysia), Adam Wong Yoon Khang (Universiti Teknikal Malaysia Melaka)

7B-3) Research Trends of WDM QKD and Satellite QKD in Quantum Communication 227
Seungho Yoon and Jun Heo (Korea University)

7B-4) Simulation Of Radar Cross Section of Foreign Object Debris Using High Frequency Simulation Software 231
Firdaus Hakimi, Sumiaty Ambran, Sevia Mahdaliza Idrus and Puteri Nadiah Syamimi Said Ja'afar (Universiti Teknologi Malaysia)

7B-5) Performance Evaluation of Sampling Clock Offset Compensation in OFDM VLC System 234
Trio Adiono, Erwin Setiawan, Michael Jonathan, Rahmat Mulyawan, Nana Sutisna and Infall Syafalni (Bandung Institute of Technology)

Oral Session 8A : Civil-Military ICT Convergence & System III

Chair : Prof. Gabriel Avelino Sampedro (University of the Philippines, Philippines)

8A-1) Comparison of YOLO-NAS and Roboflow 3.0 for Local Firearms Detection and Recognition 238
Cosmas Ifeanyi Nwakanma (Kumoh National Institute of Technology), Ikenna Uzoma Ajere (Federal University of Technology Owerri), Ebuka Chinaechetam Nkoro, Judith Nkechinyere Njoku, Love Allen Chijioke Ahakonye and Dong-Seong Kim (Kumoh National Institute of Technology)

8A-2) Design and Implementation of Simulator for Large-scaled Naval Combat Systems 242
Jae-Woo Kim, Gi-Hyeob Kwon and Dong-Seong Kim (Kumoh National Institute of Technology), Won-Jae Ryu (ICT-CRC), Jae-Ho Lee and Jae-Hak Yu (Hanwha Systems)

8A-3) Visualization Scheme for Naval Combat System Network Simulator 246
Seung-Min Lee, Neul-Som Lee, Min-Hui Jang, Jae-Min Lee, Tae-Soo Jun and Dong-Seong Kim (Kumoh National Institute of Technology)

8A-4) TwinMil: Semantic Segmentation-based Digital Twin Framework for Military Surveillance 250
Judith Nkechinyere Njoku, Cosmas Ifeanyi Nwakanma and Dong-Seong Kim (Kumoh National Institute of Technology)

8A-5) Modular Open System Approach (MOSA) for Defense Unmanned Systems 254
Min-Seon Lee, Jae-Min Lee, Tae-Soo Jun and Dong-Seong Kim (Kumoh National Institute of Technology)

Oral Session 8B : Optical Communication Technology II

Chair : Prof. Yonggang Kim (Kongju National University, Korea)

8B-1) Deep Learning-Based Classification of Laguerre-Gaussian OAM Beams in Atmospheric Turbulence and Thermal Convection 258
Young Jae Moon and Yeon Ho Chung (Pukyong National University)

8B-2) Mathematical Detection Modeling for OAM-based MIMO Free Space Optical System 260
Merhawit Berhane Teklu and Yeon Ho Chung (Pukyong National University)

8B-3) Orbital Angular Momentum States Beamforming for Multiuser Optical Wireless Network 262
Siti Hasunah Mohammad and Yeon Ho Chung (Pukyong National University)

8B-4) Precoding Scheme for Indoor Channel Modelling in OAM-based MIMO Communication System 266
Merhawit Berhane Teklu and Yeon Ho Chung (Pukyong National University)

8B-5) Beam Pointing for FSOC with Spatial Diversity 268
Hong-Seol Cha (Korea University), Byungju Lim (Pukyong National University), Young-Chai Ko (Korea University)

Oral Session 9C : Blockchain and security for ICT convergence

Chair : Prof. Sophiayati Yuhaniz (Universiti Teknologi Malaysia, Malaysia)

9C-1) Transactions Time Synchronization for Time-Critical Blockchain 269
Ahmad Akmaluddin Mazlan, Azizul Azizan, Noor Hafizah Hassan, Abdul Ghafar Jaafar and Mohd Syahid Mohd Anuar (Universiti Teknologi Malaysia), Mohamad Sofian ABU TALIP (Universiti Malaya)

9C-2) Study on Trends in Blockchain-Based Energy Trading 273
Da-Jung Lee, Ik-Hyun Kwon and Dong-Seong Kim (Kumoh National Institute of Technology)

9C-3) Enhanced Secure Communication in Maritime Tactical Networks Using Quantum Cryptography 277
Esmot Ara Tuli, Mohatsin Golam, Jae-Min Lee and Dong-Seong Kim (Kumoh National Institute of Technology)

9C-4) Study on development of security requirements for CMVP security level1 281
Sun-Woo Lee, Tae-Hun Kim and Jun-Young Lee (KEPCO KDN), Ieck-Chae Euom (Chonnam National University)

9C-5) Exploring network security with MTD in Mininet 285
Gilhan Choi and Yonggang Kim (Kongju National University)

Oral Session 10C : Big data and its applications

Chair : Prof. Adam Wong Yoon Khang (Universiti Teknikal Malaysia Melaka, Malaysia)

10C-1) 3D Positioning Algorithm based on Decision Tree Classification Model in Indoor UWB Environment 287
Ho Chul Lee and Dong Myung Lee (Tongmyong University), Hee Min Park (Sky-Tech Co.Ltd.), Gyoungbae Kim (Seowon University)

10C-2) Smart Farm: Acquiring and Analyzing Environmental Data 289
Hoon-Seok Jang (Korea Electronics Technology Institute)

10C-3) Design of Performance Measurement System for Non-Face-To-Face Remote Workers 293
Min Goo Lee and Yong Kuk Park (Korea Electronics Technology Institute), Kyung Kwon Jung (Dongshin University)

10C-4) Forecasting Daily Demand in the Busan Metro Using LSTM-Based Encoder-Decoder 297
San-Hyun Hwang, Jun-Hyeong Park and Jeong-Won Seo (Pukyong National University), Pil-Goo Jeong and Jung-Tae Kim (Busan Transportation Corporation), Jun-Pyo Hong (Pukyong National University)

10C-5) Leveraging Data Mining and NLP for Service Readiness: The Case of Hotel Industry 300
Moner Alajeli Farj Abouoden and Siti Sophiayati Yuhaniz (Universiti Teknologi Malaysia)

10C-6) Application of Big Data Analytics in Airside Operations To Improve Wildlife Hazard Management For The Case of Malaysia Airports 304
Tisiyanah Abd Baki, Badriyah Noordin, Norliza Mohamed and Sevia Mahdaliza Idrus (Universiti Teknologi Malaysia), Nor Azlina Mohd Isa | Malaysia Airports Holdings Berhad, Syamil Azmin (Universiti Teknologi Malaysia)

Poster Session

Poster Session

Chair : Dr. Jinkyu Bang (KIOST)

P-1) Performance Analysis of Ultrasonic-Based Localization: Impact of Receiver Spacing and Array Distances 308
JaeJun Lim, Jaehan Joo, Hunyoul Lee, Dongjhin Noh and Suk Chan Kim (Pusan National University)

P-2) Enhancing IoT connectivity: Implementation of Zigbee Direct for interoperability in multiprotocol network 310
Seonghyeon Park, Jaehan Joo, Minseok Kim and Suk Chan Kim (Pusan National University)

P-3) Covertness Analysis for the Covert Communications in Multi-antenna Two-way Relay Systems 312
Zhilin Fu (Korea University), Jihwan Moon (Hanbat National University), Sangwon Hwang and Sangmin Kim (Korea University), Di Zhang (Zhengzhou University), Inkyu Lee (Korea University)

P-4) A Localization Algorithm using k-Nearest Neighbors in Indoor UWB Environment 314
Mookyung Jung (Goodmorning Information Technology), Dong Myung Lee (Tongmyong University), SooK Kyoung Cho (Nestors Co.,Ltd.), Kwangchul Jung (Dankook University)

P-5) An Indoor Localization Algorithm using Pose Estimation Method in Obstacle Environments for Pedestrians 316
Tae Wan Kim and Dong Myung Lee (Tongmyong University), SooK Kyoung Cho (Nestors Co.,Ltd.), Kwangchul Jung (Dankook University)

P-6) Real-Time Sports Pose Estimation Training based on Skeleton Simulation 318
Geon-Woo Kim, Ji-Yoon Kweon, Ji-Soo Kim, Woo-Hyeon Kim, Jeong-Eun Kim and Kyungyong Chung (Kyonggi University)

P-7) Data Placement Policy considering File System Input/Output Characteristics for Maritime IoT Platform 322
Sungmin Koo (KIOST), Gunhee Choi and Jongmoo Choi (Dankook University), Juhyun Kim and Seungjae Baek (KIOST)

P-8) Design of Smart Urban Environmental Monitoring System using Personal Mobility Devices 326
Kyung Kwon Jung (Dongshin University), Yeon Man Jeong (Gangneung-Wonju National University), Byung Sang Choi, Ho Soon Jang and Yong Joong Kim (Korea Polytechnics Wonju-Campus)

P-9) Research On The Development Of Suction Pump Packages for Installation and Dismantling of Co2 Subsea Injection System 329
Hyun Kang and Osoon Kwon (KIOST)

Oral Sessions

Experimental Study of Light Wave Propagation in Underwater Optical Wireless Communication Channel using Visible Light Communication Transceiver

Wan Hafiza Wan Hassan
Faculty of Ocean Engineering Technology
Universiti Malaysia Terengganu,
Terengganu, Malaysia
whafiza@umt.edu.my

Kong Jia Wen
Validation Engineer PG16 Intel Malaysia
Penang, Malaysia
jia.wen.kong@intel.com

Faezah Jasman
Applied College, Princess Norah bint Abdulrahman University,
Riyadh, Saudi Arabia.
FJJasman@pnu.edu.sa

Md Rabiul Awal
Faculty of Ocean Engineering Technology
Universiti Malaysia Terengganu
Terengganu, Malaysia
rabiul.awal@umt.edu.my

Muhamad Shahir Mukhtar
Faculty of Ocean Engineering Technology
Universiti Malaysia Terengganu
Terengganu, Malaysia
p5414@pps.umt.edu.my

Sevia Mahdaliza Idrus
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Johor, Malaysia.
sevia@utm.my

***Abstract*— The latest advancements in technology have driven interest in underwater optical wireless communication (UOWC) as a key technology for underwater applications. However, UOWC faces challenges in the underwater optical channel, such as absorption, scattering, and turbulence. Previous studies have mainly focused on absorption and scattering, often overlooking the impact of underwater optical turbulence. This turbulence can cause significant signal intensity fluctuations at the receiver, resulting in poor UOWC system performance. Recent research tends to oversimplify turbulence by assuming it remains constant, even though factors like water temperature and salinity cause turbulence to vary. Therefore, this study aims to investigate the effect of salinity on UOWC through practical experiments. A customized visible light transceiver is employed in this experiment as the transmitter and receiver pair. The green and blue LED are used to generate optical signal propagation in saline water channel. The interaction between the selected light and the saline water is analyzed by varying the salt concentration levels. Subsequently, the attenuation constant, c, is estimated. The analysis reveals that the received light intensity decreases as the salt concentration in the water increases. The estimated c increases with higher salt concentrations, indicating that UOWC performance declines as water salinity rises. This finding provides valuable insights for designing optimal underwater communication devices and maximizing UOWC channel capability.**

Keywords—Underwater, Optical Wireless, Communication, Attenuation, Salinity

I. INTRODUCTION

On Earth, one of the most precious natural resources is the ocean. Future human growth will greatly benefit from ocean exploration, development, and use. Seabed observation stations can deliver a variety of detailed, timely, and marine information [1]. The tactical role of underwater moorings and vehicles is rising. As a result, it is crucial to have a reliable and secure communication system between underwater vehicles and a ground station when they are on a prolonged seaborne operation. In fact, one of the biggest technological difficulties the world's naval forces currently face is the deployment of effective communication ties with underwater vehicles [2].

As it is still necessary for a high-speed communication link to exist between the remote end and the surface, it is only logical to explore what could be done without the hassle (and price) of bulky cables. For this reason, underwater wireless communications were encouraged to be captivated and investigate. It enables new applications ranging from environmental monitoring to the gathering of oceanographic data, marine archaeology, and search and rescue missions. These applications will be made possible in conjunction with underwater sensor and vehicle technologies [3]. Fig. 1 depicted the network of sensor and underwater vehicles comprised of submarine, autonomous underwater vehicle (AUV) and remotely operated vehicle (ROV) which utilizing multiple wireless carriers.

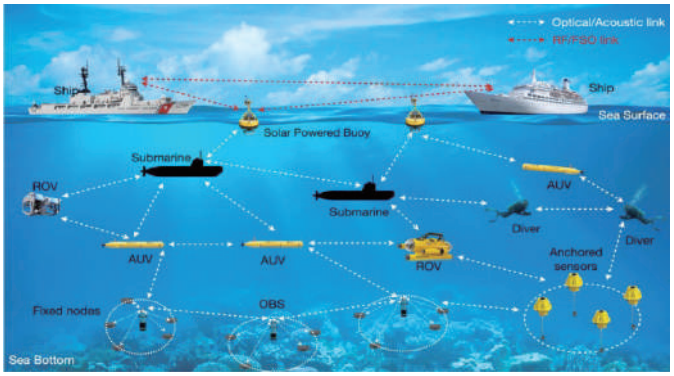


Fig. 1. The basic structure of an underwater wireless communication adapted from [4].

However, it should be noted that underwater environment have a variety of unique characteristics that set it apart from terrestrial radio propagation, where traditional communication technologies are used. In the present, there are three types of carriers for underwater wireless communication including, electromagnetic or radio frequency (RF), acoustics and optical. Each physical carrier has its advantages and disadvantages

towards efficient underwater wireless communication. Table I below shows the comparison between different underwater wireless communications media 5].

TABLE I. COMPARISON BETWEEN DIFFERENT COMMUNICATION MEDIA

Parameters	EM Waves	Acoustic Waves	Optical Waves
Propagation speed	High	Very slow	Very high
Line of sight (LOS)	Not required	Not required	Required with precise alignment of nodes
Impacts of environment (ambient noise, turbidity,temperature changes, etc.	Minimal	High	High
Acvhieveable data rates	High	Very low	Very high
Network coverage	Short ange	Very long range	Very short range
Impact on marine life	Not known	Negative	Not known

Electromagnetic waves or radio frequency (RF) travel small distances at extremely high speeds and frequencies in underwater. It does not required waves travel in a direct path from the source to the receiver [6]. Even though RF waves are the more popular and frequently employed technology in terrestrial communications due to their nature, they are not suited for usage underwater due to their significant attenuation. Additionally, due to its poor performance characteristics, such as high bit error rates, significant and variable propagation delays, short network coverage range and low bandwidth [7]. In fact, acoustic waves are now the most often use as underwater communication technology due to the low attenuation in underwater. But then, underwater acoustic communication systems are adversely affected by heavy human activity and ambient noise, but, for near-shore applications (e.g., congested environments like harbours and ports). Besides, the hard physical layer circumstances that result in poor data rate capabilities, a large propagation latency, and a high bit error rate are further characteristics of underwater acoustic channels when compared with EM waves and optical wave.

In comparison to acoustic and RF carriers, optical waves provide significantly greater transmission bandwidth, enabling data rates of up to 1 Gbps over distances spanning several tens of meters while maintaining low power and mass requirements. Therefore, underwater optical wireless communication (UOWC) has been identified as a promising technology for UWC. Despite all of the aforementioned benefits, UOWC link suffer from substantial channel impairments due to the inherent optical characteristics (IOP) of seawater, namely absorption and scattering. Absorption and scattering are the inherent physical properties of water; they cause power loss and time dispersion [i.e., inter-symbol-interference (ISI)], resulting in poorer link performance such as high path loss and high delay spread and shorter transmission spans (to less than 100 m), as documented in multiple literatures [8-10]. It is generally known that seawater has a low-loss property in blue and green lights (i.e., a transmission window of 450–550 nm) [11-12], which has served as the foundation for future UOWC development.

Turbulence in the underwater environment is also a critical aspect of the problem of optical signal intensity loss, as it is caused by the fluctuation of refractive index due to variation of salinity and temperature [13]. Turbulence field testing is difficult, expensive, and time-consuming because of the irregular and unpredictable motions in oceans. Considering that turbulence may be triggered by water salinity, leading to substantial fluctuations in receiver signal strength, this paper experimentally examines the impact of salinity on the performance of light wave propagation in UOWC.

In this work, a customized visible light communication (VLC) transceiver is set up and calibrated in order to be employed in the UOWC experimental model. Then, the intensity of the received signal under different UOWC channel conditions (varying salinity level) is measured. Finally, the received signal is analysed by estimating the attenuation constant, *c*. The remaining of this paper is organized as follows: Section II presents the methodology, Section IV discusses the obtained results and Section V concludes the work.

II. MATERIALS AND METHODS

A. *Customized Visisble Light Communication (VLC) Transceiver*

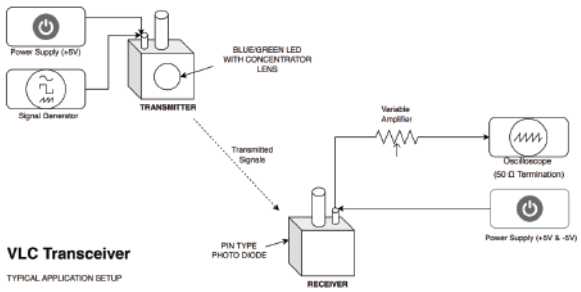


Fig. 2. Schematic drawing of the customized VLC transceiver

Fig.2 shows the schematic drawing of the customized VLC transceiver. It consists of transmitter-receiver pairs which are fabricated to be waterproof. The transmitter is made of blue and green LED with concentrator lens respectively while the receiver is a PIN type photodiode. Power supply of 5 V is required to power up the transceiver. A signal generator is connected to the transmitter, generates the carrier signal to be transmitted by the transmitter. On the other hand, an oscilloscope is attached to the receiver to graphically display and measure the received signal.

Prior to employing the customized VLC transceiver into our UOWC experimental model, this transceiver is calibrated first. The calibration process is conducted in free space environments as shown in Fig. 3.

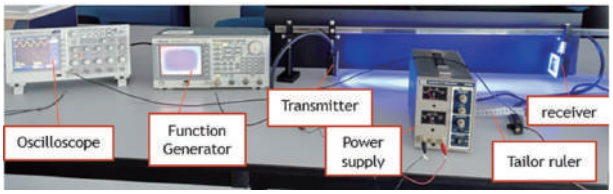


Fig. 3. Experimental set up for calibration process

The calibration process is validated by comparing the obtained frequency response with the frequency response from the datasheet as shown in Fig, 4. It is concluded that the transceiver is calibrated as the obtained frequency response is in agreement with the datasheet [14]. Therefore, the customized VLC transceiver is then employed in underwater environments as presented in the following sub-section.

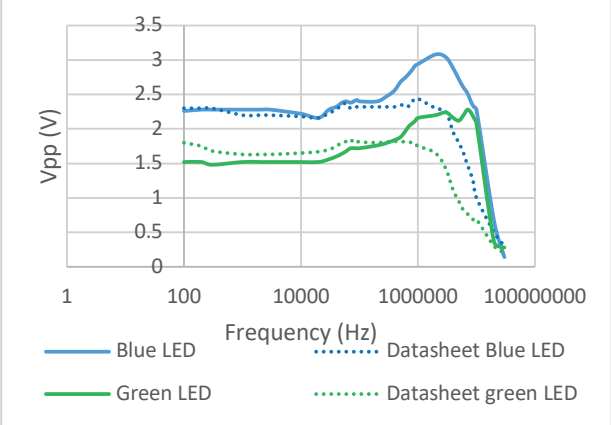


Fig. 4. Frequency response of the VLC transceiver in free space.

B. Experimental Set-up

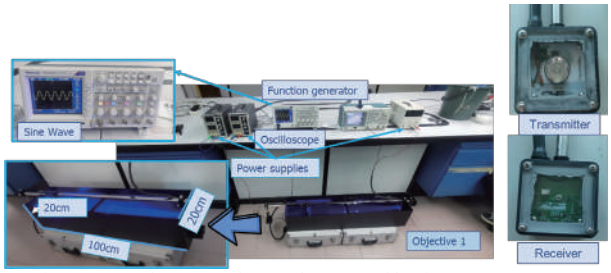


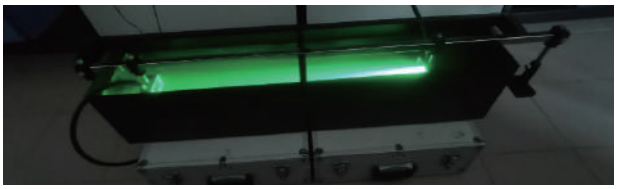
Fig. 5. Complete experimental set up of basic UOWC system

The experimental setup for UOWC transceiver includes a black tank with dimensions of 1 m length by 0.2 m wide by 0.2 m height, function as a channel reservoir and the transmitter and receiver are placed inside the water tank as shown in Fig.5. This experiment is carried-out to study the effect on received optical signal as a function of varying transmission link for saline water channel. For comparison, a clear water channel condition has also been studied. Two level of salt concentration have been chosen for the saline water, including 35 gm/L and 45 gm/L, as has been practiced by [15].

Other experimental set-up parameters are summarized in Table II. Fig. 6 shows the condition when the experiment is carried out using green LED as the light source for the transmitter. It should be noted that the minimum transmission link to measure the light intensity is 40 cm. The experiment has limitation to do the measurement for the distance less than 40 cm due to the constraint of the receiver part to detect the received signal when the transceiver is located too close to each other.

TABLE II. EXPERIMENTAL SET UP PARAMETER

Parameter	**Specification**
Light Source (Tx)	• Blue LED – 457nm • Green LED – 523nm
Viewing Angle	4 °
Aperture surface area	$1mm^2$
Effective measurement range	40-100cm
Water Types	• Clear water • Saltwater (35gm/L, 45gm/L salinity)


Fig. 6 : Light propagation using green LED as the light source

III. RESULTS AND DISCUSSION

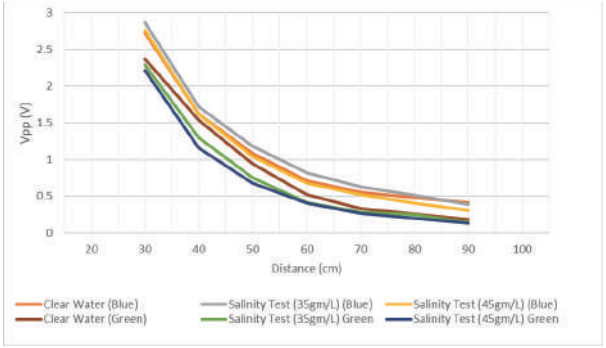


Fig. 7 Received light intensity in different condition of UOWC channel using green and blue LED.

Fig. 7 shows the plots of received light intensity in clear water and saline water with varying salt concentration of 35 g/l and 45 g/l using blue and green LED respectively. The

measure light intensity is measured in peak-to peak voltage (Vpp) as shown on oscilloscope in Fig.8.

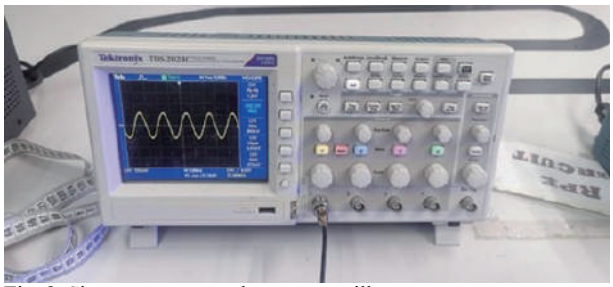


Fig. 8. Sine wave pattern shown on oscilloscope

In common, it is observed that the received intensity of the optical signal emitted by all the light sources decreased as the underwater link length increased, in agreement with the stated Beer's law [16]. It is observed that the light intensity is decreasing as the salt concentration of the UOWC channel is increasing, validating the hypothesis that the water salinity affects the performance of UOWC [15]. The salinity effect is significant when the salt concentration is reaching 45g/l. Further, an attenuation constant, *c*, is estimated to further analyse the measured light intensity. The value of *c* is estimated through linear curve fitting method as presented in Fig. 9. The slope of the linear equation deduced from the curve fitting gives the estimated value of c.

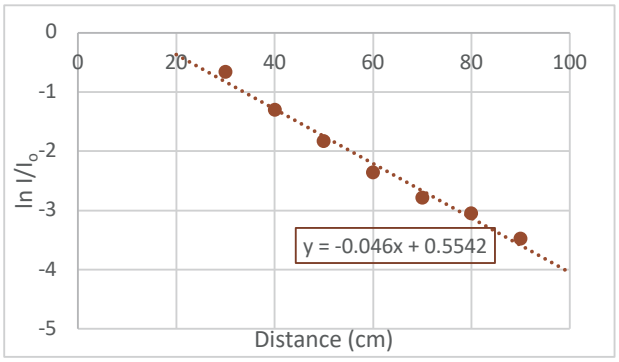


Fig. 9. Curve fitting method to estimate the attenuation coefficient (c = 0.0.46 cm^{-1}) for green LED propagates in 45 g/l of saline water.

As a result, Table III tabulates the estimated *c* for all types of water condition using blue and green LED respectively. From the table, the worst-case scenario is identified when the green LED light is propagated in 45g/l of saline water channel at which it resulted in a maximum attenuation constant of 0.046 cm^{-1}. On contrary, the minimum *c* of 0.0312 cm^{-1} is achieved when blue LED light propagates in the clear water. By theory, attenuation constant is a measure of decay of light wave in different water channels. Therefore, the minimum is the value of c suggests the better is the performance of the UOWC channel.

TABLE III. SUMMARISE OF ALL ATTENUATION COEFFICIENT, CM^{-1}

Types of water	*Clear water*	*35 gm/l salinity*	*45 gm/l salinity*
Blue LED	0.0312	0.0322	0.0359
Green LED	0.0438	0.0451	0.0460

In comparison to green LED, blue LED light wave has shown better response for all three types of of water channel. This finding is in line with the fact that the blue color spectrum attained lower absorption coefficient than the green colour spectrum as shown in the wavelength spectrum against absorption coefficient of light in pure seawater.

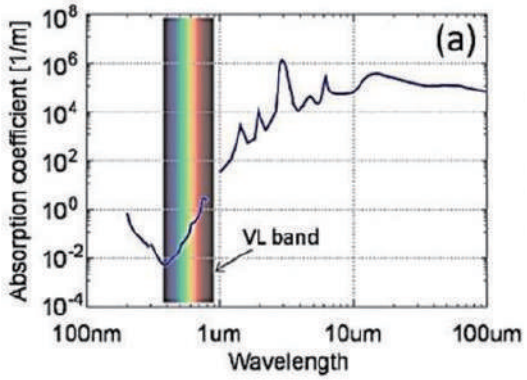


Fig. 10. The absorption coefficient of light in pure seawater.[7]

IV. CONCLUSION

This work investigated the propagation of light waves in the effect of salinity water for UOWC with the clear water channel as a comparison. The study has shown that level of salinity concentration has a reverse effect to the UOWC channel. The salinity effect is significant when the salt concentration is reaching 45g/l. In comparison to green LED, blue LED has shown better response towards saline channel condition. The worst-case condition is when green LED light source is propagated in 45 g/l saline water channel which resulted in an attenuation constant of 0.0460 cm^{-1}. In contrast, the best-case condition is achieved when blue LED is propagated in clear water channel with a attenuation constant of 0.0312 cm^{-1}. In future, it is proposed to include the effect of the water temperature and water flow into this experimental study. The findings from this study contribute to the advancement of UOWC technology by providing insight into designing optimal underwater wireless communication device.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Education Malaysia and Universiti Malaysia Terengganu for the financial assistance through Fundamental Research Grant Scheme fund with vote 59700.

.

REFERENCES

[1] Y. Chen, L. Zhang, and Y. Ling, "New approach for designing an underwater free-space optical communication system," *Front. Mar. Sci.*, vol. 9, no. August, pp. 1–14, 2022, doi: 10.3389/fmars.2022.971559.

[2] M. Lanzagorta, *Underwater Communications*, vol. 5. 2012. doi: 10.2200/S00409ED1V01Y201203COM006.

[3] M. Stojanovic, "Underwater Wireless Communications : Current Achievements and Research Challenges," *IEEE Ocean. Eng. Soc. Newsl.*, vol. 41, no. 2, pp. 1–5, 2006, [Online]. Available: http://www.ieee.org/organizations/pubs/newsletters/oes/html/spring06/underwater.html

[4] X. Wang, M. Zhang, H. Zhou, and X. Ren, "Performance Analysis and Design Considerations of the Shallow Underwater Optical Wireless Communication System with Solar Noises Utilizing a Photon Tracing - Based Simulation Platform," *Electronics*, vol. 10, no. 5, p. 632, 2021.

[5] G. Hattab, M. El-tarhuni, M. Al-ali, T. Joudeh, and N. Qaddoumi, "An Underwater Wireless Sensor Network with Realistic Radio Frequency Path Loss Model," *Int. J. Distrib. Sens. Networks*, vol. 2013, 2013, [Online]. Available: http://dx.doi.org/10.1155/2013/508708

[6] M. Furqan, A. Dushantha, N. K. Jayakody, Y. Alexandrovich, and C. Soféine, *Recent Advances and Future Directions on Underwater Wireless Communications*. Springer Netherlands, 2019. doi: 10.1007/s11831-019-09354-8.

[7] G. S. Spagnolo, L. Cozzella, and F. Leccese, "Underwater Optical Wireless Communications: Overview," *Sensors (Switzerland)*, vol. 20, no. 8, 2020, doi: 10.3390/s20082261.

[8] Ali, M. F., Jayakody, D. N. K., Chursin, Y. A., Affes, S., & Dmitry, S. (2020). Recent Advances and Future Directions on Underwater Wireless Communications. Archives of Computational Methods in Engineering, 27(5), 1379–1412. https://doi.org/10.1007/s11831-019-09354-8

[9] K. Y. Islam, I. Ahmad, D. Habibi, and A. Waqar, "A survey on energy efficiency in underwater wireless communications," *J. Netw. Comput. Appl.*, vol. 198, p. 103295, 2022, doi: https://doi.org/10.1016/j.jnca.2021.103295.

[10] S. Zhu, X. Chen, X. Liu, G. Zhang, and P. Tian, "Recent progress in and perspectives of underwater wireless optical communication," *Progress in Quantum Electronics*, vol. 73, no. July. Elsevier Ltd, p. 100274, 2020. doi: 10.1016/j.pquantelec.2020.100274.

[11] S. Q. Duntley, "Light in the sea," Jounal of the Optical Society of America, vol. 53, no. 2, p.214, 1963. doi: https://doi.org/10.1364/josa.53.000214

[12] Z. Zeng, S. Fu, H. Zhang, Y. Dong, and J. Cheng, "A Survey of Underwater Optical Wireless Communications," *IEEE Commun. Surv. Tutorials*, vol. 19, no. 1, pp. 204–238, 2017, doi: 10.1109/COMST.2016.2618841.

[13] M. Elamassie and M. Uysal, "Vertical Underwater Visible Light Communication Links: Channel Modeling and Performance Analysis," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 10, pp. 6948–6959, 2020, doi: 10.1109/TWC.2020.3007343.

[14] Rocksolid Technology," VLC Transceiver Data Sheet, September 2022.

[15] S. Kumar, S. Prince, J. Venkata Aravind, and S. Kumar G, "Analysis on the effect of salinity in underwater wireless optical communication," *Mar. Georesources Geotechnol.*, vol. 38, no. 3, pp. 291–301, Mar. 2020, doi: 10.1080/1064119X.2019.1569739.

[16] W.H.W.Hassan, M. S. Sabril, F. Jasman, and S. M. Idrus. "Experimental Study of Light Wave Propagation for Underwater Optical Wireless Communication (UOWC)." J. Commun. 17, no. 1 (2022): 23-29.

Adjusting Repetitions for Error Reduction in Underwater Optical Wireless Communication

Huicheol Shin
Korea Institute of Ocean Science and Technology
South Korea
shc0305@kiost.ac.kr

Seungjae Baek
Korea Institute of Ocean Science and Technology
South Korea
baeksj@kiost.ac.kr

***Abstract*—In this paper, we analyze the impact of underwater environmental factors such as turbidity on the communication performance of underwater wireless optical links established between underwater sensor node deployed on the seabed and unmanned surface vehicles (USV) on the sea surface. Furthermore, to reduce the error rate of UOWC, we implement a deep reinforcement learning (DRL) algorithm that adjusts the coding rate of the repetition code based on the environmental information from sensors attached to the USV. The DRL algorithm collects turbidity data in real-time to calculate the bit error rate (BER) and determine the number of data repetitions to meet the required packet error rate (PER). To evaluate the performance of the proposed algorithm, simulations in a virtual environment with irregular variations in turbidity and water tank experiment with optical model based UOWC system were conducted, both of which confirmed satisfying the required communication performance.**

***Keywords*—Underwater optical wireless communication, deep reinforcement learning, repetition code, coding rate**

I. INTRODUCTION

In recent times, there has been growing interest in underwater optical wireless communication (UOWC) as an effective solution for various applications like monitoring ocean pollution, exploring the depths of the ocean, predicting disaster, and collecting geological data from underwater environments [1-3]. This technology enables the detection, gathering, and processing of underwater data with high transmission speeds, low latency, and exceptionally wide bandwidth. However, the challenging conditions underwater, including turbidity, turbulence, and temperature variations, contribute to issues such as absorption, scattering, and signal attenuation, posing significant hurdles for UOWC system. Consequently, these environmental factors adversely affect the performance of UOWC by reducing transmission distances and increasing error rates, thereby compromising the reliability of underwater communication networks [4]. To address these challenges, many researchers have recently conducted research aimed at applying various wavelengths, adjusting modulation techniques, and adapting coding rates [5, 6]. The authors in [5] analyzed and compared the bit error rate (BER) based on various modulation techniques. In [6], the author analyzed the transmission loss of optical signals based on the wavelength of light. The author confirmed that the blue and green spectra in the wavelength range of 450-550nm are relatively less affected by attenuation. In this paper, we implement a deep reinforcement learning (DRL) algorithm to adjust the coding rate to reduce the error rate of the UOWC link. The UOWC network is considered in scenario such as transmitting data from underwater sensor (US) fixed on the seabed to unmanned surface vehicle (USV) on the sea surface. The proposed algorithm determines the optimal data repetition (i.e., coding rate) to achieve the required communication performance of the UOWC link (i.e., required packet error rate (PER)). The DRL model on the USV calculates the BER based on the collected turbidity data and learns the optimal coding rate to achieve the required PER considering the irregularly changing attenuation characteristics. To analyze the performance of the proposed algorithm, simulation is conducted in a virtual environment with irregularly varying turbidity, and water tank experiment with the applied optical modem and the proposed algorithm is performed to evaluate performance in real underwater environment.

II. SYSTEM MODEL

Fig. 1 represents the system model of the UOWC considered in this study for data transmission. To acquire collected data, the receive node (i.e. USV) moves to the location of the transmit node (i.e. US) using global positioning system (GPS) information. Afterwards, the USV transmits acoustic signal to establish the UOWC link and performs initial synchronization [7]. The reason for considering acoustic communication is that, unlike optical communication, which can be prone to misalignment issues because of its strong straightforwardness nature, acoustic communication offers slower transmission speeds but wider transmission ranges, leading to increased reliability in signal transmission. the communication link is

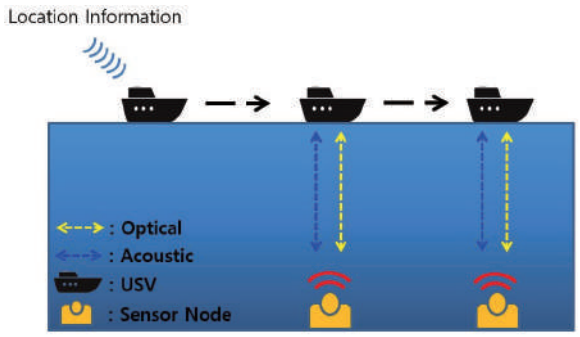


Fig. 1. UOWC system model

established, USV analyzes the BER based on the turbidity information and requests data transmission by selecting the coding rate.

To measure the BER of the established UOWC link, the received optical signal power is needed. When transmitting an optical signal, the received power at the receive node can be calculated as follows [5]:

$$P_R = P_T \frac{A_r}{4\pi d^2} e^{-c(\lambda)d}, \quad (1)$$

Where P_T is power of the optical signal transmitted by the transmit node, A_r is the aperture size of the receive node, $\boldsymbol{d}$ is distance of the UOWC link and $c(\lambda)$ is attenuation coefficient according to wavelength λ. Specifically, the attenuation coefficient $c(\lambda)$ can be defined as the sum of the absorption coefficient $a(\lambda)$ and scattering coefficient $b(\lambda)$, and it depends on the wavelength of light and the type of water.

The received optical signal is converted into an electrical signal through the photodetector integrated into the optical modem installed on the USV. In this paper, a photodetector employing SiPM, which has high sensitivity characteristic, is considered, and calculates the photocurrent based on the SiPM as follows:

$$I_P = \frac{P_R E G q}{hc/\lambda}, \quad (2)$$

where $\boldsymbol{E}$ is the photo detection efficiency, $\boldsymbol{G}$ is the gain of SiPM, $\boldsymbol{q}$ is the electric charge, $\boldsymbol{h}$ is Planch constant and $\boldsymbol{c}$ is the light speed. Subsequently, when the photo current is obtained, the signal-to-noise ratio (SNR) of the received signal can be calculated as follow:

$$SNR = \frac{I_P}{\sqrt{2qFBG(I_P + (2I_D + I_B))}}, \quad (3)$$

where F is the noise figure, B is the bandwidth, I_D is the dark current noise and I_B is the background noise current, as a result, BER of the UOWC link based on derived SNR, and the BER is determined differently depending on the modulation technique. In this study, BER is computed using pulse position modulation (PPM) technique as follows:

$$BER_{M-PPM} = \frac{1}{2} erfc\left(\frac{1}{2\sqrt{2}}\sqrt{SNR\frac{M}{2}\log_2 M}\right), \quad (4)$$

The calculated BER represents the communication performance of the UOWC link, and the error rate can be reduced by applying error correction code (ECC). One of the ECC method, called repetition code, involves transmitting data repeatedly to reduce error rate. The n-bit signal with ECC is referred to as a packet. In this work, we analyze the performance of UOWC link based on variations in the coding rate, such as the number of repetitions in the repetition code, and implement a DRL algorithm to determine the optimal coding rate that satisfy the required communication performance (i.e., PER_{req}).

III. PROPOSED ALGORITHM

The objective of this study is to minimize the efficiency loss of data capacity caused by repetition code while satisfying the PER_{req} for maintaining a stable UOWC link. In this regard, we formulate the problem based on Markov decision process (MDP) and propose a DRL framework.

In our MDP formulation, the receive node is the decision maker with an agent that selects the coding rate. The agent is composed of a tuple (S, A, r), where S is the state space, A is the action space, and r is the reward function for action. First, the state space in a time slot is t defined as follows:

$$S(t) = \{s_{his}(t-1), turbidity(t-1)\}, \quad (5)$$

where $s_{his}(t-1)$ is the historical information on the actions and rewards experienced by the agent and $turbidity(t-1)$ is the environment information at time slot $t-1$. To limit the size of the state space, S_{his} is limited to a size of m, and as learning progresses, S_{his} for each time slot is defined as follows:

$$s_{his}(t-1) = \{a(t-m), r(t-m), \ldots, a(t-1), r(t-1)\}, \quad (6)$$

Secondly, action $a(t) = \text{coding rate} \in A(t)$ at time slot t is an element in the action space $A(t)$, representing the set of the number of repetitions that can be chosen when the underwater sensor transmits data.

$$A(t) = \{1,3,\ldots,15,17\}, \quad (7)$$

Finally, the reward function r is determined based on the transmission of data performed according to the selected action, and can be calculated when the error rate of the received signal meets the PER_{req}.

$$r(t) = \begin{cases} L - N, & \text{PER} < \text{PER}_{\text{req}} \\ 0, & \text{otherwise} \end{cases}, \quad (8)$$

where L is a predefined constant and N is the number of repetition. Therefore, the reward function of the agent can obtain a higher value as the minimum number of repetitions performed that satisfies the PER_{req}.

IV. PERFORMANCE EVALUATION

A. Network environment

To analyze the performance of the proposed algorithm, we considered an UOWC environment with irregular variations in turbidity. This was done because the attenuation coefficient is correlated with turbidity [8]:

$$c(\lambda) = c_{clean} + (0.076 \times turbidity), \quad (9)$$

where C_{clean} represents clean water and has a value of $0.3m^{-1}$.

Fig. 2 illustrates the changes in turbidity of the generated UOWC network and corresponding variations in communication performance. As the turbidity value increases, it can be observed that the BER significantly increases when transmitting data. Therefore, the objective is to select the optimal coding rate with the adaptation of repetition code to achieve a PER lower than PER_{req} when the required communication performance in terms of the demanded error rate cannot be satisfied in a specific underwater environment.

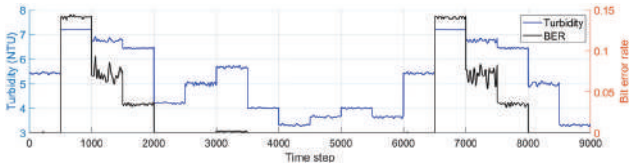


Fig. 2. BER variation according to turbidity

B. Simulation result

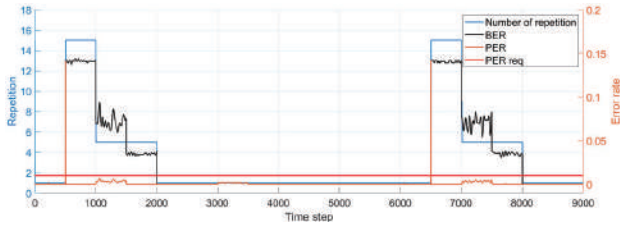


Fig. 3. PER variation according to turbidity

Fig. 3 shows the simulation results of the proposed DRL-based coding rate adjustment algorithm. When transmitting signal without applying the repetition code, it can be seen that the BER does not meet the required communication performance ($\text{PER}_{\text{req}} = 0.01$) at a specific turbidity value, but the proposed algorithm adjusts the coding rate of the repetition code according to the turbidity change, maintaining the PER lower than PER_{req}. As a result, the proposed algorithm allows for the selection of the minimum coding rate that guarantees both communication performance and efficient transmission of data capacity.

After the simulation of the proposed algorithm, we established a real UOWC environment for performance evaluation through integration with an optical modem, as illustrated in Fig. 4.

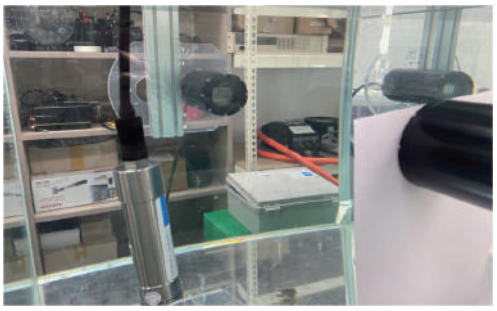


Fig. 4. UOWC link implementation

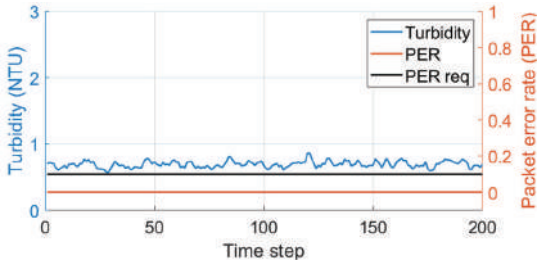


Fig. 5. Communication performance of proposed algorithm

The performance of the proposed algorithm was analyzed in a UOWC environment constructed in a water tank. Fig. 5 shows a graph of maintaining communication performance that satisfies the required performance ($\text{PER}_{\text{req}} = 0.1$) of UOWC link using the DRL model. Based on the simulation result, the proposed algorithm analyzes the channel quality based on the measured turbidity data and selects the optimal coding rate to maintain the PER lower than PER_{req}, thus ensuring the communication performance of the UWOC link.

V. CONCLUSION

In this study, we analyzed the channel quality of the UOWC link based on turbidity data, which is underwater environmental information, and implemented a DRL-based coding rate adjustment algorithm that minimizes the number of repetition while reducing the error rate. As a result, the proposed DRL model was able to maintain communication performance that satisfies PER_{req} by selecting the optimal coding rate. In the future, we plan to perform performance analysis in scenarios where the turbidity values in the water tank vary widely. In addition, we plan to analyze channel quality according to various environmental information (e.g. temperature, illumination, and depth) and compare the performance of DRL algorithms adopting various ECC schemes.

ACKNOWLEDGEMENT

This research was supported in part by the project tilted 'Development of Polar Region Communication Technology and Equipment for Internet of Extreme Things (IoET)' funded by the Ministry of Science and ICT (MSIT).

REFERENCES

[1] N. Saeed, A. Celik, T.Y. Al-Naffouri and M.-S. Alouini, "Underwater optical wireless communications, networking, and localization: A survey," *Ad Hoc Networks*, vol. 94, pp. 1-34, 2019.

[2] G.S. Spagnolo, L. Cozzella, and F. Leccese, "Underwater optical wireless communications overview," *Sensors*, vol. 20, pp. 1-14, 2020.

[3] Z. Zeng, S. Fu, H. Zhang, Y. Dong, and J. Cheng, "A survey of underwater optical wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 204–238, 2017.

[4] I. N'doye, D. Zhang, M.-S. Alouini, and T.-M. Laleg-Kirati, "Establishing and maintaining a reliable optical wireless communication in underwater environment," *IEEE Access*, vol. 9, pp. 62519-62531, 2021.

[5] M. A. Ali, "Comparison of modulation techniques for underwater optical wireless communication employing APD receivers," *Resarch Journal of Applied Scienecs, Engineering and Technology*, vol. 6, pp. 707-715, 2015.

[6] S. Q. Duntley, "Light in the Sea," *J. Opt. Soc. Am.*, vol. 53, no. 2, pp. 214-233, Feb 1963.

[7] J. Wang, W. Shi, L. Xu, L. Zhou, Q. Niu, and J. Liu, "Design of optical-acoustic hybrid underwater wireless sensor network," *Journal of Network and Computer Applications*, vol. 92, pp. 59-67, 2017

[8] K. Rachael, "Turbidity and light attenuation in coastal waters of the great barrier reef," PhD thesis, James Cook University, 2015.

Aerodynamically Considered Antenna Housing with Dielectric Lens for Extreme Environments

Jinkyu Bang
Maritime ICT Mobility Research Department
KIOST
Busan, South Korea
jinkyu.bang@kiost.ac.kr

Man-Woong Heo
Ocean Space Development Energy Research Department
KIOST
Busan, South Korea
mwheo@kiost.ac.kr

Seungjae Baek
Maritime ICT Mobility Research Department
KIOST
Busan, South Korea
baeksj@kiost.ac.kr

Jae Hee Kim
Electrical Electronics and Communication Engineering
KOREATECH
Cheonan, South Korea
jaehee@koreatech.ac.kr

***Abstract*—This paper introduces an antenna housing design intended for use in extreme environments. Typically, antenna housing for wireless communication is designed in a flat-shaped structure to minimize distortion of the radiation pattern. However, while a curved structure offers advantages in environments with significant wind impact, it can also lead to decreased antenna gain due to phase distortion resulting from the housing shape. To address this issue, this study proposes an antenna housing structure for extreme environments that incorporates both the aerodynamic drag and the dielectric lens.**

***Index Terms*—Antenna, drag coefficient, antenna housing, dielectric lens**

I. Introduction

The purpose of antenna housing is to protect the antenna elements and feed networks that comprise the antenna from the effects of the external environment. Generally, antenna housing structures are designed in a flat shape, parallel to the antenna aperture, to minimize distortion of the antenna radiation pattern. However, flat-shaped housings have the disadvantage of presenting a large effective cross-sectional area susceptible to strong winds. This is particularly problematic in extreme environments like Antarctica, where winds can exceed 40 m/s [1], potentially causing significant damage to the antenna system. To mitigate this issue, the design of the antenna housing could be altered to either a curved or a tilted structure. However, such modifications may lead to phase differences in the path of electromagnetic waves due to discrepancies between the antenna aperture and the housing shape, potentially reducing the antenna gain in the required communication band due to phase distortion. In this study, we propose a new approach to address these issues. By developing an antenna housing structure that maximizes antenna gain while being robust against strong winds, we aim to design a structure suitable for extreme environments. Our study simulates the aerodynamic drag on four types of housing structures and seeks to implement a design that incorporates a dielectric lens to counteract the gain reduction caused by the selected housing structure.

II. Analysis of Initial Structures

A. Effects of antenna housing design on drag

In this study, the effects of antenna housing shapes on aerodynamic drag were investigated by solving a 3D Reynolds-averaged Navier-Stokes (RANS) analysis. Four housing shapes, namely, flat, dome, pyramid, and dome with dimples, as shown in Fig. 1, were considered for the designs. To evaluate the aerodynamic drag of the antenna housings, the drag coefficient (C_d), as defined in (1), was used as a performance function.

$$C_d = \frac{2F_d}{\rho v^2 A} \tag{1}$$

where, F_d is the drag force, and ρ, v, and A indicate the density of working fluid, flow speed, and projected area of the antenna, respectively.

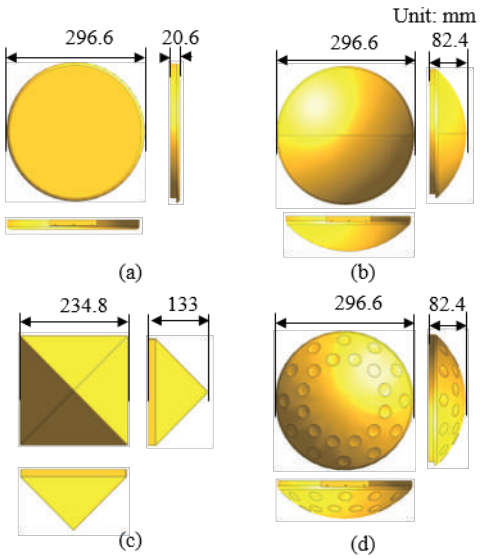


Fig. 1. Antenna housing structures: (a) flat structure (b) dome structure (c) pyramid structure (d) dome structure with dimples

The drag coefficient (C_d) of the antenna housing structures, evaluated through RANS analysis, is listed in Table I.

Compared to the flat structure, the (C_d) values for the dome, pyramid, and dome with dimples were reduced to 43.0 %, 61.8 %, and 41.9 %, respectively. It was confirmed that the drag force decreased by 2.6 % by applying dimples to the dome housing structure.

TABLE I
DRAG COEFFICIENT: SIMULATION RESULTS

Items	Antenna housing			
	Flat	Dome	Pyramid	Dimples
Wind speed $[m/s]$	40	40	40	40
C_d	0.450	0.193	0.278	0.189
C_d/C_dflat	1.0	0.430	0.618	0.419

Pressure contours on the antenna housings are represented in Fig. 2. It can be found that the antenna housing of the dome with dimples has a larger low-pressure area compared to others.

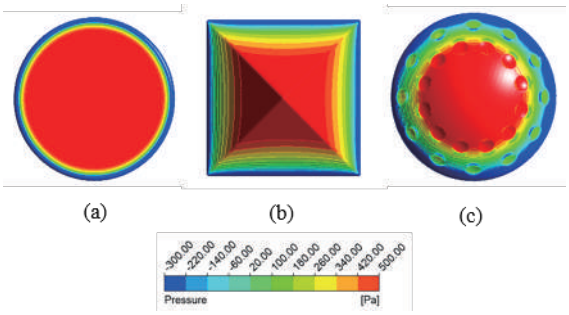


Fig. 2. Pressure contours on the antenna housing: (a) flat structure (b) pyramid structure (c) dome structure with dimples

B. Antenna performance of initial housing structures

Fig. 3 depicts the antenna structure to be used with the housing, taking into account the drag coefficient. The antenna type is a microstrip patch antenna, operating at a frequency of 5.8 GHz and utilizing FR4 material with a relative permittivity of 4.4 and a thickness of 1.6 mm. For gain enhancement, the antenna is designed as a 2×2 array, with each element of the patch measuring 11×11 mm and the spacing between elements being 14.86 mm. Electromagnetic (EM) analysis is necessary to verify the radiation performances of the structures combining the antennas with their respective housings, as depicted in Fig. 1 and Fig. 3. The pyramid structure, which had the highest drag coefficient under antenna housing conditions, was excluded from the analysis. The housing material used for EM analysis was polycarbonate (PC) with a relative permittivity of 3.5. The simulation for antenna analysis was conducted using $Sim4Life$, a full 3D EM simulation tool based on the finite-difference time-domain (FDTD) method [2].

Fig. 4 shows the results of the antenna reflection coefficient (S11) under each housing structure. Compared to the case with only the antenna, there is a tendency for the resonance frequency to shift upwards due to the housing. However, since

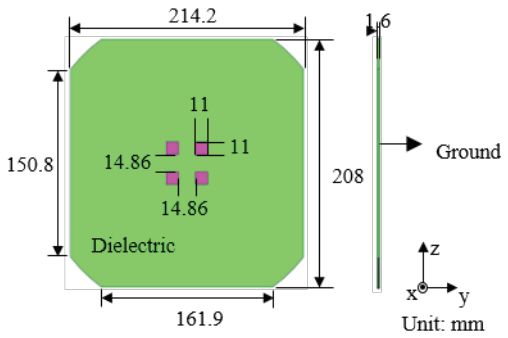


Fig. 3. Antenna structure: 2×2 microstrip array patch antenna

the S11 is below -7 dB in the target frequency band around 5.8 GHz, it is evident that the antenna resonance is well-formed. Fig. 5 illustrates the radiation patterns of the E-plane for each housing condition at the frequency of 5.8 GHz. Given their similarity to the E-plane trends, the H-plane radiation pattern results have been omitted. As can be observed from the patterns, the flat-shaped housing structure exhibits the highest antenna gain at the central position (0 degree). However, in the cases of the dome-shaped housing and the dome-shaped housing with dimple structures, it can be seen that the antenna gains at the central position decrease by more than 10 dB compared to the flat-shaped housing structure. These reductions are due to the phase path differences caused by the curvature of the dome-shaped housing. This analysis underscores the trade-offs between housing design and antenna performance, particularly regarding the aerodynamic drag coefficient and antenna gain or directivity. It highlights the necessity of carefully considering housing curvature in the design process or devising additional improvement methods to minimize performance losses.

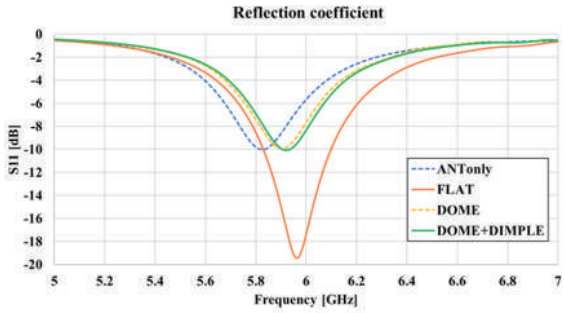


Fig. 4. S11 in antenna bandwidth under each housing condition

III. PROPOSED HOUSING STRUCTURE AND RESULTS

In Fig. 5, it was observed that the antenna gain significantly attenuates at the center of the housing, where the highest antenna gain must be required, especially in the dome-shaped structure which exhibits the best performance in terms of drag coefficient due to the curvature of the housing. In this section, to solve this issue, it is proposed to add a dielectric lens structure at the center of the dome-shaped housing with

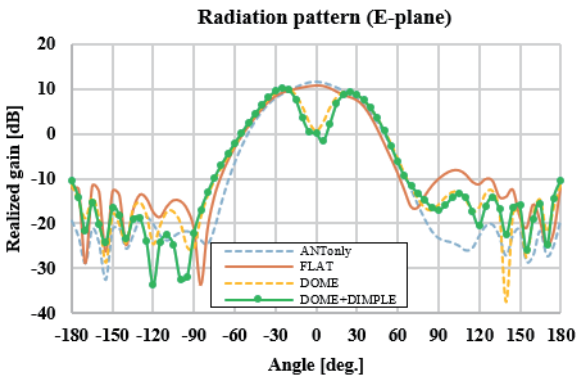


Fig. 5. Radiation patterns for each housing condition: E-plane

dimple structures, which has the same dielectric properties as the housing.

A. Dielectric lens design

Various research on dielectric lenses for antenna gain improvement, such as the design of dielectric lenses, has been conducted [3], [4]. Dielectric lens antennas operate similarly to optical lenses [5]. The design of the lens takes into account the boundary conditions between two media and the path difference of propagation at the lens boundary, where the field impinges perpendicularly. The lens structure is designed in a convex shape, considering the dome housing structure, and is set with the same relative permittivity of 3.5 as the housing material (PC).

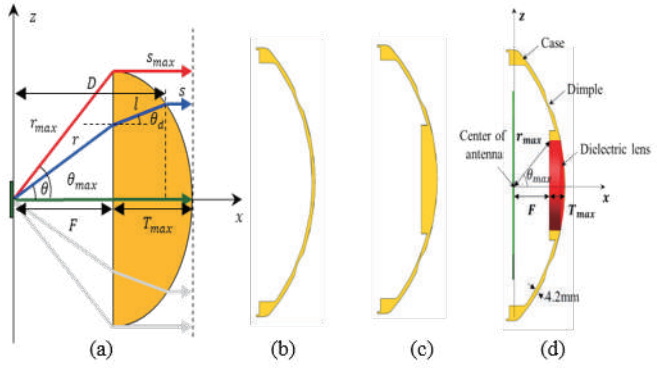


Fig. 6. Cross-sectional views of the housing structure with integrated dielectric lens: (a) dielectric lens structure (b) housing structure (dome-shaped housing with dimples) (c) housing structure with dielectric lens (d) dielectric lens parameters reflected in the housing structure

Fig. 6 (a) illustrates the structure for the dielectric lens design [3]–[5]. Using [3]–[5], and assuming the antenna's central position as the focal point, the direct distance from the focal point to the dielectric lens is denoted as F, and the maximum distance to the center of the convex lens is represented as $T_{\max}$. Considering the dimple positions of the dome housing structure, the maximum angle of the dielectric lens is denoted as $\theta_{\max}$, with the distances traveled before and after passing through the dielectric under this condition represented as $r_{\max}$ and $s_{\max}$, respectively. Assuming an arbitrary angle of incidence as θ, the distances traveled before and after passing through the dielectric at this angle are denoted as r and s, respectively, while the internal travel distance within the dielectric and the refracted angle are represented as l and θ_d, respectively. To ensure that the propagation paths are identical after passing through the lens, (2) must be satisfied, where n represents the reflection coefficient between air and the dielectric lens.

$$r + nl + s = r_{max} + s_{max} = F + nT_{max} \tag{2}$$

Equation (2) can be re-expressed in terms of distance and angle as (3):

$$\frac{F}{\lambda_0 \cos\theta} + \frac{D-F}{\lambda_d \cos\theta_d} + \frac{T_{max} - (D-F)}{\lambda_0 \cos\theta} = \frac{F}{\lambda_0 \cos\theta_{max}} + \frac{T_{max}}{\lambda_0} = \frac{F}{\lambda_0} + \frac{T_{max}}{\lambda_d} \tag{3}$$

In (3), λ_0 is the wavelength in air, and λ_d is the wavelength inside the dielectric lens. Using (3) and Snell's law, the lens thickness can be determined from (4), and the trace of the lens curvature can be found from (5) under the given conditions. Here, ϵ_d represents the dielectric constant of the dielectric lens.

$$T_{max} = \frac{F}{(\sqrt{\epsilon_d} - 1)}\left(\frac{1}{\cos\theta_{max}} - 1\right) \tag{4}$$

$$D = F\left(\frac{\epsilon_d}{\sqrt{\epsilon_d - \sin^2\theta}} + \frac{1}{\cos\theta_{max}} - 1\right)/\left(\frac{\epsilon_d}{\sqrt{\epsilon_d - \sin^2\theta}} - 1\right) \tag{5}$$

These equations are crucial for designing the dielectric lens to work efficiently with the dome-shaped antenna housing. By correctly calculating T_{max} and D, the lens will effectively focus electromagnetic waves onto the antenna, enhancing its radiation performance despite the curvature of the dome housing. This approach requires precise adjustment of the lens' properties and dimensions to ensure optimal performance.

Fig. 6 (b)–(d) shows cross-sectional views of a dome-shaped housing incorporating a dimple structure with designed dielectric lens conditions. Fig. 6 (b) displays the cross-section of the conventional dome-shaped housing, Fig. 6 (c) shows the cross-section of the dome-shaped housing including the designed dielectric lens, and Fig. 6 (d) presents the dielectric lens structure and lens design parameters as applied in Fig. 6 (c). To eliminate the variations in dielectric thickness caused by the dimple structure, setting r_{max} to 70.47 mm, θ_{max} to 44.8 degrees, and fixing F to 50 mm allows for the calculation of T_{max} as 23.5 mm according to (4). While the maximum radius of the lens cross-section is calculated to be 49.65 mm, considering the curvature margin at the lens edge, it is designed to be 55 mm.

B. Antenna performance of proposed housing structure

Fig. 7 depicts the realized gain across frequencies for housing conditions that include a dielectric lens. While the antenna gain for both the dome-shaped housing and the dome-shaped housing with dimple structures is around 10.3 dB, the inclusion of a dielectric lens in the housing structure boosts the antenna gain to over 14 dB, indicating an improvement of more than 3 dB in antenna directivity due to the dielectric lens. Fig. 8 presents the radiation patterns of the E-plane for each housing condition at a frequency of 5.8 GHz. The H-plane radiation pattern results are excluded due to their similarity to the E-plane trends. Fig. 8 reveals that the dome-shaped housing with the dielectric lens structure exhibits the highest antenna gain at the center position (0 degree), demonstrating an improvement of over 3 dB compared to the antenna without any housing structures. Fig. 9 illustrates the real part of the electric field distribution at 5.8 GHz for (a) a dome housing with a dimple structure and (b) a dome housing with a dimple structure and a dielectric lens. In Fig. 9 (a), the dome housing curvature does not focus the electric field at the center. Conversely, Fig. 9 (b) shows that adding a dielectric lens centers the electric field, improving antenna gain. These results highlight the effectiveness of integrating a dielectric lens into the antenna housing, enhancing performance by mitigating curvature limitations.

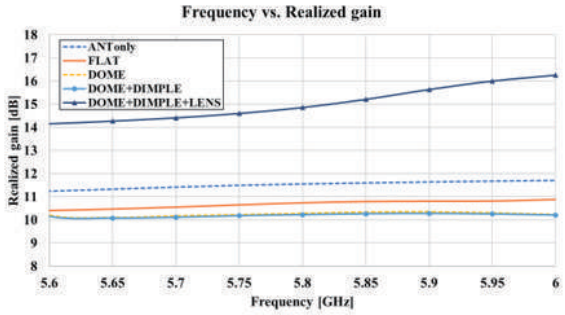


Fig. 7. Realized gain in antenna bandwidth under housing conditions

IV. Conclusion

In this study, we proposed an antenna housing structure designed to minimize the impact of strong winds in extreme environments and to maximize antenna radiation characteristics. To assess the effects of strong winds in extreme environments, we analyzed the drag coefficient of the housing structures. For dome-shaped housing structures, which have low drag coefficients, the antenna gains are diminished at the center due to the curvature of the housing. To improve this, we designed a dielectric lens shape and incorporated it into the housing structure. This study presents a method for designing a dome-shaped housing with dimple structures that includes a dielectric lens, and the effectiveness of the proposed method has been demonstrated through EM simulations. The EM simulation results show an improvement of more than 3dB in antenna gain in frequency bandwidth with the proposed

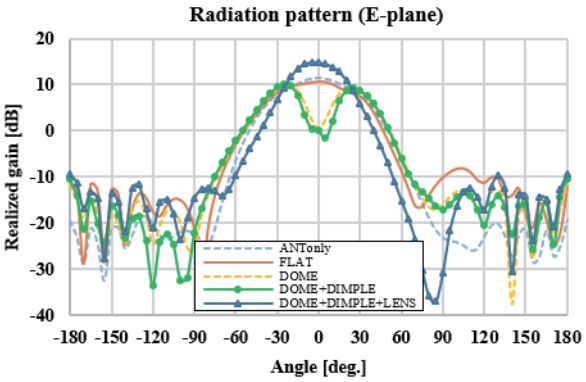


Fig. 8. Radiation patterns for each housing condition including dielectric lens structure: E-plane

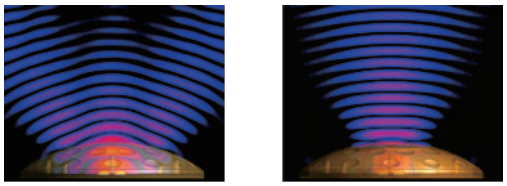


Fig. 9. Real part of electrical field distribution: (a) dome housing with dimple (b) dome housing with dimple and dielectric lens

housing structure that includes a dielectric lens, compared to traditional flat-shaped housing structure. The results of this study provide important guidelines for the design and optimization of antenna housing structures in extreme environments and are expected to make a significant contribution to improving antenna radiation performance in actual extreme conditions.

Acknowledgment

This research was supported by Sim4Life for Science Program by ZMT.

References

[1] H. Kwon, S.-J. Park, S. Lee, S.-J. Kim and B.-M. Kim, "A Numerical Simulation of Blizzard caused by Polar Low at King Sejong Station, Antarctica," The 11th Antarctic Meteorological Observation, Modeling & Forecasting Workshop, 2016.

[2] Sim4Life by ZMT, https://www.zmt.swiss.

[3] S.-M. Moon, H.-K. Kim, I.-K. Kim, S.-J. Lee, T.-S. Kim and H.-C. Lee, " The Design of Ka-band Lens Antenna for Navigation Radar on Helicopter," Journal of the Korea Institute of Military Science and Technology, vol. 7, no.4, pp. 53–60, December 2004.

[4] B. Seo, I. Yoon, and J. Oh, "Practical Design Considerations for Compact Array-Fed Huygens' Dielectric Lens Antennas," Sensors, no. 3, 538, 2019.

[5] C.A. Fernandes, J.R. Costa, E.B. Lima, Dielectric Lens Antennas. In Handbook of Antenna Technologies; Chen, Z.N., Ed.; Springer: Berlin/Heidelberg, Germany, 2016, pp. 1001–1064.

Reusable Fish Biologger

Juhyun Kim
Maritime ICT·Mobility
Research Department (KIOST),
Marine Technology and
Convergence Engineering (UST)
KIOST, University of Science and
Technology (UST)
Busan, Korea
jhk@kiost.ac.kr

Chulsu Park
Maritime ICT·Mobility
Research Department
MET-Solutions
Busan, Korea
fehpark@met-solutions.co.kr

Seungjae Baek*
Maritime ICT·Mobility
Research Department
KIOST
Busan, Korea
baeksj@kiost.ac.kr

Abstract*— A biologger is a data storage device that is attached to a living organism to collect and record environmental and ecological data such as location, temperature, light intensity, and humidity. Marine biologgers used for fish and marine mammals are designed to be watertight and pressure resistance, with an integrated battery and electronic components. However, this integrated structure makes reuse after retrieval impossible and poses challenges in conducting sufficient pre-testing to maximize actual data collection period after the target is released. Therefore, in this paper, we propose a fish biologger with a replaceable battery. The proposed biologger is divided into a system part consisting of sensors, storage devices, and antennas, and a battery part, which includes the battery and detachment unit. The replaceable battery biologger has been proven to collect various marine environmental data over long periods in actual marine environments and can be reused by replacing the battery when necessary.

Keywords—Biologger,i reuse,i detachable,i satellitei communications

I. Introduction

Recently, there has been a growing interest in biologgers for acquiring marine environmental information[1]. A biologger is a data storage device that is attached to living organisms to collect and record environmental and ecological data such as location, temperature, light intensity, and humidity. Marine biologic biologgers designed for attaching to fish migrating at various depths are designed with an integrated structure of electronic components and batteries to ensure water-tightness and pressure resistance. However, such an integrated structure renders the device non-reusable after recovery and presents challenges in conducting sufficient pre-deployment tests to maximize data collection periods.

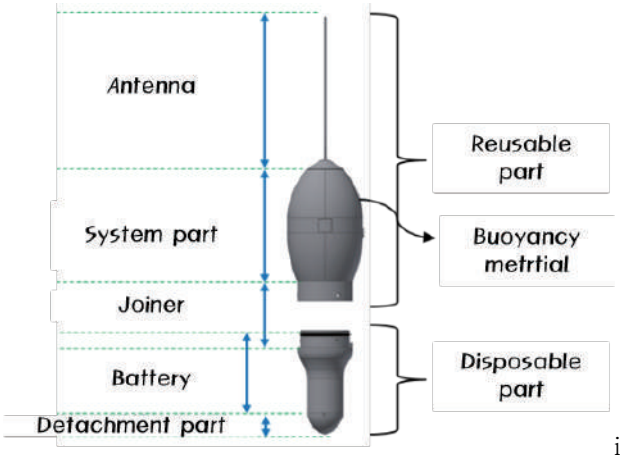
Consequently, this paper proposes a battery-replaceable fish biologger for acquiring marine environmental information. The proposed biologger is divided into two parts: a system part, which includes sensors, storage devices, and antennas; and a battery part, which consists of the battery and a detachable unit. After a preset time, it floats to the surface and transmits the collected data via satellite communication. The replaceable battery biologger minimizes power consumption through hardware and software optimization. In addition, it has been proven to collect various marine data over long periods in actual marine environments and can be reused by replacing the battery when necessary.

The structure of this paper is as follows: section 2 describes the functionality and detailed design of each component of the fish biologger. section 3 presents the experimental results of the proposed biologger. Finally, section 4 concludes the paper.

II. Design of Fish Biologger

The structure of the biologger proposed in this paper is as shown in Fig. 1. The biologger is mainly divided into reusable and disposable parts. The reusable part includes the antenna, system part, and buoyancy material, and the system part includes a basic system part, a sensor part, and a communication part. The disposable part consists of the battery and the detachment unit.

The basic system part manages the operation through biologger control, the sensor part contains various sensors for collecting marine environmental information, and the communication part is responsible for satellite communication through the antenna at the sea surface. The buoyancy material is responsible for securing buoyancy for the device to rise to the sea surface after removal from the water before satellite communication, and for keeping the antenna on the surface for stable satellite communication at the sea surface. The battery supplies power to the biologger, and the detachment unit is used for attaching or detaching to the fish.



i

Fig. 1. Biologger geometry

The antenna is made of a shape-memory alloy to increase the success rate of satellite communication in harsh marine environments. For satellite communication, a module from Collecte Localisation Satellites (CLS) was selected and customized to enable ARGOS satellite communication capabilities for the device[3]. Fig. 2 illustrates the satellite communication module mounted on the biologger.



Fig. 2. Biologger with ARGOS communication board

The system part has an interface capable of receiving data from various sensors. Fig. 3 is a block diagram of the system part, which mainly consists of the basic system prat, sensor part, and communication part

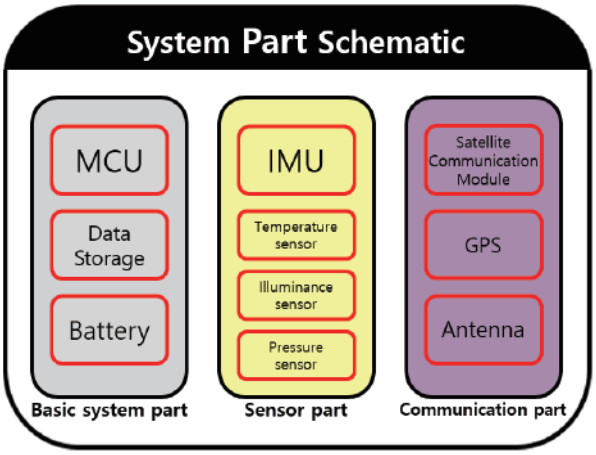


Fig. 3. Schematic of the system part internals

The basic system part is comprised of storage devices and an MCU (Micro Control Unit), specifically the STM32L451CEU6, known for its capability in low-power operation [4].

The sensor part includes IMU, temperature sensors, light sensors, and pressure sensors to identify the movement path and habitat characteristics of the target object. In the communication part, the power circuit of the satellite communication part and the RF amplifier circuit were designed to enable high-power transmission and stable data transmission for satellite communication.

The buoyancy material used was BZ Grade [5] from Microsphere Syntactic Systems, and to ensure stable buoyancy, the biologger buoyancy and residual buoyancy were calculated and designed based on 3D modeling data so that the biologger floated to the sea surface after detachment and the antenna was exposed to the ground to enable satellite communication.

The battery was selected as a Li/MnO2 (Lithium Manganese Dioxide) cylindrical battery (CR123A Lithium Battery), and was designed through a low-power operation algorithm[6]. Specifically, the battery voltage sensor utilizes the characteristics of the battery's discharge curve to predict the remaining lifespan, which enables data transmission after automatic detachment of the biologger in low-power situations.

The detachment mechanism for emergence from the detachment unit after a set period of time, or during a battery low voltage event, utilizes anodic oxidation caused by passing current through the cathode plate and anode wire as shown in Fig. 4.



Fig. 4. Oxidation of detachable parts

The joiner, a key component designed for the effective connection of the disposable and reusable parts, allows for easy separation when necessary. Fig. 5 shows the configuration of the coupling part. The coupling part is made of semi-permanent titanium metal to minimizes external exposure, and its waterproof function is designed through an O-Ring structure. Additionally, a design method that reduces the internal space of the coupling part to minimize internal pressure was applied.

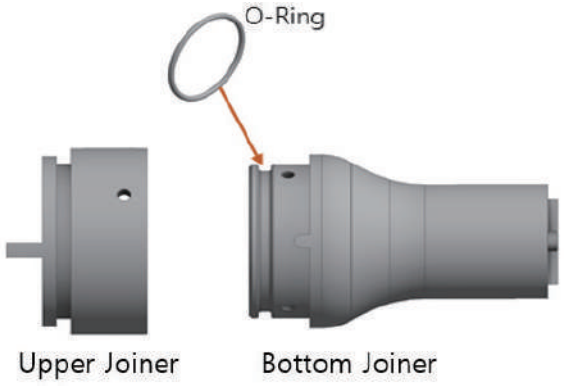


Fig. 5. Joiner Diagram

III. Experiment Result

The biologger proposed in this paper has been validated for performance through tests guaranteeing durability to withstand pressures of 100 bars at depths exceeding 1,000 meters, an operational temperature range of -20 to 50 degrees Celsius, resistance to deformation and corrosion by seawater, and an operational time of over six months. Additionally, stability was confirmed by conducting wave tank experiments to simulate real sea area conditions and check for damage to the biologger.

Experiments in real sea areas were carried out to verify the location tracking capability of the developed biologger. As shown in Fig. 6, the biologger was attached to a ship and traveled back and forth along the route from Geoje Island to Pohang to Mukho Port in the East Sea. It was confirmed that

the proposed biologger could collect various data through the sensor part of the system during operation and then transmit the collected data to the ARGOS satellite from the sea surface.

As shown in Fig. 7, the transmitted irradiance/depth data and the sunset and sunrise data were verified. In addition, it was possible to confirm that the result of estimating the location from the transmitted data was consistent with the ship's travel route as shown in Fig. 8



Fig. 6. Real-sea experiments



Fig. 7. Extract sunset and sunrise data



Fig. 8. Location estimation based on incoming data

IV. Conclusion

In this study, we developed a recyclable fish biologger that can be separated into a reusable part and a disposable part. The proposed biologger is divided into two parts: a reusable part including antenna, system part, and buoyancy material, and a disposable part consisting of battery and detachable part. The biologger was tested to meet the requirements of durability, operating temperature range, resistance to deformation and corrosion by seawater, and operating time. In addition, the performance in the actual marine environment was verified through tidal tank experiments, and the data collection and satellite communication functions were verified through actual marine tests.

Acknowledgment

This research was supported in part by the project titled 'Development of Polar Region Communication Technology and Equipment for Internet of Extreme Things (IoET)', funded by Ministry of Science and ICT (MSIT), Korea

References

[1] H. Chung, J. Lee and W. Lee, “A Review: Marine Bio-logging of Animal Behaviour and Ocean Environments” Ocean Science Journal (2021) 56:117–131
DOI: 10.1007/s12601-021-00015-1

[2] S. Koo, S. Baek, H. Kang and A. Park, “Hybrid Battery System based Fish Biologger” Journal of the Korea Academia-Industrial cooperation Society Vol. 22, No. 11 pp. 870-876, 2021
DOI: 10.5762/KAIS.2021.22.11.870

[3] CollecteLocalisation satellite. (2020), ARTIC R2 User Datasheet v1.1. https://cdn.sparkfun.com/assets/c/0/8/d/4/ENA303_ARTIC_R2_User_Datasheet_1v10.pdf

[4] STMicroelectronics. (2020), STM32L451CEU6 Datasheet. https://www.mouser.kr/datasheet/2/389/stm32l451cc-1851057.pdf

[5] Microsphere Syntactic Materials (20..), BZ Grade Datasheet. https://esyntactic.com/wp-content/uploads/2021/06/BZ-Grade-Data-Sheet-2021.pdf

[6] Panasonic. (2022), CR123A Datasheet. https://api.pim.na.industrial.panasonic.com/file_stream/main/fileversion/2500

A Sonar Micro-Doppler Analysis Study for the Estimation of Diver Motion Parameters.

Min Kim
Maritime ICT & Mobility Research
KIOST
Busan, South Korea
skymin@kiost.ac.kr

Doyoung Kim
Maritime ICT & Mobility Research
KIOST
Busan, South Korea
dykim@kiost.ac.kr

Jihyeong Lee
Maritime ICT & Mobility Research
KIOST
Busan, South Korea
wlgud0818@kiost.ac.kr

Jaeho Choi
Maritime ICT & Mobility Research
KIOST
Busan, South Korea
jaeho@kiost.ac.kr

Seungjae Baek
Maritime ICT & Mobility Research
KIOST
Busan, South Korea
baeksj@kiost.ac.kr

***Abstract*—In radar technology, micro-Doppler (m-D) signature have enabled the distinction between different types of targets, and the identification of indoor human behaviors and estimation of locomotion parameters. Despite significant advances in radar-based m-D signature applications, the potential of m-D signature in sonar signal processing, especially for underwater micro-motion recognition such as submarine propulsion systems or diver stroke styles, remains underexplored. To address this gap, we propose a new scheme for analyzing underwater diver motion parameters using sonar m-D signatures. In experiments performed using a constructed Doppler sonar system, we observed that the proposed scheme can perform effective motion parameter estimation of a diver in indoor water tank.**

***Index Terms*—Doppler sonar system, micro-Doppler signature, Sonar signal processing**

I. Introduction

The Doppler effect states that the carrier frequency of the wave, denoted as f_c, is shifted due to the movement of the target relative to the wave source, denoted as v_{los} [1]. The amount of frequency shift is referred to as the Doppler frequency (f_d), given by $f_d = \frac{2v_{los}}{\lambda_c}$, where λ_c represents the wavelength of the wave. When the target involves time-varying motion, denoted as $v_{los}(t)$, such as rotation, or vibration, the Doppler frequency is also a function of time, $f_d(t)$. This micro-Doppler (m-D) frequency is used as a signature for analyzing the motion of the target or classifying targets with micro-motion [1].

Various methods based on radar m-D signatures have been devised [2]–[6]. For instance, in [2], m-D signatures of the ballistic missile were analyzed to distinguish between ballistic missiles and decoys. Another approach [3] proposed a classification scheme for unmanned aerial vehicles, leveraging m-D signatures due to different flight systems, such as fixed wing or rotary wing, and variation in blade numbers. In addition, research employing IoT compact radar sensors has explored diverse applications [4]. Authors of [5] demonstrated the identification of indoor human behaviors, including sitting down, falling, and detecting disease-induced behaviors, using radar m-D signatures. In [6], m-D signatures were employed to estimate human gait parameters for biomechanical analysis. Consequently, radar m-D studies have been actively pursued for a myriad of purposes.

While sonar signal processing has focused on producing high-resolution 2-D and 3-D images for target detection and classification, leveraging the shorter wavelength of underwater sound waves compared to electromagnetic waves in air, this emphasis has somewhat overshadowed the potential of m-D signatures. These signatures are essential for identifying micromotion in targets such as submarine, or vessel propulsion systems, and divers, yet they have been relatively underexplored. Recent studies have begun to tap into this potential, using sonar m-D signatures for applications. In [7], sonar m-D signatures were used to estimate the revolution of the propulsion system. Another research [8] classified stroke styles of swimmer by using sonar m-D signatures and deep learning algorithm. Despite these promising developments, there remains a significant disparity in the sophistication of signal processing techniques between Doppler sonar and radar fields, highlighting a critical area for further research and development.

Motivated by the aforementioned issues, this study proposes a new approach for the analysis of underwater diver motion parameters using sonar m-D signatures. To achieve this, we commence by constructing a Doppler sonar system dedicated to capturing raw signals while monitoring divers within predefined scenarios. Subsequently, the m-D signatures derived from the divers are analyzed via a cadence-velocity diagram (CVD) alongside a phase extraction technique. This process enables the estimation of diver motion parameters, including stroke rate, length of the diver's lower body, and dynamic speed fluctuations. By employing sophisticated signal processing techniques for m-D analysis on the dataset acquired from the Doppler sonar system, the proposed approach lays

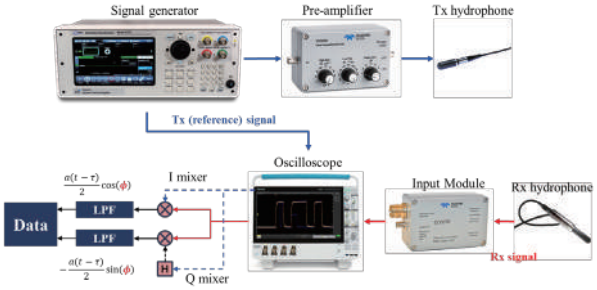


Fig. 1. Doppler sonar system.

the groundwork for using micro-Doppler based methods in addressing a diverse range of challenges. These efforts not only contribute to the advancement of sonar signal processing but also enrich the toolkit available for underwater target classification and behavior recognition.

II. Proposed Method

A. Doppler Sonar System

Traditional sonar systems employ pulse waveforms with the low pulse repetition frequency (PRF) to generate range and angular profiles. It is important to note that the lower the PRF, the greater unambiguous range. However, the dynamic range of the Doppler frequency domain is constrained by the PRF, necessitating a sufficiently high PRF in Doppler sonar systems to prevent the aliasing problem. Therefore, to analyze m-D signatures from underwater targets without encountering the aliasing problem, we construct a continuous-wave (CW) Doppler sonar system, as shown in Fig. 1. This system is consist of a signal generator, an oscilloscope, a Tx pre-amplifier, an Rx input-module, and two hydrophones, with f_c is 50 kHz.

B. Signal Model

We assume that a diver away from the Doppler sonar, as shown in Fig. 2. The raw signal after baseband conversion returned by the diver can be represented as follows:

$$s_b(t) = I_b(t) - Q_b(t) = \sigma_{tot} \exp j\theta(t), \tag{1}$$

where

$$\theta(t) = \frac{4\pi}{\lambda_c}\{R_0 + Rt + w_d(t) * \hat{r}_{tot}(t)\}, \tag{2}$$

$$w_d(t) = \sum_{k=0} \delta(t - \frac{k}{f_s}), \tag{3}$$

where $I_b(t)$ and $Q_b(t)$ are the real and imaginary parts of $s_b(t)$, respectively; σ_{tot} is the amplitude of the received signal; R_0 is the distance from the sonar to the initial location of the diver; $R(t) = \int v(t)dt$; $v(t)$ is the time-varying velocity; $\hat{r}_{tot}$ is the micro motion of the diver in a period $(0 \leq t \leq T_s = 1/f_s)$; and f_s is a stroke rate, measured in strokes per a second.

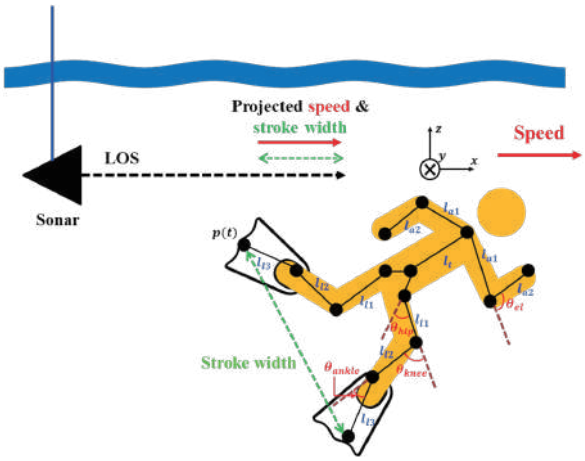


Fig. 2. Analysis of the geometry for an underwater diver based on a sonar sensor.

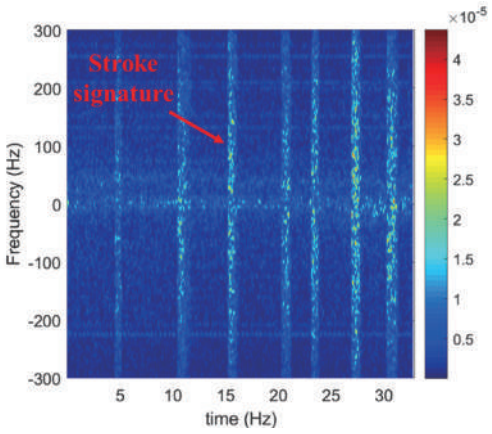


Fig. 3. The Spectrogram of $S_B(f,t)$.

C. Diver Parameters Estimation

To estimate f_s, short-time Fourier transform (STFT) is applied to $s_b(t)$ in 1 as follows:

$$S_B(f,t) = STFT[s_b(t)], \tag{4}$$

where $STFT[\bullet]$ is the STFT operator; The spectrogram of $S_B(f,t)$ is depicted as in Fig. 3. The analysis of m-D signatures is essentially the analysis of periodic signatures associated with Doppler frequencies of the target. Consequently, the CVD of $S_B(f,t)$ is generated via FFT as:

$$CVD_{SB}(f,k) = FFT[S_B(f,t)]_t, \tag{5}$$

where $FFT[\bullet]$ is the FFT operator. Because the energy of f_s is predominantly concentrated along the k-domain, f_s can be calculated on spectrum (Fig. 4), which is obtained by projecting $CVD_{SB}(f,k)$ on the k-domain, as follows:

$$f_s = \max[g(k)], \tag{6}$$

where

$$g(k) = \frac{1}{F_s}\int_{-F_s/2}^{F_s/2} CVD_{SB}(f,k)df, \tag{7}$$

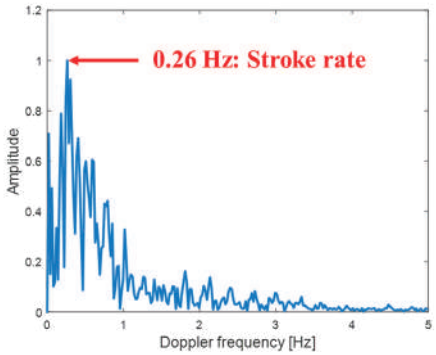


Fig. 4. The projected spectrum, $g(k)$ in 7

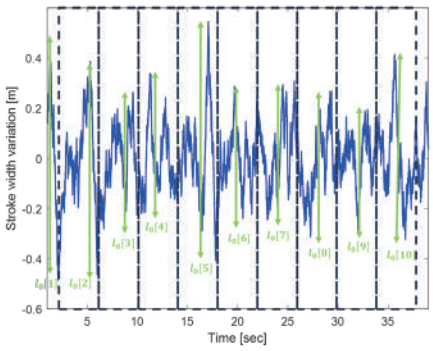


Fig. 5. The distance change per each stroke cycle in (12)

where F_s is the sampling frequency.

It is difficult to estimate speed of a diver solely using spectrogram only because the energy of the reflected sonar signal is predominantly from the diving fins. On the other hand, the phase component in 2 is significantly influenced by diver's travel distance, $R(t)$. Therefore, an intuitive method to calculate the diver's time-varying speed involves deriving it from $R(t)$. Accordingly, we begin by extracting the phase term as follows:

$$\theta(t) = \arctan \frac{I_b(t)}{Q_b(t)}, \tag{8}$$

$\theta(t)$ can be considered as a linear combination of $R(t)$ and the diver's micro motion, $w_d(t) * \hat{r}_{tot}(t)$. In order to decompose these components, we apply a moving average (MA) filter as:

$$\hat{R}(t) = R_0 + R(t) \approx \frac{\lambda_c}{4\pi}\{\theta(t) * h_{MA}(t)\}, \tag{9}$$

$$\hat{r}_{micro}(t) = w_d(t) * \hat{r}_{tot}(t) \approx \frac{\lambda_c}{4\pi}\{\theta(t) - \theta(t) * h_{MA}(t)\}, \tag{10}$$

where $h_{MA}(t)$ denotes the MA filter. The time-varying speed along the sonar line of sight (SLOS) (Fig. 2) is estimated by differentiating of $\hat{R}(t)$ as:

$$v(t) = \frac{\partial \hat{R}(t)}{\partial t}, \tag{11}$$

Subsequently, the stroke width along the SLOS, affected by the diver's micro-motion, can be inferred. The maximum distance change per each stroke cycle (see Fig. 7) is calculated as:

$$l_0[n] = \max[r_n(t)] - \min[r_n(t)], \quad n = 0, 1, 2, ..., N-1, \tag{12}$$

where

$$r_n(t) = rect\left(\frac{t - (n + 0.5)T_s}{T_s}\right) \times \hat{r}_{micro}(t), \tag{13}$$

In order to determine the length of the lower body, a model of the lower body of a diver is defined as shown in Fig. 2. The position of the end of the fin, denoted as $p(t)$, is determined by a factor including the degrees of flexing of the hip, knee, and ankle($\theta_{hip}(t)$,$\theta_{knee}(t)$ and $\theta_{ankle}(t)$, respectively), and the lengths of the thigh, shin, and fin (l_{l1},l_{l2} and l_{l3}, respectively). The time-varying locations of the end of fin along SLOS can be expressed as follows:

$$p_{LOS}(t) = p(t) \cdot \bar{l}_{los} = \|\bar{i}_x\| \cdot p_x(t) + \|\bar{i}_z\| \cdot p_z(t), \tag{14}$$

where

$$p(t) = [l_{l1}\alpha_1(t) + l_{l2}\alpha_2(t) + l_{l3}\alpha_3(t), \quad y_0 \\ ,l_{l1}\beta_1(t) + l_{l2}\beta_2(t) - l_{l3}\beta_3(t)], \tag{15}$$

$$\bar{l}_{los} = [\bar{i}_x, \quad 0, \quad \bar{i}_z], \tag{16}$$

where $\alpha_1(t) = \cos\theta_{hip}(t)$; $\alpha_2(t) = \cos\theta_{knee}(t) - \theta_{hip}(t)$; $\alpha_3(t) = \cos\theta_{ankle}(t) - \theta_{knee}(t) - \theta_{hip}(t)$; $\beta_1(t) = \sin\theta_{hip}(t)$; $\beta_2(t) = \sin\theta_{knee}(t) + \theta_{hip}(t)$; and $\beta_3(t) = \sin\theta_{ankle}(t) - \theta_{knee}(t) - \theta_{hip}(t)$.

$l_0[n]$ is determined from the trajectories of $p_{LOS}(t)$. Therefore, the relationship between $l_0[n]$ and $p_{LOS}(t)$ can be expressed as follows:

$$\begin{aligned} \bar{l}_0 &= \|\max[p_{LOS}(t)] - \min[p_{LOS}(t)]\| \\ &= \|\bar{i}_x\|(l_{l1} \cdot \gamma_1 + l_{l2} \cdot \gamma_2 + l_{l3} \cdot \gamma_3) \\ &\quad + \|\bar{i}_z\|(l_{l1} \cdot \gamma_4 + l_{l2} \cdot \gamma_5 + l_{l3} \cdot \gamma_6), \end{aligned} \tag{17}$$

where $\bar{l}_0 = \frac{1}{N}\sum_{n=0}^{N-1} l_0[n]$; $\gamma_m = \alpha_m(t_{max}) - \alpha_m(t_{min})$; $\gamma_m = \beta_m(t_{max}) - \beta_m(t_{min})$; and $m = 1, 2, 3$.

To address the underdetermined system in (17), we assume that $l_{l1} = l_{l2}$, $l_{l3} = 0.65$, and that $\alpha_m(t)$ and $\beta_m(t)$ adhere to the flexing model defined in [9]. Based on these assumptions, we can estimate the length of the lower body as follows:

$$\begin{aligned} \bar{l}_{low} &= l_{l1} + l_{l2} \approx 2l_{l1} \\ &= 2\frac{\bar{l}_0 - l_{l3}(\|\bar{i}_x\|\gamma_3 - \|\bar{i}_z\|\gamma_6)}{\|\bar{i}_x\|\gamma_1 + \|\bar{i}_x\|\gamma_2 + \|\bar{i}_z\|\gamma_4 + \|\bar{i}_z\|\gamma_5}. \end{aligned} \tag{18}$$

III. Experimental Results

To assess the performance of the proposed method in analyzing diver motion parameters, we utilized a 5 kHz Doppler sonar system, as shown in Fig. 1. The experimental setup involved an indoor 3D water tank where the diver executed a

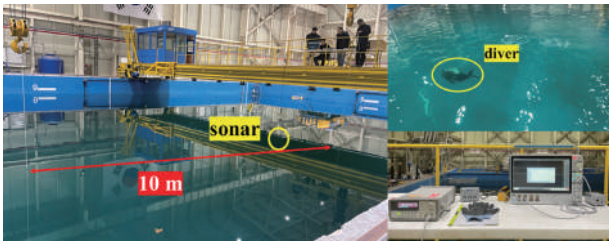
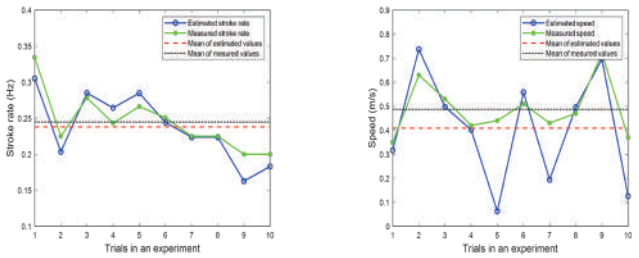


Fig. 6. Experimental setup: Indoor water tank environments for a diver's locomotion experiment; and Measurement system.



(a) (b)

Fig. 7. The experiment results. (a) The stroke rate estimation, and (b) projected speed estimation.

straight-line movement towards a point 10 meters away from the hydrophones as depicted in Fig. 6. To verify the accuracy of the proposed method, we employed a video sensor capable of determining the diver's speed and stroke rate via image analysis. We conducted 10 repeated trials to estimate three key parameters: 1) stroke rate, as detailed in (6); 2) speed along the SLOS as in (11); and 3) the lower body length, specified in (18), (see Fig. 7).

The root-mean-square error (RMSE) is a widely recognized metric for quantitatively assessing the accuracy of the estimation methods. Thus, we calculated the RMSE to compare the diver motion parameters determined by the vision sensor with those estimated by the proposed method using the Doppler sonar system. Specifically, the RMSE of the stroke rate is defined as:

$$RMSE = \sqrt{\frac{1}{L}\sum_{i=1}^{L}(\bar{f}_s(i) - \hat{f}_s(i))^2}, \qquad (19)$$

where $\bar{f}_s(i)$ and $\hat{f}_s(i)$ represent the i^{th} measured and estimated stroke rates, respectively, and L is the number of repetitive experiments. Furthermore, the RMSEs for the speed and the lower body length can also be calculated using (19).

The RMSE values for the experimental results are summarized in Tables I. The stroke rate, speed, and step length were estimated accurately enough to compare with the values measured by the vision sensor. However, when estimating the length of the lower body, the RMSE values and standard deviations were relatively high, resulting in inaccurate estimations. However, upon repeating the experiment, the mean values converged closer to the actual values.

TABLE I
EXPERIMENTS RESULTS FOR THE SECOND VOLUNTEER

	Estimated by the proposed method		Measured by a vision sensor		RMSE
	mean	std	mean	std	-
Stroke rate (Hz)	0.2381	0.047	0.2448	0.0407	0.0197
Speed (m/s)	0.4086	0.2315	0.487	0.1154	0.1658
The lower body length (m)	1.0668	0.4066	1.1	-	0.3872

IV. CONCLUSION

In this study, we proposed a new scheme to estimate diver motion parameters using Doppler sonar system. Contributions of this study are as follows: 1) construction of Doppler sonar system for utilizing underwater micro-Doppler signature; 2) analysis of underwater target motion via the application of signal processing techniques; 3) estimation of various diver motion parameters. Consequently, we achieved efficient and accurate estimations, not only for the stroke rate, but also for speed and lower body length, compared to conventional approaches.

ACKNOWLEDGMENT

This research was supported in part of the project titled 'Development of Polar Region Communication Technology and Equipment for Internet of Things (IoET)', funded by Ministry of Science and ICT (MSIT)

REFERENCES

[1] V. Chen, The Micro-Doppler Effect in Radar 2nd Edition. Artech House, 2019.
[2] I. Choi, M. Kim, J. Choi, J. Park, S. Park and K. Kim, "Robust Cardiac Rate Estimation of an Individual," in IEEE Sensors Journal, vol. 21, no. 13, pp. 15053-15064, 1 July1, 2021.
[3] I. -O. Choi, S. -H. Park, M. Kim, K. -B. Kang and K. -T. Kim, "Efficient Discrimination of Ballistic Targets With Micromotions," in IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 2, pp. 1243-1261, April 2020.
[4] K.-B. Kang, J.-H. Choi, B.-L. Cho, J.-S. Lee, and K.-T. Kim, "Analysis of micro-Doppler signatures of small UAVs based on Doppler spectrum," IEEE Trans. Aerosp. Electron. Syst., vol. 57, no. 5, pp. 3252–3267, 2021.
[5] Bo Wang, Hao Zhang, and Yong-Xin Guo, "Radar-Based Soft Fall Detection Using Pattern Contour Vector",IEEE Internet of Tings Journal, Vol. 10, No. 3, pp.2519–2527, 2023.
[6] A.-K. Seifert,M.G.Amin, andA.M. Zoubir, "Towardunobtrusive in-homegait analysisbasedonradarmicro-doppler signatures," IEEE Trans.Biomed.Eng.,vol.66,no.9,pp.2629–2640,2019.
[7] Y. Wu, M. Luo and S. Li, "Measurement and Extraction of Micro-Doppler Feature of Underwater Rotating Target Echo," OCEANS 2022 - Chennai, Chennai, India, 2022.
[8] H. Kulhandjian, N. Ramachandran, M. Kulhandjian and C. D'Amours, "Human Activity Classiffication in Underwater using sonar and Deep Learning", Proc. of ACM Intl. Conf. on UnderWater Networks and Systems (WUWNet), October 2019.
[9] Wojtków, Magdalena, and Anna Nikodem. "Biomechanics of diving: the influence of the swimming speed on the kinematics of lower limbs of professional divers." Acta of bioengineering and biomechanics vol. 19, no. 4, pp. 117-125, 2017.

Relative Precise Positioning Method for Two Moving Vehicles and UGV-UAV System based on Moving Baseline RTK

Gyeongmin Kim
Department of Aerospace Engineering and Convergence Engineering for Intelligent Drone
Sejong University
Seoul, Republic of Korea
kim.gm2921@gmail.com

Hyojung Yoon
Korea Institute of Aviation Safety Technology
Incheon, Republic of Korea
hyojjvv@gmail.com

Eunyeong Lee
LG Uplus
Seoul, Republic of Korea
leeey@lguplus.co.kr

Wonwoo Park
Department of Aerospace Engineering and Convergence Engineering for Intelligent Drone
Sejong University
Seoul, Republic of Korea
wallhouse@sju.ac.kr

Byungwoon Park
Department of Aerospace Engineering and Convergence Engineering for Intelligent Drone
Sejong University
Seoul, Republic of Korea
byungwoon@sejong.ac.kr

***Abstract*— In this paper, we propose a relative precise positioning solution called MB-RTK (Moving Baseline RTK), utilizing an RTK (Real-Time Kinematics) solution and the velocity of a reference station. Traditional RTK methods do not account for the movement of the reference station, leading to errors when the reference station is in motion. The proposed MB-RTK method compensates for these movements by incorporating the velocity of the reference station and eliminates errors caused by the movement of the reference station, thereby eliminating errors caused by its movement. Unlike other MB-RTK research, this method provides a position solution synchronized to the current time while requiring only a small amount of data. Additionally, we develop an MB-RTK based UGV (Unmanned Ground Vehicle)-UAV (Unmanned Aerial Vehicle) system. This system estimates the relative position between two moving vehicles using the baseline vector estimated by MB-RTK, allowing the UAV to control its position relative to the moving UGV. Consequently, the UAV can follow the UGV without additional control commands. The effectiveness of the control system was validated through flight experiments.**

Keywords—GNSS, Relative Precise Positioning, RTK, Moving Baseline RTK, UAV

I. Introduction

Following the Fourth Industrial Revolution, the unmanned vehicle market is expanding, and related research is progressing actively. In particular, businesses using drones, a representative UAV (Unmanned Aerial Vehicle), are receiving significant attention. However, drones have inherent limitations such as short flight time and limited operating range due to battery constraints. To overcome these limitations, the combination of UAVs with UGVs (Unmanned Ground Vehicles) has been proposed. Actual implement of this concept requires a process to estimate the relative positions of the two moving vehicle as accurately as possible.

The technique currently used to calculate the relative position between two or more moving vehicles is RTK (Real-Time Kinematics) technology [1], which calculates the precise absolute position of each moving vehicle centered on a reference station on the ground with a known exact position[1]. Based on this, the control system calculates the relative position and uses it for control and mission performance. However, this method can only be used within 10 km of a fixed reference station [2], [3], and it requires a process of converting from an absolute coordinate system to a relative coordinate system [1]. Additionally, in the case of a UGV-UAV combined system, there is the disadvantage that the location of the UGV must be continuously transmitted to the UAV as the UGV moves, to calculate the relative position.

In this paper, we propose a method of estimating the precise relative position between two moving vehicles by having one moving vehicle serve as a reference station. In this process, we use the RTK solution provided by commercial GNSS receivers, making it possible to apply the proposed method even without the need for expensive equipment. Based on this approach, we also developed a UGV-UAV system and verified its performance.

II. Methodology

A. Concept of Moving Baseline RTK

RTK is a precise relative navigation technology that estimates a cm-level accurate baseline vector by double differencing carrier phase measurements between a known reference station and a user to eliminate common errors. Since the measurements of the reference station must be delivered to the user, it uses a correction message that follows the international standard, RTCM (Radio Technical Commission for Maritime Services). The time it takes for the message to be transmitted is called the age of correction, dt [4]. Because RTK technology assumes that the reference station is stationary, most commercial GNSS receivers only consider user movements during dt and provide a position synchronized to the current time. This RTK solution is known as synchronous RTK [5], and we refer it to as conventional RTK (C-RTK) in this paper.

In contrast, MB-RTK (Moving Baseline RTK) technology provides correction messages in situations where the reference station is not fixed but moving. Therefore, it must reflect not only the user's movement during dt but also the movement of the reference station. In this paper, we propose MB-RTK technology that reflects the movement of the reference station

using C- RTK solution and the velocity of the reference station [6].

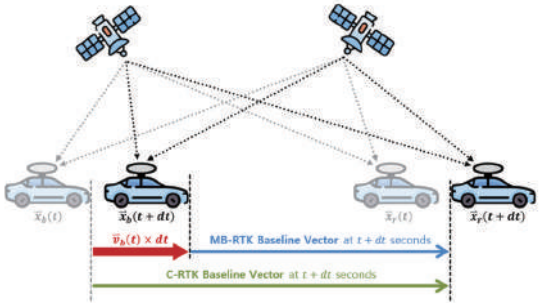


Fig. 1. Concept of Moving Baseline RTK

In Fig. 1, the green arrow represents the baseline vector when the reference station is fixed, which corresponds to the C-RTK baseline vector provided by commercial GNSS receivers. MB-RTK is designed to be utilized even when the reference station is moving. The actual baseline vector at the current time, $t + dt$, is represented by the blue arrow. In other words, the baseline vector that MB-RTK should estimate is the baseline vector computed by C-RTK minus the movement of the reference station, as shown in (1). The movement of the reference station can be estimated as the product of its velocity, $\vec{v}_b(t)$ and time, dt.

$$\vec{b}(t+dt)_{MB-RTK} = \vec{b}(t+dt)_{C-RTK} - \vec{v}_b(t) \times dt \quad (1)$$

B. Time-Differenced Carrier Phase Method

To implement MB-RTK through velocity compensation, a method to estimate the accurate velocity of the reference station is required. In this paper, the TDCP (Time-Differenced Carrier Phase) method was applied for this purpose.

TDCP is a technique that estimates the user's velocity with the cm/s level through time differences between carrier phase measurements [7]-[12]. After time differencing, the ambiguity term in carrier phase measurements is removed when there are no cycle slips [13] [14], and the final equation to estimate the change in the user position $\Delta\vec{R}_u$ is as follows:

$$\begin{bmatrix} -\hat{e}^1(t) & 1 \\ -\hat{e}^2(t) & 1 \\ \vdots & \\ -\hat{e}^n(t) & 1 \end{bmatrix} \begin{bmatrix} \Delta\vec{R}_u \\ \Delta B \end{bmatrix} = \begin{bmatrix} \lambda\Delta\tilde{\Phi}^1 - SVDoppler^1 + dLOS^1 \\ \lambda\Delta\tilde{\Phi}^2 - SVDoppler^2 + dLOS^2 \\ \vdots \\ \lambda\Delta\tilde{\Phi}^n - SVDoppler^n + dLOS^n \end{bmatrix} \quad (2)$$

where $\lambda\Delta\tilde{\Phi}$ is the compensated TDCP measurement, which accounts for known errors such as satellite clock error, ionospheric and tropospheric errors. The term $SVDoppler$ represents a change in the range and is proportional to the average Doppler frequency shift caused by satellite-receiver relative motion along the line-of-sight. The term $dLOS$ accounts for changes in the relative satellite-receiver geometry due to the changing orientation of the line-of-sight vector [7]. If the number of satellite n is more than 4, equation (2) can be solved for the $\Delta\vec{R}_u$ using the least squares method.

The TDCP method has the advantage of using only GNSS measurements without requiring additional sensors such as an IMU.

C. Correction Message for MB-RTK

To deliver the velocity of the reference station computed by TDCP to the user, we customized a correction message based on the structure of RTCM MT1006, which provides location information of the reference station. We named it MT4006, using one of the unassigned message types in RTCM. This name was tentatively chosen to demonstrate the feasibility of enabling the proposed MB-RTK algorithm with minimal modification to existing systems. In this paper, the correction messages, including MT4006, are referred to as Extended RTCM [15].

The primary feature of the proposed MB-RTK system in this paper is that it can obtain the relative position solutions synchronized with the current time. In other MB-RTK researches [16]-[18], double-difference measurements are used only for baseline vector estimation, resulting in position estimates synchronized to the time the correction message was generated, rather than the current time. However, the method proposed in this paper reflects the movement of the reference station by compensating for the result of C-RTK, which is already synchronized with the current time, thereby providing the position solution synchronized with the current time. Additionally, the proposed method has the advantage of reducing the amount of data required to implement MB-RTK because it requires additional velocity data rather than full measurements from the reference station. This is especially suitable for control systems that require high-rate position solutions.

III. Algorithm Validation

To verify the performance of the proposed MB-RTK method and the designed correction message, experiments were conducted in situations where the length of the baseline vector was fixed. Additionally, to assess the feasibility of the UAV-UGV application, we also conducted dynamic tests in scenarios where the baseline between the two vehicles was variable.

A. Fixed Baseline Dynamic Test

To evaluate the proposed system performance, a dynamic test with a fixed baseline length was conducted, where two antennas were installed on one car, with each antenna serving as a reference station and a user, as shown in Fig. 2. While the vehicle is moving, the user's direction changed relative to the reference point on the same car, keeping the length of the baseline vector constant, but causing the direction of the baseline vector to constantly change.

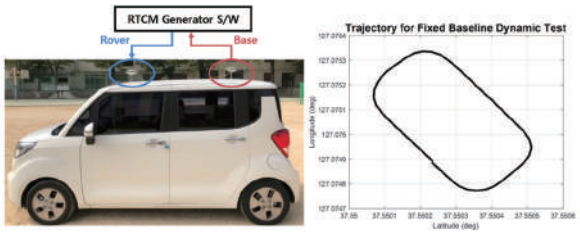


Fig. 2. Experiment Configuration for Fixed Baseline Dynamic Test

As shown in Fig. 3 and Table. 1, the MB-RTK baseline vector, which considered the movement of the reference station by velocity compensation, had an RMS horizontal error of 0.053m. In contrast, the C-RTK baseline vector, which did not consider the movement of the reference station, had an RMS horizontal error of 7.4m. This indicates that the proposed MB-RTK method, using the velocity of the reference station, can accurately estimate the baseline vector by reflecting the movement of the reference station.

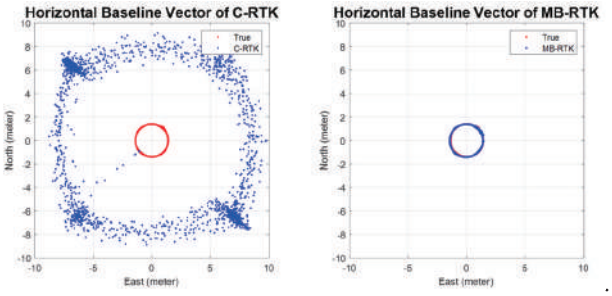


Fig. 3. Horizontal Baseline Vector for Fixed Baseline Dynamic Test

TABLE I. ERROR OF THE BASELINE VECTOR FOR FIXED BASELINE DYNAMIC TEST

	C-RTK (w/o Velocity Compensation)		MB-RTK (w/ Velocity Compensation)	
	Horizontal	***Vertical***	***Horizontal***	***Vertical***
RMS (m)	7.3993	0.0434	0.0530	0.1292
MEAN (m)	7.2775	0.0096	0.0487	0.1214

B. Two Vehicle Dynamic Test

To evaluate the performance with two moving vehicles whose baselines are variable, we conducted a flight experiment with two drones, each equipped with an antenna, and calculated the baseline vector. The experiment was divided into the two cases: Case #1 where the leader drone serving as the reference station hovered and Case #2 where the leader drone moved.

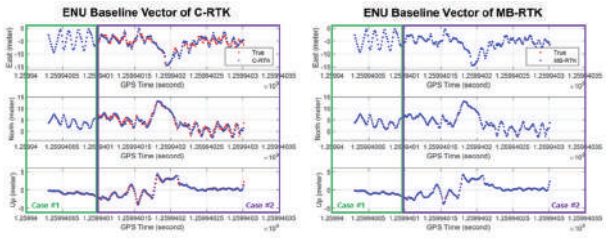


Fig. 4. ENU Baseline Vector for Two Vehicle Dynamic Test

Fig. 4 compares the ENU baseline vector results of C-RTK without velocity compensation and MB-RTK with velocity compensation. Both graphs show almost the same performance in Case #1, where the leader drone did not move vigorously. However, in Case #2, after the leader drone started moving, C-RTK baseline vector had a large error, while the MB-RTK baseline vector was almost the same as the true baseline vector. This demonstrates that the C-RTK method does not account for the movement of the reference station, resulting in significant performance variation depending on whether the reference station moves. In contrast, the proposed MB-RTK method shows consistent performance regardless of the reference station's movement.

Table 2 shows the real-time error of the MB-RTK baseline vector by applying the designed correction message to the same flight experiment data. The RMS horizontal error of 16.71cm achieved through the transmission of the customized Extended RTCM message in an outdoor communication environment did not significantly degrade compared to the 13.28 cm error observed in indoor repeat tests using the GNSS data reradiated by LabSat in a wide communication bandwidth environment. These results confirms that the baseline vector between two dynamic vehicles can be calculated with cm or dm-level accuracy in real-time.

TABLE II. REAL-TIME ERROR OF THE BASELINE FOR TWO VEHICLE DYNAMIC TEST BY APPLYING THE EXTENDED RTCM

	C-RTK (w/o Velocity Compensation)		MB-RTK (w/ Velocity Compensation)	
	Horizontal	***Vertical***	***Horizontal***	***Vertical***
RMS (m)	0.7345	0.1842	0.1671	0.0581
MEAN (m)	0.0029	0.0070	0.0036	-0.0061

IV. MB-RTK BASED UGV-UAV SYSTEM AND TEST RESULT

We developed a UGV-UAV system that performs control processes based on the relative position estimated by MB-RTK. Fig. 5 shows the structure of the developed system, which consists of a UGV serving as a reference station and a UAV as a user. In this system, the UGV is designated as the 'Leader', and the UAV as the 'Follower'.

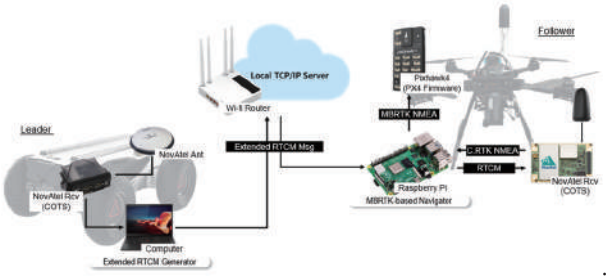


Fig. 5. Hardware Configuration of MB-RTK based UGV-UAV System [19]

The Leader estimates its own velocity using TDCP, creates the Extended RTCM including the MT4006 message, and delivers it to the Follower. The Follower, which receives correction messages through the TCP/IP server, calculates the MB-RTK solution through velocity compensation and provides the solution to its flight control system. At this time, the Follower's control mode is set to Hovering, which maintains the current state such as a position.

When the Leader moves and the relative position between the two moving vehicles changes, the Follower's flight control system operates as if it operates in RTK mode. The Follower drone interprets that its position change as being due to an external force rather than the Leader moving and performs Hovering control. In other words, the Follower moves in response to the Leader's movement without requiring additional control commands such as updating a new setpoint.

As a result of the flight experiment with the implemented UGV-UAV system, it was shown that the Follower UAV moves along the Leader UGV without any control commands, as illustrated in Fig. 6.

V. CONCLUSION

In this paper, we introduce the Moving Baseline RTK method to estimate an accurate baseline vector based on the velocity of the reference station and the position solution of conventional RTK. This method can provide a position synchronized to the current time and has the advantage of

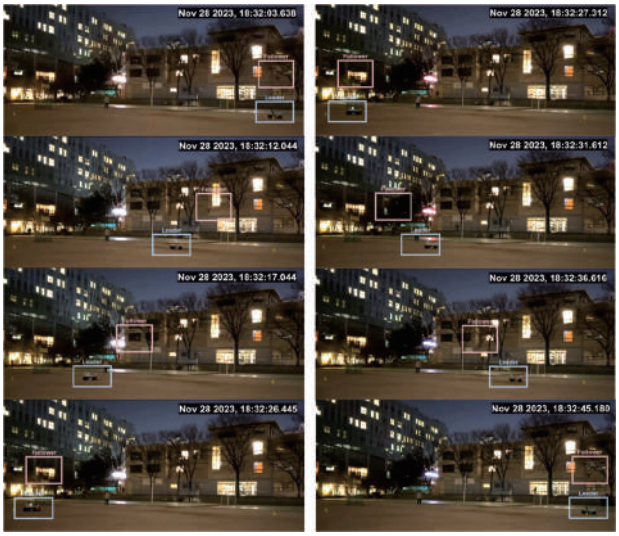


Fig. 6. Capture of Video for UGV-UAV System Fllight Test (Blue Box: Leader, Red Box: Follower)

reducing the amount of data required, as only the velocity data of the reference station is used in this process. When applying reference station velocity compensation in actual car and drone experiments, more than 95% of the errors caused by not compensating for the movement of the reference station were removed. Real-time experiments also confirmed that the error was reduced by 77.26%. In addition, we developed an MB-RTK based UGV-UAV system and established a relative position-based control system. The system was confirmed to allow the UAV to follow the movement of the UGV without additional control commands in actual flight experiments. In the future, we plan to develop an automatic landing system on a real-time moving vehicle based on the developed UGV-UAV system.

ACKNOWLEDGMENT

This research was supported by Unmanned Vehicles Core Technology Research and Development Program through the National Research Foundation of Korea (NRF) and Unmanned Vehicle Advanced Research Center (UVARC) funded by the Ministry of Science and ICT, the Republic of Korea (No. 2020M3C1C1A01086407). This work was supported by Future Space Navigation & Satellite Research Center through the National Research Foundation funded by the Ministry of Science and ICT, the Republic of Korea (2022M1A3C2074404). This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2024-RS-2024-00437494) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] S. Moon, D. Kim, and D. Lee., "Outdoor swarm flight system based on the RTK-GPS," *Journal of KIISE*, vol. 47, no. 3, pp. 328-334, Dec. 2016.

[2] C. Kee, B. Park, J. Kim, A. Cleveland, M. Parsons, and D. Wolfe, "A guideline to establish DGPS reference station requirements," *Journal of Navigation*, vol. 61, no. 1, pp. 99–114, Dec. 2007.

[3] B. Park, and C. Kee, "The compact Network RTK method: An effective solution to reduce GNSS temporal and spatial decorrelation error," Journal of Navigation, vol. 63, no. 2, pp. 343–362, Feb. 2010.

[4] HEXAGON, "*Application Note : Positioning Modes of Operation*. (APN-051 Rev 1)," 2013.

[5] B. Park, D. Yoon, J. Song, and C. Kee, "Latency compensation by compact RTK under harsh communication environment of land transportation," *Proceedings of the 27th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2014)*, pp. 933-939, Sep. 2014.

[6] H. Yoon, E. Lee, C. Lim, and B. Park, "Moving base precise relative position for drone swarm flight using conventional RTK and NMEA data," Proceedings of the 33rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2020), pp. 674-697, Sep. 2020.

[7] P. Freda, A. Angrisano, S. Gaglione, and S. Troisi, "Time-differenced carrier phases technique for precise GNSS velocity estimation," GPS Toolbox, vol. 19, pp. 335-341, Dec. 2014

[8] A. Angrisano, G. Cappello, S. Del Pizzo, and S. Gaglione, "Time-differenced carrier phase technique for precise velocity estimation on an android smartphone," Sensors, vol. 22, no. 21, pp. 8514-8532, Nov. 2022.

[9] Y. Lee, and B. Park, "Nonlinear regression-based GNSS multipath modelling in deep urban area," *Mathematics*, vol. 10, no. 3, pp. 412, Jan. 2022.

[10] Y. Lee, Y. Hwang, J. Ahn, J. Seo, and B. Park, B. "Seamless accurate positioning in deep urban area based on mode switching between DGNSS and multipath mitigation positioning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 6, pp. 856-5870, Jun. 2023.

[11] Y. Lee, P. Wang, and B. Park, "Nonlinear regression-based GNSS multipath dynamic map construction and its application in deep urban areas," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 5, pp. 5082-5093, May. 2023.

[12] D. Lee, Y. Lee, Y., and B. Park, "Carrier phase residual modeling and fault monitoring using short-baseline double difference and machine learning," *Mathematics*, vol. 11, no. 12, pp. 2696, Jun. 2023.

[13] Y. Lee, C. Lim, and B. Park, "Cycle slip detection to provide stable precision positioning under severe ionospheric conditions," *Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021)*, pp. 335-363, Sep. 2021.

[14] J. Yun, C. Lim, and B. Park, "Inherent limitations of smartphone GNSS positioning and effective methods to increase the accuracy utilizing dual-frequency measurements," *Sensors*, vol. 22, no. 24, pp. 9879, Dec. 2022.

[15] G. Kim, W. Park, C. Lim, and B. Park, "Proposal and Real-Time Application of Extended RTCM Message for Moving Baseline RTK," KSAS 2023 Spring Conference, pp. 1170-1171, Apr. 2023.

[16] H. Shim, Y. Bae, J. Hwang, C. Kee, and I. Koo, "Single Frequency RTK Relative Navigation for Autonomous Formation Flying Mission of SNUGLITE-III CubeSat," Proceedings of the 36th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2023), pp. 1474-1486, Sep. 2023.

[17] T. Suzuki, S. Kojima, K. Ohno, N. Miyamoto, T. Suzuki, K. Asano, T. Komatsu, and H. Kakizaki, "Automatic Operation of an Articulated Dump Truck: State Estimation by Combined QZSS CLAS and Moving-Base RTK Using Multiple GNSS Receivers," Proceedings of the ION 2024 Pacific PNT Meeting, pp. 65-74, Apr. 2024.

[18] X. Hou, K. Fang, Z. Wang, Q. Li, and J. Fang, "Research on Ambiguity Solution Integrity Monitoring for Moving Base RTK," Proceedings of the 2020 International Technical Meeting of The Institute of Navigation, pp. 468-486, Jan. 2020.

[19] G, Kim, W. Park, and B. Park, "Moving baseline RTK-based ground vehicle-drone combination system," Proceedings of the 2024 International Technical Meeting of The Institute of Navigation, pp. 630-636, Jan. 2024.

Performance Analysis of Multiple High-Altitude Platform Stations According to Architecture

Jaeyeol Lee*, Won-Jae Lee‡, Jongtae Lee‡ and Jae-Hyun Kim†
*Department of Space Survey Information Technology, Ajou University, Suwon, 16499, South Korea
‡Department of Artificial Intelligence Convergence Network, Ajou University, Suwon, 16499, South Korea
†Department of Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea
{jaeyel98, ljmwon, jtlee830, jkim}@ajou.ac.kr

***Abstract*—Recently, there has been a significant increase in research efforts aimed at realizing the goals of 6G technology. In this context, high altitude platform stations (HAPS) are receiving considerable attention as potential solutions for achieving objectives such as hyper-connectivity and expanded spatial coverage. HAPS operate for extended periods using solar panels at an altitude of 20 km within the stratospheric environment, thereby providing stable communication services to the ground. This capability is crucial in specific scenarios, such as hotspots or disaster situations, when a base station is present. Additionally, the flexibility of deploying ground communication services is enhanced by the choice between transparent and regenerative architectures. This paper we explain the transparent and regenerative architectures of HAPS. We also analyze the signal to interference plus noise ratio (SINR), data rate, and propagation delay according to the architecture of HAPS and the gains from repeaters.**

***Index Terms*—High altitude platform station (HAPS), transparent, regenerative, signal to interference plus noise ratio (SINR), data rate**

I. Introduction

Recently, with the advancement of communication technology, platforms that provide services in a line-of-sight (LoS) environment have been extensively studied. Among these, high altitude platform stations (HAPS) and satellites are receiving significant attention. HAPS operate in a circular path at an altitude of 20 km within the stratosphere, providing services to fixed points with minimal wind exposure. Additionally, HAPS can maintain extended flight times by equipping their wings with solar panels to charge the batteries. In contrast, satellites offer broad coverage to ground users. However, they face several disadvantages, including higher round-trip time (RTT), lower communication performance due to their greater altitude, and significant Doppler shifts due to their high velocity. Conversely, HAPS provide high-quality services to the ground due to their lower altitude and offer flexible services for specific scenarios. However, they offer narrower coverage compared to satellites. Owing to these characteristics, significant research on HAPS is being conducted. In [1], the authors study how satellites and HAPS can provide ground services using free-space optical (FSO) links. In [2], they focus on multi-mode approaches involving transparent, regenerative, and reconfigurable intelligent surface (RIS) architectures, tailored to various communication scenarios. Furthermore,

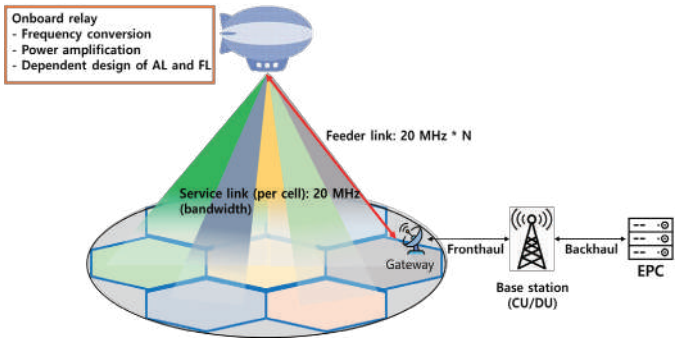


Fig. 1: HAPS transparent architecture.

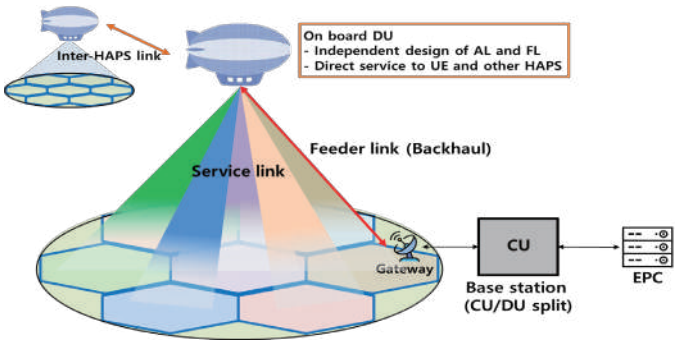


Fig. 2: HAPS regenerative architecture.

research is being conducted on resource allocation using reinforcement learning within a heterogeneous architecture that integrates and utilizes low Earth orbit (LEO) satellites, HAPS, and unmanned aerial vehicles (UAVs) [3]. HAPS can be classified into transparent and regenerative architectures, depending on the presence of an onboard base station. However, research on performance analysis according to HAPS architecture is still insufficient. In this paper we explain the two architectures of HAPS, determined by whether they are equipped with a base station. We also analyze the communication performance based on these architectures.

II. Two Architecture of HAPS

The architecture of HAPS can be classified into transparent or regenerative types, depending on whether there is a base sta-

tion onboard. The transparent architecture is equipped with a repeater that facilitates communication services on the ground by receiving signals from ground gateways connected to the ground evolved packet core (EPC), distributed unit (DU), and centralized unit (CU), as illustrated in Fig. 1. In contrast, the regenerative architecture includes an onboard base station that operates on the HAPS, as shown in Fig. 2. The base station's radio unit (RU), DU, and CU can be functionally divided and placed onboard the HAPS. Consequently, network performance is affected based on the division of functions within the base station. When incorporating more base station functions onboard, communication performance is enhanced. However, it increases power consumption and weight. In the transparent architecture, considerations must be given to both the feeder link (the link between the gateway and the HAPS) and the service link (the link between the HAPS and the user). Additionally, the bandwidth of the feeder link should be designed considering the number of cells on the HAPS. The regenerative architecture allows for the independent design of feeder and service links, considering communication performance and flight time, based on the HAPS architecture and the specific situation.

III. System Model

In the system model, we assume that the number of HAPS is denoted as $P = \{p_1, p_2...p_i\}$, each comprising cells denoted as $N = \{n_1, n_2....n_j\}$. We assume a uniform distribution of users, represented as $M = \{m_1, m_2...m_k\}$ within the system. In this study, we compare communication performance based on coverage, focusing only on the downlink. We analyze the downlink communication performance of HAPS, both with and without an on board base station.

A. HAPSMobile Antenna Pattern

In this paper, we employed the antenna patterns of HAPSMobile and the 3rd generation partnership project (3GPP). The HAPSMobile pattern is based on recommendations from the international telecommunication union radiocommunication sector (ITU-R) [4]. HAPSMobile utilizes a planar patch array antenna, and the beamwidths in the vertical (θ) and horizontal (ϕ) directions can be calculated separately. At an antenna angle of ψ, the antenna pattern is defined as follows:

$$G(\Psi) = \begin{cases} -3(\Psi/\Psi_b)^2; & 0 \le \Psi \le \Psi_1 \\ L_N; & \Psi_1 < \Psi \le \Psi_2 \\ X - Alog_{10}(\Psi); & \Psi_2 < \Psi \le \Psi_3 \\ L_F; & \Psi_3 < \Psi \end{cases}, \tag{1}$$

where ψ_b is the half-power beamwidth, L_N is the near-side lobe level, and L_F is the far-side lobe level. X is calculated as $7442/10^{0.1G_m}$, $\psi_1 = \psi_b\sqrt{-L_N/3}$, $\psi_2 = 3.745\psi_b$, and $\psi_3 = 10^{X-L_F/60}$. Using $G(\psi)$, the antenna gain G is defined as follows:

$$G = max(G_v + G_h, L_F) + G_p, \tag{2}$$

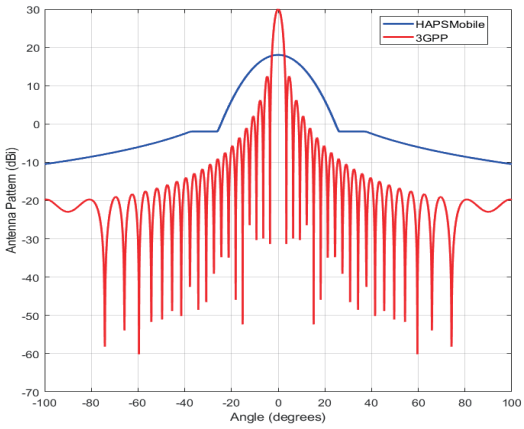


Fig. 3: HAPSMobile and 3GPP antenna pattern.

where G_v is the vertical antenna gain and G_h is the horizontal antenna gain. The maximum antenna gain G_p is defined as follows:

$$G_p = 10log_{10}(\frac{80^2}{\theta_{3dB}\phi_{3dB}}), \tag{3}$$

where θ_{3dB} and ϕ_{3dB} represent the vertical and horizontal 3 dB antenna angles, respectively.

B. 3GPP Antenna Pattern

The 3GPP antenna patterns are specified in TR 38.811 [5]. The following normalized antenna gain pattern corresponds to a typical reflector antenna with a circular aperture. The antenna pattern $G(\theta)$ is defined as follows:

$$G(\theta) = \begin{cases} 1 & \text{for } \theta = 0 \\ 4\left|\frac{J_1(ka\sin\theta)}{ka\sin\theta}\right|^2 & \text{for } 0 < |\theta| \le 90^\circ \end{cases} \tag{4}$$

where J_1 is the Bessel function, a is the radius of the circular aperture of the antenna, k is calculated as $2\pi f/c$, f is the frequency, c is the speed of light, and θ is the angle between the cell boresight and the user direction. The value of a is calculated as $10c/f$. The HAPSMobile and 3GPP antenna patterns are shown in Fig. 3.

C. Channel Model

In this paper, we consider the channel model between the HAPS and the user based on the free space path loss (FSPL). The FSPL between the k-th user and the i-th HAPS is defined as follows:

$$L_{k,i} = 20\log_{10}\left(\frac{4r_{k,i}\pi}{\gamma}\right), \tag{5}$$

where $r_{k,i}$ is the distance between the k-th user and the i-th HAPS, and γ is the wavelength. The downlink received signal of the k-th user, located in the j-th cell of the i-th HAPS is defined as follows:

$$R_{k,j,i} = P_t + G_{k,j,i} - L_{k,i} + G_{user}, \tag{6}$$

where P_t is the transmission power and $G_{k,j,i}$ is the antenna gain for the k-th user located in the j-th cell of the i-th HAPS. The interference of a k-th user, located in the n_k-th cell of the p_k-th HAPS, is defined as follows:

$$I_k = 10log_{10}(\beta \sum_{j=1, j\neq n_k}^{N} 10^{(\frac{R_{k,j,p_k}}{10})} + \sum_{i=1, i\neq p_k}^{N_{HAPS}} \sum_{j=1}^{N} 10^{(\frac{R_{k,j,i}}{10})}), \tag{7}$$

where β is the activation factor and p_c is the center HAPS. The interference is calculated as the sum of signals from all cells except the same cell. The signal-to-noise ratio (SNR) and the signal-to-interference-plus-noise ratio (SINR) are defined as follows:

$$SNR_k = 10log_{10}(\frac{10^{(\frac{R_{k,n_k,p_k}}{10})}}{\sigma^2}), \tag{8}$$

$$SINR_k = 10log_{10}(\frac{10^{(\frac{R_{k,n_k,p_k}}{10})}}{\sigma^2 + 10^{\frac{I_k}{10}}}), \tag{9}$$

where σ^2 is the noise power. The noise power is defined as follows:

$$\sigma^2 = -174 + 10log(B) + NF, \tag{10}$$

where B is the bandwidth, and NF is the noise figure.

IV. Performance Evaluation

In this paper, our system model is depicted in Fig. 4. It includes seven HAPS, each consisting of seven cells. It is assumed that 3,500 users are uniformly distributed among these cells. Each HAPS, positioned at an altitude of 20 km, connects to a gateway that serves as a feeder link. The system's structure, encompassing both HAPS and gateways, uses R to denote the radius of the HAPS, r to denote the radius of a single cell, and g to denote the distance between the centers of different cells. We utilize Matlab for performance analysis. The transparent architecture accounts for both feeder and service links, each allocated a bandwidth of 10 MHz, totaling 20 MHz for the architecture. Conversely, the regenerative architecture considers only the service link, assigning it a bandwidth of 20 MHz. In this paper we evaluate the transparent and regenerative architectures. In the transparent architecture, repeaters are implemented according to the model specified in 3GPP TR 25.956 [6]. When the repeater gains have 70 dB and 90 dB, with a noise figure of 5 dB, and a repeater gain has 105 dB with a noise figure of 7 dB. The regenerative architecture uses the 3GPP and HAPSMobile antenna patterns. Assumptions for SNR and SINR thresholds are based on experimental data from HAPSMobile, with β set at 0.5 for SINR calculations. The detailed parameters of the feeder and service link are presented in Tables I and II.

TABLE I: Service link simulation parameters

Parameters	Value
Frequency	2 GHz
Propagation	Free space loss
Transmission power	43 dBm
Bandwidth	20 MHz
Antenna gain	[4, 5]
Noise figure	-3 dB
Required SNR	8 dB
Required SINR	-7 dB
L_N	-20 dB
L_F	-30 dB

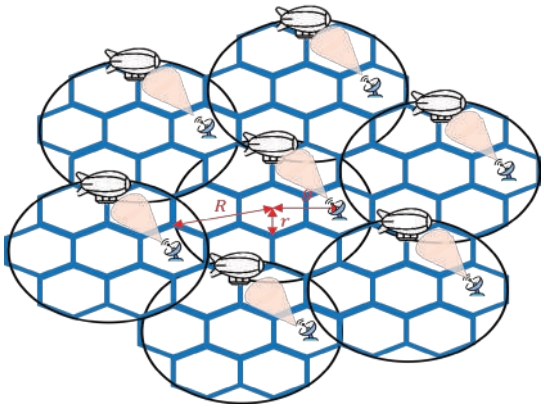


Fig. 4: System model.

A. Analysis of SINR According to HAPS Architecture

Fig. 5 depicts the architecture of HAPS and the SINR according to coverage. It illustrates that as the distance between users decreases, interference increases until reaching 40 km, after which it gradually decreases. Beyond 80 km, the SNR threshold is not satisfied, resulting in an outage. Hence, the maximum coverage is limited to 80 km. In the case of a transparent architecture, the SINR threshold is not satisfied regardless of coverage when the repeater gain is either 70 dB or 90 dB. With a repeater gain of 105 dB, significant interference occurs when the distance between users is less than 35 km, preventing the threshold from being satisfied. However, when the coverage satisfies 35 km, the SINR is observed to satisfy the threshold. Similarly, in the case of a regenerative architecture using the HAPSMobile antenna pattern, it is noted that the SINR meets the threshold when the coverage satisfies the condition of approximately 40 km.

TABLE II: Feeder link simulation parameters [7]

Parameters	Value
Frequency	3.65 GHz
Gateway transmit power	43 dBm
Gateway antenna gain	32.3 dBi
Gateway noise figure	3 dB
Gateway location	X-axis (45 km, 0, 0)
Feeder link FSPL	-137.7 dB
Repeater noise figure	Repeater gain 70, and 90: 5 dB Repeater gain 105: 7 dB

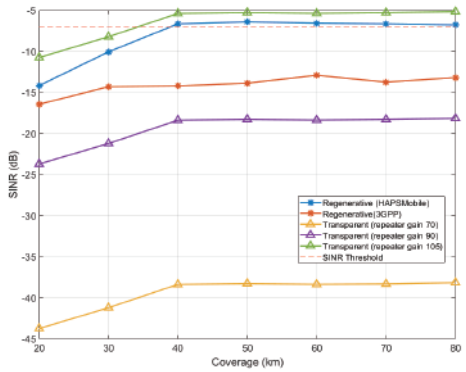


Fig. 5: SINR according to HAPS coverage and architecture.

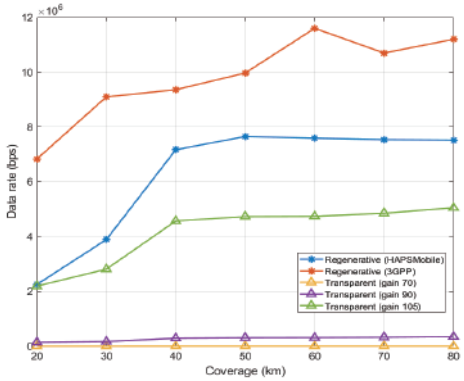


Fig. 6: Data rate according to HAPS coverage and architecture.

However, in the regenerative architecture using the 3GPP antenna pattern, the threshold is not satisfied because the antenna pattern, designed for both HAPS and satellite use, is dynamic with respect to the boresight angle.

B. Analysis of Data Rate According to HAPS Architecture

Fig. 6 depicts the data rates according to the architecture of HAPS and coverage. Similar to Fig. 5, it is observed that within a transparent architecture, repeater gains of 70 dB or 90 dB result in lower data rates. The regenerative architecture with the 3GPP antenna pattern exhibits the highest data rate due to its characteristics. This pattern achieves high antenna gain up to the 3 dB point. However, it receives lower antenna gain beyond that, resulting in significant differences in data rates between high and low gain regions. In our experiments, we calculate the average data rate of all users. We find that the 3GPP antenna pattern generally shows lower data rates in some regions but supports high data rates in others. For a transparent architecture with a repeater gain of 105 dB and a regenerative architecture using the HAPSMobile antenna pattern, the data rates are approximately 4.5 and 7.5 Mbps, respectively, for coverage satisfying 40 km. Although the SINR values are similar, the regenerative architecture only considers the service link and can utilize approximately twice the bandwidth compared to the transparent architecture, which includes both the service link and the feeder link. As shown in Table III, the propagation delay in the transparent architecture is observed to be approximately twice that of the regenerative architecture, primarily due to the time taken for the transparent architecture to receive the feeder link from the gateway.

V. Conclusion and Future Work

The architecture of HAPS can be classified into transparent or regenerative types, depending on whether there is a base station onboard. In this paper we described the architectures of HAPS and conducted performance analyses of these two structures. It was observed that, except for a repeater gain of 105 dB in the transparent architecture, the SINR thresholds were not satisfied. However, with a repeater gain of 105 dB, similar SINR values were achieved compared to those in the regenerative architecture. Nonetheless, in the transparent architecture, both the data rate and propagation delay are lower than in the regenerative architecture. This is due to the simultaneous design of the feeder link and service link. Since the required data rate varies depending on the scenario, it is crucial to design HAPS operations accordingly.

In future work, we will develop the design requirements for HAPS communication equipment based on various scenarios. Additionally, we will design to integrate a network that combines HAPS and satellites.

TABLE III: Propagation delay according to HAPS architecture

HAPS Architecture	Propagation Delay
Regenerative	0.28 ms
Transparent	0.44 ms

Acknowledgment

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean Government (MSIT) (2022-0-00704, Development of 3D-NET Core Technology for High-Mobility Vehicular Service)

References

[1] T. V. Nguyen, H. D. Le, N. T. Dang, and A. T. Pham, "On the Design of Rate Adaptation for Relay-Assisted Satellite Hybrid FSO/RF Systems," *IEEE Photonics Journal*, vol. 14, no. 1, pp. 1-11, Feb. 2022.

[2] S. Alfattani, W. Jaafar, H. Yanikomeroglu, and A. Yongacoglu, "Multimode High-Altitude Platform Stations for Next-Generation Wireless Networks: Selection Mechanism, Benefits, and Potential Challenge," *IEEE Vehicular Technology Magazine*, vol. 18, pp. 20-28, Jul. 2023.

[3] A. H. Arani, P. Hu, and Y. Zhu, "HAPS-UAV-Enabled Heterogeneous Networks: A Deep Reinforcement Learning Approach," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1745-1760, Jul. 2023.

[4] Y. Shibata, et al., "System Design of Gigabit HAPSmobile Communications," *IEEE Access*, vol. 8, pp. 157995-158007, Aug. 2020.

[5] 3GPP TR 38.811, "Study on New Radio (NR) to Support Non-Terrestrial Networks (Release 15)," Sep. 2020.

[6] 3GPP TR 25.956, "Universal Terrestrial Radio Access (UTRA): Repeater Planning Guidelines and System Analysis (Release 17)," Jun. 2020.

[7] Y. Xing, F. Hsieh, A. Ghosh, and T. S. Rappaport, "High Altitude Platform Stations (HAPS): Architecture and System Performance," *VTC2021-Spring*, pp. 1-6, Jun. 2021.

An LSTM-based Enhanced Mobility Prediciton Mechanism

Kamrul Hasan and Seong Ho Jeong
Dept. of Information and Communications Engineering
Hankuk University of Foreign Studies, Korea
kamrul@hufs.ac.kr, shjeong@hufs.ac.kr

***Abstract*—Accurate mobility prediction is essential for optimizing the usage of the network resources in dynamic environments. Existing prediction mechanisms often struggle with precise GPS location prediction due to the proximity of sequential GPS values, making it challenging to accurately forecast the next location based on current data. This paper presents an enhanced mobility prediction mechanism utilizing a Long Short-Term Memory (LSTM) network within a grid-based model. The grid-based approach mitigates the issues of closely spaced GPS values by transforming continuous location data into discrete grid cells, enhancing the LSTM's ability to capture movement patterns. Our extensive simulations show that the predicted movements closely resemble actual movements, demonstrating significant improvements over traditional methods. The proposed mechanism offers a robust solution for mobility prediction, with potential applications in network optimization and intelligent mobility management systems.**

Keywords—Mobility prediciton; LSTM; Grid-based Model

I. INTRODUCTION

In the rapidly evolving landscape of mobile networks, accurate mobility prediction has become increasingly vital for optimizing network resources and enhancing service quality, such as handover prediction in ultra-dense networks, seamless content delivery, and traffic pattern analysis [1]. Traditional methods for predicting GPS locations face significant challenges due to the high precision requirements and the minimal distance between sequential GPS values [2]. This proximity complicates the task of predicting the next location precisely, as minor inaccuracies can lead to substantial deviations in predicted paths. To address these issues, we propose an enhanced mobility prediction mechanism that leverages Long Short-Term Memory (LSTM) networks within a grid-based model, designed to improve prediction accuracy by effectively handling the inherent challenges of precise GPS location prediction.

The grid-based model transforms continuous location data into discrete grid cells, simplifying the complex patterns of movement into more manageable segments. This transformation allows the LSTM network to better capture and learn from spatial-temporal dependencies, leading to more accurate predictions. Our extensive simulations demonstrate that this approach significantly enhances the precision of mobility predictions, aligning predicted movements closely with actual movements. This paper details the design and implementation of the proposed mechanism, presents comprehensive evaluation results, and discusses the potential implications for network optimization and intelligent mobility management systems. By addressing the limitations of existing solutions, our approach offers a robust and effective tool for future advancements in mobility prediction technology. The rest of the paper is organized as follows: Section II describes the related work, Section III details our proposed LSTM-based mobility prediction model, Section IV presents simulation results, and Section V concludes our paper.

II. RELATED WORKS

Mobility prediction has been extensively studied due to its importance in optimizing network resources and improving the quality of service in mobile environments. Traditional approaches often rely on probabilistic models [3] and Markov Chains to predict user movements based on historical data [4]. While these methods provide a foundational understanding, they struggle with accurately forecasting closely spaced sequential GPS values, which are common in real-world scenarios [2]. This limitation arises because such models fail to capture the intricate temporal dependencies and nuances of human mobility patterns, resulting in predictions that often lack precision.

In recent years, the adoption of machine learning techniques, particularly Long Short-Term Memory (LSTM) networks, has shown promise in enhancing prediction accuracy [5]. LSTMs are well-suited for handling temporal sequences, making them a popular choice for mobility prediction tasks. By addressing the limitations of both traditional probabilistic models and existing machine learning approaches, our grid-based LSTM-model offers a robust solution for precise mobility prediction.

III. AN LSTM-BASED MOBILITY PREDICITON MODEL

The LSTM mechanism is particularly suitable for time-series or sequentially dependent datasets, providing more precise and accurate prediction results. Usually, the historical user movement data, e.g., GPS locations are used for mobility prediction. Since the GPS locations are uniquely identifiable and inter-position linear distances can be measured, which is useful for prediction. The dataset is divided into couple of grids according to the maximum and minimum distance of the GPS locations. Figure 1 illustrates the operation of three consecutive LSTM cells, $Cell_1$, $Cell_2$, and $Cell_3$.

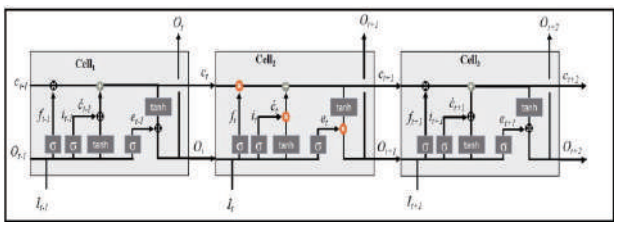


Figure 1: An LSTM networks contains three cells

Figure 1 shows that $Cell_1$ has three inputs, candidate state, c_{t-1}, previous cell output, O_{t-1}, and input gate, I_{t-1}, and two outputs, candidate state, c_t, and cell output, O_t. However, we only set I_{t-1} as input. The LSTM-based prediction model only

takes input from the set of GPS locations, l, which is defined as follows.

$$l = \{(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)\} = \{l_1, l_2, \ldots\ldots\ldots, l_n\} \ldots\ldots (1)$$

where $l_1 = (x_1, y_1)$, $l_2 = (x_2, y_2)$, and $l_n = (x_n, y_n)$ represent the sequence of GPS locations.

The number of inputs for l depends on the number of LSTM cells in the network. Based on Figure 1, a total of three inputs of l are required, such as $\{l_1, l_2, l_3\}$, and the output becomes $\{l_2, l_3, l_4\}$, where the expected predicted position is l_4 when the input length is set to three. Similarly, the expected predicted position is the next increased number of the input length. For example, if the input length is five and the inputs are $\{l_1, l_2, l_3, l_4, l_5\}$, the output becomes $\{l_2, l_3, l_4, l_5, l_6\}$, where the expected predicted position is l_6. During the training phase, the output of an LSTM cell is used for the validation check of the next LSTM cell input. This helps the LSTM network to optimize the prediction accuracy. Figure 2 describes the detailed workflows of the proposed model.

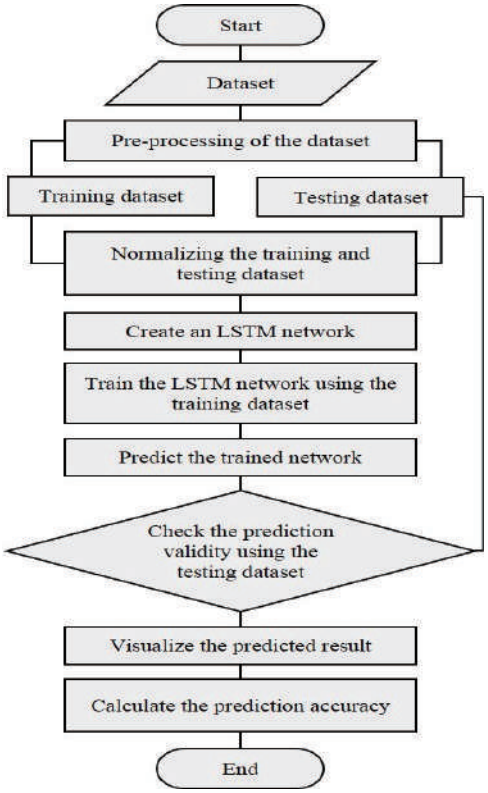


Figure 2: Work flow of the mobility prediction

IV. Simulation Results

We used three months of movement data and a Python-based simulation environment to create, train, and test the LSTM network. Figure 3 represents the simulation results. In the simulation, real movement data is represented in blue color, while the predicted movements are shown in orange color. Our results reveal that the predicted movements closely follow the real movements, indicating high prediction accuracy. This accuracy is significantly improved compared to the findings in [6], largely due to the implementation of our grid-based LSTM model. By transforming continuous location data into discrete grid cells, our model effectively mitigates the inaccuracies that typically arise in GPS location predictions, resulting in a more reliable mobility prediction mechanism.

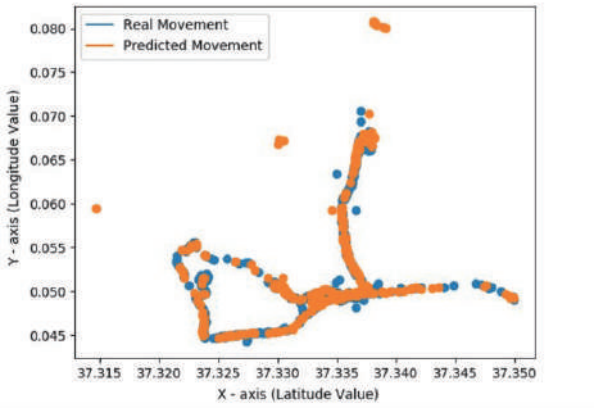


Figure 3: Real vs predicted movement

V. Conclusion

In this paper, we propose a grid-based LSTM model to enhance the accuracy of mobility prediction. We simulated our proposed model using a real dataset and demonstrated that our predicted movements are more accurate than existing prediction results. Our proposed grid-based mobility prediction model can predict user mobility more precisely and accurately than other models, including the regular LSTM-based mobility prediction model.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00484). This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-RS-2022-00156353) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

References

[1] Sun, W., Wang, L., Liu, J., Kato, N., & Zhang, Y. (2020). Movement aware CoMP handover in heterogeneous ultra-dense networks. IEEE Transactions on Communications, 69(1), 340-352.

[2] Chekol, A. G., & Fufa, M. S. (2022). A survey on next location prediction techniques, applications, and challenges. EURASIP Journal on Wireless Communications and Networking, 2022(1), 29.

[3] Lin, L., Li, J., Chen, F., Ye, J., & Huai, J. (2017). Road traffic speed prediction: A probabilistic model fusing multi-source data. IEEE Transactions on Knowledge and Data Engineering, 30(7), 1310-1323.

[4] Chen, S., Zhang, J., Meng, F., & Wang, D. (2021). A Markov chain position prediction model based on multidimensional correction. Complexity, 2021, 1-8.

[5] Ergen, T., & Kozat, S. S. (2017). Efficient online learning algorithms based on LSTM neural networks. IEEE transactions on neural networks and learning systems, 29(8), 3772-3783.

[6] N. Samaan and A. Karmouch, "A mobility prediction architecture based on contextual knowledge and spatial conceptual maps," in IEEE Transactions on Mobile Computing, vol. 4, no. 6, pp. 537-551, Nov.-Dec. 2005.

Deep Reinforcement Learning in Wireless Communications: Challenges and Solutions

Seungnyun Kim* and Byonghyo Shim†
*Laboratory for Information & Decision Systems, Massachusetts Institute of Technology, Cambridge, United States
†Department of Electrical and Computer Engineering, Seoul National University, Seoul, Republic of Korea
Email: *snkim94@mit.edu, †bshim@snu.ac.kr

***Abstract*—The proliferation of internet-connected mobile devices necessitates a shift from traditional mathematical model-based wireless communication systems to an intelligent architecture that incorporates artificial intelligence. Reinforcement learning (RL), particularly its enhancement through deep reinforcement learning (DRL), has emerged as a promising solution for making autonomous and effective decisions in wireless networks. DRL addresses complex, non-convex problems within dynamic environments by calculating the optimal policy through a deep neural network (DNN), thus enabling efficient decision-making for applications such as spectrum access and resource allocation. This paper focuses on overcoming the practical challenges of implementing DRL in wireless communications by integrating conventional system knowledge to enhance the learning process.**

***Index Terms*—6G, wireless communications, deep learning (DL), deep reinforcement learning (DRL).**

I. Introduction

Recently, various internet-connected mobile devices are penetrating to our daily lives [1]–[4]. In order to meet complicated and non-structured requirements of future applications, wireless communication systems must depart from the conventional mathematical model-based principle and move toward an intelligent system [5], [6]. Recently, a reinforcement learning (RL), mimicking human learning behavior in deep learning (DL), has received a great deal of attention. Essence of RL is learning through trial-and-error interactions. To be specific, an agent sequentially makes decisions, observes the results, and then automatically adjusts its strategy (referred to as policy) to achieve the desired goal. However, since the agent has to explore and gain knowledge of the entire system, it takes a long time to find the optimal policy in the traditional RL. To address this issue, a deep learning has been embedded in RL, opening a new era of deep reinforcement learning (DRL).

The DRL-based wireless communications paradigm provides a new perspective on the way we make decision. In wireless communications, network entities (e.g., BS, mobile device) need to make quick and autonomous decisions to achieve different goals (e.g., throughput maximization, energy minimization). Unfortunately, such problems are non-convex in general and related to fast time-varying network information (e.g, channel state). Whereas, DRL has proven to be an effective tool to address difficult decision-making problems in wireless communications. With the aid of powerful DNN, DRL agent calculates the value function for each decision and establishes the optimal policy to choose the decision with maximum value. As a result, DRL has been employed in many wireless communication applications such as channel access [7], network slicing [8], scheduling [9], and resource allocation [10].

However, there still exist practical huddles to overcome for the successful implementation of DRL in wireless communication applications. The primary purpose of this paper is to address the practical issues in the DRL-based wireless communications architecture and provide tractable solutions with advanced DRL techniques. The key message in our work is to make the most of the system knowledge in wireless communications to enhance the learning process of DRL. DRL is known to be model-free technique which means that the explicit system model is not required for its training. However, it is worth noticing that the collaboration of DRL with the conventional system knowledge can address the aforementioned practical issues and further improve its performance.

This work was supported, in part, by Institute of Information & communications Technology Planning & Evaluation (IITP) under the artificial intelligence semiconductor support program to nurture the best talents (IITP-2023-RS-2023-00256081) grant funded by the Korea government (MSIT) and by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) under Grant RS-2023-00252789 and Grant RS-2023-00208985.

II. Basics of Deep Reinforcement Learning

A. Reinforcement Learning

RL is a goal-oriented algorithm that learns a policy to achieve the desired goal from trials and errors. RL can be understood using the notions of agent, environment, state, action, and reward. In the learning process of RL, the agent observes the current state s_t from the environment, takes the action a_t, and then the environment returns the next state s_{t+1} and the immediate reward r_t as a feedback [11]. The policy π is a strategy that the agent employs to determine the next action based on the current state. The goal of the agent is to learn the optimal policy π^* maximizing the expected cumulative reward as

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid \pi\right] \tag{1}$$

where γ is a discount factor ($0 < \gamma < 1$) to give less weight to the future reward.

In order to find the optimal policy π^* in (1), the value function $\mathcal{V}^{\pi}(s)$ has been introduced. $\mathcal{V}^{\pi}(s)$ represents the expected reward obtained by following the policy π from each state s as

$$\mathcal{V}^{\pi}(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 = s, \pi\right] \tag{2}$$

Once the value function is available, the optimal policy can be easily obtained by choosing the action that maximizes $\mathcal{V}^{\pi}(s)$ of next state. Unfortunately, this is not possible since the transition probability $P(s'|s,a)$ is usually unknown to the agent. Thus, we construct another value $\mathcal{Q}^{\pi}(s,a)$ referred to as the quality function. Similar to $\mathcal{V}^{\pi}(s)$, $\mathcal{Q}^{\pi}(s,a)$ represents the expected value obtained by following policy π from each state s and action a as

$$\mathcal{Q}^{\pi}(s,a) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 = s, a_0 = a, \pi\right] \tag{3}$$

Accordingly, the optimal policy is greedily following the action that maximizes $\mathcal{Q}^{\pi}(s,a)$.

Q-learning is one of the most effective and widely used approach to find the optimal policy [12]. To actually learn $\mathcal{Q}^{pi}(s,a)$, we exploit the Markov property and Bellman equation that represents the optimal Q-value function $\mathcal{Q}^*(s,a)$ into the following recursive form [13]:

$$\mathcal{Q}^*(s,a) = \mathbb{E}\left[r(s,a) + \gamma \max_{a' \in \mathcal{A}} \mathcal{Q}^*(s',a')\right] \tag{4}$$

This means that $\mathcal{Q}^*$ can be recursively updated from the current estimates of Q-value. In particular, the Q-value function is updated according to the following rule:

$$\mathcal{Q}_{t+1}(s,a) \leftarrow \mathcal{Q}_t(s,a) + \alpha_t\Big(r(s,a) + \gamma \max_{a' \in \mathcal{A}} \mathcal{Q}_t(s',a') - \mathcal{Q}_t(s,a)\Big) \tag{5}$$

where α_t is the learning rate. The key idea behind this update equation is to minimize the temporal difference (TD) error between the target Q-value (i.e., $r(s,a)+\gamma \max_{a' \in \mathcal{A}} \mathcal{Q}_t(s',a')$) and the current Q-value (i.e., $\mathcal{Q}_t(s,a)$).

B. Deep Reinforcement Learning

One of the main drawback of conventional RL is that it cannot handle high-dimensional problems. In the traditional RL, the Q-values are stored and updated by using the Q-table of size $|\mathcal{S}| \times |\mathcal{A}|$ where $\mathcal{S}$ and $\mathcal{A}$ are the state and action spaces, respectively. However, since the sizes of state and action spaces increase exponentially with the size of the task, it is computationally burdensome to compute the Q-values of all possible state-action pair in a recursive update process. Recently, a deep reinforcement learning (DRL) has risen up as a breakthrough technique to overcome this curse of dimensionality. The key idea of DRL is to use deep neural network (DNN) to approximate the optimal policy or value functions. As a result, DNN can efficiently deal with high-dimensional observations. One of the most famous techniques of DRL is the deep Q-network (DQN) which uses DNN as a function approximator for Q-value (i.e., $\mathcal{Q}^*(s,a) \approx \mathcal{Q}(s,a;\theta)$) [14]. In the DQN, the network parameter θ of DQN is updated to minimize the TD error in (5).

It is worth mentioning that when the non-linear function approximator such as DNN is used, the Q-learning can be unstable or even might diverge. The reason is that a small change of Q-values may significantly change the policy, and thus the data distribution and the correlation between the current Q-value and the target Q-value might be tangled. DQN addresses this fundamental instability problem by using two mechanisms: 1) experience replay and 2) target Q-network. First, in the experience memory technique, the agent stores the transition (s_t, a_t, s_{t+1}, r_t) in the memory pool. Then, the agent randomly selects the samples (i.e., minibatch) from the memory pool to train DQN. By doing so, the number of interactions with the environment is reduced substantially and also the temporal correlations between the experience memories can be removed. Second, in the target Q-network technique, the DQN agent separately trains the current Q-value and the target Q-value. While the current Q-network is continuously trained at each epoch, the target Q-network is trained rather slowly and kept fixed for a large period of time. In this way, the correlation between the current Q-value and the target Q-value is reduced.

III. Challenges and Solutions for DRL-Based Wireless Communications

DRL successfully inherits the advantages of both reinforcement learning and deep learning. As a result, DRL has a wide range of applications in practice such as video game, autonomous vehicle, robotics, and wireless communications. However, since the tasks in wireless communications are in general complicated, large-scale, and requires sophisticated control, applying DRL in wireless communications still has many challenges to overcome. In this section, we discuss such practical issues and present tractable solutions with advanced DRL techniques. We also present a collaborate technique exploiting the system knowledge of wireless communications to enhance the learning process of DRL.

A. Domain Selection Problem

When applying DRL to wireless communications, the first thing to consider is to determine the basic elements of DRL (i.e., agent, environment, action, state, and reward). In fact, deciding which types of input and feedback the agent should pay attention to, so-called domain selection, is one of the most important problems in the application of DRL. The domain selection requires human decisions based on the knowledge or theories about the problem. Applications such as go game can ignore this problem since the environments are man-made, and thus strictly limited. On the other hand, in wireless communications, the domain selection is a serious problem since various types of approaches are available to achieve the desired goal. In this subsection, we provide useful guidelines

to address the domain selection problem in the DRL-based wireless communications. In particular, we focus on setting the action, state, and reward.

Action represents the decision that the agent makes to accomplish the task. Unfortunately, determining the action is not trivial in wireless communications due to the variety of solutions. For example, in the dynamic power control problem in heterogeneous network, the action stands for allocating the transmission power into multiple base stations (BSs). To do so, one can set the action to determine the cooperating group of BSs for each user and then equally allocate the transmission power. Another way is to quantize the power level (i.e., low, medium, high) and set the action to decide the power level of each BS. Hence, in the DRL-based wireless communication architecture, the agent needs to select the most effective action among all possible choices. Naturally, one might be tempted to use sophisticated action with high potential gain such as continuous level of power allocation. However, such action inevitably has a large action space which can seriously disturb the learning process of DRL. To be specific, the agent suffers from a huge computational overhead to explore the large action space and find the optimal action maximizing the value function. In fact, the learning process of DRL might even diverge when the action space is large. Thus, one should carefully set the action to achieve both desired performance within a tractable time.

State represents the current situation of the environment. In particular, the state includes various network information and service requirements (e.g., channel state or user demand) which work as input of the value and the policy function. Consequently, the network structure of DRL depends on the type of state. For example, in the dynamic power control problem, the state consists of the mobile user demand and the current working state of remote radio heads (RRHs). Since this information can be represented in an independent vector form, the deep Q-learning (DQL) with fully-connected neural network (FNN) is employed in this work. On the other hand, in the dynamic channel allocation problem in the multi-beam satellite system, the state includes the set of current and new arrived users and the channel allocation matrix. In this case, the state has a spatial correlation due to the co-channel interference and thus it can be represented in an image-like fashion. Thus, a convolutional neural network (CNN) is adopted to extract useful spatial features of the state. Further, if the state has sequential properties (e.g., time-varying channel state) the recurrent neural network (RNN) or the long short-term memory (LSTM) can be employed for DQL training.

Reward is the feedback that the environment provides in order to measure the success or failure of the agent's action. For example, in the spectrum access problem, the reward is given by the current data rate. However, in wireless communications, the reward is not strictly a sore function of action due to various network constraints such as user demand or transmit power limit. Thus, one should find the appropriate reward function which takes such constraints into account. One effective way to address this issue is to add a penalty term at the reward. For example, the penalty term for the user demand constraint can be a linear function of the difference between the real data rate and the user demand. When using the penalty term at the reward, it is worth mentioning that the weight of penalty (i.e., linear coefficient of linear function) greatly affects the learning process of DRL. To be specific, if a huge penalty is given for the irrelevant action which fails to satisfy the constraints, then the trained policy would strongly avoid such irrelevant action. However, in this case, the learning speed decreases due to the fluctuation of value function. On the other hand, if a small penalty is given for the irrelevant action, then the learning speed will increase but the optimization constraints might not be satisfied. Thus, it is important to find a proper penalty weight through various experiments.

B. High-Dimensional Action Space

In wireless communications, high-dimensional or even continuous control is required in various applications such as resource management, interference cancellation, and network slicing. Unfortunately, in the DRL-based wireless communication architecture, it is very difficult to support such delicate control due to the huge inefficiency in exploring the large action space. This problem is even more serious in the popularly used value-based approaches (i.e., Q-learning) since the agent has to find the optimal action maximizing the Q-value among the infinite number of candidates. In this subsection, we introduce two approaches to support the continuous action in DRL-based wireless communications: 1) actor-critic method and 2) action elimination network.

As mentioned, the value-based DRL techniques cannot be straightforwardly applied to high-dimensional actions. Recently actor-critic method which combines the benefits of policy gradient and value-based approaches [15] has grown in popularity as an effective means to support the continuous action. The general methodology of actor-critic method is to use a learned Q-value for the baseline to reduce the variance of expected reward. In other words, the "actor (policy)" learns by using the feedback from the "critic (Q-value)" which is calculated from the Q-learning. One of key development in the recent actor-critic method is a deterministic policy gradient (DPG) technique. When compared to the stochastic policy whose output is a probability distribution over the action space, the output of deterministic policy is a specific action. Thus, the deterministic policy gradient only integrates over the state space whereas the stochastic policy gradient integrates over both state and action spaces. As a result, DPG requires fewer samples in the action space exploration which means that DPG can efficiently find the optimal policy even in the large action space. Later work introduced deep DPG (DDPG) that exploits a deep neural network to train both policy and value networks. In fact, the DDPG algorithm has become one of the most popular and powerful DRL techniques for supporting the continuous action.

The second method is the action elimination network (AEN) technique which incorporates the system knowledge in

wireless communications to remove the irrelevant actions in prior [16]. The key idea of AEN is to exploit the property that many of the actions are redundant or irrelevant. For example, in the BS-user association problem, the associations which fail to satisfy the user demand or associate with too many users can be regarded as irrelevant action. Such actions can be identified by using the mathematical system model well-established in numerous previous works. Based on this observation, the environment sends the elimination signal which indicates the feasibility of the current action to the agent. Then the agent, who recognizes the elimination signal as an auxiliary reward, trains AEN as a function approximator for the elimination signal. As a result, the agent can distinguish the irrelevant action by itself and then efficiently explore the desirable action space.

C. Dynamic Network Information

One of major issues in wireless communications is to efficiently solve complex network optimization problems which consist of various network information such as channel state, user demand, or cell status. Naturally, in the DRL-based wireless communication architecture, this information is included in the state. However, a fundamental challenge arises from the time-varying properties of the network information. To be specific, since the transition of network information depends on the physical environments, it is in general independent with the action taken by the agents. Hence, the agent cannot guarantee that the current chosen action will lead to a desirable state maximizing the cumulative reward which means that the far-sighted policy is intractable. To address this issue, it is obvious that the correlation between the state transition and the action should be increased. It is worth mentioning that we can exploit the system knowledge to use an alternative form of network information as a state which has a high correlation with the action. For example, in the dynamic power control problem, by using the fact that the data rate is a function of the useful signal and the interference, one can use them as a state instead of the channel. These information includes both channel state and allocated transmit power and thus the correlation with the action is very high. As a result, the agent can choose the optimal power allocation that leads to good state (i.e., strong useful signal and weak interference) and finally maximizes the cumulative reward.

D. Off-Line Training Problem

Like most of the machine learning techniques, DRL requires a huge amount of training data to achieve its full potential. Specifically, DRL requires experience replays in a form of (s_t, a_t, s_{t+1}, r_t) so that the agent can explore the memory pool to find the optimal policy. Unfortunately, in wireless communications, it is very difficult to obtain such a large amount of training data due to practical problems such as cost or time. One simple way to address this issue is to exploit a synthetic data set generated by using the conventional mathematical system model. Although the mathematical system model is an approximation of the real system, we can infinitely generate the synthetic data from it without any pressure of practical issues. Moreover, by using the well-developed traditional analytic techniques based on this system model, we can obtain a sub-optimal policy in which a supervised learning is applicable. For example, in dynamic power control problem, the authors presented an off-line Q-learning that exploits the experience replays generated from the conventional techniques to train Q-value in advance. By doing so, the agent can quickly learn a policy which has a similar performance with the conventional techniques. Also, during the testing phase, this sub-optimal policy can be further improved by exploiting the real data. Thus, the agent can develop the optimal policy adapted for the practical system.

REFERENCES

[1] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[2] S. Kim, J. Moon, J. Wu, B. Shim, and M. Z. Win, "Vision-aided Positioning and Beam Focusing for 6G Terahertz Communications," *IEEE J. Sel. Areas Commun.*, 2024.

[3] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G Internet of Things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2021.

[4] Y. Ahn, W. Kim, and B. Shim, "Active user detection and channel estimation for massive machine-type communication: Deep learning approach," *IEEE Internet Things J.*, vol. 9, no. 14, pp. 11 904–11 917, Jul. 2021.

[5] H. Lee, B. Lee, H. Yang, J. Kim, S. Kim, W. Shin, B. Shim, and H. V. Poor, "Towards 6G hyper-connectivity: Vision, challenges, and key enabling technologies," *J. Commun. Netw.*, vol. 25, no. 3, pp. 344–354, Jun. 2023.

[6] S. Kim, J. Son, and B. Shim, "Energy-efficient ultra-dense network using LSTM-based deep neural networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4702–4715, Jul. 2021.

[7] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 2, pp. 257–265, Jun. 2018.

[8] K. Suh, S. Kim, Y. Ahn, S. Kim, H. Ju, and B. Shim, "Deep reinforcement learning-based network slicing for beyond 5G," *IEEE Access*, vol. 10, pp. 7384–7395, Jan. 2022.

[9] J. Moon, S. Kim, H. Ju, and B. Shim, "Energy-efficient user association in mmWave/THz ultra-dense network via multi-agent deep reinforcement learning," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 692–706, Jun. 2023.

[10] X. Wang, Y. Zhang, R. Shen, Y. Xu, and F.-C. Zheng, "DRL-based energy-efficient resource allocation frameworks for uplink NOMA systems," *IEEE Internet Things J*, vol. 7, no. 8, pp. 7279–7294, Aug. 2020.

[11] R. S. Sutton, A. G. Barto *et al.*, *Introduction to reinforcement learning*. MIT press Cambridge, 1998, vol. 135.

[12] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[13] R. E. Bellman and S. E. Dreyfus, *Applied dynamic programming*. Princeton university press, 2015, vol. 2050.

[14] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

[15] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in *Advances in neural information processing systems*, 2000, pp. 1008–1014.

[16] H. Ju, S. Kim, Y. Kim, and B. Shim, "Energy-efficient ultra-dense network with deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6539–6552, Aug. 2022.

Spectum Status and Issues in Korea

Seung-Hoon Hwang
Dongguk University
Seoul, Korea
shwang@dongguk.edu

***Abstract*—This paper summarizes spectrum status and issues in Korea. Both 5G and 6G status are described. First, spectrum strategies and 5G status in Korea is explained. Also, digital spectrum plan of Korean government is summarized.**

Keywords—Spectrum, Korea, 5G, 6G

I. INTRODUCTION

In Korea, 5G forum was renamed by 6G forum in 2024. The objective of the spectrum committee is to investigate of issues on frequency spectrum ranging over Low-, Mid-, and High-bands for 5G/6G mobile communication system to support both legacy and new services. Regarding the governance, Prof. Seung-Hoon Hwang is the committee chair and three vice-chairs from SKT, KT, Samsung electronics and Ericsson LG. More than 30 members are attending from more than 15 member companies. The Spectrum Committee collaborates and exchanges information with the government, forums, academia, and industry. To this end, it holds regular meetings every two months and publishes issue reports every year. In addition, workshops and seminars are held irregularly. In January, we held a 6G open symposium in collaboration with the KICS to share WRC23 results and exchange information on frequency issues.

In December of 2023, 6G framework was published by ITU[1]. From the timeline provided by ITU and 3GPP, 6G is officially researched and developed from 2024. Out of many items, spectrum is the first item which is interested. Therefore, in this paper, the spectrum status and issues in Korea is described. Note that this paper is based on the presentation material for NGA and 6G forum joint workshop, May 2024[2].



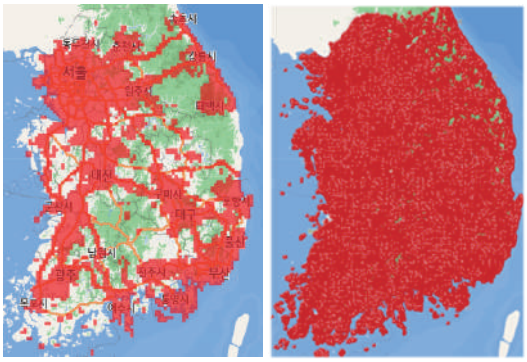
Fig.1 Spectrum committee in 6G forum

II. SPECTRUM STRATEGY AND 5G STATUS IN KOREA

In this section, history of spectrum strategy in Korea is introduced. Also, 5G status in Korea is described.

Regarding the spectrum plan, there were mobile spectrum 1.0 and 2.0 in 2012 and 2013. These provided LTE frequencies such as 1.8 GHz and 2.6 GHz. In 2017, 5G frequencies such as 3.5 GHz and 28 GHz were supplied through the K-ICT spectrum. Through the 5G+ spectrum released in 2019, 6 GHz unlicensed band and 4.8 GHz e-UM 5G frequency were supplied for convergence and digital transformation. Additionally, the digital spectrum plan will be announced based on the Digital Bill of Rights in May this year.

The 5G base station deployment status is described in Fig.2. The red color shows the base station construction area. It is shown that the 5G coverage, which was built around large cities in 2022, is built nationwide in the first half of 24. Based on this national network, it is expected to continue to evolve into voice over NR, NR carrier aggregation, and open RAN in the future.



<Aug. 2022> <The 1st half, 2024>

Fig.2 5G base station deployment status [3]

Local 5G is called e-UM 5G in Korea. It is also called Private 5G, non-public network. It is a network that is specialized for a specific purpose in a local area. The e-Um is named after the initials of the three usage scenarios of 5G, and it also means connecting in Korean. A bandwidth of 600 MHz can be used in the 28 GHz band and a bandwidth of 100 MHz in the 4.7 GHz band. Related applications include smart factory, remote operation, AR/VR value, and education using AR/VR.

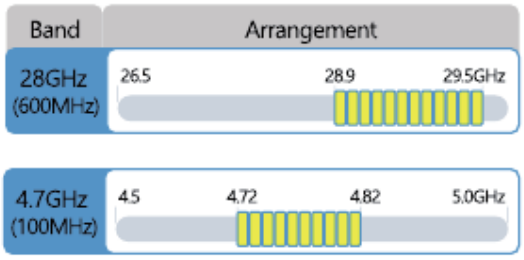


Fig.3 Spectrum for e-Um 5G in Korea [3]

Wi-Fi 4/5 supports 20-80 MHz bandwidth, Wi-Fi 6 supports 160 MHz bandwidth, and Wi-Fi 7 supports 320 MHz. Korean government plans to expand the use of Wi-Fi 6

and promote evolution to the next-generation standard Wi-Fi 7 within 24 years. According to the WRC-23 decision, wireless LANs may be supported in the 6 GHz band.

III. Digital Spectrum Plan (Tentative)

The digital spectrum plan is being planned by MSI, Korea. A public hearing was held in January, 2024. Then, the plan will be announced in May, 2024. Here, key strategies are summarized.

Strategy 1 is to improve the efficiency of all mobile communication frequencies. To this end, the government plans to promote balanced supply of low, medium and high bands, open mobile communication frequencies to all areas, including non-mobile communication companies, and promote supply in consideration of broadband utilization.

Strategy 2 is to secure 6G frequencies to lead the global market. To this end, it is said to promote international cooperation and response such as ITU and 3GPP, and to discover and secure balanced frequencies in low, middle, and high bands. In addition, it is planning to discover 6G candidate bands in new bands discovered and studied by ITU and existing bands in use as mobile communication. For WRC-27, the Korean government is planning to discover 6G frequencies for 4.4-4.8 GHz, 7.125-8.4 GHz, and 14.8-15.35 GHz, and also to pursue satellite frequency discovery research at the existing 3G/4G frequency of 698-2.7 GHz. In particular, it is necessary to support research on how to utilize 6G frequencies in the 7.125-8.4 GHz band while protecting existing business.

Figure 4 shows the 6G frequency in the agenda of the WRC-27 IMT, and the three aforementioned frequency bands are marked with red boxes, which shows that it is also in line with Korea's position.

Finally, Strategy 4 is considering industrial and life wireless service innovation using unlicensed frequencies, expanding the channel width for Wi-Fi 7 for high-speed near-field communication and introducing an automatic frequency adjustment system for joint use with other services.

IV. Conclusion

In this paper, the spectrum status and issues in Korea was described. The spectrum strategies have been developed until now. The MSIT of Korea will announce the digital spectrum plan soon.

For your information, figure 5 shows the direction of 6G R&D in Korea. From this year to 28, 327 million dollars will be invested. Among the advanced wireless technologies, the development of the upper midband, the AI low power technology among the energy-saving technologies, and the high-efficiency technology are considered good items for the two countries to cooperate on frequency.



Fig.5 6G R&D plan in Korea [2]

Acknowledgment

This paper includes most of material which was presented at NGA and 6G forum joint workshop, May 2024. Also, note that the presentation material was helped by Dr. Jaewoo Lim in National Radio Research Agency.

References

[1] 6G framework and overall objectives of the future development of IMT for 2030 and beyond, ITU

[2] Presentation material, NGA and 6G forum joint workshop, May 2024.

[3] Presentation material, 6G open symposium, Feb. 2024

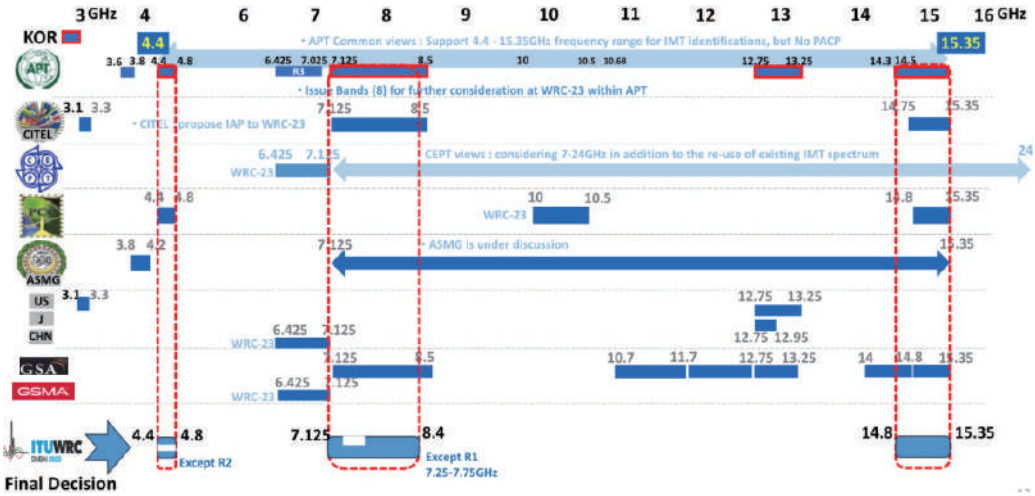


Fig.4 WRC-27 IMT agenda for future IMT-2030 [3]

Improved Human Echolocation Click Detection and Ranging with Gammatone Filter Banks

Nur Luqman Saleh, Alyani Ismail, & Aduwati Sali
Dept. of Comp. and Comm. Systems Engineering
Faculty of Engineering, Universiti Putra Malaysia.
Email: nurluqman,alyani,aduwati@upm.edu.my

Nur Emileen Abdul Rashid
Microwave Research Institute
Universiti Teknologi Mara, Malaysia.
Email: emileen98@uitm.edu.my

***Abstract*—Human echolocation, involving object detection via sound reflections, poses challenges for signal processing. This paper proposes an improved pulse compression technique based on the summation of gammatone filter banks for detecting and estimating the range of human echolocator mouth-clicks. Conventional matched filter techniques suffer from higher side-lobe levels, which lead to false alarms and poor detection. This approach extends previous work by incorporating multi-phase correlation and summation within the gammatone filterbank scheme. Compared to the conventional matched filter technique, the proposed method achieves a significant improvement in signal power exceeding -8 dB. Additionally, simulations demonstrate reliable accuracy in the estimation of the target range using the proposed technique. These findings suggest that the multi-phase summation of gammatone filterbanks offers a promising approach for enhancing both detection and range estimation in applications utilizing human echolocation mouth-click signals.**

***Index Terms*—Human echolocator, Signal processing, Gammatone filter, Bio-inspired, Radar.**

I. Introduction

Similar to bats and dolphins, some blind individuals navigate using echolocation from mouth clicks. Researchers explore how they process echoes for object detection. This paper proposed an improved pulse compression technique inspired by the multi-phase correlation of gammatone filter banks (GFs) and summation, using simulations to assess its potential for improved object detection and range estimation in human echolocation.

Building upon the work by *N.L. Saleh et al.* [1], this paper aims to further exploit the potential of multi-phase correlation and summation of gammatone filter bank (GFs) banks within the pulse compression process. This novel approach leverages simulations to investigate its capabilities for enhanced object detection and range estimation in human echolocation.

As this paper is in its early stages, the specific goal is to conduct a proof-of-concept using simulations. The paper is organized as follows: Section II presents a concise description of the required simulation techniques. Section III details the simulation results and discussion, while Section IV provides concluding remarks.

II. Enhanced Pulse Compression with Gammatone Filter Banks

To understand how human echolocators analyze their echolocation signals, it is essential to examine the human auditory system's processing mechanisms. The human inner ear's basilar membrane acts like a filter bank, with different regions sensitive to specific frequencies. This tonotopic organization is similar to GFs based on the Equivalent Rectangular Bandwidth (ERB) concept [2]. The ERB ensures a uniform frequency distribution along the cochlea. GFs are like multiple bandpass filters, each tuned to a distinct center frequency within a specific bandwidth, given in (1).

$$\begin{aligned} GF_i(t) &= at^{n-1}e^{-2\pi d_i t}\cos(2\pi cf_i t + \theta) \\ d_i &= 1.012 \times 2\pi \times ERB_i \\ i &= 1, 2, 3, ..., 10 \end{aligned} \tag{1}$$

where cf_i is the center frequency on each channel, a is a constant value controlled by the gain, θ is the phase of the cf_i, d_i is the decay factor and n is the filter order equal to 4 (it is sufficient to provide a good fit for the human auditory shape) [3], [4].

Figure 1 depicts the proposed mouth-click pulse compression method. The details of the multi-phase correlation and summation of GFs scheme are discussed in [5].

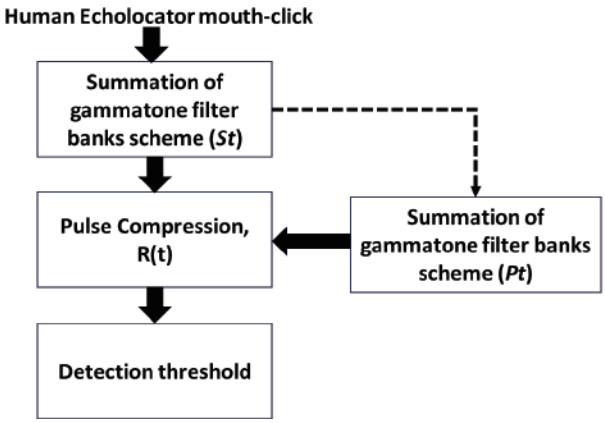


Fig. 1: Proposed detection scheme utilizing pulse compression based on the summation of gammatone filters for the detection and estimation of human echolocator mouth-clicks

The coefficients from the transmit and backscatter simulation mouth-click (pulses), denoted as $S(t)$ and $P(t)$ respectively, are given in Eq. (2). During the pulse compression

phase, these signals ($S(t)$ and $P(t)$) are substituted into Eq. (2) to extract the obstacle information, represented by $R(t)$, along with the time delay τ. The section III will describe the evaluation of this approach by comparing its output to that of a matched filter (MF).

$$R(t) = \int_{-\infty}^{\infty} S(t) \cdot P(t-\tau)dt \tag{2}$$

where $R(t)$ is the received signal, $S(t)$ is the transmitted pulse, $P(t)$ is the received pulse, and τ is the time delay between the transmitted and received pulses.

III. Simulation Setup

The simulation utilized synthetic human echolocator mouth-clicks derived from the mathematical formulas in [6]. To replicate a verifiable signal, three main frequencies were defined based on the mouth-click data in [7]. The sampling rate was set to 44.1 kHz, double the maximum acoustic frequency. A single obstacle was positioned 0.3 meter away within the simulation setup. For this preliminary investigation, the influence of radar cross-section, propagation effects, and noise are not considered. These complexities will be incorporated in future refinements of the model.

IV. Results and Discussion

The local maxima peak from $R(t)$ is detected and visible at approximately 0.3 meter, as illustrated in Fig. 2 (a). This peak exhibits a higher correlation coefficient, indicating a greater likelihood of correct target detection, consistent with radar principles. In contrast, the local maxima peak from the MF is significantly cluttered, as depicted in Fig. 2 (b). This clutter may lead to false alarms after normalization due to higher side lobes relative to the peak maxima. The impact of these side lobes on false alarm rates will be further discussed in the next paragraph.

Figure 2 further strengthens the case for the proposed method. In comparison to the MF output, the $R(t)$ signal shows a notable reduction in side lobes (MLLs) and a significant improvement in SLLs (by more than -8 dB). This enhanced performance adheres closely to radar detection principles, which typically require SLLs to exceed -3 dB. These results indicate that the $R(t)$ processing scheme effectively detects diverse waveform signals such as human echolocation mouth-clicks, minimizing signal interference from side lobes

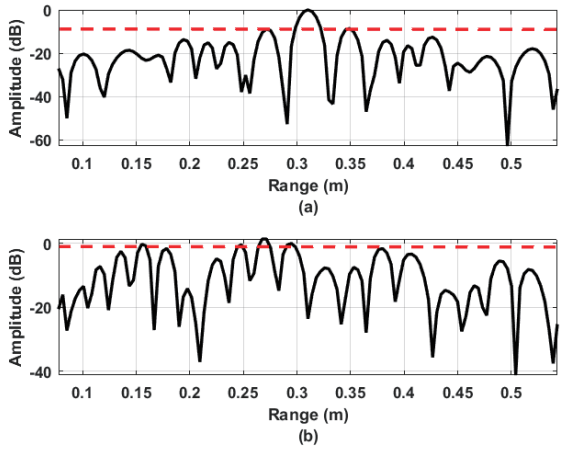


Fig. 2: Normalized pulse compression output

Therefore, this study highlights the potential of human echolocation mouth-clicks for bio-inspired sensor design, particularly in complementing existing radar and sonar systems. Operating within acoustic frequency components, human echolocation demonstrates its ability to provide reliable obstacle information. This capability is particularly valuable in miniature hardware design, using low-frequency waveforms to achieve effective detection capabilities.

V. Conclusion

In conclusion, this study explored a novel pulse compression technique inspired by the human auditory system. The proposed method utilizes multi-phase correlation and summation of gammatone filter banks for processing human echolocator mouth-click signals. Our results demonstrate that this approach offers significant advantages compared to conventional matched filter techniques. The proposed method achieved a substantial reduction in side lobe levels (exceeding -8 dB), enabling reliable target range estimation in simulations. These findings suggest the potential of multi-phase summation of gammatone filter banks for bio-inspired sensor design, particularly in applications requiring enhanced detection and range estimation with low-frequency signals, inspired by human auditory processing systems. Moreover, findings from this paper could lead to future exploration of assistive technologies that empower blind individuals to leverage echolocation for navigation.

References

[1] R. S. A. Raja Abdullah, N. L. Saleh, N. E. Abdul Rashid, and S. M. S. Ahmad, "Bio-inspired signal detection mechanism for tongue click waveform used in human echolocation," Electronics Letters, vol. 53, no. 22, pp. 1456–1458, Oct. 2017, doi: 10.1049/el.2017.3096.

[2] M. Slaney, "An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank," 1993. [Online]. Available: https://api.semanticscholar.org/CorpusID:60158999

[3] J. Holdsworth, I. Nimmo-Smith, R. Patterson, and P. Rice, "Implementing a gammatone filter bank," Annex C of the SVOS Final Report: Part A: The Auditory Filterbank, vol. 1, pp. 1–5, 1988.

[4] R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, "Complex Sounds and Auditory Images," in Auditory Physiology and Perception, Elsevier, 1992, pp. 429–446. doi: 10.1016/b978-0-08-041847-6.50054-x.

[5] N. L. Saleh et al., "Utilizing gammatone filter coefficient to improve human mouth-click signal detection using a multi-phase correlation process," Measurement, vol. 224, p. 113887, Jan. 2024, doi: 10.1016/j.measurement.2023.113887.

[6] X. Zhang et al., "Human echolocation: waveform analysis of tongue clicks," Electronics Letters, vol. 53, no. 9, pp. 580–582, Apr. 2017, doi: 10.1049/el.2017.0454.

[7] L. Thaler, S. R. Arnott, and M. A. Goodale, "Neural Correlates of Natural Human Echolocation in Early and Late Blind Echolocation Experts," PLoS ONE, vol. 6, no. 5, p. e20162, May 2011, doi: 10.1371/journal.pone.0020162.

Exploring Machine Learning Models for Real-Time Maritime Vessel Monitoring on Interlinked Camera-Mounted Buoys

Michael Ralph A. Estrada*, Margrette B. Yebes†, Emilio Escobillo III‡, Gabriel Avelino Sampedro§
*†‡Gokongwei College of Engineering, De La Salle University, Manila, Philippines
§College of Engineering, University of the Philippines, Diliman, Philippines
Correspondence Email: garsampedro@ieee.org

***Abstract*—Undocumented maritime vessels traversing the Philippines pose significant threats, from potential smuggling of contraband and illegal harvesting of its natural resources due to the country's archipelagic nature; it gives additional challenges to law enforcement to foresee activities on the bodies of water. In this paper, the proponent's objective is to develop a device to monitor territorial waters. Using Buoys equipped with real-time monitoring cameras with built-in algorithms for image processing and detection of ships passing through, the paper aims to determine which algorithm has the highest accuracy. Based on the results of the three algorithms, SSD had the highest score vs Yolov8 and Faster RCNN failed to yield any results.**

***Index Terms*—Computer Vision, Maritime Monitoring, Data Transmission, Deep Learning, RF Transmission.**

I. Introduction

The archipelagic configuration of the Philippines, comprising over 7,000 islands, presents unique security challenges, especially in policing its extensive and fragmented maritime boundaries. This geographical setup facilitates unauthorized entries and exits, making the region susceptible to various maritime-related crimes such as smuggling and illegal fishing. Additionally, the sprawling nature of these territories complicates surveillance efforts, which are crucial for safeguarding national interests and marine resources. The practice of illegal, unreported, and unregulated (IUU) fishing, as highlighted by the Philippine IUU Fishing Assessment Report 2022, is particularly rampant, accounting for an estimated 27 to 40 percent of the total fish catch in 2019, which translates to a substantial economic loss of approximately P62 billion (US$1.3 billion).

Historically, the initiatives to monitor these activities have been limited by technological constraints and the vast area needing surveillance. Previous methods have relied heavily on manned patrols and static surveillance systems, which are resource-intensive and limited in coverage and efficacy. There has been a significant reliance on systems that capture video and imagery to detect and track marine vessels; however, these systems often face challenges such as high operational costs, limited range of detection, and data transmission issues. The effectiveness of these systems is further hampered by the slow pace of technological adaptation and integration into existing maritime security frameworks, leaving a considerable gap in real-time monitoring capabilities.

In response to these challenges, this study proposes the deployment of interlinked camera-mounted buoys equipped with advanced object detection algorithms, specifically SSD, RCNN, and YOLOv8, to provide real-time, autonomous maritime vessel monitoring. This solution aims to bridge the gap in current maritime surveillance by leveraging cutting-edge computer vision and deep learning technologies to enhance detection accuracy and operational efficiency. The buoys will be strategically placed across critical maritime zones to monitor vessel movements and activities autonomously. However, the scope of this research is limited to the development and preliminary testing of the object detection capabilities of these buoys in a controlled environment, acknowledging that further field testing and integration with existing systems will be necessary to fully realize the potential of this innovative surveillance approach.

II. Review of Related Literature

A. Computer Vision Techniques

Computer vision involves analyzing and interpreting real-world scenes using cameras. This field addresses both theoretical and practical challenges in designing and implementing algorithms. This paper will review various algorithms applicable to the discussed application.

B. Existing Technology

In 2009, a system was introduced whose primary function was to capture video and imagery to detect and track marine vessels near the platform automatically; data is processed and transmitted to a ground control station via a bi-directional RF satellite link. The design is considered to be low-cost, easy to deploy and recover, and discreet [1].

C. Research Gap

The existing study utilized an algorithm that involves anisotropic diffusion [2]. There are more deep learning object detection algorithms that are constantly updated and could be a potential candidate for an alternative model compared to the one utilized in the previous study. The previous study also utilized a satellite link and incorporated the Bottom Ocean

§Corresponding author

Profiling Station [3]. This design aspect could be explored due to more emerging technologies in data transmission.

III. Proposed System

There will be 2 devices aimed to be developed in the study: the Vessel Detection Device and the Receiving Hub. The Vessel Detection would have the Camera and Microprocessors have a trained Object Detection Algorithm that would take videos and snap pictures to transmit using the transmitter. The transmitter will send the image to the 2nd device; the receiving hub will then process the image and display the image to the person in charge. Fig. 1 illustrates the setup of the system.

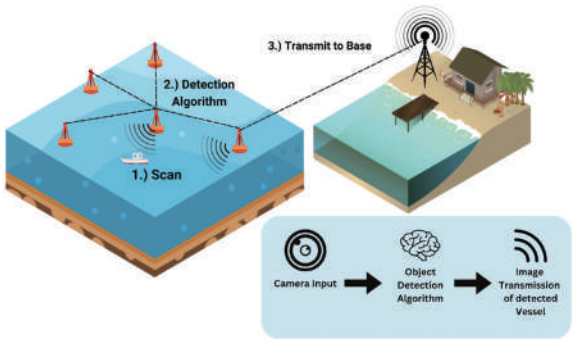


Fig. 1. Detection Process from Camera Scan to Receiving Hub via RF Transmission

A. Ship Detection

The experiment would utilize multiple object detection models to test which model would suit this research. They will be trained with the same dataset and will utilize Intersection over Union (IoU) to measure the accuracy of detection. The proponents will identify the model with the least detection error. When having the same least error value in detection, the proponent will identify which model utilizes the least computing power.

B. Data Transmission

Tentatively, the proponents are eyeing DigiXbee RF modules as a means of transmitter and receiver, specifically the Xbee Pro S2C, because of its capability to transmit up to 3.2 kilometers, and typically, coastal territories would range more than the component's capable range but interconnecting all the transmitters to one another would mean make it feasible. Further research is needed to ensure proper image data transmission to the receiving hub.

C. Equations

The error in object detection will be computed by Intersection over Union (IoU), also known as the Jaccard Index. It is described to be:

$$IoU = \frac{A \bigcap B}{A \bigcup B} \tag{1}$$

It can be interpreted that the closer the value of 1 is, the more accurate the prediction of the tracked object. The paper currently focuses on the object detection algorithm. The table presents evaluation metrics for the three models. The criteria used for evaluation are mAP (mean Average Precision) at different IoU thresholds: 50-95, 50, and 75.

TABLE I
Evaluation Metrics from different Models

Criteria	Model		
	YOLOv8	***R-CNN***	***SSD***
mAP 50-95	0.440	-	0.87
mAP 50	0.657	-	0.81
mAP75	0.464	-	0.87

IV. Performance Evaluation

SSD consistently performs better across all evaluation metrics compared to Yolov8 and Faster-RCNN. It has the highest mAP scores at all thresholds (50-95, 50, and 75), indicating better overall performance in object detection tasks.

In terms of mAP50, this is at an IoU threshold of 0.5. It represents how well the model predicts the bounding boxes around the detected object with a minimum of 50% overlap with the ground truth boxes. The SSD did relatively well compared to the other candidates, with 0.87, while Yolov8 garnered 0.657, and Faster RCNN failed to yield any results

In mAP75, the mAP at a higher IoU threshold of 0.75, the SSD showed 0.87, indicating that the object detection model performs better than the other models when evaluated under a stringent 75% IoU threshold.

V. Conclusion

The preliminary testing within this study has demonstrated that SSD is the most effective object detection model for our proposed system. The model outperforms others in accuracy across various metrics, particularly at the standard detection threshold (IoU of 0.5), making it a potential solution for real-time maritime vessel monitoring. While the results are encouraging, they also highlight the importance of continuous improvement and technology adaptation to meet the dynamic challenges of maritime surveillance.

Future work should focus on expanding the training dataset to improve model robustness and accuracy, especially under varied conditions. Integrating more advanced models and exploring new deep-learning frameworks can enhance detection capabilities. Optimizing the buoys' data transmission systems is crucial for reliable and timely surveillance data delivery, particularly over large distances and in adverse weather conditions.

References

[1] H.-M. Hu, Q. Guo, J. Zheng, H. Wang, and B. Li, "Single image defogging based on illumination decomposition for visual maritime surveillance," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2882–2897, 2019.

[2] A. Abdullah Yahya, J. Tan, B. Su, K. Liu, and A. N. Hadi, "Image edge detection method based on anisotropic diffusion and total variation models," *The Journal of Engineering*, vol. 2019, no. 2, pp. 455–460, 2019.

[3] L. Langebrake, C. Lembke, R. Weisberg, R. Byrne, D. Russell, G. Tilbury, and R. Carr, "Design and initial results of a bottom stationing ocean profiler," in *OCEANS '02 MTS/IEEE*, vol. 1, 2002, pp. 98–103 vol.1.

Deep Learning Model for Anomaly Detection in Water Supply Pipes Based on Sensor Data

Jinhyung Cho
School of Computer Engineering
Dongyang Mirae University
Seoul, Republic of Korea
cjh@dongyang.ac.kr

Hwansoo Kang
School of Computer Engineering
Dongyang Mirae University
Seoul, Republic of Korea
hskang@dongyang.ac.kr

Heejin Lee
School of Computer Engineering
Dongyang Mirae University
Seoul, Republic of Korea
heejinlee@dongyang.ac.kr

Yoorim Cho
Department of Computer Science
University of Sheffield
Sheffield, United Kingdom
kayla3cho@gmail.com

Yudoo Kim
School of Computer Engineering
Dongyang Mirae University
Seoul, Republic of Korea
kimyudoo@dongyang.ac.kr

Abstract— The primary aim of this research is to devise a robust anomaly detection model for pipelines, enabling the differentiation of leakage status based on sensor data integrated into intelligent water distribution systems. To accomplish this objective, a series of experiments employing various deep learning models were conducted to craft an optimal anomaly detection framework. Through meticulous analysis of the training data, significant feature variables influencing leakage were identified. Additionally, leveraging a fundamental multi-class classification model, coupled with CNN-based training, facilitated the development of a sophisticated deep learning architecture tailored for anomaly detection in pipeline systems.

Keywords—Deep Learning, Anomaly Detection, Water Supply Pipe, Sensor Data, Smart Pipe.

I. Introduction

The task environment for conducting this study is a self-diagnostic and state monitoring integrated smart pipe system, which is divided into three specific development items, each combining the technologies of the respective items into one to aim for the development of a smart pipe operational solution. Firstly, the development involves receiving sensor data from smart pipes capable of self-diagnosis and state monitoring to create an integrated solution for efficient operation and monitoring, incorporating artificial intelligence technology to support intelligent decision-making. Secondly, through the development of self-diagnostic and state monitoring sensor system devices, a pipeline-integrated sensor system capable of monitoring performance and state changes such as physical, chemical, and structural alterations of conduits, as well as fluid transport characteristics, is developed. Thirdly, by integrating sensor system devices into the joints, manholes, and valve chambers of the pipelines to make them usable, the smart pipe is refined with the goal of enhancing performance compared to traditional pipes through performance and relevant verification, seismic resistance, corrosion resistance to chlorine, and longevity.

The specific objective of the sensor data-based deep learning model for anomaly detection in water supply pipes, focused on in this study, is to develop a digital twin-based integrated management system for efficient use of smart pipes. This involves collecting and analyzing various sensor data to automatically predict leaks and damages in water supply pipes through the development of an deep learning model[1].

II. Trends in Sensor Data based Anomaly Detection Models

An analysis of trends in sensor data anomaly detection models aimed to introduce applicable AI models. Verified models such as VGGNet, InceptionNet, ResNet, and TadGAN have been used in various multi-class classification and time-series data anomaly detection techniques[2,3]. It was suggested that models comparing original and reconstructed data could effectively detect anomalies. Lifelong learning, involving periodic model training with sensor data, was deemed suitable for learning comprehensive data features but posed challenges in training time and computational resources as data volume grows.

To enhance anomaly detection rate(performance) without training the model, a method utilizing AE(AutoEncoder)based RaPP(Reconstruction as Pre-Processing) was introduced[4]. Traditional AE consists of an Encoder, which compresses input data, and a Decoder, which reconstructs compressed data[5]. Through the Reconstruction process of reconstructing compressed data using the Decoder, the model performs anomaly detection by comparing the reconstructed output data with the input data using Reconstruction error. However, due to the structure of traditional AE, which primarily focuses on minimizing the difference between input and output, the performance may deteriorate if the hidden layers, excluding the input and output layers, are not fully utilized. Consequently, it was concluded that utilizing AE-based RaPP, which leverages intermediate layers, is favorable as following Fig.1.

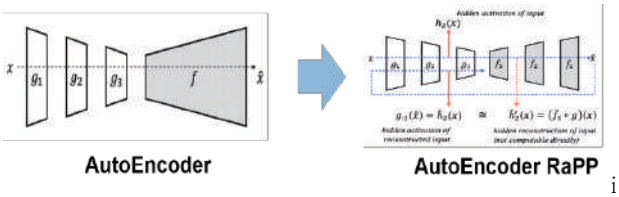


Fig. 1. AE-based RaPP

III. Model Construction and Performance Analysis

In order to construct a sensor data-based anomaly detection model, preliminary research was conducted using dummy sensor data and leakage detection data from upper and lower pipelines. Artificial intelligence training data capable of distinguishing leakage status based on leakage detection data obtained from sensors installed in water supply pipes were

supplied for research, and a study was conducted to classify five leakage statuses using a multi-class classification model. The fundamental idea of the artificial intelligence model involved utilizing multivariate sensor data and comparing input and output values using an LSTM Auto Encoder model as following Fig.2.

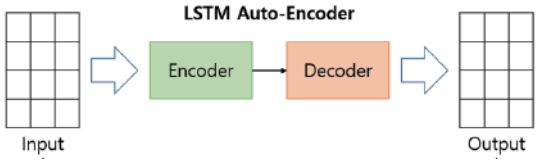


Fig. 2. LSTM Auto Encoder model

For model training, the following basic conditions were fixed: Learning Rate = 0.001, Batch_size = 128, Input_size = 40, Output_size = 40, Max_iter = 100,000. The Latent size represents the size of the hidden vector generated by a single LSTM unit; increasing the size results in a larger hidden vector output, while the Window size determines how many timestamps are grouped together as input, allowing for the use of multiple timestamps as a single input.

Data analysis was conducted by inputting training data into a data visualization module, and basic multi-class classification model was used to conduct CNN network-based artificial intelligence model training. Three types of models were experimented with: a classification model consisting of Fully Connected Layers, employing Dropout technique to prevent overfitting, using ReLU function as the Activation function, and varying the number of Hidden Layers during training. During training, the model's hyper-parameters were set as follows: Learning Rate = 0.001, Batch_size = 32, Num_epoch = 20, and the accuracy change of the trained model was as following Fig.3.

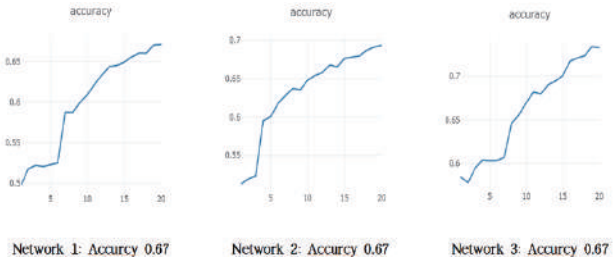


Fig. 3. The Accuracy Variation of the Training Model

Experimental results indicated that in the CNN network-based artificial intelligence model utilizing water supply pipeline leakage detection data, as the number of hidden layers increased, the accuracy tended to improve. Expanding the first experiment, an experiment was conducted to confirm the accuracy results when changing the number of Hidden Layers and Activation functions, with the same model training environment as the first experiment. Test cases included 10 experiments with a combination of 12, 15, 18, 21, and 24 Hidden Layers and two Activation functions: ReLU and Leaky ReLU.)

Experimental results showed that the combination of Hidden Layer = 21 and Activation function = ReLU exhibited the optimal performance among the combinations of Hidden Layers and Activation functions as following Fig.4. As the number of Hidden Layers increased, the accuracy tended to improve, but beyond a certain threshold, overfitting occurred, leading to a decrease in accuracy.

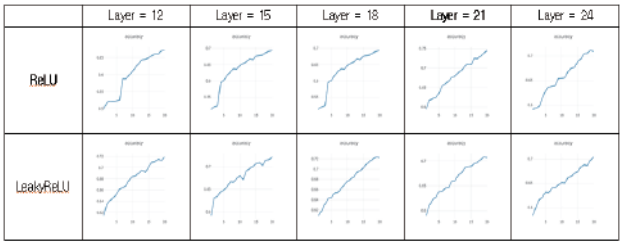


Fig. 4. The experimental results of combining Hidden Layers and Activation functions

IV. Conclusions

Four experiments were conducted using two sets of dummy sensor data. Anomaly detection utilizing CNN -based artificial intelligence models exhibited a tendency for accuracy to increase as the number of Hidden Layers increased. When using the ReLU activation function, better results were observed. However, blindly increasing the number of Hidden Layers led to overfitting, resulting in decreased accuracy. It was concluded that a case study should be conducted to find the optimal number of Hidden Layers depending on the training data or adjust the Learning rate hyper-parameter to improve performance in areas where overfitting occurs. For anomaly detection using LSTM Auto Encoder, adjusting the Window size showed the most significant impact on the model's performance among the three conditions: Latent size, Window size, and Layer num. Experimentation to find the optimal conditions using these three parameters revealed that the condition with Latent size = 20, Window size = 9, and Layer num = 2 exhibited the best performance. Similarly, it was concluded that a case study should be conducted to find the optimal conditions depending on the training data. Ultimately, it was observed that while the selection of model type is crucial, optimizing hyper-parameter combinations within the model is also highly important for both performance and training efficiency.

Acknowledgment

This work was supported by Korea Environment Industry & Technology Institute(KEITI) through Project for developing innovative drinking water and wastewater technologies Program, funded by Korea Ministry of Environment(MOE)(RE202101601).

References

[1] Y. H. Kim, J. Lee. “The Research of Smart Pipe Construction Location Recommendation Techniques Using Underground pipe information,” KIICE Conference Proceedings, vol. 27, no. 2, pp. 346-347

[2] Y. Kim and N. Kim, “Implementation of Integration Management System for Smart-pipe” in Proceedings of the 9th International Conference on Next Generation Computing, 2023.

[3] P. D. Hong, “A Real-time Monitoring Agent Design for Digital Twin-based Smart Pipe Integrated Management System” in Proceedings of the Korean Institute of Information and Commucation Sciences Conference, Gunsan, pp. 292-294, 2021.

[4] Ahmed, Mohiuddin, et al. “An investigation of performance analysis of anomaly detection techniques for big data in scada systems.” EAI Endorsed Trans. Ind. Networks Intell. Syst. 2.3 (2015): e5.

[5] Ethem Alpaydin, Introduction to machine learning, 4th ed., The MIT Press, 2020.

Towards Time-Aware Decentralized Federated Learning for Bearing Fault Detection

Made Adi Paramartha Putra*, Nengah Widya Utami*, I Gede Juliana Eka Putra*, Nyoman Karna†
*Faculty of Information Technology and Design, Primakara University, Denpasar, Indonesia.
†School of Electrical Engineering, Telkom University, Bandung 40257, Indonesia.
{adi*, widya*, gedejep*}@primakara.ac.id, aditya@telkomuniversity.ac.id

***Abstract*—Federated Learning (FL) presents a promising alternative, bolstering system reliability by decentralizing the learning process. In our research, we advocate for a Decentralized FL (DFL) approach to collaborative bearing fault detection. DFL stands out for its elimination of a single point of failure inherent in centralized FL setups. By embracing the decentralized FL paradigm, vulnerabilities to attacks in collaborative frameworks can be mitigated. Our proposed DFL model incorporates continual learning techniques to streamline communication and reduce overhead. Experimental results showcase the efficacy of decentralized collaborative learning, achieving a commendable accuracy rate of 96.08% and significantly reducing learning time by up to 37.52%.**

***Index Terms*—Decentralized federated learning; bearing fault detection; industrial iot;**

I. Introduction

Federated Learning (FL) introduces a fresh approach to AI model training within a decentralized framework, specifically tailored to address the constraints of centralized learning within the Industrial Internet of Things (IIoT) sphere. In FL, training is distributed among client devices, ensuring both privacy and the development of robust models without the necessity of sharing client data [1]. However, FL implementation must carefully consider factors such as device heterogeneity, access management, security, and data distribution [2].

In prior research, we introduced a technique for client selection driven by accuracy, prioritizing clients based on centralized evaluation [3]. Nonetheless, the centralized nature of FL, with the aggregation process managed by an FL server, poses vulnerability to attacks. Compromise of the FL server could exploit the FL process by adversaries. Hence, eliminating the FL server is advisable. This entails shifting the aggregation process to be conducted by each participant in the FL process. Decentralized FL (DFL) proves beneficial in this context, enabling participants to gather models from others and carry out local aggregation processes, which are then shared within the FL process itself.

Numerous studies have explored fault detection using the FL method. For instance, [4] discusses adaptive control for fault detection. While state-of-the-art studies focused on improving prediction accuracy, they often overlooked the consideration of overall training time. Therefore, this study incorporates time-aware DFL to provide an effective and efficient FL system in industrial environments. The key contributions of this paper are summarized as follows:

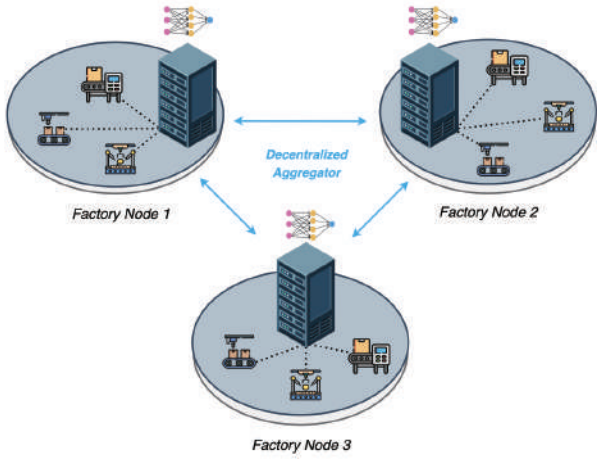


Fig. 1. The illustration of DFL in an industrial network incorporating multiple factories.

- We propose a time-aware DFL that takes into account the maximum allocation time for each participant to respond to the continual FL process.
- We conduct initial evaluations by comparing the accuracy and the total training time required to complete the FL process.

II. Proposed System

Let's denote the factories depicted in Fig.1 as n groups, represented by *F(N)* = 1, 2, 3, ..., n, where each factory depends on others to execute the DFL mechanism. In each communication round, the total number of aggregation processes can be determined by n, as every factory must complete aggregation before other participants can proceed with the learning process. The proposed DFL begins with factory node 1 initializing it. Subsequently, factory node 2 employs the updated local model to train based on its dataset, which can be represented as:

$$\omega_l^n \leftarrow \omega_g - \eta \nabla \ell \left(\omega_g; b\right). \tag{1}$$

Here, the updated local model ω_l^n from factory n is obtained using the global model ω_g and the training outcomes on a batch b of local data β, employing a learning rate η.

Once training is completed, factory node 2 initiates aggregation by merging parameters from factory node 1 with its own trained parameters to create a new aggregated parameter

TABLE I
THE SIMULATION PARAMETER CONFIGURATIONS FOR THE PROPOSED CONTINUAL DFL.

Parameter	Symbol	Value
Communication Round	t	50
Total Factory	K	10
Batch Size	b	32
Local Epochs	e	1
Threshold time (s)	τ	300
Aggregation Technique	n/a	FedAvg
Model Configuration	n/a	MLP

for the subsequent factory training. This aggregation process is represented by:

$$\omega_{g_{t+1}} \leftarrow \sum_{n=1}^{N} \frac{\eta\, n}{\eta} \omega_{l_{t+1}^n} \,. \tag{2}$$

This iterative process continues until all available factories have trained their datasets within the DFL communication rounds. It is important to note that each factory only has τ time to conduct the training process. If it fails, the subsequent factory will take over the training and exclude the node from the specific FL round.

The effectiveness of the proposed system was evaluated using an industrial IoT dataset, specifically bearing fault detection data sourced from Case Western Reserve University (CWRU) [5]. This dataset comprises information collected from a machinery motor equipped with a torque transducer, dynamometer, and control electronics. The data was recorded at two different sampling rates: 12,000 and 48,000 samples per second. The comprehensive parameter configuration for the continual DFL system is outlined in Table I.

III. RESULTS AND DISCUSSION

For the assessment phase, this study undertakes a comparison using the MLP model. Initially, the comparison is made between the centralized and decentralized approaches employing the MLP model. Figure 2 depicts the performance comparison of the two approaches explored in this study. It is evident from the illustration that the performance of centralized and decentralized FL is comparable. Through our simulations, we notice that the performance of CFL is marginally superior in the initial three communication rounds, whereas for subsequent federations, DFL attains higher accuracy.

Lastly, to illustrate the advancement of the proposed continual DFL, the evaluation emphasized learning time. The scenario was devised to simulate device failure and demonstrate the efficacy of the time threshold in this study. A total of 10 nodes were examined across 50 communication rounds, with the outcomes outlined in Table II. The results highlight the effectiveness of the proposed DFL in terms of total training time, showcasing a 37.52% enhancement. This improvement is reinforced by the threshold time, which disregards malfunctioning nodes within the DFL system.

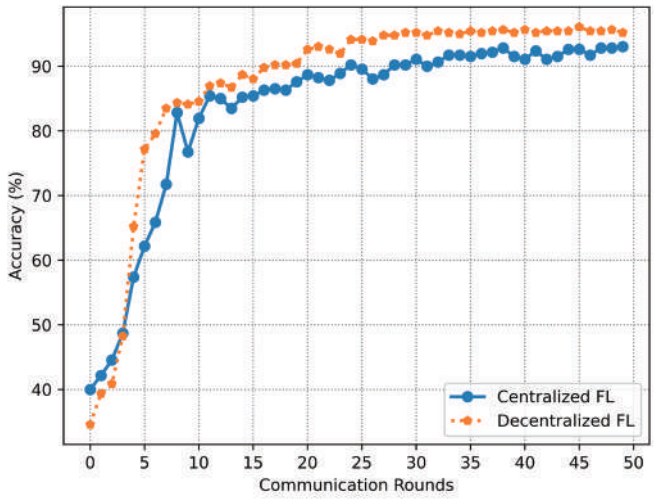


Fig. 2. The proposed continual DFL performance among two different approaches with 50 communication rounds.

TABLE II
TIME COMPARISON BETWEEN TRADITIONAL METHODS AND THE PROPOSED DFL.

Method	Total Training Time
Traditional DFL	1355.450 s
Proposed DFL	**846.931 s**

IV. CONCLUSION

This study presents the utilization of decentralized FL approach to remove single points of failure raises from centralized approach. The main focus in this work is the use case scenario in industrial environment, especially the bearing fault detection. The proposed method introduces threshold time for each factory to finish the training and update process. The results observed from CWRU dataset shows significant performance of the proposed method. The decentralized FL approach with threshold time achieving a satisfactory accuracy of 96.08% and significantly reduce the overall training time by up to 37.52%.

REFERENCES

[1] M. Aledhari, R. Razzak, R. M. Parizi, and F. Saeed, "Federated learning: A survey on enabling technologies, protocols, and applications," *IEEE Access*, vol. 8, pp. 140 699–140 725, 2020.

[2] L. Li, Y. Fan, and K.-Y. Lin, "A survey on federated learning," in *2020 IEEE 16th International Conference on Control & Automation (ICCA)*, 2020, pp. 791–796.

[3] M. A. P. Putra, A. R. Putri, A. Zainudin, D.-S. Kim, and J.-M. Lee, "Acs: Accuracy-based client selection mechanism for federated industrial iot," *Internet of Things*, vol. 21, p. 100657, 2023.

[4] S. Zhou, J. Bai, and F. Wu, "Decentralized fault detection and fault-tolerant control for nonlinear interconnected systems," *Processes*, vol. 9, no. 4, 2021. [Online]. Available: https://www.mdpi.com/2227-9717/9/4/591

[5] X. Chen, B. Zhang, and D. Gao, "Bearing fault diagnosis base on multi-scale CNN and LSTM model," *Journal of Intelligent Manufacturing*, vol. 32, no. 4, pp. 971–987, 2021. [Online]. Available: https://doi.org/10.1007/s10845-020-01600-2

Interactive Tablet Menu for Information-underprivileged Class

※ This study was conducted as a result of the SW-centered university support project of the Ministry of Science and ICT and the National IT Planning and Evaluation Institute (2021-0-01393)

Seo-il Han
Dept. AI Computer Engineering, Kyonggi University
Suwon, South Korea
xhdsxhd@kyonggi.ac.kr

Kwansang Na
Dept. AI Computer Engineering, Kyonggi University
Suwon, South Korea
ksna001@kyonggi.ac.kr

Yoon-Jong Lee
Dept. AI Computer Engineering, Kyonggi University
Suwon, South Korea
zoniblack@naver.com

Yoon-Seo Koh
Dept. AI Computer Engineering, Kyonggi University
Suwon, South Korea
icon0615@kyonggi.ac.kr

Min-Ju Jeon
Dept. AI Computer Engineering, Kyonggi University
Suwon, South Korea
mingmingmon@kyonggi.ac.kr

Tae-Yun Rho
Dept. AI Computer Engineering, Kyonggi University
Suwon, South Korea
iantyrho13@kyonggi.ac.kr

Han-Eum Lee
Dept. AI Computer Engineering, Kyonggi University
Suwon, South Korea
leehaneum160924@kyonggi.ac.kr

Su-Min Goo
Dept. AI Computer Engineering, Kyonggi University
Suwon, South Korea
tnals22@kyonggi.ac.kr

***Abstract*—This study deals with the digitalization trend of the food service industry and the accessibility problem of the information-underprivileged class. Mainly, the goal is to combine tablet menus and AI interactive chatbots to make it easier for users to order and choose food. To this end, voice input technology and AI models are used to interact with users and are specially designed to meet the needs of the information-underprivileged class. In addition, a backend system and an administrator webpage are implemented to support efficient service operation. This approach aims to promote social integration by making digital services more inclusive and having equal opportunities for all users.**

Keywords—Information-underprivileged Class, OpenAI, GPT, Chat Completions API, Whisper, Function Calling, Tablet Menu

I. INTRODUCTION

In recent years, the food service industry has undergone rapid digitization, with the utilization of tablet menus witnessing a significant surge. Particularly, there has been a notable increase in the adoption of tablet menu systems for ordering and payment, driven by the rise in contactless transactions and the quest for labor cost reduction, especially in the aftermath of the COVID-19 pandemic. However, alongside the proliferation of such technologies, there is a growing concern regarding accessibility for vulnerable populations. Recent studies indicate that a portion of the population, particularly those deemed information-vulnerable, refrains from using digital technologies due to difficulties in usage. For instance, a survey revealed that 33.8% of the information-vulnerable population, including the elderly, cited "lack of knowledge or difficulty in use" as the reason for their avoidance of device usage [1]. Additionally, an evaluation of 20 public and private sector kiosks in Seoul and Gyeonggi-do found that none of the devices supported Braille display or voice guidance technology for the visually impaired [2].
There are Braille kiosks as a solution to the problem, but products such as Braille kiosks for the visually impaired are expensive, making it burdensome for general restaurant industry to place them. Previous studies related to this only include kiosk-related products, such as 'Voice recognition ordering system for the disabled and the weak using kiosks' [3] and 'Design and implementation of order-payment system using artificial intelligence speakers' [4]. Moreover, the problem of system cost still exists.
Therefore, this study aims to improve accessibility and convenience for information-vulnerable groups and develop a tablet menu board that is easy to distribute. Related technologies that can be used include LLM (Large Language Model) for natural language conversation, STT (Speech To Text), and TTS (Text To Speech) for voice input and output. This study uses Assistant which is OpenAI's API as an appropriate FM (Foundation Model). The OpenAI Assistant used in this study can apply the functions of OpenAI's artificial intelligence model. It uses GPT (Generative Pre-trained Transformer) as an LLM, creates images through DALL-E, and provides STT and TTS functions through Audio API [5]. Additionally, they are easy to port directly to Android-based tablets since they support a variety of languages.

II. INTERACTIVE TABLET MENU DESIGN AND IMPLEMENTATION DIRECTION

A. App Design

The UI screen was constructed using XML, which makes it easy to visually understand the overall layout structure and design, and allows user interface elements to be defined in a hierarchical structure. Interaction between the app's business logic and user interface was implemented using Kotlin. Retrofit was used to simplify communication with the server and enable efficient processing of data requests and responses. Recycler View items are managed by ViewModel, and the screen is updated by detecting changes in data [6].

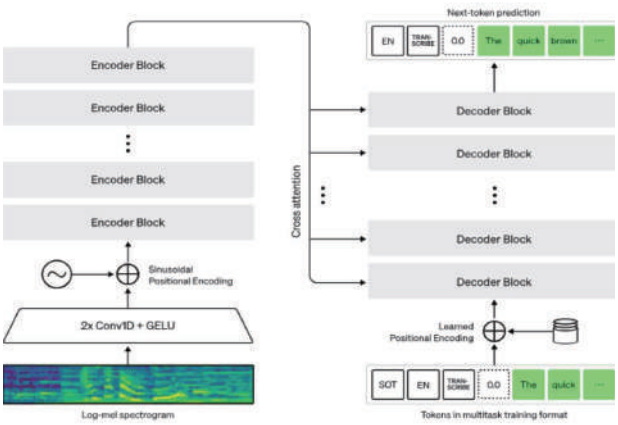
B. Voice Interactive Chatbot

This study was based on use in a virtual restaurant for ease of design. GPT-based interactive chatbots to be inserted into tablets are largely divided into voice input, store customized information output, menu selection, and ordering in implementation.

First, voice input uses Whisper provided by OpenAI. Whisper is an Automatic Speech Recognition (ASR) system trained on 680,000 hours of multilingual and multitasking supervision data collected from the web.
Whisper divides input audio into 30-second segments, converts them into Log-Mel spectrograms, passes them through an encoder-decoder stack to perform tasks such as language identification, syntactic-level timestamps, multilingual speech transcription, and English-to-foreign language speech translation. Despite Clover Speech, a Korean-specialized STT model, demonstrating superior performance in Korean. However, we opted for Whisper due to its relatively lower usage fees and better mobile portability.

Secondly, it's necessary to give tailored information for each store. For instance, when a menu request is received, only information about the food items sold at that specific store should be provided, while other details should not be disclosed. OpenAI's Chat Completions API allows us to provide guidelines for the Assistant's behavior or personality through parameters like 'message' or a JSON file, enabling us to deliver store-specific information to users effectively.

Fig. 1. Whisper Architecture [7]



Lastly, we apply OpenAI's Function Calling feature to implement menu selection and ordering. Function Calling allows the model to call one or more functions defined in the user's query and function parameters. Through this, we can call functions related to menu selection to facilitate menu browsing and ordering. Additionally, calling the function to retrieve the current cart items allows users to review their current order status.
This implementation enables users to interact with the chatbot solely through voice dialogue, allowing them to receive menu descriptions, make menu selections, and check orders without the need for a waiter.

C. Design and Implementation of Administrator Webpage

The technology stack employed in the webpage design is as follows. It was implemented to operate in a React environment with Client-Side Rendering (CSR) [8]. Tailwind CSS was utilized to implement a responsive webpage [9], while Daisy UI was introduced to create a user-centric and intuitive UI/UX environment [10]. Through this series of technologies, the system enables administrators to smoothly operate it.
Below are the features implemented using the technology stack. Key information can be easily monitored in real-time through the administrator page. Instead of the client periodically polling the server for information, Server-Sent-Event (SSE) communication is used, where the server forcefully pushes events. This reduces server load and conserves resources. Furthermore, order statistics can be analyzed to provide meaningful information to administrators. By providing various information necessary for restaurant operation, the business processes of owners (administrators) can be effectively automated.

D. Backend Implementation

For an effective system, Spring Boot, Spring Security, JWT (JSON Web Tokens), MySQL, and Swagger were selected to implement the backend system.
Spring Boot shortened the time required for initial setup and enhanced development productivity. Spring Security, combined with JWT (JSON Web Tokens), simplified the authentication process [11]. Moreover, by utilizing JWT, secure data transmission and authentication were ensured, and permissions were differentiated between information accessible only to store owners and information accessible to anyone, thereby strengthening authorization management [12]. The characteristic of Java, 'platform independence', ensured service stability [13]. MySQL leveraged the advantages of open source to reduce licensing costs and enable efficient data management. The use of Swagger alongside Spring Boot automatically generated API (Application Programming Interface) documentation, aiding in stable web and app integration [14].
These technologies were designed to consider backend system performance and scalability. With query optimization and cloud-based architecture, a more efficient server environment can be provided.

II. IMPLEMENTATION OF AI CHATBOT INTERACTIVER MENU

An invisible button exists on the initialization screen of the tablet. When the left area of the screen is touched, it switches to the regular tablet menu mode, and when the right area is touched, it switches to the AI chatbot mode. In the main screen of the regular tablet menu, users can view food photos and prices for each menu item received from the server. Upon selecting a food photo, the basket fragment is displayed. Users can increase, decrease, or remove the quantity of selected food on the screen. Upon pressing the order button, an order confirmation dialog appears, and after receiving user confirmation, the order is completed, and the list of ordered items is saved in the order history. In the order history, users can press the pay button, which displays a payment dialog, and after confirmation, the payment is completed. Additionally, the order history is reset, and the end-use screen is displayed for 5 seconds before returning to the initialization screen. Users can switch to the AI chatbot mode screen by pressing the AI chatbot button.

In the AI chatbot mode screen, users can input via voice and receive appropriate responses via voice and screen output. Users can switch to the regular tablet mode through the menu button on the current screen. This allows users to either use voice conversation features or place food orders in regular tablet menu mode according to their needs. Pressing the order history button displays it in the same visual formats the regular tablet menu mode.

IV. CONCLUSION

In this study, to address the accessibility issues faced by information-vulnerable groups in the food service industry, we combined artificial intelligence-based conversational interface (GPT) with voice-based interfaces. This system provides users with familiar and intuitive real-time conversational capabilities, thereby making it easier and more convenient for information-vulnerable groups to utilize digital services in the food service industry. As a follow-up study, it is anticipated that improving the accuracy of speech recognition due to surrounding noise and implementing more robust services by using auxiliary devices such as braille cases rather than just tablets will be possible. This will play an important role not only in improving services but also in enhancing social integration and creating an environment where all citizens can enjoy equal opportunities, going beyond simple service improvements.

REFERENCES

[1] S. M. Park. (2022). Key Findings of the 2021 Seoul Citizens' Digital Competency Survey. Seoul Digital Foundation. Report Publisher: Seoul Digital Foundation. p.21

[2] S. R. Ahn. (2022). Survey on the Utilization Status of Kiosks (Self-service Information Terminals). Report by the Market Surveillance Team of the Market Research Division at the Korea Consumer Agency. Publisher: Korea Consumer Agency. p.38

[3] H. J. Lee, C. H. Hong, S. H. Cho, C. W. Yoon, D. W. Kim, S. H. Choi. (2021). "Ordering system for the disabled and the weak using a KIOSK with speech recognition technology". Korea Institute of information and Communication Engineering Proceedings of the 2021 Spring Academic Conference. pp.544 - 546

[4] D. H. Kim, B. H. Kim, C. H. Ban. (2019). Design and Implementation of Order Settlement System Using Artificial Intelligence Speaker. The Journal of the Korea Institute of Electronic Communication Sciences, 14(6), pp.1181 - 1186.

[5] OpenAI API, https://platform.openai.com/docs/

[6] Github. (2024). testRepo. Github Repository, https://github.com/tn-als/testRepo

[7] OpenAI, Whisper, https://openai.com/research/whisper

[8] React, Client, https://react.dev/reference/react-dom/render

[9] Tailwind, Responsive Design, https://tailwindcss.com/docs/responsive-design

[10] DaisyUI, Components, https://daisyui.com/docs/config/

[11] H. S. Son. 『Implementation of framework for efficient system development』Master's thesis at Korea University. 2018. pp.6 - 8

[12] H. J. Kim. (2019). A Study on efficient authentication method using JWT in a electronic document sending system. Sungkyunkwan University. pp.11 - 30

[13] H. S. Son. 『Implementation of framework for efficient system development』Master's thesis at Korea University. 2018. p.3

[14] Swagger, OpenAPI, https://swagger.io/docs/specification/about/

Impact of Beam Split on BER Performance of IRS-aided THz Communications

Sanjeev Sharma[1], and Yeon Ho Chung[2]
[1]IIT (BHU) Varanasi, India (sanjeev.ece@iitbhu.ac.in) [2]PKNU, Busan, Korea (yhchung@pknu.ac.kr)

***Abstract*—Intelligent reflecting surfaces (IRS) and multiple-input and multiple-output (MIMO)-based design are considered for mitigating the effects of molecular absorption in terahertz (THz) systems operating at high-frequency bands. However, the phenomenon of beam splitting occurs at both the transceiver and IRS in the THz system. In this paper, time-delay (TD) and delay-phase (DP) based precoding at the transmitter (Tx) and IRS are introduced to overcome the effect of beam splitting. The average bit error rate (BER) of the THz system is analyzed by considering the impact of transmitting antennas, reflecting elements, and in-phase and quadrature (IQ) mismatch for a multi-carrier system. Beam splitting results in a reduction in BER performance of up to 20 dB, as observed in the study.**

***Index Terms*—Beam Split, BER Performance, IQ Mismatch, IRS-aided MIMO THz System, Time-Delay and Delay-Phase Precoding**

I. Introduction

Terahertz (THz) communication is considered a promising technology for the next-generation wireless networks due to viability of large bandwidth. THz communication can support the ultra-high data rate with ultra-massive connectivity in the wireless network for various applications such as holographic communication, immersive virtual reality, and digital twins. As compared to millimeter Wave (mmWave) (30-100 GHz) for 5G, THz band (0.1-10 THz) can support the 10 GHz or even more channel bandwidth for data transmission. However, the transmission of THz signals is significantly hindered by substantial propagation attenuation resulting from the exceedingly high carrier frequencies.

In THz communication, maintaining a line-of-sight (LoS) path is essential for successful data transmission. If the LoS path is blocked or severely faded, reliable communication becomes impossible in THz bands. Recently, intelligent reflecting surfaces (IRS) have been introduced to create virtual LoS paths in the network or manipulate the channel between the transmitter (Tx) and receiver (Rx) in THz communication. Furthermore, IRS has the capability to create additional signal propagation paths, effectively resolving blockage issues through proactive manipulation of the wireless environment. Therefore, leveraging IRS in THz communication presents an attractive prospect for the advancement of future 6G wireless systems. Similarly, massive multiple-input multiple-output (MIMO), which employs a large antenna array to produce directional beams with high array gains, offers a solution to mitigate the considerable signal attenuation inherent in THz communications. Thus, THz massive MIMO holds significant promise for future next-generation wireless communications. Thus, a substantial number of antennas at the Tx and a considerable number of elements at the IRS are utilized to generate highly directional and energy-focused beams. This strategy aims to counteract the significant signal propagation attenuation experienced in THz communications.

In IRS-aided MIMO THz (IMT) communication, the integration of extensive bandwidth, numerous IRS elements, and massive antennas leads to significant beam splitting. As a result, the conventional phase-shifter-based hybrid precoding architecture becomes ineffective. Therefore, the literature proposes a delay-phase (DP) precoding (frequency-dependent) architecture to alleviate the impact of beam splitting in IMT communication.

Contributions: This paper investigates the bit error rate (BER) performance of IMT communication. We compare various DP precoding techniques for IMT communication. Our observations indicate that double beam splitting (at IRS and Tx) can significantly reduce the BER performance of IMT communication. We analyze the impact of the number of IRS elements, Tx antennas, sub-carriers, IQ mismatch, and IRS location on BER. The presented results suggest that the large bandwidth of THz exacerbates the beam splitting effect and degrades the BER performance significantly.

Related Works: IRS-aided MIMO THz systems are analyzed in the literature [1–3]. Researchers are focusing on BER analysis [4], sumrate [1, 2, 5], signal detection [6, 7], beam split effect [8, 9], IRS-aided design [8, 9], coverage analysis [7], outage probability (OP), and channel modeling [6] in the THz communication. While beam split effects are commonly analyzed for sumrate analysis in the literature, their impact on BER analysis in IMT system design is not considered or has received less attention. A summary of works related to beam split effects is provided in Table I.

II. IMT Communication System Model

Tx is equipped with N_{t} transmit antennas, incorporating both a time-delay (TD) network and phase shifts (PS). Modulated data undergoes precoding by a digital precoder before being transmitted through the radio-frequency (RF) chains, as depicted in Fig. 1. The IRS panel comprises $N = N_1 \times N_2$ elements arranged horizontally and vertically. The direct LoS between Tx and Rx is obstructed due to severe fading or obstacles, with a virtual LoS established between them via the IRS. To mitigate the effect of beam splitting in IMT communication, frequency-dependent precoding is implemented at the Tx, while DP shifting is applied at the IRS panel. Rx employs maximum likelihood signal detection and is equipped with a single antenna.

The commonly employed orthogonal frequency division multiplexing (OFDM) scheme with M sub-carriers is consid-

TABLE I: A Summary of Representative Works on Beam Split Effect in THz System

Reference	Work	Contribution
[1, 2, 5]	Sumrate analysis	Beam split effect analysis and time-delay (TD) module design in THz
[10]	mean square error analysis	Near field channel estimation impact in THz
[11]	Array gain analysis	Beam forming design in THz
[12]	Sumrate analysis	Hybrid beamformers design using deep learning approach in THz
[13]	Sumrate analysis	Multiuser precoding design using deep learning in THz
[3]	Energy efficiency and sumrate analysis	Impact of beam split effect and possible TD based solution in THz
Proposed work	BER analysis	Impact of beam split and IQ mismatch in IRS-aided massive MIMO THz

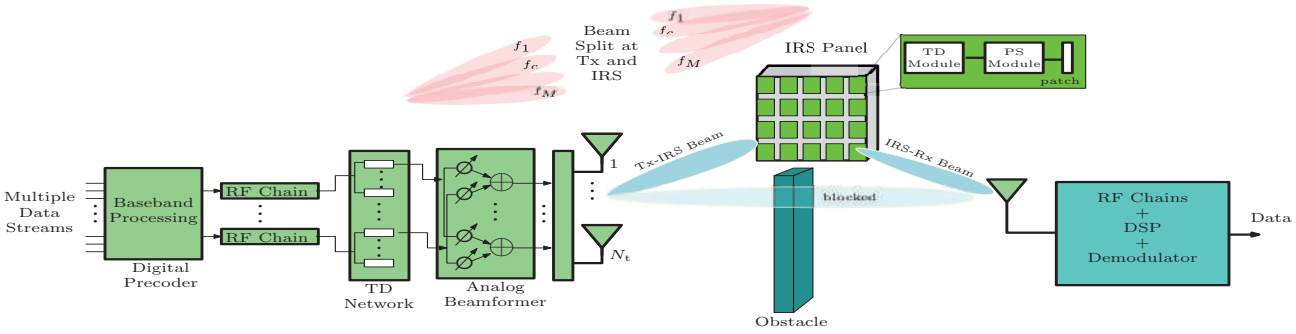


Fig. 1: An IRS-aided massive MIMO THz (IMT) communication system block diagram with DP and TD precoding.

ered. The received downlink signal, denoted as y_m at the mth sub-carrier ($m = 1, \ldots, M$), can be expressed as

$$y_m = \mathbf{h}_m^{\mathrm{T}} \boldsymbol{\Theta} \mathbf{G}_m \mathbf{w} x_m + n_m \in \mathbb{C}. \tag{1}$$

The expression $\mathbf{w}x_m \in \mathbb{C}^{N_\mathrm{t} \times 1}$ signifies the transmitted signal at the Tx. Here, $x_m \in \mathbb{C}$ denotes the transmitted information symbol at the mth sub-carrier, ensuring unit energy, i.e., $E[|x_m|^2] = 1$, while $\mathbf{w} \in \mathbb{C}^{N_\mathrm{t} \times 1}$ represents the precoding vector. Additionally, n_m represents the additive white Gaussian noise (AWGN) with zero mean and variance σ^2. The matrix $\mathbf{G}_m \in \mathbb{C}^{N \times N\mathrm{t}}$ represents the Tx-RIS channel response. The vector $\mathbf{h}_m^{\mathrm{T}} = [h_{m,1}, h_{m,2}, ..., h_{m,N}] \in \mathbb{C}^{N \times N_\mathrm{t}}$ denotes the channel from the IRS to the Rx at the mth sub-carrier.

The response of each IRS element to the incident signal is expressed as $a_n e^{j\theta_n}$, $n = 1, 2, ..., N$, where θ_n is the phase and a_n denotes the reflection coefficient. Thus, the overall IRS panel response can be written as

$$\boldsymbol{\theta} = [a_1 e^{j\theta_1}, a_2 e^{j\theta_2}, ..., a_N e^{j\theta_N}]^{\mathrm{T}}. \tag{2}$$

The IRS matrix is denoted as $\boldsymbol{\Theta} = \mathrm{diag}[\boldsymbol{\theta}]$, where the expression diag$[\boldsymbol{\theta}]$ represents a diagonal matrix constructed with elements selected from the vector $\boldsymbol{\theta}$.

Channel matrix $\mathbf{G}_m$ between Tx-IRS panel is

$$\mathbf{G}_m = g_1(f_m, d_1) e^{-\mathrm{J}2\pi\tau_1 f_m} \mathbf{a}(f_m, \eta_1^r, \zeta_1^r) \mathbf{b}^{\mathrm{H}}(f_m, \phi), \tag{3}$$

where τ_1 is the path delay from Tx to IRS. Path gain $g_1(f_m, d_1) \in \mathbb{C}$ is expressed as

$$g_1(f_m, d_1) = \frac{c}{4\pi f_m d_1} e^{-\frac{1}{2}\tau_{\mathrm{abs}}(f_m) d_1}, \tag{4}$$

where c denotes the speed of light, and d_1 is the distance from the Tx to the IRS panel. $\tau_{\mathrm{abs}}(f_m)$ is the molecular absorption coefficient at the mth sub-carrier frequency f_m. The elevation angle and the azimuth angle of the angle-of-arrival (AoA) of the RIS panel are denoted as η_1^r and ζ_1^r, respectively. The transmit angle of the Tx is denoted as ϕ. Furthermore, in the channel matrix, we assume that a LoS path exists between the Tx and the IRS panel.

The steering vector $\mathbf{a} \in \mathbb{C}^{N \times 1}$ of IRS panel having uniform planar array (UPA) is expressed as

$$\mathbf{a}(f_m, \eta, \zeta) = \frac{1}{\sqrt{N_1 N_2}} \left[1, ..., e^{j2\pi \frac{f_m}{c}(N_1 - 1) d \sin(\eta) \cos(\zeta)}\right]^{\mathrm{T}} \otimes \left[1, ..., e^{j2\pi \frac{f_m}{c}(N_2 - 1) d \sin(\eta) \cos(\zeta)}\right]^{\mathrm{T}} \tag{5}$$

The steering vector $\mathbf{b}(f_m, \phi) \in \mathbb{C}^{N_\mathrm{t} \times 1}$ of the uniform linear array (ULA) is expressed as

$$\mathbf{b}(f_m, \phi) = \frac{1}{\sqrt{N_\mathrm{t}}} \left[1, e^{j2\pi \frac{f_m}{c} d \sin(\phi)}, ..., e^{j2\pi \frac{f_m}{c}(N_\mathrm{t} - 1) d \sin(\phi)}\right]^{\mathrm{T}}.$$

Similarity, channel vector $\mathbf{h}_m$ between IRS and RX is expressed as

$$\mathbf{h}_m = g_2(f_m, d_2) e^{-\mathrm{J}2\pi\tau_2 f_m} \mathbf{a}(f_m, \eta_2^t, \zeta_2^t) \in \mathbb{C}^{N \times 1}, \tag{6}$$

where τ_2 is the path delay from IRS to Rx. Path gain $g_2(f_m, d_2)$ is expressed as

$$g_2(f_m, d_2) = \frac{c}{4\pi f_m d_2} e^{-\frac{1}{2}\tau_{\mathrm{abs}}(f_m) d_2}, \tag{7}$$

where d_2 is the distance from the IRS panel to the Rx. η_2^t and ζ_2^t are the elevation angle and the azimuth angle of the angle-of-departure (AoD) at the IRS panel, respectively. These angles satisfy $\eta \in \mathbb{U}(0, \pi/2)$, $\zeta \in \mathbb{U}(-\pi, \pi)$, and $\phi \in \mathbb{U}(-\pi/2, \pi/2)$, where $\mathbb{U}[a, b]$ denotes the uniform distribution between a and b.

A. Delay-phase Precoding at IRS

In the DP architecture, the frequency-dependent reflection coefficient of the nth IRS element is characterized by $e^{j(\theta_n - 2\pi f_m \tau_n)}$, where τ_n represents the time delay. By appropriately adjusting the time delay of each element, the beams

produced across various carriers can be synchronized with the desired target or physical direction. This synchronization effectively eradicates the beam split phenomenon in IMT communication. Nonetheless, the utilization of TD modules typically entails significantly greater power consumption and hardware expenses when contrasted with systems employing solely phase shifters [1].

To mitigate high hardware complexity and power consumption, a sub-connected architecture for IRS is preferred. Each component within the sub-array is equipped with dual-layer phase shifters and is connected to a shared TD module via a circulator [1]. Initially, the signal received at each component passes through the first-layer phase shifter, aiming to create constructive interference within the sub-array. Next, the signal is adjusted by the shared TD module. Finally, the signal passes through the second-layer phase shifter to complete the beamforming process of the IRS.

Let $Q_1 \times Q_2$ be the sub-arrays of an IRS panel, with each sub-array connected to a TD module. Each sub-array consists of $K_1 \times K_2$ elements, where $K_1 = N_1/Q_1$ and $K_2 = N_2/Q_2$. The index for elements in each sub-array is denoted as $k = (k_1, k_2)$, where $k_1 = 1, 2, ..., K_1$ and $k_2 = 1, 2, ..., K_2$. Each sub-array is identified by index $q = (q_1, q_2)$, where $q_1 = 1, 2, ..., Q_1$ and $q_2 = 1, 2, ..., Q_2$. Channel models are expressed as [1]

$$\begin{aligned}\mathbf{G}_m &= \frac{g_1(f_m, d_1)}{\sqrt{N_1 N_2}} e^{-\mathrm{J}2\pi\tau_1 f_m} \mathbf{F}(f_m, \eta_1^r, \zeta_1^r) \\ \mathbf{H}_m &= \frac{g_2(f_m, d_2)}{\sqrt{N_1 N_2}} e^{-\mathrm{J}2\pi\tau_2 f_m} \mathbf{F}(f_m, \eta_2^t, \zeta_2^t),\end{aligned} \tag{8}$$

where steering matrix $\mathbf{F}$ is expressed as

$$\mathbf{F}_{n_1,n_2}(f_m, \eta, \zeta) = e^{\pi\zeta_m(n_1 \sin\eta\cos(\zeta) + n_2 \sin\eta\sin(\zeta))}. \tag{9}$$

Here, $n_1 = q_1 K_1 + k_1$ and $n_2 = q_2 K_2 + k_2$. Phase shift matrices ($\mathbf{\Theta}_1^q, \mathbf{\Theta}_2^q$ representing the first and second layers), time delay vector $\mathbf{p}_m$, and channel matrices ($\mathbf{G}_m^q, \mathbf{H}_m^q$) are generated for each sub-array. The received signal y_m^q, after traversing from the transmitter, first-layer, and second-layer phase shifters of the IRS panel, can be formulated as [1, 9]

$$y_m^q = \mathrm{sum}\left(\mathbf{H}_m^q \odot \left(\mathbf{\Theta}_2^q p_m^q \left[\mathrm{sum}\left(\mathbf{\Theta}_1^q \odot \mathbf{G}_m^q x_m\right)\right]\right)\right), \tag{10}$$

where the TD vector $\mathbf{p}_m$ is

$$\mathbf{p}_m = e^{-j2\pi f_m \mathbf{t}}, \quad \mathbf{t} = [t_1, t_2, ..., t_q, ..., t_{Q_1 \times Q_2}]^{\mathrm{T}}.$$

B. Time-Delay Precoding at Tx

Suppose the Tx is partitioned into P sub-arrays, each linked to a TD module. Let $t_{\mathrm{tx}} \in \mathbb{R}$ represent the TD, and $\mathbf{w}_{u,p}$ denotes the phase shift of the pth sub-array ($p = 1, ..., P$) at the Tx. Consequently, the frequency-dependent transmit precoding of the Tx at the mth sub-carrier can be described as $\mathbf{W}_m = \mathbf{W}_u \mathbf{p}_m^{\mathrm{tx}} = \mathrm{blkdiag}\left([\mathbf{w}_{u,1}, ..., \mathbf{w}_{u,P}]\right) . e^{-j2\pi f_m [t_1^{\mathrm{tx}}, t_2^{\mathrm{tx}}, ..., t_P^{\mathrm{tx}}]^{\mathrm{T}}}$, where blkdiag denotes the block diagonal matrix. Further, the TD vector $\mathbf{p}_m^{\mathrm{tx}}$ and the phase shifting matrix $\mathbf{W}_u$ adhere to

$$p_p^{\mathrm{tx}} = \frac{(p-1)(\zeta_m - 1) N_{\mathrm{t}} \sin(\phi)}{2 f_m P}, \text{and } [\mathbf{w}_{u,1}^{\mathrm{T}}, ..., \mathbf{w}_{u,P}^{\mathrm{T}}]^{\mathrm{T}} = \mathbf{b}(f_m, \phi).$$

III. Results and Discussions

We analyze the average BER performance of the IMT communication system. Total system bandwidth is B Hz and sub-carriers are generated as

$$f_m = f_c + \frac{B}{M}\left(m - 1 - \frac{M-1}{2}\right) \quad \text{Hz}$$

Here, f_c denotes the carrier frequency, and in simulations, we consider f_c=100 GHz. The molecular absorption coefficient is $\tau_{\mathrm{abs}}(f_m) = 5.157 \times 10^{-4}$ dB/m. The noise power is assumed to be $\sigma^2 = -95$dBm. Perfect knowledge of the channel state information is assumed in simulations.

Fig. 2 illustrates the impact of beam split on BER performance at the IRS panel in the IMT system, considering average gain per sub-carrier. Optimal precoding denotes that each element has a separate DP unit, while IRS precoding denotes that only one DP module is attached to a sub-array of IRS elements. Conventional IRS design uses only frequency-independent precoding, here by considering center frequency only. Conventional IRS design exhibits degraded performance compared to optimal and IRS precoding methods due to beam split effect.

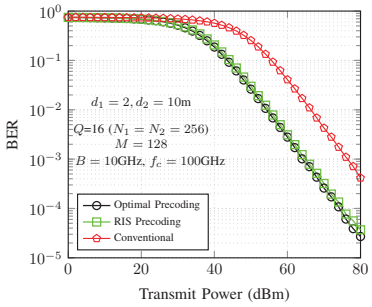


Fig. 2: Average BER performance of IMT system for $N_{\mathrm{t}} = 1$ and TD modules=16 (N=256 × 256).

In Fig. 3, the average gain per sub-carrier of the IMT system is plotted for $M = 128$. It can be observed that the gain of the conventional precoding scheme reduces as we move away from the center frequency of the beam due to the beam split effect. Optimal precoding demonstrates almost similar gain for all the sub-carriers, as observed in Fig. 3, due to DP precoding at the IRS and TD at the transmitter to compensate for the beam split phenomenon in the IMT. Only the IRS or the Tx exhibits improved gain compared to the conventional scheme, as the TD scheme is utilized either at the IRS or the Tx. Therefore, employing frequency-dependent precoding at both the Tx and IRS is essential in IMT.

The impact of the number of transmit antennas on BER performance is illustrated in Fig. 4, considering the average channel gain per sub-carrier. Optimal precoding exhibits the best performance due to the reduced beam split effect in the IMT system. TD-based compensation is used in optimal, RIS, and BS precoding methods. The beam split effect is higher for $N_{\mathrm{t}} = 64$ compared to $N_{\mathrm{t}} = 16$. Additionally, the impact of IRS elements on BER performance is illustrated in

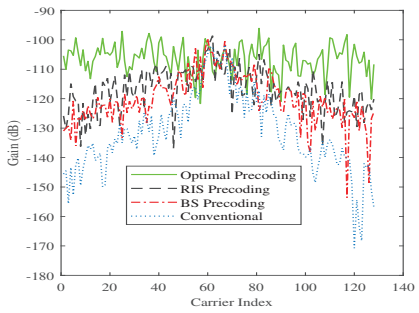


Fig. 3: Gain vs sub-carrier index plot in IMT system with $N_{\mathrm{t}} = 256$ and TD modules=16 (N=256 × 256).

Fig. 5, considering $N_{\mathrm{t}} = 256$. The number of TD modules significantly affects the system performance, as observed in Fig. 5.

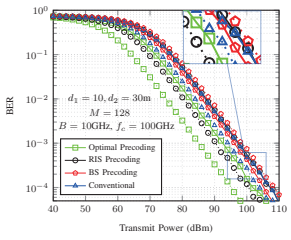


Fig. 4: Average BER performance of IMT system for $Q = 8(N{=}64 \times 64)$ with TD modules=8. Solid and dotted lines denote the performance for $N_{\mathrm{t}} = 16$ and $N_{\mathrm{t}} = 64$ antennas, respectively.

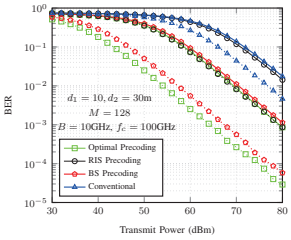


Fig. 5: Average BER performance of IMT system for $N_{\mathrm{t}} = 256$ with TD modules=16. Solid and dotted lines denote the performance for $Q = 16$ (N=256 × 256) and $Q = 8(N{=}64 \times 64)$, respectively.

A. IQ Mismatch

At higher carrier frequencies, the IQ mismatch significantly degrade the performance of IMT systems. IQ mismatch transmitted signal is expressed as

$$x_m^{\mathrm{iq}} = \alpha_1 x_m + \alpha_2 x_m^{\star}$$

where $x_m^{\star}$ is the complex conjugate. IQ imbalance coefficients such that $\alpha_1 + \alpha_2^{*} = 1$, and are given as

$$\alpha_1 = 0.5(1 + g_{\mathrm{t}} e^{-j\phi_{\mathrm{t}}}), \text{ and } \quad \alpha_2 \quad = 0.5(1 - g_{\mathrm{t}} e^{j\phi_{\mathrm{t}}}),$$

where g_{t} and ϕ_{t} are transmitter gain and phase mismatch, respectively. The impact of IQ mismatch at the Tx on BER performance is analyzed in Fig. 6. The BER performance deteriorates as IQ mismatch increases, ranging from 1 dB to 10 dB, as observed in Fig. 6.

IV. Conclusion

We have examined the BER performance within IMT communication. Our study investigated the implications of beam splitting in THz communication and compares various TD precoding methods in IMT communication. We further analyzed the effects of factors such as the number of IRS elements, Tx antennas, IQ mismatch, and IRS location on BER. Furthermore, our analysis highlights that the wide bandwidth of THz exacerbates the beam splitting effect, leading

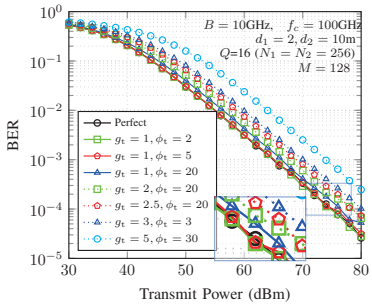


Fig. 6: Average BER performance of IMT system using optimal precoding scheme for $N_{\mathrm{t}} = 1$ and TD modules=16 (N=256 × 256).

to degradation in BER performance. However, our results demonstrate that the judicious application of DP precoding can effectively mitigate the beam splitting challenge within IMT communication, as evidenced in results.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the the Korean Government (MSIT 2023R1A2C2006860)

References

[1] R. Su, L. Dai, and D. W. K. Ng, "Wideband precoding for RIS-aided THz communications," *IEEE Trans. on Commun.*, vol. 71, no. 6, pp. 3592–3604, 2023.

[2] J. Gao, C. Zhong, G. Y. Li, J. B. Soriaga, and A. Behboodi, "Spatially sparse precoding in wideband hybrid terahertz massive MIMO systems," *IEEE Trans. on Wireless Commun.*, 2023.

[3] J. Tan and L. Dai, "THz precoding for 6G: challenges, solutions, and opportunities," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 132–138, 2022.

[4] V. Sharma, S. Sharma, and V. Bhatia, "Performance analysis of intelligent reflecting surface assisted MIMO terahertz system," in *2022 IEEE Int. Conf. on Adv. Netw. and Telecommun. Syst. (ANTS)*. IEEE, 2022, pp. 1–5.

[5] S. Kim, J. Park, J. Moon, and B. Shim, "Fast and accurate terahertz beam management via frequency-dependent beamforming," *IEEE Trans. on Wireless Commun.*, 2023.

[6] H. Du, J. Zhang, K. Guan, D. Niyato, H. Jiao, Z. Wang, and T. Kürner, "Performance and optimization of reconfigurable intelligent surface aided THz communications," *IEEE Trans. on Commun.*, vol. 70, no. 5, pp. 3575–3593, 2022.

[7] M. H. Kumar, S. Sharma, K. Deka, and M. K. Sharma, "RIS-assisted user pairing OMA system for THz communications," in *2023 Nat. Conf. on Commun. (NCC)*. IEEE, 2023, pp. 1–6.

[8] Q. Xue, C. Ji, S. Ma, J. Guo, Y. Xu, Q. Chen, and W. Zhang, "A survey of beam management for mmwave and THz communications towards 6G," *IEEE Commun. Surv. & Tut.*, 2024.

[9] G. Sun, W. Yan, W. Hao, C. Huang, and C. Yuen, "Beamforming design for the distributed RISs-aided THz communications with double-layer true time delays," *IEEE Trans. on Veh. Technol.*, 2023.

[10] S. Tang, Z. Dong, W. Zhou, D. Cai, E. Zhou, and P. Lan, "Beam splitting sensing for terahertz near-field channel estimation," in *2023 3rd Int. Conf. on Intell. Commun. and Comput. (ICC)*. IEEE, 2023, pp. 141–146.

[11] Y. Jiang, Z. Zhou, X. Li, and Y. Gong, "Beamforming design for RIS-aided THz wideband communication systems," in *2023 IEEE 24th Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*. IEEE, 2023, pp. 296–300.

[12] A. M. Elbir, K. V. Mishra, and S. Chatzinotas, "Terahertz-band joint ultra-massive MIMO radar-communications: Model-based and model-free hybrid beamforming," *IEEE J. of Sel. Topics in Signal Process.*, vol. 15, no. 6, pp. 1468–1483, 2021.

[13] Z. Liu, H. Ke, Y. Zhang, X. Zhao, Y. Liu, and M. Jin, "A deep learning-based hybrid precoding with attention mechanism for THz massive MU-MIMO systems," in *ICC 2023-IEEE Int. Conf. on Commun.* IEEE, 2023, pp. 5639–5644.

Analysing of 90-100 GHz Radar using Triangular Trihedral Corner Reflector for Airport Runway Safety

Nur Aqilah Yusri
School of Electrical Engineering
Universiti Teknologi Malaysia
81300, Johor Bharu, Malaysia
nuraqilahyusri96@gmail.com

Sevia Mahdaliza Idrus Sutan Nameh
School of Electrical Engineering
Universiti Teknologi Malaysia
81300, Johor Bharu, Malaysia
sevia@utm.my

Sumiaty Ambran
Malaysia-Japan International Institute Of Technology, UTM Kuala Lumpur
54100, Kuala Lumpur, Malaysia
sumiaty.kl@utm.my

Muhammad Al Farabi Muhamad Iqbal
School of Electrical Engineering
Universiti Teknologi Malaysia
81300, Johor Bharu, Malaysia
alfrabi@utm.my

Nobuhiko Shibagaki
Hitachi Kokusai Electric Inc Tokyo, Japan
shibagaki.nobuhiko@h-kokusai.com

Shunichi Futatsumori
Electronic Navigation Research Institute (ENRI), Tokyo, Japan
futatsumori@mpat.go.jp

***Abstract*—Triangular trihedral corner reflectors (TTCR) are passive radar targets that provide the best solution in characterizing the radar system performance. The Radar Cross Section (RCS) of the TTCR plays an essential role for estimating the characterizing parameters and hence the backscattering properties. RCS characterization for three sizes of TTCR considering airport runway at clear weather condition is addressed. This work introduces an experimental measurement of 90-100 GHz millimeter-wave radar detection. These measurements were carried out at runway 2, Kuala Lumpur International Airport for 93.1GHz using Frequency Modulated Continuous Wave (FMCW) radar. The characterization method for radar detection system performance considering radar incident angle and transverse slope, the reflectivity characteristic and experimental site measurement and findings are presented in this paper.**

Keywords—triangular trihedral corner reflector, radar detection, FMCW radar, radar cross section, airport runway.

I. INTRODUCTION

In today's public safety and defence, Foreign Object Debris (FOD) detection system is entering a new phase at airport operation. Any object that found in an inappropriate location at the airport that may harm the aircraft or airfield personnel are known as FOD [1]. The economic cost of FOD to the airline industry has been estimated over three to four billion USD per year [2][9]. FOD has emerged as one of the most serious threats to the aviation sector and needs to be removes instantly. According to the International Civil Aviation Organization's (ICAO) recommendations, runways must be explicitly inspected four times a day [3]. However, this inspection procedure is not applicable since manual inspection often results in higher time consumption and cost [11][16]. Hence, automated FOD inspection and detection system that can perform 24/7 detection is a promising solution. For example, a Frequency-Modulated Continuous Wave (FMCW) radar integrated with a millimetre-wave has been proposed in [4][5][6]. Advantages of a millimetre-wave radar include high detection performance, high resolution, and high weather resistance [6], [12-14].

While conducting the evaluation on a real airport, the surrounding environment must be considered to prevent deterioration in the performance of the FOD detection system by receiving unwanted signals. The environment of the airport, influenced by the existing ground structure and runway slope may affect the backscattered signal of RCS and interrupt the airport operation in detecting debris. The transverse runway slope is designed to promote water drainage during heavy rain and according to recommendation of ICAO Annex 14 is given 1.5 per cent for the runway width of 60 meters. The incident angles from the radar to the runway can further be measured to compute radar echo of FOD target properties study. The triangular trihedral corner reflector (TTCR) was chosen as the passive radar target because of its common design, rigid structure and designability which is suitable for higher radar frequency band [7]. The target material is made from aluminium material and promotes a good reflectivity which specifically give higher backscatter reflection. Additionally, in radar-based system, FOD detection can be accomplished by analysing the radar reflection [10].

In this paper, we conducted a characterization evaluation method of 90-100 GHz millimetre wave at runway 2, Kuala Lumpur International Airport (KLIA). A simple target of triangular trihedral corner reflector with relatively three sizes 0.1 m, 0.2 m and 0.3 m is proposed. This study will do an evaluation and analysis of the RCS of different size of corner reflector placed at different location along the runway. All measurements were taken under clear weather to prevent weakening of the echo signal. This paper is further divided into four sections. Section 2 describes the experimental work. In Section 3, measurement results are explained and discussed. Lastly, section 4 concluded the paper.

II. EEPERIMENTAL MEASUREMENTS

The FOD characterizing method was perform on a busy airport at runway 2, Kuala Lumpur International Airport (KLIA) involving the radar antenna unit 2-3 (RAU 2-3) tower, runway slope and triangular trihedral corner reflector (TTCR) that act as the FOD target for optimum small object detection. The FMCW radar detection system on RAU 2-3 operates at 93.1GHz. The RAU 2-3 tower is built 187 meters away from the runway centreline considering the existing ground structure and slope. For the measurement site, the FOD under tested was placed 90° aligned from the RAU 2-3 tower. Nine identical-sized of triangular trihedral corner reflector was

positioned in a 3x3 arrangement with a 10-meter interval as illustrate in Fig 2. The nine-placement points are labelled as P1 to P9 and was measured once at a time. The location P1 to P3 are placed near RAU 2-3 and P4 to P6 are placed on the runway centreline along with P7 to P9 are placed away from RAU 2-3 tower. Placement point label P5 was set as the centre point. The first measurement starts with a 0° angle of trihedral reflector facing the radar. The radar then scans 360 degrees to collect data. The TTCR are then rotated for twelve turning points from 30° to 330° as explain in Fig 2. The experiment is repeated for three sizes of TTCR. Fig 1. shows the experiment site of RAU 2-3 and positioning TTCR for all measurement in satellite map view.

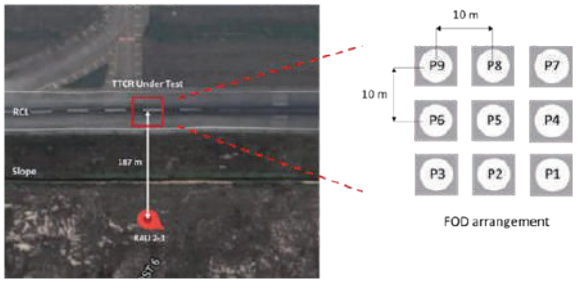


Fig. 1. Experiment site of RAU 2-3 and positioning TTCR for all measurement in satellite map view

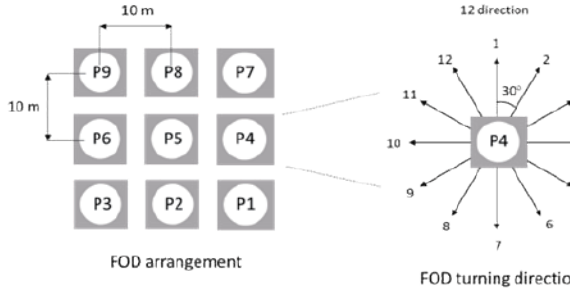


Fig. 2. FOD arrangement and twelve turning points.

The measurement emphasizes the nine different placements to measure the radar cross section (RCS) of the TTCR versus the runway slope. Radar cross section is measured by the backscatter signal to the radar receiver, with respect to incident radar power density and generally depends on the frequency of operation, aspect angle and polarization [8]. However, measuring the reflected signal of the trihedral reflector depends on the wavelength and the edge length of reflector. Thus, in this project three different sizes of trihedral reflector were selected 0.1 m, 0.2 m, and 0.3 m respectively. The measurement site from the side view of RAU 2-3 tower and placement of FOD target are illustrated in Fig 3. The arrangement for three range placements of TTCR on runway are 177 m, 187 m, and 197 m from the RAU 2-3 tower. The characterization evaluation is conducted to validate the RCS value of TTCR as optimum small object detection against the incident radar power density. Fig 3 shows the TTCR measurement setup on the runway from side view.

III. Results and Discussion

The characterization evaluation for a proper RCS measurement conducted for different sizes of triangular trihedral corner reflector was completed. The RCS

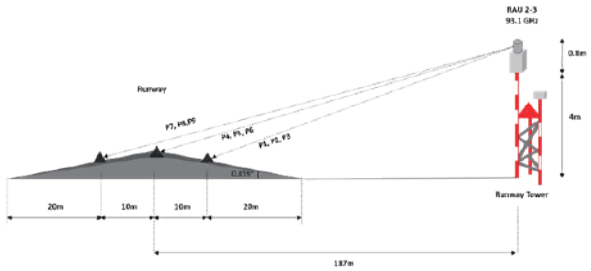


Fig. 3. TTCR measurement setup on runway (side view).

measurement was analysed, and data was tabulated in Table I, II, and III for three sizes of TTCR, respectively. The value of RCS measurement was observed increased from 177 m (P1, P2, P3) to the runway centreline at 197 m (P4, P5, P6) and decreases when distance is measured at 197 m from RAU 2-3 (P7, P8, P9). Some measurement data was found incoherent when turning the TTCR at angle 90°, 270° and 300° but only affected on 0.1 m and 0.2 m sizes of trihedral reflector. Due to twelve turning angle of TTCR, the size and shape varies from the radar point of view.

TABLE I. RCS MEASUREMENT VALUE FOR 0.1 M OF TTCR AT EACH SELECTED POSITION.

FOD Distance to RAU 2-3	Average RCS at 12 Turning Point (dBsm)											
	0°	30°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°
177 m (P1, P2, P3)	-30.29	-43.21	-64.62	-76.45	-63.10	-52.18	-61.44	-57.38	-60.37	-60.37	-63.45	-34.16
187 m (P1, P2, P3)	-26.25	-36.02	-73.69	-74.23	-70.50	-56.83	-67.09	-55.31	-58.02	-77.92	-72.97	-36.36
197 m (P1, P2, P3)	-36.67	-40.76	-62.66	-78.15	-66.87	-48.98	-60.89	-53.61	-60.27	-77.71	-75.59	-37.47

TABLE II. RCS MEASUREMENT VALUE FOR 0.2 M OF TTCR AT EACH SELECTED POSITION.

FOD Distance to RAU 2-3	Average RCS at 12 Turning Point (dBsm)											
	0°	30°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°
177 m (P1, P2, P3)	-20.97	-32.92	-58.39	-71.74	-61.85	-49.63	-62.56	-58.06	-63.47	-70.27	-72.93	-35.59
187 m (P1, P2, P3)	-15.51	-25.86	-62.34	-77.29	-72.15	-59.93	-61.25	-58.08	-53.42	-68.25	-75.53	-32.89
197 m (P1, P2, P3)	-19.91	-28.40	-74.39	-71.93	-59.10	-51.51	-61.73	-60.83	-60.92	-54.99	-71.31	-33.78

TABLE III. RCS MEASUREMENT VALUE FOR 0.3 M OF TTCR AT EACH SELECTED POSITION.

FOD Distance to RAU 2-3	Average RCS at 12 Turning Point (dBsm)											
	0°	30°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°
177 m (P1, P2, P3)	-36.87	-40.71	-40.77	-39.54	-39.57	-37.57	-36.30	-36.94	-39.38	-36.37	-41.58	-34.66
187 m (P1, P2, P3)	-40.57	-40.62	-43.36	-44.49	-42.07	-34.90	-44.96	-40.25	-35.67	-49.10	-49.00	-38.61
197 m (P1, P2, P3)	-36.69	-40.98	-53.12	-41.30	-40.19	-39.23	-41.90	-38.17	-35.57	-42.59	-52.26	-43.99

The average RCS value for all sizes of trihedral reflector considering the twelve turning point and distance from RAU 2-3 was plotted in a radar chart form. As shown in Figure 4, 5 and 6 each size of the TTCR on selected location created a pattern. The plotted graph demonstrates a pattern between the RCS values toward the position of FOD on the runway slope. Despite from that, each size of trihedral reflector for measurement distance 177 m (P1, P2, P3), 187 m (P4, P5, P6) and 197 m from RAU 2-3 (P7, P8, P9) shows almost identical pattern but increased in size of reflector at 0.3 m gives significantly different pattern than 0.1 m and 0.2 m sizes. Besides, it was observed that the evaluation gives a concise pattern in determining the FOD target position. From the radar chart, the RCS value for measurement distance 177 m (P1, P2, P3), 187 m (P4, P5, P6) and 197 m from RAU 2-3 (P7, P8, P9) show insignificant different but concise pattern. It is clearly noticed that the runway slope and existing ground structure can impact the radar detection performance.

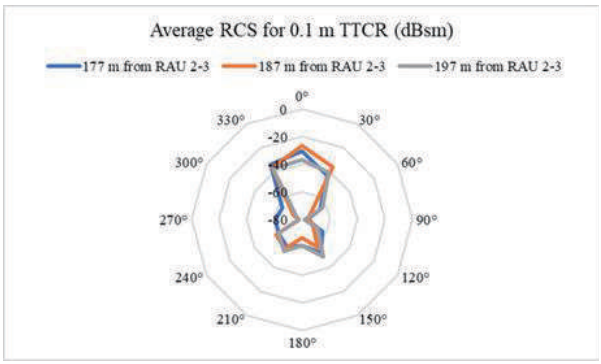


Fig. 4. Radar chart of average RCS for 0.1 m TTCR.

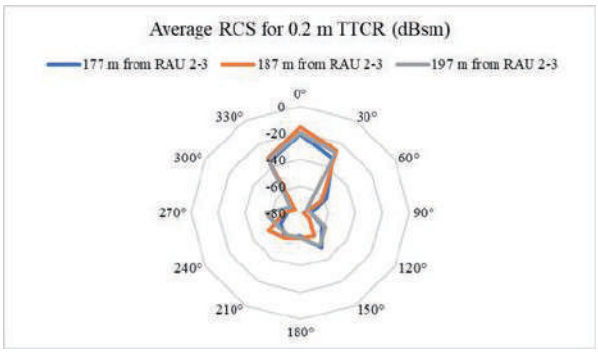


Fig. 5. Radar chart of average RCS for 0.2 m TTCR

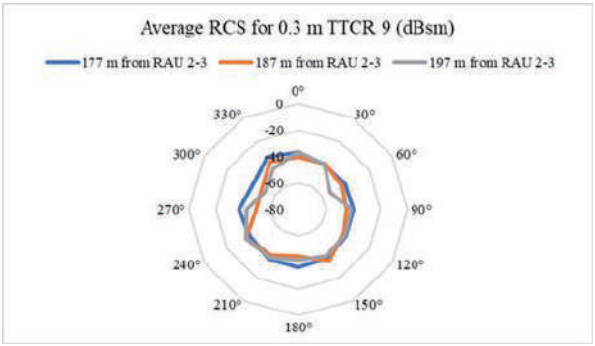


Fig. 6. Radar chart of average RCS for 0.3 m TTCR

IV. Conclusion

This work presents an RCS measurement for the radar detection system performance considering the radar incident angle and transverse slope at a physical airport runway by performing a FMCW radar characterization method. The triangular trihedral corner reflector that was selected as the FOD target is vital in estimating the RCS parameters and backscattering properties. Therefore, the characterization evaluation of trihedral reflector was performed at nine positions on the runway with twelve turning angles. Then a radar chart pattern was created from the characterization evaluation. The finding has proved that runway slope can affect the RCS value of the FOD target between the RAU 2-3 and runway that may deteriorate the performance of radar detection system.

Acknowledgment

This work was supported as a part of a collaborative project entitled "FODDS Field Trial at KLIA" funded by the Ministry of Internal Affairs and Communications, Japan. The authors would like to thank Malaysia Airport Holding Berhad, Hitachi Kokusai Electric Inc, Waseda University, NICT Japan and iSmartUrus Sdn Bhd for their cooperation in the field experiments.

References

[1] Zhang, H. (2015). The current status and inspiration of FOD industry in international civil aviation. Civ. Aviat. Manag, 295, 58-61.

[2] SRI, I. (2008). The economic cost of FOD to airlines.

[3] Patterson Jr, J. (2008). Foreign object debris (FOD) detection research. International Airport Review, 11(2), 22-7.

[4] Kawanishi, T., Kanno, A., & Yamamoto, N. (2017, July). Concept of sensor over fibres and its application to high resolution millimetre-wave radar. In 2017 19th International Conference on Transparent Optical Networks (ICTON) (pp. 1-4). IEEE.

[5] Kawanishi, T., Kanno, A., & Yamamoto, N. (2018, July). 90-GHz linear-cell systems for public transportation systems. In 2018 20th International Conference on Transparent Optical Networks (ICTON) (pp. 1-4). IEEE.

[6] Kawanishi, T., Kanno, A., Yamamoto, N., Yonemoto, N., Shibagaki, N., & Kashima, K. I. (2017, July). Optical fiber network-connected distributed mm-wave radar system. In 2017 IEEE Photonics Society Summer Topical Meeting Series (SUM) (pp. 197-198). IEEE.

[7] Algafsh, A., Inggs, M., & Mishra, A. K. (2016, November). The effect of perforating the corner reflector on maximum radar cross section. In 2016 16th Mediterranean Microwave Symposium (MMS) (pp. 1-4). IEEE.

[8] N. Malachias, I. Kakavas, S. M. Said Al Harthi, and A. Said Al Saidi, "Design and Experimental Evaluation of a Novel Type Radar Reflector for use in the Marine Environment.," 2019

[9] X. Wang, Z. Lan, Z. Li, Y. Lin, and Q. Xiao, "A Practical Application of Runway Foreign Object Debris Detection System at the Airport," 2018 Int. Conf. Microw. Millim. Wave Technol. ICMMT 2018 - Proc., no. 1, pp. 3–5, 2018.

[10] P. Ni, C. Miao, H. Tang, M. Jiang, and W. Wu, "Small Foreign Object Debris Detection for Millimeter-wave Radar Based on Power Spectrum Features," Sensors, vol. 2020, no. 1, 2020.

[11] Alathari, Bashar, Mohammed Falih Kadhim, Salam Al-Khammasi, and Nabeel Salih Ali. "A framework implementation of surveillance tracking system based on PIR motion sensors." Indonesian Journal of Electrical Engineering and Computer Science 13, no. 1 (2019): 235-242.

[12] Y. K. S. Kim, M. P. K. Park, and M. K. G. Hong, "Fixed and Moving Automatic FOD Detection Test using Radar and EO Camera," vol. 24, no. 6, pp. 479–484, 2020.

[13] F. Qin, X. Bu, Y. Liu, X. Liang, and J. Xin, "Foreign object debris automatic target detection for millimeter-wave surveillance radar," Sensors, vol. 21, no. 11, 2021.

[14] S. Futatsumori, K. Morioka, A. Kohmura, K. Okada, and N. Yonemoto, "Design and Field Feasibility Evaluation of Distributed Type 96 GHz FMCW Millimeter-Wave Radar Based on Radio-Over-Fiber and Optical Frequency Multiplier," J. Light. Technol., vol. 34, no. 20, pp. 4835–4843, 2016.

[15] Kulemin, G. P. (2003). Millimeter-wave radar targets and clutter. Artech House.

[16] Fizza, G. (2021). Review on Foreign Object Debris Detection Technologies and Advancement for Airport Safety and Surveillance. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(3), 1431 -1436.

Performance Comparison using various LDPC coded and CSK modulated Schemes

1st Daekyeong Kim 2nd Hong-Yeop Song
dept. of Electrical and Electronic Engineering
Yonsei University
Seoul, Republic of Korea
{daky33, hysong}@yonsei.ac.kr

***Abstract*—In this paper, the performance of LDPC encoded and CSK modulated signals are compared. 500-bit messages are encoded with LDPC codes of length 600, 700, 800, 900, 1000 and they are modulated by CSK with order 6, 7, 8, 9, 10 respectively. From the simulation result, the frame error rate are analyzed considering trade-off between data rate and error rate performance.**

***Index Terms*—LDPC code, cyclic shift keying, frame error rate**

I. Introduction

The code shift keying (CSK) is a higher-order modulation used in communication systems employing direct-sequence spread spectrum (DSSS) technique. CSK modulation maps several bits to some cyclically shifted version of pseudo-random noise (PRN) code representing one modulation symbol. By using CSK, the data rate can be increased while maintaining signal bandwidth since chip rate is not changed.

In quasi-zenith satellite system (QZSS), CSK modulation of order 8 is used for centimeter-level augmentation services (CLAS) [1]. As a forward error correction code, Reed-Solomon (RS) code over $GF(2^8)$ is used together with CSK modulation for L6 band signal in QZSS. Each RS coding symbol representing 8-bit in navigation message is mapped to one of 256-ary CSK modulation symbols.

The LDPC codes, proposed by Gallager and rediscovered by MacKay, are error correction codes approaching the channel capacity under iterative decoding [2], [3]. With robust error correction capability, they are used widely including data channel in 5G NR and some modernized GPS signals. To provide more reliability in navigation message, CSK modulation schemes using low-density parity-check (LDPC) code are recently studied [4].

C/N_0 is a measure commonly used to evaluate performance in satellite navigation system. C/N_0 is independent of bandwidth while C/N is not. E_b/N_0 is a measure used to evaluate performance considering energy per information bit. When compared to C/N_0, E_b/N_0 is independent on the data rate. Let R_b be denoted as the data rate, then relationship between E_b/N_0 and C/N_0 (in dB scale) can be written simply as

$$C/N_0 = E_b/N_0 + 10log_{10}(R_b) \quad (1)$$

II. Performance Comparison

The frame error rate (FER) of several schemes using LDPC code and CSK modulation is analyzed with respect to E_b/N_0 and C/N_0. For 500 bit message, LDPC codes of length $n = 600, 700, 800, 900, 1000$ are constructed using PEG algorithm. Let $m = 6, 7, 8, 9, 10$ be denoted as modulation order of CSK respectively. Then, every m-bits of LDPC codes of length n are mapped to one of CSK modulation symbols resulting 100 symbols respectively. And, Weil code of length 10230 is cyclically shifted according to CSK modulation symbols. The overall schemes are summarized in Fig. 1.

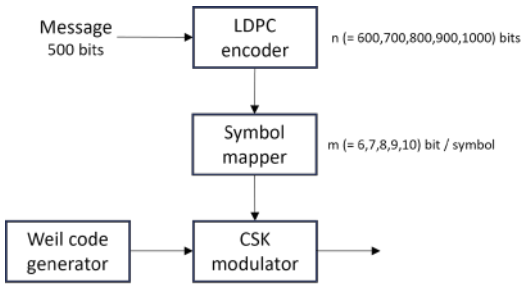


Fig. 1. Overall signal structure

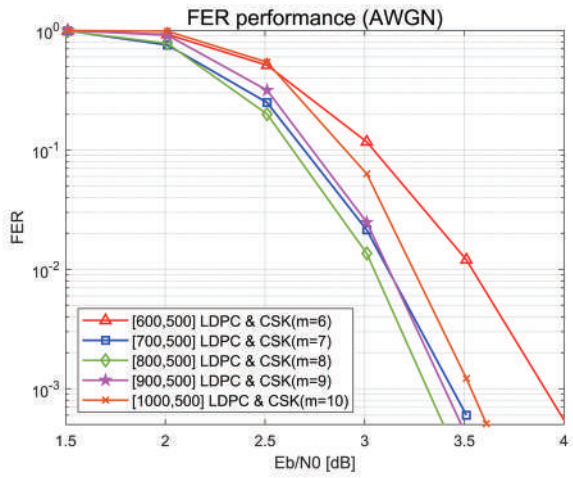


Fig. 2. Comparison of performance over E_b/N_0

Figure 2 shows the FER vs E_b/N_0 performance of each schemes in AWGN channel. Generally, error correction codes are expected to perform better with higher redundancy. And, being regarded as a orthogonal modulation, performance of CSK improve as modulation order increases. But it is observed in Fig. 2 that scheme with the largest redundancy and highest order does not shows best FER performance. This is due to the difference in effective data rate at the input of CSK modulator.

Now, consider a constant transmission of 500 bit message in 500 bps. As Weil code of length 10230 is modulated by each symbol with 100 sps, the spread signals are transmitted with the chip rate of 1.023 Mcps. The overall trend in performance concerning C/N_0 does not change with that of E_b/N_0 as data rate is constant. But, there is practical difference of data rate at the stage between the output of the LDPC encoder and the input of the CSK modulation stage as in Fig. 3.

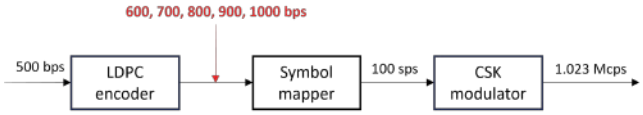


Fig. 3. Transmission rate

Thus, performance of CSK modulations over C/N_0 is compared separately with uncoded case. Figure 4 shows the FER vs C/N_0 performance of uncoded CSK modulation with varying data rate. It can be observed in the C/N_0 plot that the

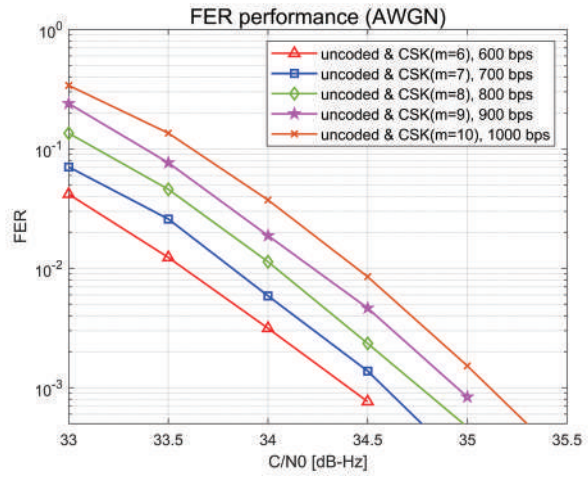


Fig. 4. Comparison of performance of uncoded CSK over C/N_0

performance worsens as the modulation order increase at the expense of increasing the data rate.

In the LDPC coded and CSK modulated scheme, increasing the modulation order of CSK comes at the expense of increasing the redundancy of LDPC codes. Thus, required C/N_0 at FER 10^{-3} are plotted to compare LDPC encoded and CSK modulated signal in Fig. 5. In Fig. 5, scheme 1 through 5 correspond to LDPC code of length 600, 700, 800, 900, 1000 or CSK modulation of order 6, 7, 8, 9, 10 respectively. The required C/N_0 decrease as redundancy of LDPC codes

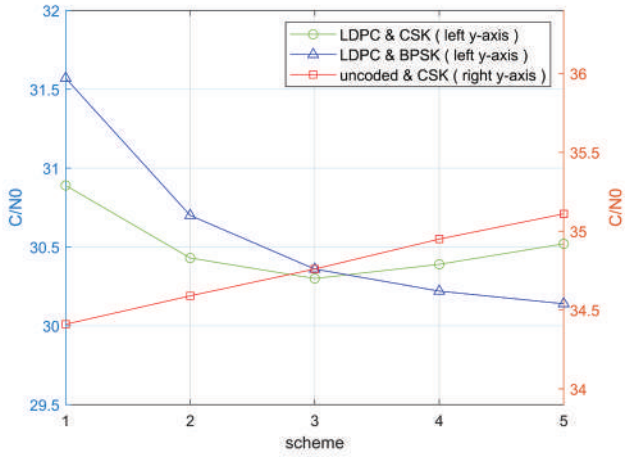


Fig. 5. Comparison of the required C/N_0 at FER 10^{-3}

increase to a certain point. But after that, the performance degrade as the expense of increasing effective data rate at CSK modulator affects more.

III. Conclusion

In this paper, frame error rate performance of various LDPC encoded and CSK modulated is compared. The 500 bits messages are encoded with LDPC codes of length 600, 700, 800, 900, 1000 and modulated by CSK with order 6, 7, 8, 9, 10 respectively. In the simulation, FER performance shows best when using LDPC code with length 800 and CSK modulation with order 8. Some trade-off relationship between LDPC code redundancy and CSK modulation order within the considering schemes are explained with regard to FER performance.

There exist much more considerations to analyze the performance of several LDPC coded and CSK modulated schemes First, the LDPC codes used in this paper are not optimized but the performance of LDPC codes are significantly varies with construction and optimization method. And, LDPC code and CSK modulation is considered as affecting the overall performance individually even though there may exist optimized combinations of LDPC code and CSK modulation.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) Grant by the Korea Government through Ministry of Sciences and ICT (MSIT) under Grant RS-2023-002090000.

References

[1] Quasi-Zenith satellite system interface specification centimeter level augmentation service, IS-QZSS-L6-005, Cabinet Office, Sept. 2022.
[2] R. Gallager, “Low-density parity-check codes,” IRE Transactions on Information Theory, vol. 8, no. 1, pp. 21–28, 1962.
[3] D. J. C. MacKay, “Good error-correcting codes based on very sparse matrices,” IEEE Transactions on Information Theory, vol. 45, no. 2, pp. 391–431, Feb. 1999.
[4] H. Cho, J.M. Ahn, J.H. Noh et al., Some new LDPC-coded orthogonal modulation schemes for high data rate transmissions in navigation satellite systems, ICT Express (2024).

Complexity analysis to obtain maximum performance of successive cancellation (SC) decoder for Reed-Muller codes with dynamic frozen bits

Woo Yong Lee and Keunyoung Kim
Mobile Communication Research Division,
Telecommunication & Media Research Laboratory
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, Republic of Korea
{wylee, kykim12}@etri.re.kr

Dongwook Cho

Korea Association of University, Research Institute and Industry
Sejong-si, Republic of Korea

ducho@auri.go.kr

***Abstract*— The wireless communication environment in extreme cold areas consisting of seawater and glaciers is limited in terms of power consumption and the size and weight of equipment, but the condition for unmanned exploration is long-distance, high-speed communication. The wireless communication system for operating unmanned robotic exploration in Antarctica requires a high-speed transmission rate of more than 10Mbps over long distances of more than 50km. At this time, we want to redesign the beam forming technique to concentrate limited power in a specific direction and the encoding technique to optimize channel gain to satisfy the long-distance communication requirements of mobile objects. Recently, interest in channel coding research has increased due to new applications requiring short-length data transmission. Additionally, it has been proven that there are codes that can achieve channel capacity in low complexity encoding and decoding algorithms. Meanwhile, the interest is growing as to whether theoretical performance can be achieved with a low-complexity decoding algorithm even with finite-length codes. In this environment, recent research results have proven that the Reed-Muller (RM) code with dynamic frozen bits can achieve theoretical performance while using a successive cancellation list decoding technique in a binary erasure channel. This paper analyzed the possibility of further improving complexity in the successive cancellation list decoding technique for RM codes using dynamic frozen bits. It is expected that these results can be used as a conceptual basis for an encoder and decoder design that can be applied to a wireless communication system for operating unmanned robotic exploration in Antarctica.**

Keywords*— Channel code, decoding complexity dynamic frozen bits, successive cancellation

I. INTRODUCTION

Polar codes use a successive cancellation list (SCL) decoder[1], and channel code research is actively taking place from 128 to 1024 bits by adding a high-speed outer code [2]. Reed-Muller (RM) codes are closely related to polar codes [3], but surpass them in maximum a posteriori (MAP) decoding performance [4], so some decoders proposed for polar codes have also been used for RM codes. The authors of [5] analyzed which linear codes can be viewed as polar codes with dynamic frozen bits. In addition, it has been proven that theoretical performance can be achieved with a successive cancellation list decoding technique in a binary erasure channel (BEC) by applying dynamic frozen bits to the RM code [6]. When information bits were used as frozen bits in these codeword designs, the distance between codewords was analyzed and the decoding error probability of the designed codes was analyzed [7].

In this paper, we analyze the computing complexity of the decoder in the situation of maximum channel code performance improvement in the successive cancellation (SC) list decoding technique for RM codewords using partial codes, that is, dynamic frozen bits.

II. ANALYSIS OF DECODING COMPLEXITY OF RM CODE WITH MAXIMUM EXTENDED CODEWORDS

Any linear block code can be composed of an invertible matrix **A** as follows.

$$A = B^m F^{\otimes m}. \quad (1)$$

Here, $\mathbf{B}^m$ is the bit reversal permutation matrix. $F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. $\otimes m$ is the Kronecker product operation for m related matrixes. Index $i \in \{1, \dots, n\}$, the codeword **C** for the input vector u with n elements u_i can be expressed as **uA**. In the case of a polar code, k rows are selected from the generator matrix **A** to minimize decoding errors. The remaining n-k inputs are set to $u_j=0$ and these bits are called frozen bits.

Polar codes or RM codes, which are linear block codes designed using density evolution, have a small minimum distance between codewords. To solve this problem of codewords, it was found that it is not necessary to set all frozen bits to $u_j=0$ [5]. For $j < i$, the frozen bit u_i is a function of predefined information bit u_j. This does not affect the performance of the bit sub-channel linearly converted by **A** and the performance of the SC decoder. For a linear code **C** ($n=2^m$, k, d) with a parity-check matrix H, if $uAH^T=uV^T=0$ is satisfied, a codeword with the length of an appropriate number of rows can be obtained. This has the effect of expanding the minimum distance between codewords.

For $1 \le j \le n-k$, when $i_j = \max\{t \in \{1, \dots, n\} | V_{j,t} = 1\}$, the frozen bit index set $\Im$ can be expressed as $\{t \mid \exists\, j : i_j = t\}$. Thus, $\Im = \{t \mid \exists\, j: i_j = t\}$ and $S_j = \{t \mid V_{j,t} = 1, t < i_j\}$, the limiting conditions of the dynamic frozen bit u_j are as follows[6].

$$u_j = \sum_{t \in S_j} u_t, \quad j \in \Im \quad (2)$$

If S_j is an empty set, it represents an existing static frozen bit, and $|Sj| \ge 1$, it means dynamic frozen bit. Meanwhile, the information bit index for use as a frozen bit is $I_f = \{t \in \{1, \dots, i_j\}, t \notin \Im \mid V_{j,t} = 1, 1 \le j \le n-k\}$. The dynamic frozen bit index $D_f = \{i \in \Im,\ S_i \neq \emptyset\}$. The estimated value $\hat{u}_i$ for bit u_i during the SC decoding process for the output to the channel is as follows by expanding the recent research results [7].

$$\hat{u}_i = \begin{cases} \underset{u_i \in \{0,1\}}{\operatorname{argmax}} P\left(u_i \middle| y_1^n, \hat{u}_1^{i-1}\right), & for\ i \notin \Im, i \notin I_f \\ u_i, & for\ i \in \Im, S_i = \emptyset \\ \underset{u_i \in \{0,1\}}{\operatorname{argmax}} P\left(u_i, \sum_{t \in S_i} \hat{u}_t \middle| y_1^n, \hat{u}_1^{i-1}\right), & for\ i \in D_f, i \in I_f \end{cases} \quad (3)$$

In this time, the error probability $p(u_i)$ of the i-th u_i bit of the SC decoder is as follows, considering the maximum posterior probability decoding for the matching codeword allocation bits.

$$p(u_i) = \begin{cases} 0 \text{ for } i \in \Im,\ S_i = \emptyset \\ \left(\prod_{t \in \{I_f, D_f\}} \epsilon_t \right)^{\frac{1}{|I_f|+|D_f|}} \text{for } i \in D_f, i \in I_f \\ \epsilon_i \quad \text{otherwise} \end{cases} \quad (4).$$

Meanwhile, in the case where the first codeword distance extension ends, the relative distance remaining after subtracting the used code from among the codes with a codeword distance less than 2tcd is $\sum_{i=1}^{tcd} \frac{\binom{n}{tcd-i}}{2^i} - \left\{ K - \sum_{k=tcd}^{n} \binom{n}{k} \right\}$, and if the secondary codeword distance can be further expanded, the number of codewords assigned to the minimum distance (2tcd) is $\binom{n}{tcd} + K - \sum_{k=tcd}^{n} \binom{n}{k}$ [7]. Therefore, for $j_{min} = \log\left(cd_j(tcd)\right) > 0$, the maximum number of codewords that can be determined by frozen bits is given by the following equation (5).

$$\sum_{k=n-j_{min}+1}^{n} \binom{n}{k} + \left\lfloor 2^{tcd-j_{min}} \left\{ \sum_{i=1}^{tcd} \frac{\binom{n}{tcd-i}}{2^i} - K + \sum_{k=tcd}^{n} \binom{n}{k} \right\} - K + \sum_{k=tcd+1}^{n} \binom{n}{k} \right\rfloor \quad (5)$$

Therefore, the above number is the number of codewords $i \in \Im$, $S_i = \emptyset$ that are purely allocated as frozen bits. As an example, if the number of code words to which information bits are allocated is K = 7 in an RM code with a length of N = 16, the largest j exists for $1 \le j < tcd$ (= 3) and is 1, so the minimum distance between code words can be assigned to be 10 (= 23 + 21). At this time, from the above formula, the number of frozen bits is

$$\binom{n}{k} + \left\lfloor 2^{tcd-j} \left(\sum_{i=1}^{tcd} \frac{\binom{n}{tcd-i}}{2^i} \right) - \hat{R} \right\rfloor - \left\{ \binom{n}{r} + \hat{R} \right\}$$
$$= \binom{4}{4} + \left\lfloor 2^2 \left(3 + 1 + \frac{1}{8} - 2 \right) - 6 \right\rfloor = 3.$$

Therefore, when assigning the minimum distance between codewords to be 10, using u_1, u_2, and u_3 as frozen bits corresponds to being the maximum frozen number.

For $\epsilon_i \ll 1$, if $\left(\prod_{t \in \{I_f, D_f\}} \epsilon_t \right)^{\frac{1}{|I_f|+|D_f|}} \approx 0$, then the error probability $p(u_i)$ of the u_i bit can be simplified.

$$p(u_i) = \begin{cases} 0 \ \ for\ i \in \Im \text{ or } i \in I_f \\ \epsilon_i \quad otherwise \end{cases} \quad (6)$$

Assuming SC-list decoding for random variables (RV) to analyze decoding complexity [6], the average value E(B) of the number B of the unbounded list is as follows.

$$E(B) = \sum_{i \in \Im^c} \epsilon_i = \sum_{i \in D_f} \epsilon_i + \sum_{i \in [N-K+1,\, N]} \epsilon_i \quad (7)$$

The average value E(B) of the number of lists B refers to the decoding complexity and means that it has increased by $2^{|D_f|\epsilon_i}$ times compared to the preceding result in [6], and $|D_f|$ can be represented as the following equation (8).

$$|D_f| = N - K - \sum_{k=n-j_{min}+1}^{n} \binom{n}{k} - \left\lfloor 2^{tcd-j_{min}} \left\{ \sum_{i=1}^{tcd} \frac{\binom{n}{tcd-i}}{2^i} - K + \sum_{k=tcd}^{n} \binom{n}{k} \right\} - K + \sum_{k=tcd+1}^{n} \binom{n}{k} \right\rfloor \quad (8)$$

III. SC LIST DECODING COMPLEXITY AT THE MAXIMUM DISTANCE BETWEEN CODE WORDS

A BEC channel was assumed with linear channel code lengths of 4 (red), 6 (blue), and 7 (green). If the channel code is designed with the maximum distance between code words, the code rates are β= 0.3164 (n=4), 0.17798 (n=6), and 0.1334 (n=7), respectively. At each code rate, significant performance improvement was confirmed in terms of linear code error probability and outage probability of the proposed method [7]. Calculating the decoding complexity for the linear code error probability of the existing method at each code rate is as shown in the dotted line in Figure 1. Meanwhile, assuming the maximum code word distance at the code rate of each linear code, when analyzing the decoding complexity for errors, the result is as shown in the solid line in Figure 1. Therefore, we confirmed the results that the decoding complexity for the error probability of the code does not significantly increase at each coding rate. Moreover, when obtaining maximum decoding performance, we obtained results showing that there is almost no difference in decoding complexity in the area of error probability of the code we will use (≈1%).

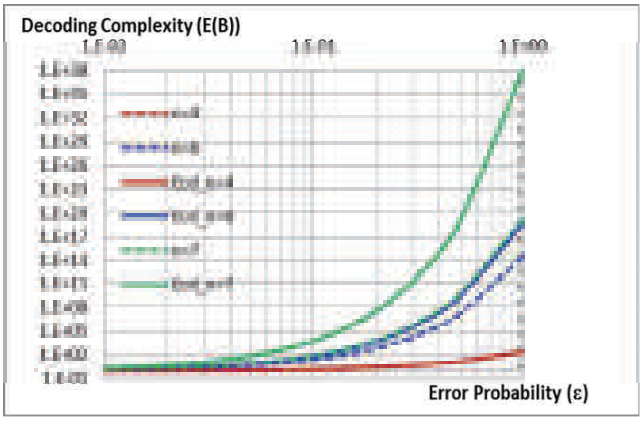


Figure 1. Decoding complexity versus error probability for linear codes with code lengths of 4 (red), 6 (solid), and 7 (green) (dotted line for existing method, solid line for proposed method).

ACKNOWLEDGMENT

This paper is a research conducted with the support of the Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the government (Ministry of Science and ICT) in 2024. [No. 2021-0626, Development of Polar Region Communication Technology and Equipment for Internet of Extreme Things (IoET)].

REFERENCES

[1] I. Tal and A. Vardy, "List decoding of polar codes," IEEE Trans. Inf. Theory, vol. 61, no. 5, pp. 2213–2226, May 2015.

[2] M. C. Coskun, G. Durisi, T. Jerkovits, G. Liva, W. Ryan, B. Stein, and F. Steiner, "Efficient error-correcting codes in the short blocklength regime," Elsevier Physical Communication, vol. 34, pp. 66–79, Jun. 2019.

[3] E. Arıkan, "Channel polarization: A method for constructing capacity achieving codes for symmetric binary-input memoryless channels," IEEE Trans. Inform. Theory, vol. 55, no. 7, pp. 3051–3073, July 2009.

[4] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "From polar to Reed-Muller codes: A technique to improve the finite-length performance," IEEE Trans. Commun., vol. 62, no. 9, pp. 3084–3091, Sep. 2014.

[5] P. Trifonov and V. Miloslavskaya, "Polar subcodes," IEEE J. Sel. Areas Commun., vol. 34, no. 2, pp. 254–266, Feb. 2016.

[6] M. C. Coskun, J. Neu and H. D. Pfister, "Successive Cancellation Inactivation Decoding for Modified Reed Muller and eBCH Codes," in Proc. IEEE ISIT 2020, Jun. 2020.

[7] Woo Yong Lee and Keunyoung Kim, "An effect analysis of code-word distance extension in Reed-Muller codes with dynamic frozen bits," in Proc. ICTC 2022, pp. 552–554, Oct. 2022.

Differentiated Traffic Class Mean Delay of Degradation Attack in XG-PON

F.M. Atan[1,2], N. Zulkifli[2], S.M Idrus[2], A.Azizan[3] , A.M. Zin[1,2], N.A. Ismail[1,2]
[1]*Lightwave Communication Research Group (LCRG), Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*
[2]*Center of Electrical Engineering, Universiti Teknologi MARA Cawangan Johor, 81750 Pasir Gudang, Johor, Malaysia.*
[3]*Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*
fadila2533@uitm.edu.my

***Abstract*— The Next-Generation Gigabit Passive Optical Network (NG-PON) enhances bandwidth, data rate, and allows for ongoing network expansion. Given the closed-loop nature of NG-PON, security measures are typically concentrated on the downstream channel transmission due to its broadcast nature. As a result, the upstream point-to-point direction has been assumed to be secure. However, because Time Division Multiple Access (TDMA) is used as a transmission mode in the upstream channel, each packet is assumed to go from end to end, leading to a less stringent approach to its security. This paper introduces a secured DBA mechanism for XG-PON to address this specific threat. The Security-Enhanced DBA (SE-DBA) algorithm includes a detection phase to identify abnormal user behavior. Once an attacker is identified, a mitigation phase follows, reducing the bandwidth of the identified users in the next bandwidth cycle. Network simulations, both standard and attacked, were conducted using a network testbed based on OMNET++. The results show that SE-DBA improves the mean delay to all Transmission Containers (TCONT) compared to the GIANT DBA. Specifically, the simulations show a performance increase of 5%, 30%, and 58% in upstream delay for TCONT2, TCONT3, and TCONT4, respectively, compared to the unprotected DBA scheme. This indicates the effectiveness of the proposed SE-DBA mechanism.**

Keywords—XG-PON, DBA, Degradation Attack, Delay, TCP

I. INTRODUCTION

Information and communication technology (ICT) has become an integral part of our daily lives, closely intertwined with our routines. The exponential growth in internet traffic and bandwidth-intensive applications is fuelling the penetration of fiber networks into the access network segment. The ever-growing network demand necessitates a solution to support its expansion[1] . The Gigabit Passive Optical Network (GPON) is thriving due to its capacity to provide the customer with a higher bandwidth network while reducing the cost per bit by a significant margin. With a greater number of users and network components, security concerns become a much more important factor to consider. Due to the enormous volume of traffic and the shift to cloud networking, the security risks associated with high bandwidth networks are becoming more significant[2].

The Dynamic Bandwidth Allocation (DBA) is integral in GPON operation. The DBA facilitates in distributing bandwidth fairly among all connected users[3]. The exact implementation of DBA in PON depends on the vendor and the specific implementation, but the basic idea is to monitor the bandwidth usage of each Optical Network Unit (ONU) and adjust the allocation as needed to ensure that all devices have enough bandwidth to operate effectively. Despite the presence of various DBA algorithms in the literature, most of them focus on Quality of Service (QoS) improvement leaving behind the critical aspect of security [4]. A specific type of attack named Degradation Attack (DA) results in interruption on both MAC and transport layer protocol [5]. In general, a degradation attack is not aimed at a particular user, but rather at disrupting the entire network. Bandwidth availability is the primary attraction for attackers since PON's abundance of resources is especially tempting because bandwidth is provided on demand.

This paper aims to highlight the effect of DA on the network's delay performance and present a new DBA (SE-DBA) that protect against DA. A comparison of delay performance between GigaPON Access Network (GIANT) and SE-DBA shows improved performance of DBA when fortifying the network against threats, ensuring that the dynamic allocation of resources aligns with stringent security requirements

II. LITERATURE REVIEW

The basic PON architecture consists of an Optical Line Terminal (OLT) at the central office, ONUs at customer premises, and passive optical splitters connecting them. Downstream transmission involves the OLT broadcasting data to ONUs, while upstream transmission involves ONUs transmitting data to the OLT using TDMA.

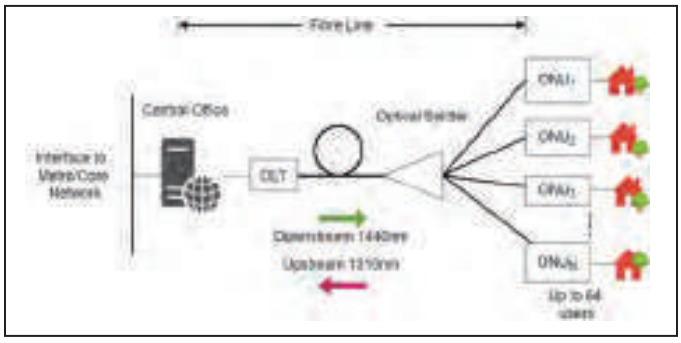


Fig. 1. Basic PON architecture

Figure 1 illustrates the basic PON architecture. The OLT broadcasts downstream traffic to all ONUs that share the same optical fiber, with encryption applied to maintained data secrecy. Upstream transmission, on the other hand, refers to the flow of data from ONU back to the OLT. The ONU uses TDMA to combine multiple data streams into a single optical signal.

The DBA design for GPON networks is a critical research area due to the challenges in adhering to strict QoS definitions of GPON standards and the complexity of DBA implementation[5]. GPON networks must serve various traffic types with varying latency requirements, including audio, video, VoIP, data, and leased lines. The traffic types are divided into four categories accordingly: TCONT1, TCONT2, TCONT3, and TCONT4. The GIANT algorithm was the first standard-compliant mechanism for GPON, and the foundational DBA algorithm physically deployed in a GPON environment[6].

Despite new numerous DBA algorithms designs, most focus on improving QoS while ignoring the security aspect of DBA. Typically, GPON is considered secure due to the implementation of various security protocols, such as data encryption, authentication, and key formation in GPON [7][8]. The DA occurs when an attacker transmits meaningless data packets outside of their allocated time slots [5]. The attack is carried out by transmitting signals beyond the authorized timeslot, effectively utilizing the timeslots of other ONUs, resulting in a collision. However, the research is being carried out using the EPON standard, which differ from GPON.

The effect of timeslot violations in GPON was recently established in [9]. The approach, however, is unable to identify the source of interference. Thus, it recommends physical protection of all passive optical hubs, which automatically raises the network setup cost. Another attack detailed in [4], and machine learning was utilized to counteract the attack. However, the attack was pre-planned, and the legitimately of the attack launched is questionable as the method is unable to determine whether actual attack occurred, or the ONU simply has a high volume of traffic for a specific duration.

In summary, DBA methods optimize resource utilization, indirectly enhancing security by promoting stable and efficient network operation. However, for specific security needs, additional security mechanisms can be layered on top of DBA to increase the performance.

III. Methodology

A DA is a unique type of attack that targets upstream transmission by exploiting the workings of Transmission Control Protocol (TCP), the standard upper-layer protocol that runs over the GPON network. Specifically, transmissions in the upstream direction are based on scheduled time slots by OLT. This means that each ONU must adhere to its assigned transmission grants and transmit data only during the specified time. However, there is no monitoring technique in place to ensure that ONUs adhere to the assigned time slots. This fact creates the possibility of a degradation attack based on a violation of assigned time slots.

When two or more ONU are transmitting at the same time the light signals from the ONUs are mixed. Upon reaching the OLT, the mixed light signals undergo an irreversible process and cannot be rectified, leading to the occurrence of packet loss for the lawful ONUs. The loss of multiple packets will inherently evoke the TCP congestion control algorithm, i.e., a traditional upper-layer transport protocol running over the GPON network layer. Figure 2 shows DA scenario.

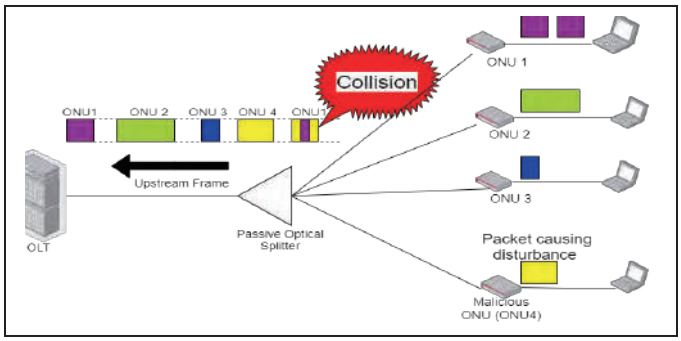


Fig. 2. Degradation Attack Scenario

The evoking of congestion control algorithm results in reduced throughput speed (the one experiencing collision) for the lawful ONU. The attack time is particularly aimed at lawful ONU's timeslot, which means the degradation attack never violates the malicious ONU's timeslot. Hence, malicious ONU continued at the same rate enjoyed without experiencing reduced throughput. The uninterrupted TCP response also meant that the malicious ONU's traffic flow (traffic request/ traffic granted) was done in a constant manner. It allows the malicious ONU to place a higher bandwidth demand than lawful ONU since it has higher throughput. Having a dynamically assigned bandwidth further helps the malicious ONU get more bandwidth as its bandwidth request is granted at all times. The delay modelling is given as Figure 3 below:

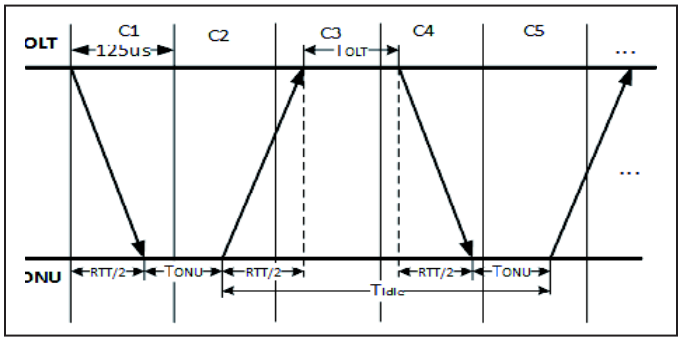


Fig. 3. Delays involved in the DBA process [10]

$$D_{US} = T_{Idle} + T_{Queueing} g = 2*(RTT + T_{OLT}) + T_{ONU} + T_{Queueing} \quad (1)$$

$$T_{OLT} = T_{DBA} + T_{Grant} = T_{DBA} + RTT + T_{OLT} + T_{ONU} \quad (2)$$

$$T_{OLT} = T_{DBA} + T_{Grant} = T_{DBA} + RTT + T_{OLT} + T_{ONU} \quad (3)$$

$$T_{Grant} = RTT + T_{OLT} + T_{ONU} \quad (4)$$

$$T_{Queue} = \lambda / (2\mu^2 (1 - P_{attack})) \quad (5)$$

$$D_{USattack} = (1-P_{attack})\times D_{US} + P_{attack}\times D_{TCP} \quad (6)$$

The expressions for D_{US}, T_{idle} , T_{OLT} ,$T_{DBACycle}$ are shown in (1) to (5) [95]. $T_{Grant,}$,T_{idle} , $D_{US,}$ and RTT should be decreased to lower T_{OLT}. RTT is mostly determined by the length of optical fibre from the OLT to the ONU and is therefore uncontrollable, whereas TOLT can be decreased by reducing T_{Grantt} at least one GPON cycle. Nevertheless, the effect of RTT can be decreased by distributing available bandwidth efficiently and fairly among ONUs, hence lowering the DUS . Equation (6) is the upstream delay during attack which takes into consideration the properties of delay introduced by the TCP component.

The SE-DBA introduces two steps detection and mitigation algorithms, Detection algorithm's principle involves identifying the ONU with no collisions while all other ONUs has a higher number of collisions. Since a specific condition is needed to cause a collision and evoke the TCP congestion control as mentioned above, it is highly unlikely to deduce this condition other than from a network attack. After receiving the Alloc ID, the mitigating process checks the network's DBA. The DBA process connotes a bandwidth request made by all ONUs, regardless of traffic condition. The preceding section described how all bandwidth requests are accumulated in the BWmap following the OLT's polling and scheduling mechanism[11].

TABLE 1 Simulation Parameters in OMNeT++

Parameter	Values
Upstream / Downstream Line	2.5/10 Gbps
Upstream / Downstream Traffic load	0.5 to 1.5
Traffic Arrival	Varied from 72000 frames/s to 120000 frame/s corresponding to a load variation from 0.1 to 1.6 [10]
Guard band period for uplink Channel	8 bits
Disturbance	Poisson distributed traffic of 500-bit interarrival of 2ms [5]
TCP Rules	New Reno
Average TCP Traffic Size	1476-1500 bit [5]
TCONT2	Assured, SI=5 [6]
TCONT3	Guaranteed, SI=10 [6]
TCONT4	Best-effort, SI=10 [6]

IV. RESULT & DISCUSSION

One of the most crucial measurements for DBA effectiveness is denoted by the value of delay. Since DBA processing occurs in the MAC layer, an attack in a higher layer will affect its effectiveness toolbar. During simulation TCONT 1 is not entertained due to fix bandwidth assignment. The simulation shows two conditions which are the normal and attack condition. Both normal condition for GIANT and SE-DBA shows similar performance. Which shows additional steps of detection and mitigation algorithm does not impact delay values significantly.

The result can be analyzed particularly with respect to TCONT. The additional notion of 'A' in the graph's legend shows the presence of attack. Figure 4 shows that TCONT 2 is the least affected by the attack, especially when the network is underutilized. The insignificant difference is due to the GIANT prioritizing the TCONT2. The situation is reversed, however, when comparing TCONT3 and TCONT4. The TCONT3 achieves a nearly 27% increase in mean attack delay while TCONT4 suffers 51.4% more than normal.

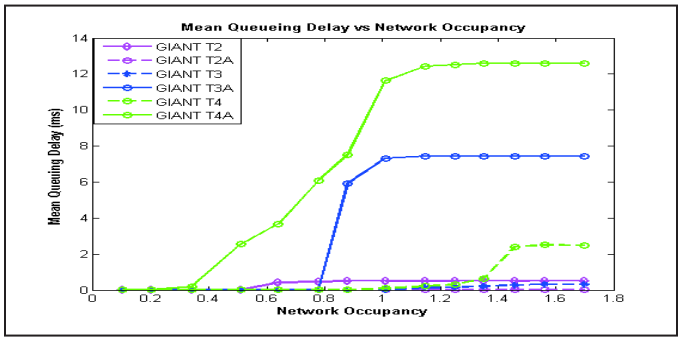


Fig. 4. Mean Queueing Delay vs. Network Occupancy for GIANT

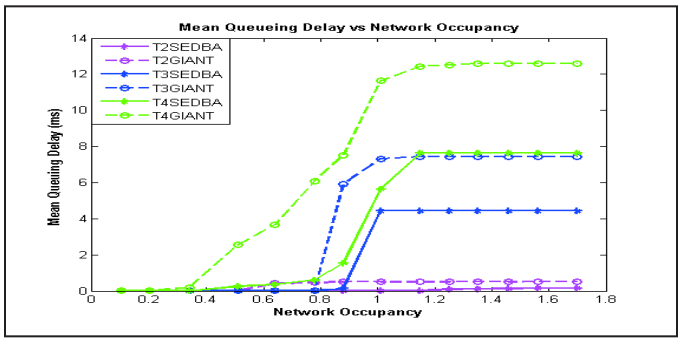


Fig. 5. Mean Queueing Delay vs. Network Occupancy for SE-DBA

Figure 5 shows the simulation results using SE-DBA. The TCONT2, which is always prioritised before, slightly varied during attack, albeit by only 5%. However, TCONT3 and TCONT4 suffer severely from the attack. It can be seen that in low network occupancy, the attack is less severe compared to high network occupancy. With SE-DBA, TCONT3 is able to reduce the mean delay up to 30%, while massive improvement of 58% is seen on TCONT4. Being the best-effort or lowest priority type of traffic, TCONT4 that is least served during a highly loaded network experiences the most delay. During the attack, the delay caused the network to have such a high delay that it exceeded the standard.

It is apparent that without the SE-DBA, bandwidth assignment is manipulated during the attack. As a result, the bandwidth to lawful ONUs is reduced, consequently increasing the delay for all TCONT. With SE-DBA, bandwidth assignment is monitored not just by request and availability, but with added

identification worthiness which assured lawful ONUs' bandwidth allocation.

V. CONCLUSION AND FUTURE WORKS

As a conclusion, this paper shows the importance of addressing the security-performance gap in DBA. This paper focus on developing adaptive DBA algorithms that not only optimize bandwidth allocation but also integrate intrusion detection and mitigation mechanisms. These enhancements will fortify the network against threats, ensuring that the dynamic allocation of resources aligns with stringent security requirements. This research applied to GPON, hence as the network evolves the DBA can be fitted to the immerging XG-PON allowing for faster internet speeds and better support for bandwidth-intensive applications. This research work also not consider the availability of many wireless devices accessing this XG-PON network. In line with this 5G and IoT era, the security and protection of DBA mechanism analysis/development for this optical-wireless network as a whole using this research framework is also an interesting research topic.

REFERENCES

[1] Cisco Systems, "Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper," *Cisco Forecast Methodol.*, pp. 2017–2022, 2019.

[2] T. Cisco and A. Internet, "Cisco: 2020 CISO Benchmark Report," 2020. Transmission convergence (TC) layer specification," *ITU-T G.987.3 Recomm.*, 2010.

[3] FSAN, "10-Gigabit-capable passive optical networks(XG-PON):

[4] R. A. Butt, M. Faheem, M. W. Ashraf, A. Khawaja, and B. Raza, "Attack-Aware Dynamic Upstream Bandwidth Assignment Scheme for Passive Optical Network," *J. Opt. Commun.*, vol. 0, no. 0, pp. 1–9, Sep. 2020

[5] S. Drakulic, M. Tornatore, and G. Verticale, "Degradation attacks on Passive Optical Networks," in *2012 16th International Conference on Optical Networking Design and Modelling, ONDM 2012*, 2012, pp. 1–6

[6] M.-S. Han, H. Yoo, B.-Y. Yoon, B. Kim, and J.-S. Koh, "Efficient dynamic bandwidth allocation for FSAN-compliant GPON," *J. Opt. Netw.*, vol. 7, no. 8, p. 783, Aug. 2008

[7] L. Malina, J. Hajny, R. Fujdiak, and J. Hosek, "On perspective of security and privacy-preserving solutions in the internet of things," *Comput. Networks*, vol. 102, pp. 83–95, Jun. 2016

[8] V. Clupek, T. Horvath, P. Munster, and V. Oujezsky, "New security improvements in next-generation passive optical networks stage 2," *Appl. Sci.*, vol. 9, no. 20, pp. 1–16, 2019

[9] T. Horvath,V. Clupek, P. Munster, and V. Oujezsky, "GPON Network with Simulated Rogue ONU," in *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 2019, pp. 1–5

[10] R. A. Butt, S. M. Idrus, K. N. Qureshi, N. Zulkifli, and S. H. Mohammad, "Improved dynamic bandwidth allocation algorithm for XGPON," *J. Opt. Commun. Netw.*, vol. 9, no. 1, pp. 87–97, 2017

[11] F. M. Atan, N. Zulkifli, S. M. Idrus, A. M. Zin, N. A. Ismail, "Security enhanced dynamic bandwidth allocation algorithm against degradation attacks in next generation passive optical networks," *J. Opt. Commun. Netw.*, vol. 13, no. 12, p. 301, Dec. 2021

Semantic Communication with Adversarial Defense for Malware Detection in Metaverse E-Commerce

Ebuka Chinaechetam Nkoro *, Cosmas Ifeanyi Nwakanma†, Jae-Min Lee *, and Dong-Seong Kim *
*IT Convergence Engineering, † ICT Convergence Research Center, Kumoh National Institute of Technology, Korea.

***Abstract*—With the proliferation of the Internet of Things, artificial intelligence, and virtual reality, securing Metaverse E-commerce against malware has become increasingly critical. This study introduces a novel framework integrating Semantic Communication (SemComm) with Deep Learning for Intrusion Detection Systems (IDS) in Metaverse E-commerce. Leveraging Stochastic Activation Pruning (SAP), the proposed adversarial defense mechanism enhances model robustness against adversarial attacks. Through comparative evaluation using the CICMalDroid2020 malware dataset, the defense model demonstrates improved resilience against adversarial samples while maintaining competitive detection rates. The results show the effectiveness of the dual-model approach, combining pruned and unpruned IDS models, in strengthening IDS in Metaverse E-commerce.**

***Index Terms*—Malware detection, metaverse security, adversarial machine learning, semantic communication, and e-commerce.**

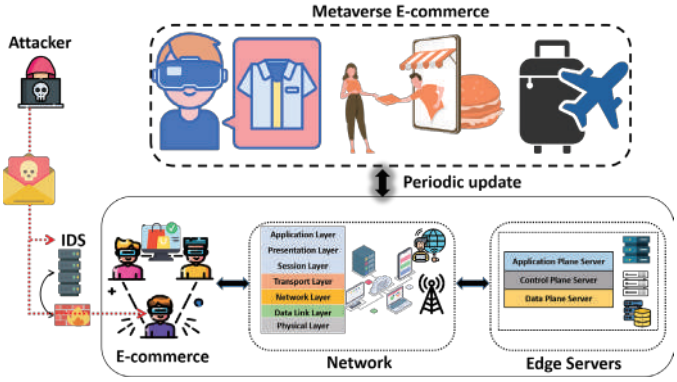


Fig. 1. General overview of Metaverse E-commerce and its susceptibility to malware attacks.

I. INTRODUCTION

The rise of the Metaverse has brought immersive and undeniable convenience in Electronic commerce (E-commerce). However, it has also opened a plethora of security challenges. The Metaverse aims to provide users with an immersive experience and interaction between physical and virtual objects in diverse domains, e.g., E-commerce [1]. Detecting and preventing malware intrusions targeting E-commerce/marketplaces in the Metaverse is particularly concerning to enhance security against frauds [2]. Compared to other platforms/ shopping scenarios, E-commerce in the Metaverse is a characteristic *immersive representation* of the Metaverse's digital economy that offers seamless, virtual integration between physical and digital online retail with in-person shopping experiences. It offers interactive 3D visuals and immersive interactions beyond traditional web stores [2], which relies on domain edge servers to host virtual commerce spaces and gamified user experience.

An attacker can successfully launch an immersive E-commerce malware attack by compromising the foundational haptic devices and its overall insecure architectures [3], forcing users to reveal personal information and compromise overall security within the Metaverse [4]. Specifically, cybercriminals exploit vulnerabilities in E-commerce platforms to launch hidden malicious malware payloads (e.g. *magecart*), compromised payment checkout pages, and even immersive phishing attacks [5], [6]. Such fraudulent attacks in Metaverse E-commerce attacks can cause physical disruptions and discomforts like poor service and quality of experience, as illustrated in Fig. 1.

Previous studies in [7]–[9] have attempted to adapt to the fast-paced malware detection mechanisms using Deep Learning (DL) Intrusion Detection System (IDS) designs, which suffer the problem of overwhelming network traffic in Metaverse applications, high resource consumption, and adversarial attacks. Recent research has shown that Semantic Communication (SemCom) systems integrated with DL can adapt to the data-intensive Metaverse network and real-time utility [10].

Motivation: Although DL-based SemComs have robust utility, a slight perturbation attack with crafted malware samples by an attacker can force the IDS to misclassify attacks [10]. As emphasized by previous works in [11], Defensive methods like adversarial training can be expensive and time-consuming. To defend the SemCom IDS from adversarial exposure and shield the IDS model from adversarial attacks, this study proposes a defensive parameter protection technique using stochastic gradients; thus, the attacker cannot obtain true gradients cheaply.

The main contributions are as follows:

- We propose a novel Metaverse E-Commerce framework with SemCom to perform Malware IDS efficiently, thus addressing the privacy issues and computational demands of previous DL-based IDS in E-commerce.
- We propose a novel adversarial defense technique parameter using a substitute model with Stochastic Activation Pruning (SAP) [11], [12] to preserve SemCom IDS model gradients from adversarial attacks.
- The utilized real-world Malware dataset validates the efficacy of the proposed SemComm adversarial defense

in terms of high detection rates and adversarial defense.

II. RELATED WORK

Traditional malware detection methods like signature-based detection, behavioral analysis, heuristic analysis, and sandboxing spyware detection methods still face some hindrances like static matching of existing signatures, false positives, continual update requirements, and evasion of sandbox environments [7]. To address the above-mentioned drawbacks, previous security studies have employed DL-based malware IDS to handle issues related to computation and accurate and accurate malware IDS [13], [14]. Although their approach showed promising results, they lacked adversarial robustness [15]–[17]. Authors in [18] employed an AT mechanism to recognize malware samples from benign in the Internet of Things cyber-physical systems. While AT offers valuable benefits for malware IDS, it primarily focuses on the specific attack patterns used to train the model and is time and resource-consuming. It might not be effective against entirely blocking the transferability attacks. Similarly, authors in [19] employed an AT scheme to defend against adversarial attacks in hardware-enabled scenarios.

Following the advent of SemComm with DL, recent research in [20] has shown its higher potential in capturing semantic information better than traditional SL methods. Within the Metaverse and VR networks, authors in [10] demonstrated that SemComm DL models can handle the data-intensive network of the Metaverse and reduce the number of bits required for nodal transmission between sensors, users, wide area networks, cloud and edge servers. They also validated the severity of adversarial perturbation attacks in vehicular Metaverse.

To further strengthen DL-based SemComm systems, substantial research in [11] has validated the efficacy of adversarial parameter protection using SAP [12] to hide model weights from an attacker's knowledge and prevent IDS model poisoning attacks. For brevity, this study aims to better detect malware using SemComm encoder models with better efficiency and adversarial robustness compared to recent and previous studies.

III. METHODOLOGY

In the context of modeling a DL SemComm IDS using a DNN with representative Malware dataset [16], the adapted equation is represented in Equation 1 as [21] :

$$p(t \mid s) = \sum_{A \in \wp(s,t)} \left(\prod_{l=1}^{L} \hat{pl}(al \mid s)\right) \tag{1}$$

Where s represents the semantic (highly important) features extracted from network traffic data, t denotes the sequence of indicators representing malicious behaviors, $\wp(s,t)$ symbolizes all possible valid alignments of these indicators with the input features, al signifies an individual alignment within these possibilities, and $\hat{p}(al \mid s)$ calculates the probability of each alignment being indicative of an intrusion, given the malware features s. This SemComm approach enhances the detection capability of the IDS by utilizing DL to interpret and predict complex malware intrusion through semantic alignments.

A. *Proposed Adversarial Defense:*

Parameter protection using SAP is the adversarial defense technique employed to enhance the IDS of Metaverse E-commerce platforms. The proposed framework as represented in Fig. 2 is achieved by integrating an unpruned Malware IDS model with a decoy-pruned SemComm model (with SAP) for malware IDS. The pruned model with SAP employs a stochastic dropout process with sparsity for malware traffic features while preserving significant activations, thus boosting the model's efficiency and robustness against adversarial attacks (exposed weights and parameters). The efficacy of SAP is further emphasized by its goal to minimize losses from adversarial actions, formalized by policy optimization in Equation 2:

$$\pi^*, \rho^* := \arg\min_{\pi} \max_{\rho} E_{\rho \sim \pi, r \sim \rho}[J(M_p(\theta), x + r, y)] \tag{2}$$

Here, ρ denotes the adversary's policy, which introduces perturbations $r \sim \rho$ within the allowable space, ensuring $\|r\|$ is within the specified range. Conversely, π represents the defender's policy, initializing and guiding the response strategy $\rho \sim \pi$. On the other hand, the unpruned IDS model operates without sparsity constraints and is more focused on accurately identifying malware samples. The dual-model strategy, encompassing both pruned and unpruned models, substantially improves the IDS security framework of Metaverse E-commerce, guaranteeing reliable defense against adversarial perturbation or the total exposure of model gradients to adversarial attacks.

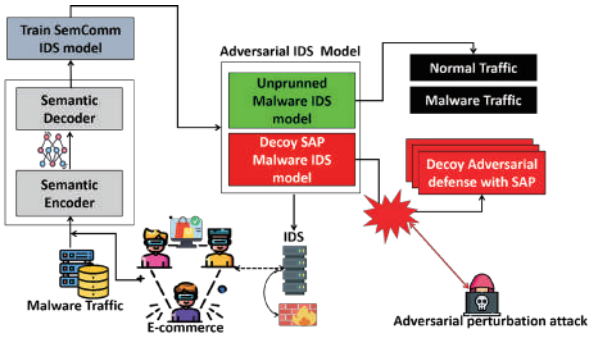


Fig. 2. Proposed SemComm IDS for Metaverse E-commerce with adversarial defense.

B. *Dataset:*

The real-world CICMalDroid2020 malware dataset consists of 11,598 samples categorized into five classes: Adware (1,253 samples), Banking malware (2,100 samples), SMS malware (3,904 samples), Riskware (2,546 samples), and Benign (1,795 samples) [16].

C. DNN SemComm:

The neural network architecture consists of an input layer determined by the number of features in the input data. Two hidden layers, each comprising 128 and 64 neurons, utilize ReLU activation functions. Dropout layers with a rate of 0.5 are applied after each hidden layer to prevent overfitting. The output layer is a Dense layer with a softmax activation function, suitable for multi-class classification tasks. Regularization techniques, including L1 and L2 regularization on kernel and bias parameters and activity regularization, are employed to prevent overfitting and encourage sparse representations. SAP is achieved with pruning implemented via $prune_low_magnitude from_tfmot.sparsity.keras$. The pruning schedule, defined as a PolynomialDecay, starts with an initial sparsity of 0.50 and reaches a final sparsity of 0.90. Fine-grained sparsity control is achieved through block size and pooling type specifications. The model is compiled with the Adam optimizer for optimization and training with a learning rate of 0.001.

D. Simulation Environment:

The experimentation used Python with the Tensorflow 2.13.0 framework within a Windows 10 environment. Computational setup: intel(R) Core(TM) i3-7100 CPU clocked at 3.90GHz, 16GB of RAM, and a Tesla K80 GPU.

IV. Performance Evaluation

The experimental results illustrate a comparative analysis between the base IDS model (without SAP) and the defense IDS model (with SAP). Several key metrics are evaluated: Accuracy, F1-Score, Precision, and Recall.

TABLE I
TABLE COMPARISON OF THE BASE MODEL VS PRUNNED MODEL

SemComm Model	Acc. %	F1-Score %	Precision %	Recall %
Base Model without SAP	89	87	88	87
Defense Model with SAP	82	80	81	79

Base SemComm IDS Model without SAP: The base model, as shown in Table I, which does not incorporate the SAP, exhibits higher values across all metrics, with an Accuracy, F1-Score, Precision, and Recall of 89%, 87%, 88%, and 87%, respectively. These metrics indicate that the base model performs robustly identifying malware accurately and consistently. The high precision and recall values suggest that the model is efficient at correctly classifying both positive and negative instances of malware without significant trade-offs.

Defense SemComm IDS Model with SAP:

The defense model, equally shown in Table I, which integrates SAP, shows slightly reduced Accuracy, F1-Score, Precision, and Recall performance of 82%, 80%, 81%, and 79%, respectively. Although there is a drop in these performance metrics compared to the base model as presented in the confusion matrix in Fig. 3, it is crucial to understand the broader context and benefits of the SAP integration, which is *Robustness against Adversarial Attacks*. The multiclass results of the Adware, Banking, SMS malware, Riskware, and Benign traffic are more accurately identified with the base model with an accuracy of 78%, 86%, 97%, 88%, and 86%, respectively, compared to that of the defense decoy model with SAP yielding a lower accuracy of 68%, 80%, 96%, 71%, and 80%, respectively.

The primary advantage of incorporating SAP is to enhance the model's resilience against adversarial samples. While this integration might result in a trade-off with slight decreases in accuracy, precision, recall, and F1-Score, the defense model is significantly more robust in adversarial environments. This robustness is not directly reflected in traditional performance metrics but is essential for maintaining reliable IDS performance under adversarial reconnaissance attack conditions.

Table II compares the baseline results of previous approaches with ours using the same dataset, thus proving its viability compared to the existing techniques regarding baseline accuracy and adversarial robustness. Table II presents a comparative analysis of our proposed contributions towards malware detection in Metaverse E-commerce of the Metaverse, juxtaposed with other relevant studies. Based on the proposed layered defense approaches and obtained results utilizing SemComm, it can be concluded that the IDS framework presented in this study significantly contributes to the detection of privacy-breaching malware attacks in Metaverse E-commerce, thus making the proposed framework a promising solution for enhancing security in the Metaverse.

TABLE II
COMPARISON WITH RELATED STUDIES.

Author	Year	Model	Dataset	Class	Adversarial Defense
[16]	2020	DNN	CICMaldriod	binary	✗
[17]	2023	LSTM	CICMaldriod	binary	✗
[15]	2023	Ensemble Trees	CICMaldriod	binary	✗
Proposed Defense	2024	SemComm DNN	CICMaldriod	binary	✓

V. Conclusion

This study presented a robust IDS framework for Metaverse E-commerce in the Metaverse, integrating SemComm with SAP to advance defenses against adversarial attacks. Despite a slight trade-off in traditional performance metrics, the defense model with SAP significantly enhances robustness and security, ensuring reliable malware detection under adversarial conditions. Comparative analysis with previous studies confirms the proposed method's superiority in maintaining high detection rates and adversarial resilience. This layered defense approach represents a promising solution for enhancing the security of Metaverse E-commerce, thus addressing privacy concerns within the Metaverse.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea

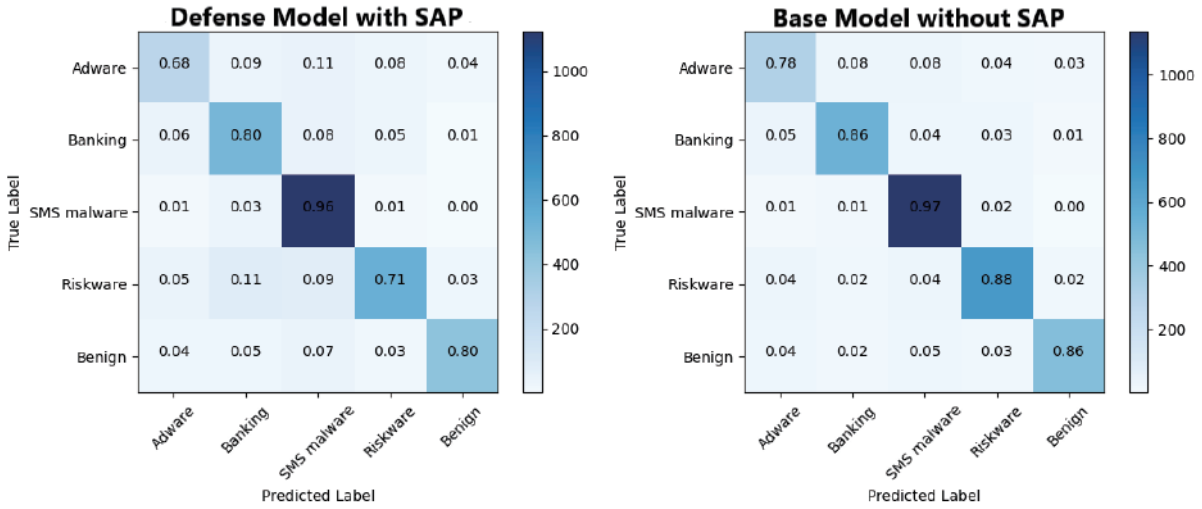


Fig. 3. Showing the confusion matrix of the multi-class malware classification of the SemmComm IDS model with and without adversarial pruning defense.

government(MSIT) (IITP-2024-2020-0-01612, 50%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 50%)

REFERENCES

[1] J. N. Njoku, C. I. Nwakanma, G. C. Amaizu, and D.-S. Kim, "Prospects and Challenges of Metaverse Application in Data-driven Intelligent Transportation Systems," *IET Intelligent Transport Systems*, vol. 17, no. 1, pp. 1–21, 2023.

[2] Y. Wang, Z. Su, N. Zhang, D. Liu, R. Xing, T. H. Luan, and X. Shen, "A survey on metaverse: Fundamentals, security, and privacy," *arXiv preprint arXiv:2203.02662*, 2022.

[3] E. C. Nkoro, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Industrial network attack vulnerability detection and analysis using shodan eye scanning technology," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 886–889.

[4] M. Vondráček, I. Baggili, P. Casey, and M. Mekni, "Rise of the metaverse's immersive virtual reality malware and the man-in-the-room attack & defenses," *Computers & Security*, vol. 127, p. 102923, 2023.

[5] A. Mutemi and F. Bacao, "E-commerce fraud detection based on machine learning techniques: Systematic literature review," *Big Data Mining and Analytics*, vol. 7, no. 2, pp. 419–444, 2024.

[6] N. Vullam, D. Roja, N. Rao, S. S. Vellela, L. R. Vuyyuru, and K. K. Kumar, "Enhancing intrusion detection systems for secure e-commerce communication networks," in *2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM)*, 2023, pp. 1–7.

[7] A. Souri and R. Hosseini, "A state-of-the-art survey of malware detection approaches using data mining techniques," *Human-centric Computing and Information Sciences*, vol. 8, no. 1, pp. 1–22, 2018.

[8] Z. Bazrafshan, H. Hashemi, S. M. H. Fard, and A. Hamzeh, "A survey on heuristic malware detection techniques," in *The 5th Conference on Information and Knowledge Technology*, 2013, pp. 113–120.

[9] K. Shin, Y. Lee, J. Lim, H. Kang, and S. Lee, "System api vectorization for malware detection," *IEEE Access*, vol. 11, pp. 53 788–53 805, 2023.

[10] J. Kang, J. He, H. Du, Z. Xiong, Z. Yang, X. Huang, and S. Xie, "Adversarial attacks and defenses for semantic communication in vehicular metaverses," *IEEE Wireless Communications*, vol. 30, no. 4, pp. 48–55, 2023.

[11] K. He, D. D. Kim, and M. R. Asghar, "Adversarial machine learning for network intrusion detection systems: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 538–566, 2023.

[12] E. C. Nkoro, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Explainable adversarial mitigation framework for zero-trust cyber warfare," in *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*, 2023, pp. 947–952.

[13] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, and S. Venkatraman, "Robust intelligent malware detection using deep learning," *IEEE Access*, vol. 7, pp. 46 717–46 738, 2019.

[14] A.-A. Mustafa Majid, A. J. Alshaibi, E. Kostyuchenko, and A. Shelupanov, "A review of artificial intelligence based malware detection using deep learning," *Materials Today: Proceedings*, vol. 80, pp. 2678–2683, 2023, sI:5 NANO 2021.

[15] R. Islam, M. I. Sayed, S. Saha, M. J. Hossain, and M. A. Masud, "Android malware classification using optimum feature selection and ensemble machine learning," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 100–111, 2023.

[16] A. Rahali, A. H. Lashkari, G. Kaur, L. Taheri, F. Gagnon, and F. Massicotte, "Didroid: Android malware classification and characterization using deep image learning," in *Proceedings of the 2020 10th International Conference on Communication and Network Security*, 2020, pp. 70–82.

[17] M. I. Sayed, S. Saha, and A. Haque, "Deep learning based malapps detection in android powered mobile cyber-physical system," in *2023 International Conference on Computing, Networking and Communications (ICNC)*, 2023, pp. 443–449.

[18] S. Patil, V. Varadarajan, D. Walimbe, S. Gulechha, S. Shenoy, A. Raina, and K. Kotecha, "Improving the robustness of ai-based malware detection using adversarial machine learning," *Algorithms*, vol. 14, no. 10, 2021.

[19] Z. Pan, J. Sheldon, and P. Mishra, "Hardware-assisted malware detection and localization using explainable machine learning," *IEEE Transactions on Computers*, vol. 71, no. 12, pp. 3308–3321, 2022.

[20] J. Kang, J. He, H. Du, Z. Xiong, Z. Yang, X. Huang, and S. Xie, "Adversarial attacks and defenses for semantic communication in vehicular metaverses," *IEEE Wireless Communications*, vol. 30, no. 4, pp. 48–55, 2023.

[21] Z. Weng, Z. Qin, X. Tao, C. Pan, G. Liu, and G. Y. Li, "Deep learning enabled semantic communications with speech recognition and synthesis," *IEEE Transactions on Wireless Communications*, 2023.

A Survey on Cybersecurity Techniques for AI based Intelligent Systems

Insoo Sohn
Dongguk University - Seoul
Seoul, Korea
isohn@dongguk.edu

Abstract - In this paper, we present technological trends related to cybersecurity in communication system intelligence, defense system intelligence, and medical system intelligence. Furthermore, a possible trend in the future security technology for AI is provided.

Keywords: Security, Artificial Intelligence, Intelligent Systems

I. INTRODUCTION

Cybersecurity is a technology that covers a wide range of area with the goal of providing security in computer systems, communication networks, and digital data that make up the infrastructure of individuals, organizations, and countries against internal or external cyber threats [1-3]. According to the Center for Strategic and International Studies (CSIS), the direct damage caused by cybercrime worldwide will reach $945 billion in 2020. In this paper, we present technological trends related to cybersecurity in communication system intelligence, defense system intelligence, and medical system intelligence. Furthermore, a possible trend, based on the survey, in the future security technology for AI is provided.

II. SECURITY FOR INTELLIGENT SYSTEMS

Unlike AI technology used to optimize individual functions in 5G, Native AI Networking technology is considered in 6G for autonomous system optimization of the enormous number of diverse communication sensor devices that exist in heterogeneous networks. AI-Native Networking technology does not take charge of some functions, but rather operates all functions of the communication system entirely by AI, for total communication system optimization. Artificial intelligence, which plays a key role in 6G, provides the possibility of expanding new cyber attack surfaces, and security technology research and development is needed to protect against the new threat.

Advanced military countries are investing enormous budgets to develop AI-based smart military weapon systems. In particular, along with intelligent combat robots that perform missions, the core of future wars is optimal tactical decision-making using AI technology based on big data on battlefield situations. Thus, AI is an essential element for strengthening future defense power in all areas for intelligent information analysis and combat performance. The goal of cyberattacks on future defense intelligence platforms is to seize control of the weapon system through precise strikes on the artificial intelligence system responsible for intelligence, to disable, target, and hijack combat robots, and to transmit contaminated information to the strategic decision-making system.

Artificial intelligence has recently been attracting attention in the field of medical intelligence, improving medical diagnosis through intelligent medical image processing and supporting rapid medical diagnosis by medical staff. Through data processing and information acquisition from medical big data, the medical artificial intelligence system shows performance close to that of a radiologist in the fields of skin cancer classification, early diagnosis, diabetic retinopathy classification, and pneumonia detection through chest X-ray. According to a recent study on classification and segmentation of medical images, state-of-the-art intelligent medical systems are significantly vulnerable to malicious cyberattacks.

Proliferation of AI based intelligent systems, in every corner of human society, is steering the humanity into a new course for the Age of Ubiquitous Intelligence. However, the humanity is also facing the threat of cyber attacks in Everything, Everywhere, and All At Once. Global investment into cybersecurity research and development programs is imperative to provide secure businesses and public services in the new era of digital world.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00252328).

REFERENCES

[1] S. Kaviani, S. Shamshiri, and I. Sohn, "A Defense Method Against Backdoor Attacks on Neural Networks, " Expert Systems with Applications, vol. 213, March 2023, 118990 based on Complex Network Theory: A Survey, " Mathematics, vol. 11, no. 2, Jan. 2023

[2] M. Mohammadi and I. Sohn, "AI based Energy Harvesting Security Methods: A Survey, " ICT Express, vol. 9, no. 6, pp. 1198-1208, Dec. 2023

[3] M. Pooyandeh and I. Sohn, "Smart Lithium-ion Battery Monitoring in Electric Vehicles: AI Empowered Digital Twin Approach, " Mathematics, vol. 11, no. 23, Dec. 2023 Mathematics, vol. 11, no. 2, Jan. 2023

PhishNet: Exploring Machine Learning Techniques for Advanced URL Phishing Detection

Sean Andrei Olores*, Renee Althea Khan†, Shanley Valenzuela‡, Gabriel Avelino Sampedro§
*†‡College of Computer Studies, De La Salle University, Manila, Philippines
§College of Engineering, University of the Philippines, Diliman, Philippines
Email: {sean_olores*, renee_khan†, shanley_valenzuela‡}@dlsu.edu.ph; garsampedro@ieee.org§

***Abstract*—Cyber threats, particularly phishing — a prevalent cybercrime exploiting human vulnerabilities to steal sensitive information — are rapidly evolving, leading to an increased number of organizations and individuals falling prey to these attacks. As cyber attackers become more sophisticated in their tactics, organizations must continually update their cybersecurity measures to stay ahead of potential threats. To help address this growing issue, this research explores the use of various machine learning models for detecting phishing URLs across a dataset of 10,000 web pages. Random Forest (RF), Support Vector Classifier (SVC), Logistic Regression (LR), and Naïve Bayes (NB) are all explored and ensembled to form the algorithm PhishNet, possibly increasing performance. This research shows that the PhishNet outperformed individual models in detecting phishing URLs with a notable accuracy of 0.9880, demonstrating the potential of machine learning in enhancing cybersecurity measures. By leveraging these advanced technologies, organizations can better protect themselves against evolving cyber threats and safeguard sensitive information.**

***Index Terms*—Phishing detection, machine learning, ensemble learning, URL classification, cybersecurity.**

I. Introduction

Phishing is a serious cybercrime where individuals are tricked into revealing personal and financial information through deceptive communications. These attacks often impersonate reputable organizations via email, websites, and other online platforms. As digital communication increases, so do phishing attempts, which are becoming more sophisticated and pose significant cybersecurity challenges. In the fourth quarter of 2019 alone, nearly five million people were victims of phishing attacks [1]. Traditional anti-phishing methods, like URL blocklisting, only identify known malicious sites, leaving newly created phishing sites undetected [2].

This research aims to improve phishing detection by developing an ensemble model, PhishNet, which combines multiple machine learning algorithms. Using a dataset from Kaggle with 10,000 web pages, we train and test our model, integrating Random Forest, Support Vector Classifier, Logistic Regression, and Naïve Bayes. The ensemble approach enhances the accuracy of individual models and adapts to evolving phishing tactics.

II. Review of Related Literature

Phishing is an escalating threat in cybersecurity as cybercriminals continually evolve their tactics. This has spurred the adoption of machine learning to enhance phishing detection. Vaitkevicius and Marcinkevicius evaluated several machine learning algorithms, noting strong performances from Neural Networks and Random Forest [2]. Burbela furthered this by demonstrating that an ensemble of Convolutional Neural Networks (CNN) and Multi Head Self-Attention (MHSA) exceeded the performance of individual models, achieving an accuracy of 98.3% [3].

Despite these advances, a significant gap remains in the models' ability to adapt to new phishing techniques and integrate real-time data, limiting their effectiveness against emerging threats. Additionally, deep learning has shown promise in improving the resilience of intrusion detection systems, especially in dynamic environments like the Internet of Things (IoT), as evidenced by Abbas et al. [4].

While current research has improved the accuracy and adaptability of phishing detection systems, the ongoing challenge is developing methods that can continuously adapt to new phishing strategies and ensure robust long-term protection.

III. Methodology

A. Dataset Description

The dataset, curated by Choon Lin Tan, consists of 10,000 web pages, equally divided into 5,000 phishing and 5,000 legitimate pages. It includes 48 attributes such as domain age, presence of IP addresses, and HTTP tokens, collected between 2015 and 2017 [5].

B. Data Processing

We performed several steps to prepare the dataset for model training:

- **Cleaning**: Checked for and confirmed the absence of duplicates and missing values.
- **Data Splitting**: Divided the dataset into 80% training and 20% testing subsets.
- **Feature Selection**: Removed constant features and those with no significant contribution to the predictions.
- **Standardization**: Applied standardization to ensure all features are on a common scale, which enhances model performance.

§Corresponding author

C. Ensemble Model Development

We developed an ensemble model, PhishNet, to leverage the strengths of multiple machine learning algorithms. The ensemble approach integrates the outputs of Logistic Regression, Naïve Bayes, Support Vector Classifier, and Random Forest using a meta-learner (Logistic Regression). This method aims to improve the robustness and accuracy of phishing detection by combining the diverse predictive capabilities of individual models.

IV. Performance Evaluation

The key metrics used to assess performance include accuracy, precision, recall, and F_1-score:

- **Accuracy (ACC)**: Proportion of total true results (both true positives and true negatives) in the data.
- **Precision (PREC)**: Proportion of true positive results in all positive cases identified by the model.
- **Recall (REC)**: Proportion of true positive results identified from all actual positives.
- **F_1-score (F_1)**: Harmonic mean of precision and recall, indicating a balance between the two.
- **Time Complexity (TC)**: Time in milliseconds in which the detection algorithm runs.

TABLE I
PERFORMANCE METRICS OF THE CLASSIFICATION ALGORITHMS

Algorithm	ACC	PREC	REC	F_1	TC
LR	0.9355	0.9379	0.9359	0.9355	0.99 ms
NB	0.8065	0.8423	0.8044	0.8025	1.57 ms
SVC	0.9655	0.9653	0.9655	0.9654	309.29 ms
RF	0.9845	0.9840	0.9845	0.9842	33.34 ms
PhishNet	**0.9880**	**0.9876**	**0.9879**	**0.9877**	**350.08 ms**

In terms of accuracy, precision, recall, and F_1-score, the proposed PhishNet algorithm performed the best. However, regarding time complexity, PhishNet performed the least, while LR performed the best. The poor performance may be attributed to how the ensembling algorithm applies all the other four algorithms. Despite being the slowest, the 350.08 ms time complexity may be neglected due to the application being time-insensitive.

In evaluating model performance, the Receiver Operating Characteristics (ROC) and Area Under The Curve (AUC) metrics provide crucial insights:

- **ROC Curve**: Displays the true positive rate (recall) versus the false positive rate at various threshold settings. A curve closer to the top-left corner indicates better performance.
- **AUC**: The area under the ROC curve, with a value of 1.00, indicates perfect discrimination between classes. The Random Forest and the ensemble model achieved an AUC of 1.00, showcasing their ability to distinguish between phishing and legitimate URLs.

This high level of performance implies that the models are exceptionally effe ctive at making accurate classifications, enhancing their reliability in detecting phishing attempts.

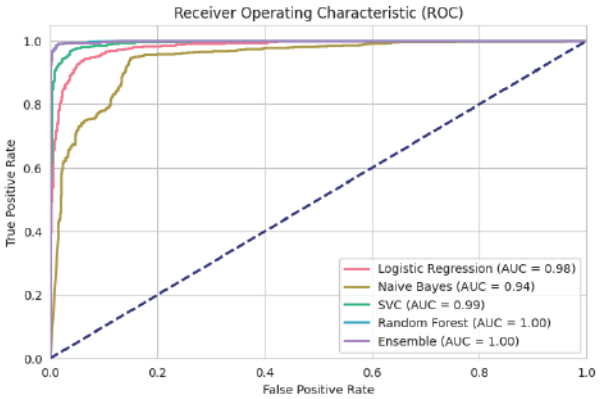


Fig. 1. Receiver Operating Characteristic (ROC) curves for the individual algorithms and the ensemble model.

V. Conclusion

This research evaluated various machine learning algorithms for their efficacy in detecting phishing URLs, culminating in developing a superior ensemble method. The study revealed that the Random Forest algorithm excels in classification accuracy. However, the ensemble approach, which combines the strengths of Logistic Regression, Naïve Bayes, Support Vector Classifier, and Random Forest, significantly enhances performance. This method not only elevates predictive accuracy but also increases the system's robustness against diverse phishing sites.

Since the model was trained and tested on a single dataset comprising a limited number of URLs, the researchers recommend further training on more datasets that arise in the future and real-time data from credible and legal sources to enhance the model's performance. Moreover, integrating the ensemble model with security systems as APIs or plug-ins can improve the protection against phishing attacks on various platforms.

Future work could focus on optimizing the ensemble process to reduce time complexity in order to make PhishNet more suitable for real-time applications. Potential optimization techniques include model pruning, parallel processing, and using more efficient algorithms for real-time use.

References

[1] APWG, "Apwg — phishing activity trends reports," https://apwg.org/trendsreports/, 2019.

[2] P. Vaitkevicius and V. Marcinkevicius, "Comparison of classification algorithms for detection of phishing websites," *Informatica*, vol. 31, no. 1, pp. 143–160, 2020.

[3] K. Burbela, "Model of detection of phishing urlsbased on machine learning," 2023.

[4] S. Abbas, S. Alsubai, S. Ojo, G. A. Sampedro, A. Almadhor, A. A. Hejaili, and I. Bouazzi, "An efficient deep recurrent neural network for detection of cyberattacks in realistic iot environment," *The Journal of Supercomputing*, pp. 1–19, 2024.

[5] C. L. Tan, "Phishing dataset for machine learning: Feature evaluation," 2018, data Mining; Machine Learning; Feature Selection; Information Security. [Online]. Available: https://data.mendeley.com/datasets/h3cgnj8hft/1

Reinforcement Learning-Based Optimal Task Offloading for Multi-Tier Cloud-Edge Computing Systems

Avilia Kusumaputeri Nugroho
Pusan National University
Busan, Republic of Korea
avilia22@pusan.ac.kr

Min Sang Yoon
Fujitsu Network Communications
Richardson, Texas, United States
minsang.yoon@fujitsu.com

Taewoon Kim*
Pusan National University
Busan, Republic of Korea
taewoon@pusan.ac.kr

***Abstract*—This work introduces online resource scheduling in a multi-tier cloud-edge computing system that integrates Proximal Policy Optimization (PPO) reinforcement learning with state stacking. By integrating PPO with state stacking, our approach allows agents to effectively balance between exploration and exploitation, thereby allowing agents to effectively capture long-term dependencies and patterns in the environment. The proposed approach aims to enhance system efficiency through integrating load balancing and optimal resource allocation. For extensive evaluation, we have implemented a customized reinforcement learning environment and evaluated the effectiveness of the proposed strategy in ensuring prompt task completion while minimizing energy consumption for task processing.**

***Index Terms*—Cloud computing, edge computing, task offloading, reinforcement learning, PPO**

I. Introduction

The edge-cloud concept has emerged due to the rapid growth of cloud and edge computing in recent years. Edge servers connect within edge-cloud networks to minimize the demand on the cloud. This not only addresses the challenge of managing data sources over long distances but also the demand for low-latency applications by bringing computation units closer to users [1]. Edge computing, particularly task offloading, enables application tasks to be transferred from user equipment (UE) to the network edge. The goal of this approach is to minimize the delay in data transmission and the amount of energy consumed. However, efficiently managing resources in dynamic edge environments presents significant challenges. Traditional integer programming solutions [2], while widely used, struggle with high time complexity and lack adaptability to dynamic conditions, making them unsuitable for real-time, online scheduling.

Deep Reinforcement Learning (DRL) integrates Reinforcement Learning (RL) with Deep Neural Networks (DNN), offering a promising approach to address various challenges across different domains, such as robotics [3], gaming [4], and autonomous vehicles [5]. DRL enables systems to achieve flexible and adaptable task offloading without relying on expert knowledge, seeking to optimize task offloading decisions by acquiring an efficient policy through interactive experiences with the environment and maximizing rewards. While DRL has shown promise in addressing task offloading challenges [6], [7], existing methods often overlook task dependencies, assuming tasks are independent.

To address these challenges, we propose an approach that integrates the Proximal Policy Optimization (PPO) [8] reinforcement learning algorithm with state stacking for online resource scheduling in edge computing. By training agents to interact with the environment and learn optimal policies, we enable adaptive resource allocation and task offloading. State stacking refers to the method of arranging a sequence of observations into stacks, thereby allowing the agent to have a historical record of observations as input. The objective of this approach is to enhance the temporal data accessible to the agent, thus enhancing comprehension of the dynamic environment. By recognizing patterns over time, agents can make better decisions based on observational history. Although state stacking enhances temporal understanding, it may necessitate additional computational resources and memory due to the utilization of multiple stacked states as input. Therefore, optimizing the state stack size becomes crucial for balancing computational efficiency and performance.

The remainder of this paper is structured as follows. Section II presents an explanation of a system design based on reinforcement learning, addressing a scheduling problem, PPO, and state stacking. Section III discussed the experimental setup and simulation results, and Section IV concludes this paper.

II. Proposed Idea

In this section, we present an explanation of a system design based on reinforcement learning, addressing a scheduling problem, and PPO. Following that, we discuss state stacking.

This research was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-RS-2023-00259967) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation) and in part by IITP under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00254177) grant funded by the Korea government(MSIT)

Corresponding author: Taewoon Kim (taewoon@pusan.ac.kr)

A. System Design Based on Reinforcement Learning

In our proposed scheduling system design, we propose a multi-tier cloud-edge computing environment for online scheduling. This system assumes unlimited resources at the cloud level and utilizes edge servers equipped with flexible containers to process various application tasks. Each edge server (ES) is assumed to host two distinct containers, each with a capacity of 50%, allowing for dynamic task processing based on situational demands. Mobile devices may send tasks to nearby edge servers for processing. Tasks are routed to the cloud or processed locally if no suitable edge server is available, based on efficiency considerations. Depending on where each task is processed, the resulting service response time may differ.

Tasks are categorized based on specific application numbers and are influenced by various factors such as cycles per bit, task data size, and delay tolerance. In our environment, tasks can be processed in multiple time states in order to represent the dynamic of the system, where tasks arrive and are processed continuously throughout time. This allows agents to explore different workload conditions and resource utilization. Additionally, our proposed approach integrates DRL with task offloading optimization, emphasizing the association between policy optimization and task offloading strategies.

Our approach revolves around a policy optimization training program that utilizes PPO and integrates state stacking. PPO agents are responsible for optimizing policies by continuously monitoring and training them to improve decision-making. To ensure stability and efficiency, we employ the clipped function to limit policy changes. We use state stacking to enhance temporal understanding and improve the agent's ability to make informed decisions. As a result, the trained optimal policy is integrated into the task scheduler within a multi-tier cloud-edge environment. The primary objective is to achieve online resource scheduling with integrated load balancing, resulting in improved system efficiency and performance.

B. Scheduling Problem

Tasks are generated from user devices, and the task T_n indexed by an increasing counter $n \in \mathbb{Z}_+$, is defined as $T_n = (K_n, C_n, D_n, \delta_n)$ where K_n denoted as the unique application number generating the task, C_n represents the CPU cycles needed to complete the task, D_n represents task data size, and δ_n represents delay tolerance.

Following that, the agent evaluates the environment state by monitoring various states such as task arrival time (t_n), task (T_n), availability of bandwidth (b_n), offloading targets (o_n), device locations (l_n), and available edge server resources (e_n), these include container status (s_n), remaining CPU capacity of each container (c_n), and the type of task application being processed by each container (a_n). The state space can be defined as:

$$S_t = \{t_n, T_n, b_n, o_n, l_n, e_n\}, \tag{1}$$

where

$$e_n = \{s_n, c_n, a_n\} \tag{2}$$

The action space is structured to accommodate decisions for local processing, edge offloading, and cloud offloading, denoted as:

$$\mathcal{A} = \{\mathcal{A}_{\text{local}}, \mathcal{A}_{\text{edge}}, \mathcal{A}_{\text{cloud}}\} \tag{3}$$

If the agent chooses to offload to the edge, there are further choices. Let E_n represent the set of edge servers, and $C_{n,m}$ denote the set of containers on edge server E_n, where m represents the index of the container on server E_n. The agent's decision-making process extends to selecting an edge server E_n and a container $C_{n,m}$ from the respective sets E_n and $C_{n,m}$. Thus, it can be defined as:

$$\mathcal{A}_{\text{edge}} = \{(E_n, C_{n,m}) \mid E_n \in E, C_{n,m} \in C\}, \tag{4}$$

where E represents the set of all edge servers, and C represents the set of all containers.

The reward function is designed to optimize performance while considering energy and time constraints. The reward function can be expressed as:

$$\text{Reward} = -(w_1 \cdot \frac{T_{n,\text{total}}}{c_1} + w_2 \cdot \frac{E_{n,\text{total}}}{c_2} + w_3 \cdot \frac{\text{Pr}_{n,\text{total}}}{c_3}) \tag{5}$$

It is defined as a weighted sum of different components, including the total task completion time ($T_{n,\text{total}}$), energy consumption ($E_{n,\text{total}}$), and price ($\text{Pr}_{n,\text{total}}$). The weights w_1, w_2, and w_3 determine the importance of each component in the overall reward calculation, where $\sum_{\forall i} w_i = 1, \forall w_i \geq 0$. The constants, c_i, are empirically computed constant factors used to normalize corresponding terms. The objective is to minimize the total cost (or maximize the negative reward) while maximizing rewards. Successful task completion within the delay constraints yields a total reward for the agent, while failure results in a penalty.

C. PPO

Policy-based DRL utilizes neural networks as policies to learn decision-making processes. In this approach, the agent interacts with the environment's state and determines actions based on estimated probability distributions. PPO ensures consistent training and policy updates by limiting changes to ensure optimal utilization of samples. The optimization objective of PPO can be defined as follows [8]:

$$L(\theta) = E_t \left[\frac{\pi_{\theta_{old}}(a_t \mid s_t)}{\pi_\theta(a_t \mid s_t)} A^{\pi_{\theta_{old}}}(s_t, a_t) \right], \tag{6}$$

where $\pi_{\theta_{old}}(a_t \mid s_t)$ represents the probability of selecting action a_t given state s_t under policy θ_{old}, $A^{\pi_{\theta_{old}}}(s_t, a_t)$ represents the advantage function under the old policy $\pi_{\theta_{old}}$, and θ_{old} denotes the parameters of the old policy.

$$L(\theta) = E_t \left[\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{old}}(a_t \mid s_t)} (Q^{\pi_{\theta_{old}}}(s_t, a_t) - V^{\pi_{\theta_{old}}}(s_t)) \right] \tag{7}$$

$$L(\theta) = E_t \left[\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{old}}(a_t \mid s_t)} (r_t + \gamma V^{\pi_{\theta_{old}}}(s_{t+1}) - V^{\pi_{\theta_{old}}}(s_t)) \right] \tag{8}$$

The equation (7) updates the policy by considering the difference between two functions: $Q^{\pi_{\theta_{old}}}(s_t, a_t)$, representing

the action value based on the old policy, and $V^{\pi_{\theta_{old}}}(s_t)$, representing the state value based on the old policy, aiming to optimize the policy by evaluating the advantage of certain actions relative to the average action in a given state under the old policy.

The equation (8) updates the policy by incorporating the immediate reward r_t received at time step t, along with the discounted value of the next state s_{t+1} and the value of the current state s_t under the old policy, aiming to maximize the expected return over time while ensuring stability and adaptability during the learning process.

The improvement in the policy lower bound must be positive, indicating that the new policy outperforms the old one. Policy updates involve a clipping function to constraint policy parameter changes, ensuring stability and adaptability. The clipping function is defined as:

$$L_{\text{CLIP}}(\theta) = E_t \Bigg[\min \Bigg(\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{old}}(a_t \mid s_t)} A^{\pi_{\theta_{old}}}(s_t, a_t), \\ \text{clip} \left(\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{old}}(a_t \mid s_t)}, 1-\epsilon, 1+\epsilon \right) A^{\pi_{\theta_{old}}}(s_t, a_t) \Bigg) \Bigg] \tag{9}$$

The clipping function, named clip, constrains the probability ratio within the range of $1-\epsilon$ to $1+\epsilon$, where ϵ is a small, non-negative constant representing the clipping parameter. This constraint maintains stability and adaptability during policy updates in PPO.

D. State Stacking

State stacking, which we propose integrates with PPO, enhances temporal understanding by concatenating sequential observation states. This approach provides agents with a comprehensive record of previous observations, enabling them to accurately identify and understand patterns that occur in particular sequences. Therefore, it enhances their ability to comprehend and react to temporal dependencies within the environment. Since our tasks can be processed in multiple time states, there is a correlation between states, as tasks are still processed several sequential times. Thus, we propose using state stacking as an approach for agents to effectively explore different workload conditions and resource utilization, which will result in enhanced performance. State stacking can be represented as:

$$S_t = S(s_t, s_{t-1}, ..., s_{t-n+1}), \tag{10}$$

where S_t represents the combined state at time t, which is a function S of the current state s_t and the previous $n-1$ states, providing a history that helps agents perceive temporal dependencies and capture sequential patterns in the environment.

III. Evaluation

For our experimental setup, 20 edge servers and 25 devices are randomly distributed throughout a 400 m $\times$ 400 m area. Each edge server has been designed with flexible containers that have the capability to execute a variety of application tasks. We assume that each ES hosts 2 distinct containers, each with a capacity of 50%. This allows for dynamic task processing according to current demands. Each device operates with a transmission power of 500 mW and an idle power consumption of 100 mW. Additionally, we assume unlimited cloud resources for offloaded tasks, which may incur additional network delays. Tasks, including data sizes measured in Kbits, CPU cycles measured in Megacycles, and delay tolerances measured in milliseconds, are determined during simulation based on uniform distributions ranging from 600–1200, 900–1100, and 100–200, respectively.

We employ PPO state stacking for training, with a learning rate (lr) of 0.001, a discount factor (γ) of 0.99, a GAE parameter (gae_lambda) of 0.97, a PPO clip parameter ($clip_epsilon$) of 0.2, an entropy coefficient (ent_coef) of 0.01. For comparison, we have also implemented the following RL algorithms:

- *Proposed (n=4)*: The method proposed in this paper employs PPO with state stacking of 4 previous state sequences.
- *Proposed (n=8)*: The method proposed in this paper employs PPO with state stacking of 8 previous state sequences.
- *PPO* : The original formulation of the PPO algorithm.
- *DQN* (Deep Q-Network): A reinforcement learning algorithm that integrates Q-learning.

The implementations of PPO and DQN are obtained from the stable baseline implementation [9], and it is assumed n=1 unless mentioned otherwise.

In Fig. 1, all algorithms prioritize offloading tasks to the edge in order to minimize the reliance on local processing or cloud offloading. Both versions of PPO state stacking (n=4 and n=8) stand out by surpassing both PPO and DQN, achieving the highest percentage of edge offloading. PPO state stacking also has the lowest percentages in both local processing and cloud offloading. This shows that the PPO state stacking consistently optimizes the allocation of tasks to the edge, which reduces both the battery consumption of the user device and network delay.

In Fig. 2, it shows that energy consumption grows accordingly with the size of the task data. However, our proposed solution shows the lowest energy consumption compared to PPO and DQN. These results prove that both versions of PPO state stacking (n=4 and n=8) efficiently allocate tasks and utilize resources in an optimal manner.

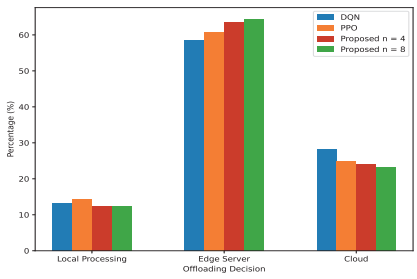


Fig. 1. The offloading decision comparison

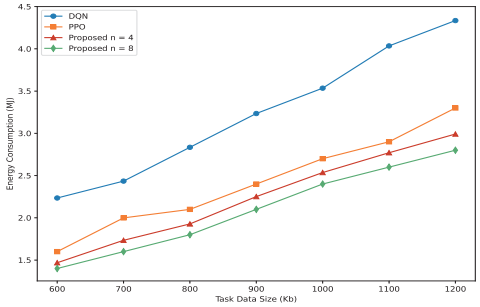


Fig. 2. The energy consumption across various task data sizes

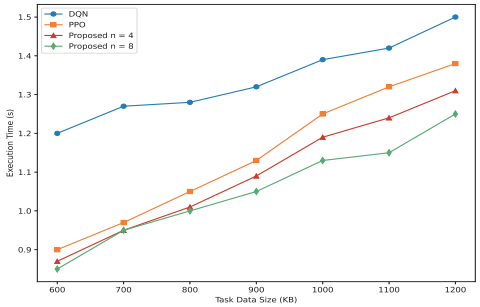


Fig. 3. The task execution time across various task data sizes

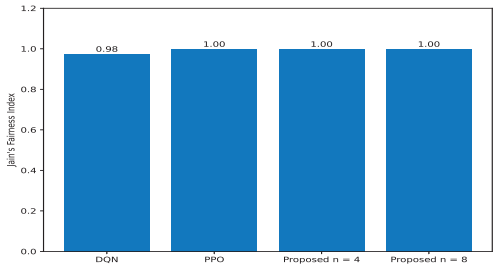


Fig. 4. The fair task allocation among containers on edge servers

In Fig. 3, the time of task execution increases as the number of tasks increases. Both versions of PPO state stacking (n=4 and n=8) achieves minimized task execution times. Furthermore, Fig. 4 illustrates the load balancing performance among containers on edge servers, evaluated using Jain's fairness index [10] as a measure of fairness. The result shows that both versions of PPO state stacking (n=4 and n=8) and PPO show the highest degrees of fairness compared to DQN.

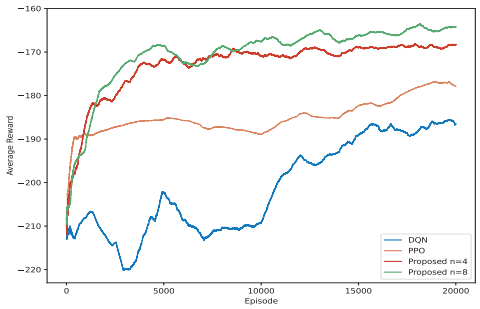


Fig. 5. The average reward during training process

The Fig. 5 shows the mean reward values obtained during the training period. The reward value shows the smoothed slope of the mean reward values. The average reward for all algorithms consistently demonstrates a gradual improvement over time, showing their ability to learn within the environment. The DQN shows an annual and gradual increase in the average reward. In contrast, the PPO shows an initial period of significant improvement, which is then followed by a slower rate of increase. This pattern indicates efficient early exploration. However, PPO state stacking (n=8) demonstrates greater overall performance, earning the highest average reward compared to other approaches. Intregating state stacking into the PPO enhances the agent's learning capabilities.

IV. Conclusion

We present the PPO state stacking approach for online resource scheduling in edge computing. Our proposed approach aims to optimize resource allocation and load balancing in order to increase system efficiency. Through evaluation, we have shown the efficiency of our method in achieving shorter task completion times while minimizing the energy consumed for task processing. Through the process of stacking states, agents capture temporal dependencies and patterns, enabling more informed decision-making regarding resource allocation and task scheduling, which results in a higher average reward. The evaluation results show that both versions of PPO state stacking (n=4 and n=8) outperform both PPO and DQN without state stacking in terms of offloading decisions, energy efficiency, and task execution time, while also maintaining a high level of fairness in load balancing.

References

[1] K. Cao, Y. Liu, G. Meng, and Q. Sun, "An overview on edge computing research," *IEEE Access*, vol. 8, pp. 85714–85728, 2020.

[2] A. K. Nugroho, S. Shioda, and T. Kim, "Optimal resource provisioning and task offloading for network-aware and federated edge computing," *Sensors*, vol. 23, no. 22, pp. 1–22, 2023.

[3] U. Patel, N. K. S. Kumar, A. J. Sathyamoorthy, and D. Manocha, "Dwa-rl: Dynamically feasible deep reinforcement learning policy for robot navigation among mobile obstacles," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6057–6063, IEEE, 2021.

[4] J. Li, S. Koyamada, Q. Ye, G. Liu, C. Wang, R. Yang, L. Zhao, T. Qin, T.-Y. Liu, and H.-W. Hon, "Suphx: Mastering mahjong with deep reinforcement learning," *arXiv preprint arXiv:2003.13590*, 2020.

[5] J. Liao, T. Liu, X. Tang, X. Mu, B. Huang, and D. Cao, "Decision-making strategy on highway for autonomous vehicles using deep reinforcement learning," *IEEE Access*, vol. 8, pp. 177804–177814, 2020.

[6] H. Li, K. D. R. Assis, S. Yan, and D. Simeonidou, "Drl-based long-term resource planning for task offloading policies in multiserver edge computing networks," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4151–4164, 2022.

[7] H. Huang, Q. Ye, and H. Du, "Reinforcement learning based offloading for realtime applications in mobile edge computing," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2020.

[8] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[9] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: Reliable reinforcement learning implementations," *Journal of Machine Learning Research*, vol. 22, no. 268, pp. 1–8, 2021.

[10] R. K. Jain, D.-M. W. Chiu, W. R. Hawe, *et al.*, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, vol. 21, 1984.

Differentiated Traffic Class Delay Modeling for XG-PON Using ANN

Nur Asfahani Ismail[1,2], Farabi Iqbal[2], Sevia Mahdaliza Idrus[2], Nor Affida M.Zin[1,2], Fadila Mohd Atan[1,2]
[1]*College of Engineering, UiTM Johor Branch Pasir Gudang Campus*
[2]*Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia*
alfarabi@utm.my

***Abstract*—Passive Optical Networks (PONs) play a crucial role in providing broadband services with high capacity and low latency. The XG-PON, a 10 Gigabit-capable Passive Optical Network, improves data speeds and scalability, effectively catering to numerous clients with varying bandwidth requirements. Dynamic Bandwidth Allocation (DBA) in XG-PON is essential for the flexible distribution of bandwidth among users, guaranteeing equitable and efficient utilization of the network. This research presents a novel methodology for predicting delays in various traffic containers in XG-PON. By employing supervised training, we show that Artificial Neural Networks (ANN) can accurately simulate upstream delays, thereby aiding in making informed decisions on bandwidth allocation. The suggested approach seeks to enhance the efficiency of DBAs by offering precise delay estimates, hence improving the overall performance of the network.**

Keywords—ANN, delay modeling, dynamic bandwidth allocation, XG-PON.

I. Introduction

Passive optical networks (PONs) are now recognised as a vital technology for providing high-capacity, low-latency services to end users in order to meet the growing demand for high-speed broadband access. The 10 Gigabit-capable Passive Optical Network (XG-PON) is a PON technology that distinguishes itself from other PON technologies by offering high data rates, scalability, and efficient service to a large number of customers with different bandwidth requirements. XG-PON utilises a tree architecture, wherein a central node called the Optical Line Terminal (OLT) is connected to several endpoints referred to as Optical Network Units (ONUs) or Optical Network Terminals (ONTs) through a branching structure, as depicted in Fig. 1. The OLT is often located at a central office or data centre, serving as the primary hub for data transmission. From the OLT a single optical fibre extends to a splitter, which divides the signal into many branches. These branches, in turn, link to the ONUs or ONTs, which are often located at the premises of residential or business users. The tree topology is advantageous for XG-PON as it efficiently utilises the fiber-optic infrastructure. Using a splitter allows for the optical signal to be divided into several outputs, which provides a cost-effective way to connect many subscribers without requiring separate fibre lines from the central office to each user. This collaborative structure reduces the overall cost of installation while enabling scalability, as additional branches can be integrated to accommodate a greater number of users.

However, because of the interconnected nature of the tree topology, it is essential to allot bandwidth in a flexible way among users to ensure fair and efficient use of the network. Dynamic Bandwidth Allocation (DBA) [1], [2] is crucial in this context because it allows the OLT to assign upstream bandwidth based on real-time demand and Quality of Service (QoS) requirements. DBA plays a vital role in PONs, especially in scenarios that involve frequent modifications or fluctuations. In order to ensure optimal network performance and adhere to the service level agreements (SLAs) for various users and applications, the DBA permits the adjustment of resource allocation based on real-time demand. Efficient DBA requires accurate analysis and prediction of network delay in order to make informed decisions regarding bandwidth allocation. According to ITU-T Recommendation, DBA for XG-PON is categorized as status reporting and traffic monitoring. The former has an advantage of efficient bandwidth allocation as it is based on demands from ONUs. Status reporting is also widely used as basis for DBA algorithms as presented in [3]–[6]. Although it is perceived as the best option for DBA, it can also lead to reporting overhead that causes inefficient bandwidth consumption [7].

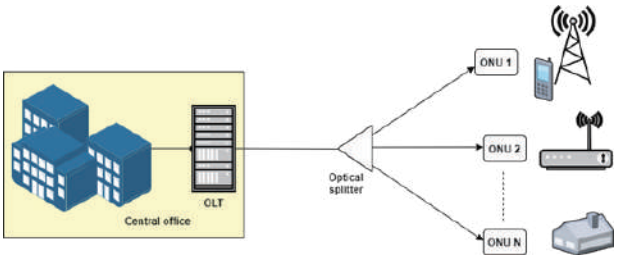


Fig. 1. PON with tree topology.

Recent development in DBA is the implementation of machine learning used to predict bandwidth allocation [8]–[10] and self-adaptive [11]. Inaccurate predictions of delays can lead to inefficient distribution of resources, causing network congestion, increased latency, and reduced QoS. The intricacy of this task is heightened in XG-PON due to the substantial volume of traffic and varied service demands, rendering it unfeasible to manually modify and allocate resources in a fixed manner. Delay modeling is essential as it allows for the identification of delay sources inside the network and enhances understanding of how these delays impact various applications. The delays may arise from several sources, including signal propagation along the fiber, processing time at the OLT and ONUs and queue delays resulting from bandwidth allocation decisions [12][13]. In this paper, we propose an upstream delay model for various traffic containers in XG-PON. Supervised training demonstrates that upstream delay can be effectively modeled using ANN to aid in bandwidth allocation decisions.

II. Methodology

Figure 2 illustrates a flowchart depicting the process of supervised training model. The selected input features are obtained from simulation. The network simulation setup is the

same as in [7]. The data collection phase is conducted using network simulation. The XG-PON network comprises one OLT and 16 ONUs. The traffic arrival process at the ONUs follows a Poisson distribution, determined by the exponential value of the frame arrival rate.

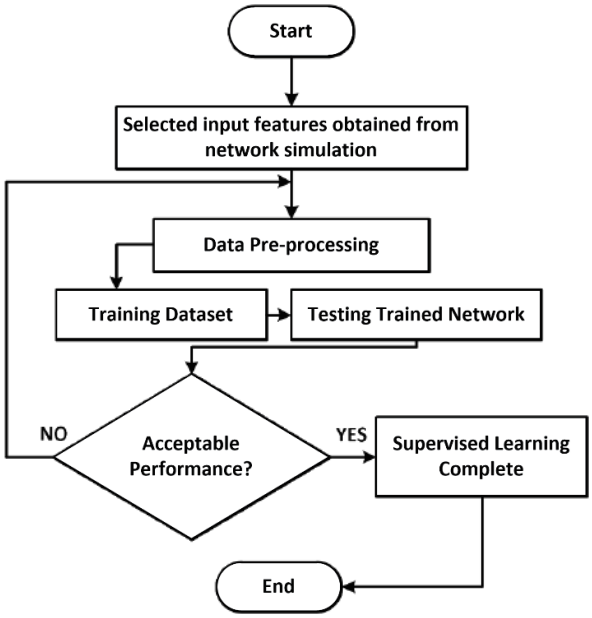


Fig. 2. Flowchart of the ANN supervised training.

A. Data Collection

Table 1 lists the description of each input features to the ANN model. The network features are formulated as X={ρ, AllocIDk,i, REPORTk, BWalloc} as the input feature vector and the target output y is therefore the average upstream delay. The simulation time is set to 1 second, equivalent to 8000 frames. A total of 5000 samples was collected for the purpose of supervised training. The data was collected for various normalized traffic load ranging from 0.1 to 1.

B. Data Preprocessing

All zeros data are removed from the datasets thus the input and output of ANN have the same length to ensure smooth training and testing process. They are split into input X and output Y variables, where X contains independent variables and Y contains the dependent variable. The datasets contain 5000 samples for each T-CONT, are split into 80% training and 20% testing and then standardized by calculating the mean and standard deviation of the training dataset. The standardized value, Z is calculated using Z=(X-μ)/σ where X is the original value, μ is the mean value and σ is the standard deviation value.

C. ANN Model

The ANN model is constructed by developing a sequential model using the feedforwardnet function, which is suitable for regression tasks. The hidden layer size is set at a single layer consisting of 70 neurons, a decision supported by the evaluation metrics. The fitness of this setting is confirmed by the fact that the residuals exhibit a random pattern and are consistently centred around zero. The sigmoid activation function is employed for the hidden layer, whereas the linear activation function is utilised for the output layer. The number of epochs and learning rate are set to 1000 and 0.01 respectively. The evaluation metrics used are root mean square error (RMSE), mean square error (MSE) and mean absolute error (MAE).

Table 1 Input features for ANN model

Network Features	Description
ρ	Network traffic load
AllocIDk,i	Distinguishes between T-CONT type k =2, 3 or 4 for each ONU i
$REPORT_k$	Buffer occupancy report sent by ONU for T-CONT k
BW_{ALLOC}	Bandwidth allocated at ONUs

III. Results

Fig. 3 – 5 show the relationship between the true upstream delay and the predicted upstream delay of T-CONT 2, 3 and 4 trained with a single layer ANN with 70 neurons. The closer the predicted data points to the target, represents better prediction. Based on the figures, the data points appear to be scattered around the target values, but with some deviations for all traffic classes. This suggests that the ANN model is mostly predicting the upstream delay with some accuracy, but there may be a tendency to underestimate the true delay for some samples.

The performance of the ANN models is presented in Table 2. T-CONT 2 and 3 exhibit lower MSE and MAE compared to T-CONT 4. The lower MSE and MAE for T-CONT 2 and 3 suggest that the models are better at predicting upstream delays for these traffic classes. T-CONT 2 and 3 may handle types of traffic with more predictable characteristics, leading to more accurate predictions and fewer errors. Conversely, the higher MSE and MAE for T-CONT 4 indicate that the model struggles more with predicting its upstream delays. This is due to T-CONT 4 has best effort bandwidth which handles more variable or bursty traffic, is inherently harder to model accurately. Fig. 6 – 8 depict the RMSE of the ANN model. Overall, the scatter plot indicates that the single layer ANN with 70 neurons is able to predict the upstream delay of XG-PON with some accuracy. However, there is room for improvement, as there are some deviations between the true and predicted values.

Table 2 Performance of ANN model

Traffic Class	MSE	MAE
T-CONT 2	2.60E-08	0.000125543
T-CONT 3	1.06E-06	0.000688872
T-CONT 4	0.011059	0.144498

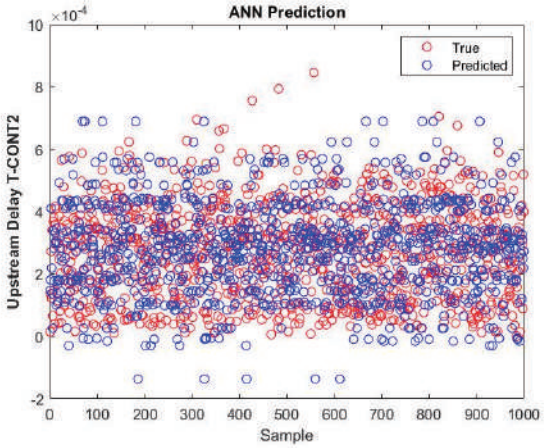


Fig. 3. ANN prediction for T-CONT 2.

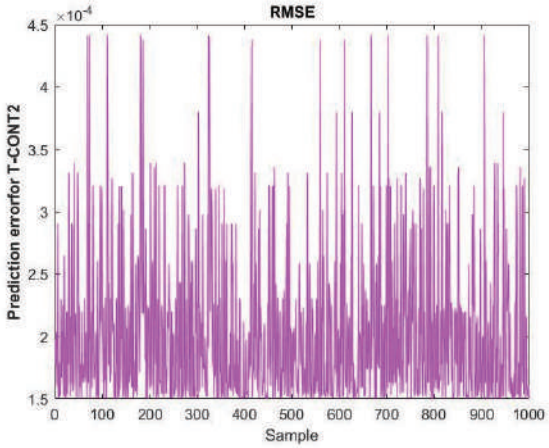


Fig. 6 RMSE plot for T-CONT 2.

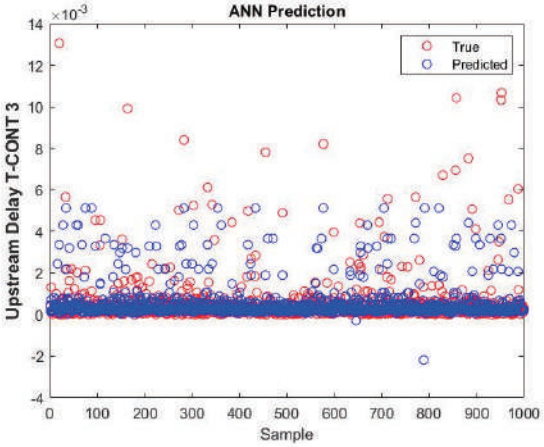


Fig. 4. ANN prediction for T-CONT 3.

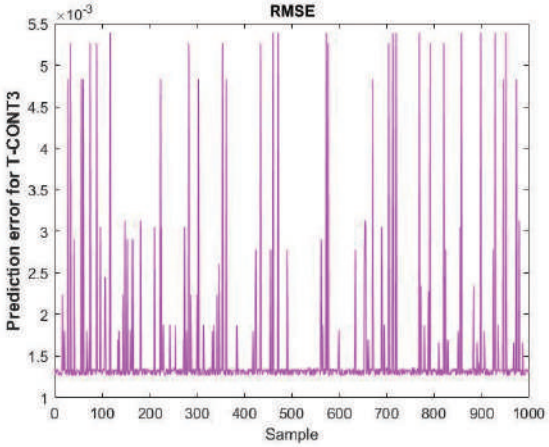


Fig. 7 RMSE plot for T-CONT 3.

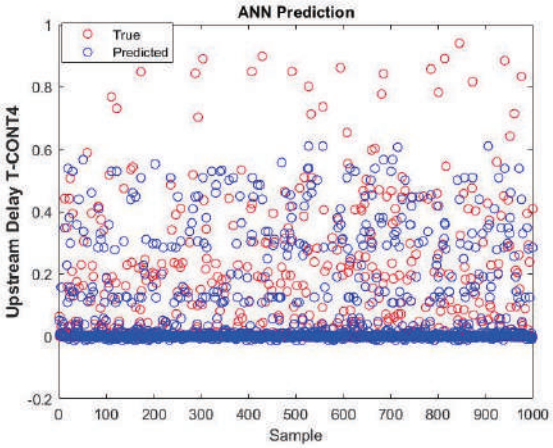


Fig. 5 ANN prediction for T-CONT 4.

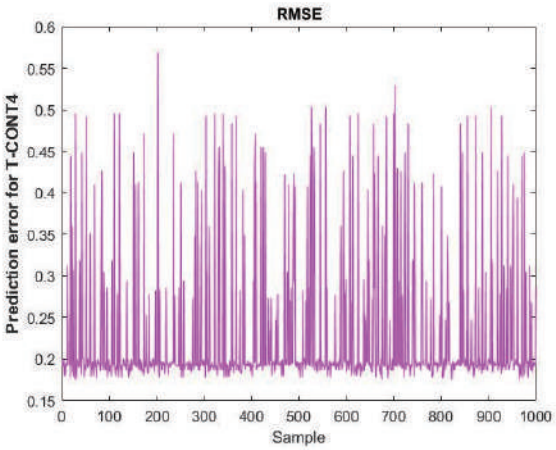


Fig. 8 RMSE plot for T-CONT 4.

IV. Conclusion

This paper introduces a method for modelling upstream delays in XG-PON for different traffic containers using ANN. The findings of our supervised training demonstrate that ANN can effectively simulate upstream delays, offering useful information for making dynamic decisions regarding bandwidth distribution. This approach aids in reducing inefficiencies in the allocation of bandwidth and minimizing network congestion, hence enhancing the QoS. The scatter

charts comparing projected and actual delays indicate that the ANN model performs rather accurately, but there is room for improvement. Subsequent research will prioritize improving the accuracy of the model and investigating supplementary characteristics to more effectively comprehend the intricacies of network delay dynamics.

ACKNOWLEDGMENT

The authors would like to thank Universiti Teknologi Malaysia (UTM) for the support of the project under UTM Institutional Grant vote 08G49. The authors would like to thank Universiti Teknologi MARA in the field of postgraduate research study.

REFERENCES

[1] D. Systems, "10-Gigabit-capable passive optical networks (XG-PON): Transmission convergence (TC) layer specification," *ITU Stand. G.987.3*, 2014.

[2] G 987, "ITU-T SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS Digital sections and digital line system – Optical line systems for local and access networks 10 Gigabit-capable passive optical networks (XG-PON): Reach extension," 2012.

[3] H. C. Leligou, C. Linardakis, K. Kanonakis, J. D.
arbitration of FSAN-compliant GPONs," *Int. J. Commun. Syst.*, vol. 19, no. 5, pp. 603–617, 2006.

[4] M.-S. Han, "Dynamic bandwidth allocation with high utilization for XG-PON," *16th Int. Conf. Adv. Commun. Technol.*, pp. 994–997, 2014.

[5] R. A. Butt, S. M. Idrus, N. Zulkifli, and M. Waqar Ashraf, "Comprehensive bandwidth utilization and polling mechanism for XGPON," *Int. J. Commun. Syst.*, vol. 31, no. 3, 2018.

[6] M.-S. Han, H. Yoo, B.-Y. Yoon, B. Kim, and J.-S. Koh, "Efficient dynamic bandwidth allocation for FSAN-compliant GPON," *J. Opt. Netw.*, vol. 7, no. 8, p. 783, Aug. 2008.

[7] R. A. Butt, S. M. Idrus, N. Zulkifli, and M. Waqar Ashraf, "Comprehensive bandwidth utilization and polling mechanism for XGPON," *Int. J. Commun. Syst.*, vol. 31, no. 3, 2018.

[8] L. Ruan, I. Dias, and E. Wong, "Machine intelligence in supervising bandwidth allocation for low-latency communications," *IEEE Int. Conf. High Perform. Switch. Routing, HPSR*, vol. 2019-May, pp. 1–6, 2019.

[9] L. Ruan, M. P. I. Dias, and E. Wong, "Towards self-adaptive bandwidth allocation for low-latency communications with reinforcement learning," *Opt. Switch. Netw.*, vol. 37, no. October 2019, 2020, doi: 10.1016/j.osn.2020.100567.

[10] L. Ruan, M. P. I. Dias, and E. Wong, "Machine learning-based bandwidth prediction for low-latency H2M applications," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3743–3752, 2019.

[11] E. Wong and L. Ruan, "(ECOC 2022) Towards 6G: Fast and Self-Adaptive Dynamic Bandwidth Allocation for Next-Generation Mobile Fronthaul," *J. Opt. Commun. Netw.*, vol. 15, no. 8, pp. 203–211, 2023.

[12] J. A. Hernández, A. Ebrahimzadeh, M. Maier, and D. Larrabeiti, "Learning EPON delay models from data: a machine learning approach," *J. Opt. Commun. Netw.*, vol. 13, no. 12, p. 322, 2021.

[13] J. A. Hernández, "Learning from data: Applications of Machine Learning in optical network design and modeling," in *2020 International Conference on Optical Network Design and Modeling (ONDM)*, pp. 1–6, 2020.

Queuing Delay Analysis Based on Statistical Quality-of-Service Guarantee for Latency-Sensitive Mobile Edge Computing

Gusun Joung, Mintae Kim, and Inkyu Lee, *Fellow*, *IEEE*
School of Electrical Engineering, Korea University, Seoul, South Korea
Email:{rntjs300,wkd2749,inkyu}@korea.ac.kr

***Abstract*—In this paper, we study the queuing delay in mobile edge computing (MEC) networks, which consist of a single-antenna base station (BS) equipped with an edge server (ES) and single-antenna user equipment (UE) that executes multiple tasks in each time slot. In particular, we derived closed-form expressions and analyzed the queuing delay-bound overflow probability under the quality-of-service (QoS) requirements of the UE. This analysis can be applied to scenarios where application tasks, transmitted with time-varying channel state information (CSI), continuously arrive at the UE in each time slot, resulting in incomplete task computations.**

***Index Terms*—Mobile edge computing, statistical QoS guarantee, queuing delay-bound overflow probability.**

I. Introduction

Motivated by the surging traffic demands spurred by the UE, MEC has emerged as a technology that distributes computation for executing tasks to ES. As the UE executes computation-intensive applications, latency performance is significantly affected by the limited computation capability of the UE. To efficiently reduce latency, the UE can wirelessly offload its tasks to proximal ES with relatively large computation capability [1].

Previous works [2], [3] proposed optimizations for end-to-end delay based on deterministic CSI, assuming that applications are completely computed within each time slot. However, in the context of next-generation communications, real-time applications that require computation-intensive and latency-sensitive processing continuously arrive at each time slot. Consequently, due to the limited resources of the UE and ES, such a situation leads to incomplete computations of tasks and the accumulation in the queue buffer. Furthermore, time-varying CSI further complicates the behavior analysis of latency in communication, computation, and queue buffer.

Therefore, in this paper, we mathematically derive and analyze the behavior of the queue buffer in practical scenarios where real-time applications continuously arrive at each time slot, resulting in cases where tasks are not fully computed and incorporating time-varying CSI.

II. System Model

For the sake of simplicity in deriving a closed-form expression, the considered MEC network consists of a single antenna BS equipped with the ES and a single antenna UE that executes a set $\mathcal{K}$ of K application tasks in each time slot $t \in \mathbb{N}$. We assume that the BS and UE are located in a fixed position. As described in the work of [4], each application task $k \in \mathcal{K}$ at the UE has a task arrival $A_k^{in}(t)$ (in units of bit per second) in time slot, and the different application tasks are computed independently in parallel. Moreover, we assume that task arrivals are independent and identically distributed over time with the average arrival rate $\lambda_k^{in} = \mathbb{E}[A_k^{in}(t)]/\tau$. The computation of each task's arrivals becomes dynamic, based on the CPU cycles required per bit, the instantaneous channel gain for each task, and the allocated communication and computing resources. The computation and communication delays for each task have a fixed slot length τ in time slot, consistent with the coherence block of the wireless channel. Following the data-partition model [5], it is assumed that in time slot, task k can either be computed locally at the UE (i.e., $o_k(t) = 0$) or be offloaded remotely to the ES (i.e., $o_k(t) = 1$). Consequently, both the UE and the ES have queue buffers to store task k for local computation and offloading, respectively.

For task k of the UE, the queue length at time slot t, denoted as $Q_k^L(t)$ (in units of bits) with $Q_k^L(0) = 0$, is given by

$$Q_k^L(t) = \max\Big\{Q_k^L(t-1) + A_k^{in}(t) - \frac{\tau f_k^L(t)}{L_k}(1 - o_k(t)) - \tau R_k^U(t) o_k(t), 0\Big\}, \tag{1}$$

in which

$$R_k^U(t) = B_k^U(t)\log_2\left(1 + \frac{P_{max}^U |h_k(t)|^2}{K N_0 B_k^U(t)}\right), \tag{2}$$

where $f_k^L(t)$ (in units of cycles per second) is the allocated CPU-cycle frequency for executing the task k at the UE. L_k is the required CPU cycles per bit for the computation of task k. P_{max}^U and N_0 are the transmit power of the UE and the power spectral density of the additive white Gaussian noise (AWGN), respectively. It is assumed that the transmit power for task k is uniformly allocated within the total available power limits. $B_k^U(t)$ is the bandwidth allocated for the offloaded task k. $|h_k(t)|^2$ is the wireless channel gain for the offloaded task k , including path loss and channel fading. It is assumed that both the UE and the ES have knowledge of the statistical distribution of the wireless channel.

For the offloaded task k of the UE, the queue length at time slot t, denoted as $Q_k^O(t)$ (in units of bits) with $Q_k^O(0) = 0$, is

given by

$$Q_k^O(t) = \max\Big\{Q_k^O(t-1) + \tau R_k^U(t) o_k(t) - \frac{\tau f_k^O(t)}{L_k} o_k(t), 0\Big\}, \tag{3}$$

where $f_k^O(t)$ (in units of cycles per second) is the allocated CPU-cycle frequency for the offloaded task k at the ES.

The theory on statistical QoS guarantees [6] provides an asymptotic approach in analyzing both queuing delay and stability for time-varying and stationary task arrival and/or service processes. Given the importance of QoS guarantee in latency-sensitive MEC systems, this approach provides statistical QoS provisioning for queue buffer, specifically in terms of guaranteeing a queuing delay-bound overflow probability. Hence, the queuing delay-bound overflow probability $\Pr\{D_k^L(t) \geq d_{th}^L\} \leq P_{th}^L$ for task k's local computation queue can be approximated [6], [7] by

$$\mathcal{C}_k^L\left(f_k^L(t), B_k^U(t), o_k(t)\right) \geq \lambda_k^{in} = \frac{-\ln P_{th}^L}{\theta_k^L d_{th}^L}, \tag{4}$$

where

$$\begin{aligned}
&\mathcal{C}_k^L\left(f_k^L(t), B_k^U(t), o_k(t)\right) \\
&= -\frac{1}{\theta_k^L t}\log \mathbb{E}\left\{ e^{-\theta_k^L \sum_{i=1}^{t}\left(\frac{\tau f_k^L(i)}{L_k}(1-o_k(i)) + \tau R_k^U(i) o_k(i)\right)} \right\} \\
&\overset{(a)}{=} \frac{1}{t}\sum_{i=1}^{t}\Bigg[\frac{\tau f_k^L(i)}{L_k}(1-o_k(i)) \\
&\quad + \tau B_k^U(i) log_2\left(\frac{P_{max}^U}{K N_0 B_k^U(i)}\right) o_k(i) - \frac{K N_0 B_k^U(i)}{\theta_k^L P_{max}^U} \\
&\quad - \frac{1}{\theta_k^L}\log\Gamma\left(1 - \frac{\theta_k^L \tau B_k^U(i) o_k(i)}{\log 2}, \frac{K N_0 B_k^U(i)}{P_{max}^U}\right)\Bigg].
\end{aligned} \tag{5}$$

In the special case of independent channel coefficients in each block and Rayleigh fading, $\mathcal{C}_k^L\left(f_k^L(t), B_k^U(t), o_k(t)\right)$ can be expressed in closed-form as (a) of equation (5). Here, $\Gamma(s,w) = \int_w^\infty v^{s-1}e^{-v}dv$ is the upper incomplete gamma function. Similarly, the queuing delay-bound overflow probability $\Pr\{D_k^O(t) \geq d_{th}^O\} \leq P_{th}^O$ for task k's offloading queue can be approximated [6], [7] by

$$\mathcal{C}_k^O\left(f_k^O(t), o_k(t)\right) \geq \mathcal{A}_k^O\left(f_k^L(t), B_k^U(t), o_k(t)\right) = \frac{-\ln P_{th}^O}{\theta_k^O d_{th}^O}, \tag{6}$$

where

$$\begin{aligned}
&\mathcal{A}_k^O\left(f_k^L(t), B_k^U(t), o_k(t)\right) \\
&= \begin{cases}
\lambda_k^{in}, \text{if } 0 \leq \theta_k^O \leq \theta_k^L, \\
\frac{\theta_k^L}{\theta_k^O}\lambda_k^{in} + \frac{\theta_k^O - \theta_k^L}{\theta_k^O t}\sum_{i=1}^{t}\left[\frac{\tau f_k^L(i)}{L_k}(1-o_k(i))\right. \\
\left. + \tau B_k^U(i)\log_2\left(\frac{P_{max}^U}{K N_0 B_k^U(i)}\right) o_k(i) - \frac{K N_0 B_k^U(i)}{(\theta_k^L - \theta_k^O) P_{max}^U}\right. \\
\left. - \frac{1}{\theta_k^L - \theta_k^O}\log\Gamma\left(1 - \frac{(\theta_k^L - \theta_k^O)\tau B_k^U(i) o_k(i)}{\log 2}, \frac{K N_0 B_k^U(i)}{P_{max}^U}\right)\right], \\
\text{if } \theta_k^O > \theta_k^L,
\end{cases}
\end{aligned} \tag{7}$$

and

$$\begin{aligned}
\mathcal{C}_k^O\left(f_k^O(t), o_k(t)\right) &= -\frac{1}{\theta_k^L t}\log \mathbb{E}\left\{ e^{-\theta_k^L \sum_{i=1}^{t}\frac{\tau f_k^O(i)}{L_k} o_k(i)} \right\} \\
&= \frac{1}{t}\sum_{i=1}^{t}\frac{\tau f_k^O(i)}{L_k} o_k(i).
\end{aligned} \tag{8}$$

Just like $\mathcal{C}_k^L\left(f_k^L(t), B_k^U(t), o_k(t)\right)$, $\mathcal{A}_k^O\left(f_k^L(t), B_k^U(t), o_k(t)\right)$ is also expressed in closed-form. Here, θ_k^L and θ_k^O provide the exponentially decaying rate of the queuing delay-bound overflow probability and are referred to as QoS exponents [7]. The larger θ_k^L and θ_k^O correspond to stricter queuing or QoS constraints, while the smaller θ_k^L and θ_k^O imply looser queuing constraints [8].

Based on the mathematical and theoretical derivation in closed-form as discussed above, the queuing delay-bound overflow probability for task k's local computation and offloading queues becomes tractable. Additionally, equations (4) and (6) provide constraints on the allocated communication and computing resources necessary to support the arrival rate of the UE, based on the specified statistical QoS requirements. Using this approach, the minimum or maximum of the allocated communication and computing resources can be derived.

III. Conclusion

In this paper, we assumed a practical scenario where real-time application tasks incorporating time-varying CSI continuously arrive at each time slot, resulting in tasks not being fully computed. In such a scenario, we derived and analyzed closed-form expressions for the queuing delay-bound overflow probability, based on specified statistical QoS requirements. We believe that our analysis of queue behavior will provide insights regarding resource allocation optimization for minimizing objective functions such as latency or energy consumption in other MEC studies that consider queue stability.

References

[1] T. X. Tran and D. Pompili, "Joint task offloading and resource allocation for multi-server mobile-edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 856–868, 2019.

[2] W. Wen, Y. Cui, T. Q. S. Quek, F.-C. Zheng, and S. Jin, "Joint optimal software caching, computation offloading and communications resource allocation for mobile edge computing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7879–7894, 2020.

[3] S. Bi, L. Huang, and Y.-J. A. Zhang, "Joint optimization of service caching placement and computation offloading in mobile edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4947–4963, 2020.

[4] C.-F. Liu, M. Bennis, M. Debbah, and H. V. Poor, "Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4132–4150, 2019.

[5] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 2017.

[6] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Auto. Control*, vol. 39, no. 5, pp. 913–931, 1994.

[7] Y. Wang, X. Tao, Y. T. Hou, and P. Zhang, "Effective capacity-based resource allocation in mobile edge computing with two-stage tandem queues," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6221–6233, 2019.

[8] M. Ozmen and M. C. Gursoy, "Wireless throughput and energy efficiency with random arrivals and statistical queuing constraints," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1375–1395, 2016.

WIP: Analysis of Distance-based Handover Parameters in LEO Satellite Communications

Jongtae Lee[1], Yujeung Ahn[1], Jaeyeol Lee[2], Wonjae Lee[1], and Jae-Hyun Kim[3]
[1]*Department of Artificial Intelligence Convergence Network, Ajou University, Suwon, 16499, South Korea*
[2]*Department of Space Survey Information Technology, Ajou University, Suwon, 16499, South Korea*
[3]*Department of Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea*
{jtlee830, dbwjd5825, jaeyel98, ljmwon, jkim}@ajou.ac.kr

***Abstract*—Low-Earth orbit (LEO) satellites, operating at high speeds, result in frequent handover (HO) for user equipment (UE). In the Earth-moving cell (EMC) environment, where cell locations change with satellite movement, HOs are even more frequent. Such frequent HOs significantly degrade the communication service quality of the UE. The 3rd generation partnership project (3GPP) proposed EventD2, a distance-based HO mechanism in EMC, to address these issues. However, in an EMC environment, the continuous movement of cells leads to various situations depending on the location and time. Therefore, EventD2 parameters must be empirically derived for all possible situations, and relevant studies have been insufficient to date. This paper proposes a method for empirical parameter derivation based on the average signal-to-interference-plus-noise ratio (SINR) of the UE and determines the effective threshold ranges for triggering through performance analysis. Also, the analysis identifies improvements in single key performance indicator (KPI) analysis and suggests future research to further narrow the effective range.**

***Index Terms*—Handover, Mobility management, LEO satellite**

I. Introduction

Non-terrestrial networks (NTN) utilizing low-Earth orbit (LEO) satellites are being increasingly recognized as a future communication solution due to their global coverage and low latency. However, their high speeds lead to frequent handover (HO), causing low-quality service due to HO interrupt times for user equipment (UE) and significant signaling overhead. Also, received signal strength (RSS) based HO triggering events are less effective due to low signal strength variations across the cell.

To resolve these challenges, the 3rd generation partnership project (3GPP) has proposed new HO triggering events based on time(r), distance, and other metrics [1]. Following this, in [2], it was decided to add distance and time(r) as new HO triggering events based on the study item. Among them, distance-based HO triggering events like EventD1 and EventD2 consider the UE's position relative to a referenceLocation, which is the cell center [3]. In the Earth-moving cell (EMC) environment, EventD2 is utilized. As shown in Fig. 1, the referenceLocation changes periodically in EMC environment. This change affects the HO for the UE and also impacts the HO due to satellite mobility. Moreover, HO decisions rely on independent thresholds for the serving and target cells and utilize the distance between the UE and each referenceLocation. Hence, determining the HO parameters for EventD2, considering satellite mobility in EMC environments, should be based on empirical findings.

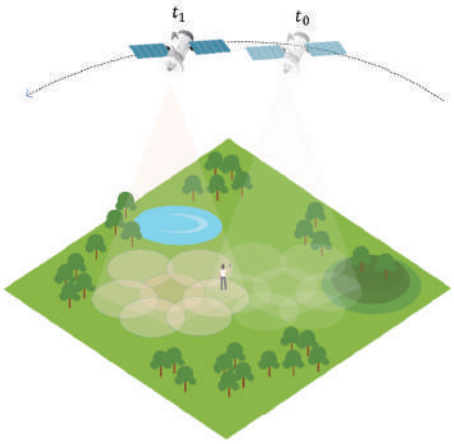


Fig. 1. Intra-Satellite HO Scenario in EMC System

This paper aims to derive the effective threshold range for EventD2 based on the average signal-to-interference-plus-noise ratio (SINR) of the UE considering satellite mobility in EMC environments. Additionally, it suggests other key performance indicators (KPIs) to be considered for future improvements.

II. Distance-based HO in 3GPP

In this section, we present 3GPP-based EventD2 HO method and propose a method for deriving the required parameters.

A. EventD2 HO method

3GPP Release 17 presents various HO triggering events to address the issues mentioned in Section I. Among these, the distance-based HO triggering event constructs conditions based on the distance between the referenceLocation and the UE. Here, the referenceLocation refers to the center of the cell. This condition holds if the following conditions are satisfied:

$$(Ml1 - Hys > TH_1) \wedge (Ml2 + Hys < TH_2), \quad (1)$$

where $Ml1$ and $Ml2$ represent the distances between the UE and the referenceLocations of the serving cell and the target cell, respectively. Hys represents the hysteresis, and TH_1 and

TH_2 are the thresholds for $Ml1$ and $Ml2$, respectively. All parameters are measured in meters.

B. Parameter derivation method

To determine the range of TH_1 and TH_2 for initial evaluation, we first check when the SINR of the UE in the serving cell first falls below Q_{in}. Here, Q_{in} is the threshold signal value for determining radio link failure (RLF) and is set at an SINR of -6 dB [2]. To derive the maximum distance TH_{max}, the UE is connected to the serving cell, and the satellite is moved while measuring the SINR at each measurement period. The measurement results are averaged over multiple trials, and the distance at which the SINR first enters Q_{in} is determined as TH_{max}. Using the previously obtained TH_{max}, we measure the average SINR for each pair of TH_1 and TH_2 within the range of 0 km to TH_{max} km, and evaluate the effective range of TH_1 and TH_2.

III. Performance Evaluation

This section presents the system model for two simulation scenarios and analyzes the results from the simulations.

A. System model

We designed a Direct-to-Cell (D2C) intra-satellite HO scenario. The system parameters were referenced from system-level simulation study case 9 in [1], and the antenna pattern was considered from [4]. The satellite operates at a speed of 7.56 km/s at an altitude of 600 km. To consider additional interference, 2-tier cells were added, resulting in a total of 61 cells deployed. The cell radius is 50 km, and the distance between the centers of each cell is 43.3 km. The channel model is rural scenario, using the S-band frequency. One UE (handheld) is placed at the center of the central cell to determine the maximum effective range. The HO is designed to occur only once when conditions are met, ignoring Hys from (1) and considering only Ml and TH. This is because our goal is an analysis of the entire range of parameters.

B. Simulation results

Fig. 2 (a) shows the average serving cell $SINR_{avg}$. The distance at which the SINR first reaches Q_{in} is about 28 km. This means that UEs located more than 28 km away from both the serving and target cells cannot be served. Therefore, TH_1 and TH_2 are analyzed within the range of 0 km to 28 km. Fig. 2 (b) presents a heatmap of $SINR_{avg}$ for each pair of TH_1 and TH_2, displayed on a log scale for the entire range. Combinations with higher average SINR occur when TH_1 ranges from 0 km to 24 km and TH_2 from 17 km to 28 km. However, when TH_1 is between 0 km and 15 km or TH_2 is between 24 km and 28 km, the UE is positioned within the inner region of the serving cell. In this scenario, HO can occur even when the signal from the serving cell is stronger than that of the target cell. Therefore, this range is ineffective for HO. Conversely, if the UE is within the target cell, a HO event may lead to RLF due to a low RSS from the serving cell. Thus, considering trade-off relationships between average SINR and other KPIs, it is necessary to further narrow down the effective range of TH_1 and TH_2.

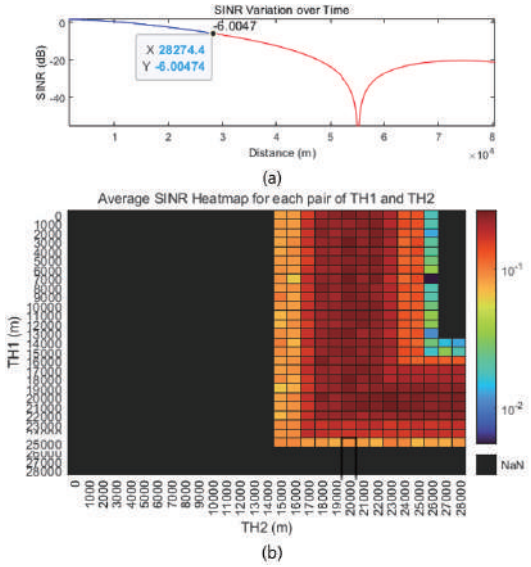


Fig. 2. Simulation results: (a) Average SINR Variation (b) Average SINR Heatmap for each threshold pairs

Conclusion and Future Works

In this paper, we derived parameters for EventD2 in LEO-based EMC scenarios, focusing on average SINR. The effective ranges for TH_1 and TH_2 were identified as 0 km to 24 km and 17 km to 28 km, respectively. However, additional KPIs are needed for more precise optimization. Two additional KPIs to be considered are the HO ping-pong (HOPP) rate and HO failure (HOF). The HOPP rate measures frequent HO where UEs return to the original cell within a certain time of stay (typically 1s). This indicates a good trade-off relationship when TH_1 is within the range of 0 km to 15 km and TH_2 exceeds 23 km or 24 km, meaning it is within the serving cell. This can further reduce the effective range. Additionally, HOF refers to RLF during HO, which can help evaluate the overall effective range during the simulation time. Future work will focus on these KPIs and optimization to derive suitable parameters through hysteresis and further analysis.

Acknowledgments

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean Government (MSIT) (2022-0-00704, Development of 3D-NET Core Technology for High-Mobility Vehicular Service)

References

[1] 3GPP TR 38.821 v16.1.0, “Solutions for NR to support non-terrestrial networks (NTN) (Release 16),” May, 2021.
[2] 3GPP TR 21.917 v17.0.0, ”Release 17 Description; Summary of Rel-17 Work Items (Release 17),” Jan. 2023.
[3] 3GPP TS 38.331 v17.8.0, ”NR; Radio Resource Control (RRC); Protocol specification (Release 17),” Jan. 2024.
[4] 3GPP TR 38.811 v15.4.0, “Study on New Radio (NR) to support nonterrestrial networks (NTN) (Release 15),” Sep. 2020.

A Review on Resource Allocation in FL-assisted IoT Networks: Overview, Research Challenges and Future Trends

Manh Cuong Ho, Dongwook Won, Thanh Phung Truong, Jaemin Kim, and Sungrae Cho*
School of Computer Science and Engineering, Chung-Ang University, Korea (South)
Email: hmcuong, dwwon, tptruong, jmkim@uclab.re.kr, srcho@cau.ac.kr

***Abstract*—Recently, Federated Learning (FL) has gained traction as a promising approach, prioritizing user data security and privacy while optimizing communication resources effectively. FL serves as a distributed machine learning technique commonly utilized in various Internet of Things (IoT) applications, facilitating collaboration between a central server and user devices to train a global model from multiple IoT devices (IoTDs). As the IoT continues to proliferate and integrate into various fields, merging it with FL models presents a notable challenge that requires significant research investment. While the constrained resources of IoTDs necessitate tailored design considerations in structuring FL models, FL confronts communication and computation challenges akin to those prevalent in IoT networks. Hence, optimizing the utilization of resources becomes imperative when contemplating the integration of IoTDs into FL frameworks. This paper undertakes a comprehensive review of the foundational principles of FL and delves into problems and goals for resource allocation within IoT systems facilitated by FL. Subsequently, it delineates the challenges encountered and outlines future research avenues about resource allocation in FL-assisted IoT networks.**

***Index Terms*—Federated Learning, Resource Allocation, IoT Networks**

I. Introduction

Machine Learning techniques have revolutionized our ability to extract insights from vast volumes of data, particularly in applications such as the IoT, where sensors and IoTDs collect, adapt, and respond to real-time information at scale. However, with the robust development of IoT applications today, data privacy and security concerns often limit the applicability of powerful decentralized learning methods in many IoT applications. To address this challenge, FL has been developed to train distributed models while keeping data localized, thus mitigating privacy and security risks [1, 13]. In FL, a global model is initialized by a central server, which is then shared with remote IoTDs for iterative training rounds. Each IoTD trains its individual model using its respective data and then sends its local parameters to the FL server. The server subsequently updates the global model from all devices and broadcasts the new model to all IoTDs. FL's distributed processing makes it suitable for smart IoT systems based on addressing privacy and security concerns. Moreover, cooperative training without direct data sharing among users in the FL system is crucial for IoT applications where sensitive sensor data, such as health, education, industry, and data collection, must remain locally stored to protect privacy [1].

Despite the benefits of integrating FL and IoT, several design challenges specific to IoT networks need careful consideration. Challenges of FL-aided IoT networks include ensuring optimizing resources and security robustness for computational and communication tasks between the central server and remote IoTDs [1, 4]. However, the heterogeneity of IoTDs and their constrained communication and computing resources pose significant obstacles to FL-assisted IoT adoption. Additionally, factors like limited resources and unbalanced data distribution constrain resource allocation, impeding satisfactory and timely FL training. Thus, overcoming these challenges is crucial for effective FL implementation in IoT settings. In recent times, numerous solutions have been proposed to tackle the aforementioned issues.

This study offers a comprehensive review of goals, challenges, and future directions of resource optimization in FL-assisted IoT networks. The structure of this paper is outlined in Section I as an overview and the FL framework, the resource allocation in the FL-aided IoT systems in Section II. In Section III, we delineate and elucidate several research challenges, while also providing insights into future trends.

II. Resource Allocation in the FL-assisted IoT networks

Presently, FL finds widespread application in numerous IoT networks, highlighting a range of contemporary concerns within these systems. In this section, a primary focus lies on the fundamentals of FL and resource allocation within IoT networks leveraging FL.

A. Federated Learning

FL is a distributed machine learning approach wherein user devices independently train themselves locally without the need to directly share their data with a central server. Through periodic updates of a shared global model by aggregating information from each local model, FL enables user devices to capture a collective global view [1]. The FL process typically involves three steps:

1) During the initial phase, the FL server initializes a global model, denoted as W_g^0, and sends this model to all user devices.

2) Upon receipt of the global model W_g^t, each user device k proceeds to generate its own model by utilizing its local data with t represents the tth iteration. Subsequently, user device k updates its model parameters W_k^t to minimize its local loss function $F_k(W_k^t)$. Following this optimization step, the locally optimal models are then transmitted to the FL server.
3) After receipt of the local model from the participating devices, the FL server proceeds to aggregate them to update a global model denoted as W_g^{t+1}. Subsequently, the new global model is broadcast once again to all the user devices within the network.

Steps are repeatedly executed until the FL server attains convergence by minimizing the global loss function $F(\mathbf{w})$ as

$$\min_{\mathbf{w}} F(\mathbf{w}) = \min_{\mathbf{w}} \sum_{k=1}^{N} e_k F_k(\mathbf{w}), \tag{1}$$

where N indicates the number of user devices, e_k is the weighting coefficient of user device k [5], $F_k(\mathbf{w})$ is the average loss function of user device k, and $\mathbf{w}$ represents the model parameter vector.

B. Resource allocation in the FL-assisted IoT networks

Within FL-aided IoT networks, resource allocation objectives commonly revolve around improving model performance metrics. This includes minimizing loss decay, maximizing accuracy, and optimizing training time. Moreover, there is a notable emphasis on optimizing costs, reducing energy consumption, maximizing utility, and minimizing latency in IoT networks based on FL [1, 4]. Certain studies address highly specific objectives such as the optimization of convergence of algorithms or the performance during training time. Besides, some new metrics are defined by authors to improve the performance in FL systems [5]. These specialized objectives underscore the diverse and nuanced nature of resource optimization efforts in FL-assisted IoT networks. However, recent research efforts have predominantly focused on addressing primary concerns such as loss decay, accuracy, training time, cost, energy consumption, utility, and latency. These critical issues have been at the forefront of exploration and optimization endeavors in the field.

On the other hand, various resource allocation problems demand meticulous attention due to the involvement of heterogeneous IoTDs, which are related to factors such as battery life, imbalanced data, and computational capability [4]. Efficient allocation of resources in FL-supported IoT systems, such as CPU frequency, bandwidth, and power, continues to pose a significant challenge. Nevertheless, device selection, joint optimization of computation and communication, and global aggregation are problems that have recently garnered more attention. Participant selection in FL is a pivotal issue wherein devices are chosen to partake in each training round. While the FL server typically randomly selects participants from a large pool of IoTDs, the presence of slow or malfunctioning devices can impede FL training progress significantly. Consequently, novel participant selection strategies are actively being researched to alleviate the training bottleneck in FL. Furthermore, given the pivotal role of joint optimization of local model computation and transmission in FL, there has been a notable surge in research exploring innovative wireless communication techniques aimed at optimizing FL efficiency. Lastly, in FL systems, global aggregation involves transmitting model parameters to the FL server for aggregation. While traditional global aggregation follows a synchronous approach, wherein aggregations take place at fixed intervals after participants complete a predetermined number of local computation rounds, there is a rising interest in exploring adaptive adjustments to the frequency of global aggregation and designing a suitable asynchronous approach. This adaptive approach holds the potential to enhance training efficiency and algorithm convergence while respecting resource constraints.

III. Challenges and Future Research Directions

As previously mentioned in Section II, recent research endeavors have primarily concentrated on tackling primary challenges such as loss decay, accuracy, training time, cost, energy consumption, utility, and latency. Additionally, certain resource allocation issues arise due to the involvement of heterogeneous devices in various kinds of FL models with varying dataset qualities, computation capabilities, and energy budgets [4].

Currently, numerous solutions are being proposed to effectively tackle challenges in FL-assisted IoT networks, with many studies leveraging mathematical techniques to address these issues. In recent times, numerous effective algorithms have been proposed not only to enhance the performance metrics of FL models but also to address issues related to energy consumption, cost, and latency [4]. In [11], Yang et al. sought to reduce the system's energy consumption while ensuring compliance with a latency constraint. Next, Pham et al. in [7] minimized the energy consumption of the system to advance sustainable FL in an unmanned aerial vehicle-based network. Li et al. in [5] focused on maximizing a novel metric referred to as the Weighted Global Proportion of Trained Mini-batches (WGPTM) to exploit the convergence potential of the new system. This was achieved through the joint optimization of transmit powers and sub-channel bandwidths. Moreover, several works have explored the utilization of a reinforcement learning approach as a key component of their solutions. In [10], Wu et al. concentrated on the maximization of efficient energy of the base station in FL networks with simultaneous wireless information and power transfer (SWIPT). Next, Li et al. [6] focused on minimizing long-term energy consumption while guaranteeing FL convergence in a network assisted by SWIPT. They achieved this through dynamic resource allocation and User Equipment (UE) scheduling. In [3], Ho et al. tackled the challenge of minimizing energy consumption of IoTDs in FL-aided IoT systems by integrating SWIPT with the multi-carrier non-orthogonal multiple access (MC-NOMA) technique. Overall, Reinforcement Learning (RL) serves as a robust decision-making technique that has demonstrated

efficacy in addressing various challenges associated with FL such as device selection and solving optimization problems [12]. Hence, it is justifiable to further investigate and explore this approach in future endeavors. Besides, the combination of mathematical techniques and RL techniques, or the integration of different RL techniques, represents potential solutions for optimizing resource allocation in FL-based IoT systems [9]. Additionally, other significant aspects were highlighted as potential future directions in the research studies. Shan et al., as described in [8], aim to concentrate on optimizing the scheduling of tasks and collaborative manner as a future direction of their work. In [2], Hao et al. explore the utilization of blockchain technology to improve architecture and enhance the security of 6G networks.

Furthermore, the utilization of FL for wireless networks presents several directions and challenges for the future. FL should be scalable and fair to accommodate large-scale learning networks. Besides, privacy and security represent critical concerns in the present landscape. While FL offers protection for local user information by transmitting only local parameters to the FL server, there remains the risk of leakage or eavesdropping on FL local models. Next, information exchange in the FL-assisted IoT network occurs between wireless devices and the server. Thus, while synchronous communication methods are straightforward, they may be interrupted among IoTDs. An appealing approach to mitigate disruption due to IoTDs in an environment characterized by heterogeneity is through asynchronous solutions. Although asynchronous approaches have been successful in dealing with delay, a system structure and convergence of algorithms need to be carefully exploited. Additionally, the effectiveness of FL learning schemes is often hampered by limited resources and a less reliable wireless link between FL servers and IoTDs. Therefore, it is crucial to investigate the integrated management of computation and communication resources to attain the efficiency of FL in IoT networks.

IV. Conclusions

In this paper, we have provided a thorough review of studies pertaining to FL and Resource Allocation within FL-assisted IoT networks. Firstly, we provided the fundamentals of FL and conducted an extensive review concerning resource optimization in FL-assisted IoT networks. Subsequently, we highlight the prevailing challenges and intricacies inherent in this field. Through meticulous scrutiny, we identify crucial issues that warrant attention and propose potential avenues for future investigation. Our aim is to shed light on the pressing concerns faced by practitioners and researchers alike, while also offering valuable insights into the directions that warrant exploration in forthcoming studies.

Acknowledgment

This work was supported in part by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2022-00156353) supervised by the IITP (Institute for Information Communications) and in part by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (RS-2023-00209125).

References

[1] Syreen Banabilah et al. “Federated learning review: Fundamentals, enabling technologies, and future applications”. In: *Information processing & management* 59.6 (2022), p. 103061.

[2] Min Hao et al. “URLLC resource slicing and scheduling for trustworthy 6G vehicular services: A federated reinforcement learning approach”. In: *Physical Communication* 49 (2021), p. 101470.

[3] Manh Cuong Ho et al. “A DDPG-based energy efficient federated learning algorithm with SWIPT and MC-NOMA”. In: *ICT Express* (2023).

[4] Ahmed Imteaj et al. “A survey on federated learning for resource-constrained IoT devices”. In: *IEEE Internet of Things Journal* 9.1 (2021), pp. 1–24.

[5] Weicai Li et al. “Multi-Carrier NOMA-Empowered Wireless Federated Learning with Optimal Power and Bandwidth Allocation”. In: *IEEE Transactions on Wireless Communications* (2023).

[6] Yang Li et al. “Dynamic user-scheduling and power allocation for SWIPT aided federated learning: A deep learning approach”. In: *IEEE Transactions on Mobile Computing* (2022).

[7] Quoc-Viet Pham et al. “Energy-efficient federated learning over UAV-enabled wireless powered communications”. In: *IEEE Transactions on Vehicular Technology* 71.5 (2022), pp. 4977–4990.

[8] Nanliang Shan, Xiaolong Cui, and Zhiqiang Gao. ““DRL+ FL”: An intelligent resource allocation model based on deep reinforcement learning for mobile edge computing”. In: *Computer Communications* 160 (2020), pp. 14–24.

[9] Shaoyang Wang et al. “Joint resource management for MC-NOMA: A deep reinforcement learning approach”. In: *IEEE Transactions on Wireless Communications* 20.9 (2021), pp. 5672–5688.

[10] Yuan Wu et al. “Simultaneous wireless information and power transfer assisted federated learning via nonorthogonal multiple access”. In: *IEEE Transactions on Green Communications and Networking* 6.3 (2022), pp. 1846–1861.

[11] Zhaohui Yang et al. “Energy efficient federated learning over wireless communication networks”. In: *IEEE Transactions on Wireless Communications* 20.3 (2020), pp. 1935–1949.

[12] Rongfei Zeng et al. “A comprehensive survey of incentive mechanism for federated learning”. In: *arXiv preprint arXiv:2106.15406* (2021).

[13] Guangxu Zhu et al. “Toward an intelligent edge: Wireless communication meets machine learning”. In: *IEEE communications magazine* 58.1 (2020), pp. 19–25.

Advancing Green Network Communication through TRIZ: Analysing Patented Technologies for Sustainable 6G Connectivity

Zulhasni Abdul Rahim 1,*
Malaysia-Japan International Institute of Technology
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
zulhasni@utm.my
ORCID: 0000-0001-7958-4114

Muhammad Saqib Iqbal 2,
Malaysia-Japan International Institute of Technology,
Universiti Teknologi Malaysia,
Jalan Sultan Yahya Petra, Kampung Datuk Keramat, 54100 Kuala Lumpur,
Wilayah Persekutuan Kuala Lumpur
ORCID: 0000-0001-7958-4114

Sevia Mahdaliza Idrus 3
Faculty of Electrial Engineering,
Universiti Teknologi Malaysia
(UTM),81310 Skudai, Johor, Malaysia.
ORCID: 0000-0001-5471-6002

***Abstract*— The article discusses the feasibility of applying the TRIZ methodology for increasing sustainability in 6G wireless networks. In this paper, an effort is made to analyse patented technologies associated with green network communication and develop innovative solutions that may help reduce the environmental footprint of 6G networks using the TRIZ patent literature review methodology. Patent literature on green network communication and research in the TRIZ methodology has assured the development of energy-efficient hardware implementation, advanced intelligent network management systems, and renewable energy sources. A case study is presented on patented technology under Cooperative Patent Classification (CPC) subclass H04W of green network for 6G and TRIZ tool called function analysis enable innovation of green network technology in 6G patented technologies. The paper, therefore, outlines findings and gives recommendations for future innovation and technological growth towards 6G green network.**

Keywords—Green network, TRIZ, 6G, Patent analytics, Sustainability

I. INTRODUCTION

The telecommunications sector is a significant enabler of connectivity in the global context. The industry enables fast communication and information interchange in the present world. However, such a pivotal role comes with an expensive literary environmental cost in that keeping up with the increased data needs requires energy-intensive infrastructure. 1. Given society's increased reliance on digital connectivity, mitigating the environmental impacts caused by network communication is significant. In this regard, the highly required are the sustainable 6G practices for the optimisation of energy use, reducing carbon emissions, and minimising electronic waste [2-3]. By applying green strategies related to optimising energy usage and reducing carbon emissions and electronic waste, the 6G-related industry will play an essential role in preserving the environment while satisfying the changing needs of users worldwide [4]. Moreover, sustainability in 6G network communication goes hand in hand with the more significant agendas on climate change mitigation and long-term environmental sustainability. The estimate thus would be that green communications networking would be first and foremost in avoiding ecological degradation but also essential to foster a more sustainable and robust 6G telecommunications infrastructure for future generations [5-6].

What would be necessary for developing countries is the evolution from 5G to 6G and how this has made network communication sustainable. There is a structured way through which innovation can be achieved in TRIZ. The first step involves the identification of technological contradictions and then finding creative ways to overcome them. Much has been done using TRIZ in many areas of technical problem-solving and innovation; however, more needs to be done to further sustainability. Such gaps can be overcome by integrating TRIZ analysis into the research of patented technologies in network communication [8-9]. An analysis of the patent literature using the principles of TRIZ will allow the global telecommunications sector to detect opportunities for enhancing its sustainability as 5G is superseded by 6G.

This approach finds innovative solutions for such global environmental protection challenges as optimising energy consumption, reducing carbon emissions, and minimising electronic waste. More than that, using TRIZ in the worldwide context provides a structured way to navigate the complexities of technological innovation while aligning with national sustainability initiatives and goals. As sub-cited, where the trend analysis of Orbit Intelligence related to a patented database by Questel [10-11] on green networks from 2000 to 2024 is presented, the growth in patent applications increased year by year up to, as per Fig.1.

The trend of green network technologies patent applications has been upward, increasing impressively and consistently over the past two decades and representing a growing interest in sustainability within the telecommunications industry. As recently as the year 2000, there were only four patents on green network technologies focusing on environmental awareness. However, with the growing concerns about climate change and resource depletion, patent applications grew, reaching as high as 127 in 2017. This dramatic increase reflects a paradigm shift toward the infusion of sustainability considerations into network communication infrastructure and operations. The trend has been flat since 2017, remaining at about 90 annually until 2022. Such persistent patent activities on this level reflect the long-term commitment of the industry to develop green network technologies and reduce the environmental impact of telecommunication tools. The trend of green network technologies patent applications will likely be maintained upward as the industry seeks balance with the dual imperatives of technological advancement and environmental sustainability. Continuing innovation and collaboration across

all sectors can better equip the telecommunications sector to further accelerate this shift to a greener, more resilient network infrastructure that can meet societal needs while safeguarding the planet for future generations, especially for 6G network technology development.

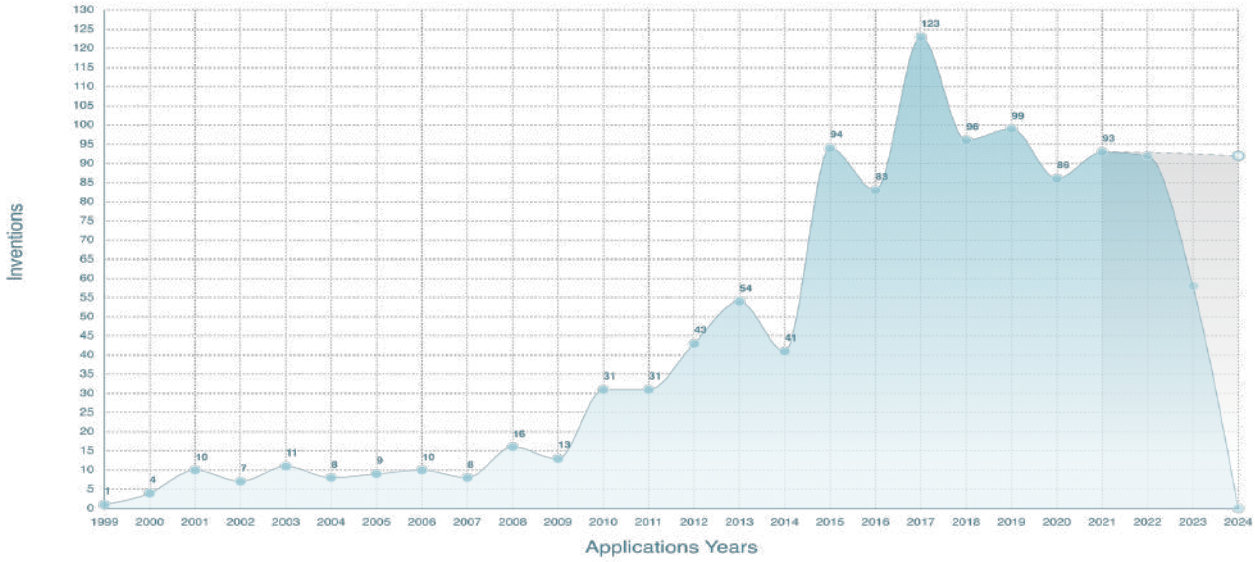


Fig. 1 Trend of patent application on green network technology

II. Patent Literature Review on Green Network Technologies

A. TRIZ-Patent Literature Review (PLR) Methodology

In this study, a specific methodology will be used to review the patented document utilising the concept of TRIZ. It is unique because the nature of patent literature differs in many aspects—for example, the structure of the document, formatting, technical and legal language, technical figures, etc. [12]. Therefore, TRIZ PLR focuses on reviewing technical concepts regarding patent claims, system-based concepts, component functionality, etc. TRIZ, in the context of green network patented technology, acts as the foundational tool for technical problem identification in patent literature reviews. This occurs by first reviewing a patented document and mostly its independent claims. It is through these claims that the review model obtains information on system functionality and root causes critical to the proprietary technologies [13]. It allows the model to generalise from the scenario covered by a particular patent to a more generic problem model.

After defining the generic problem model, the research of the TRIZ model into the scientific answers found in patent literature follows. The scientific solution is anchored in the independent claims. Systematic analysis by TRIZ of these scientific elements aims to form new insights and principles that may further help create new systems with improved value and performance.

The TRIZ-PLR framework enables structured traversal of the vast landscape of patented technology in green networks. The TRIZ methodology in this study allowed for the unravelling of hidden patterns, identifying underlying challenges, and existing scientific knowledge gulps for innovation in sustainable network communication technologies. Within this iterative process of generating and solving problems, TRIZ-PLR can expand the boundary of green network technology and, hence, contribute to a more sustainable and efficient telecommunications infrastructure.

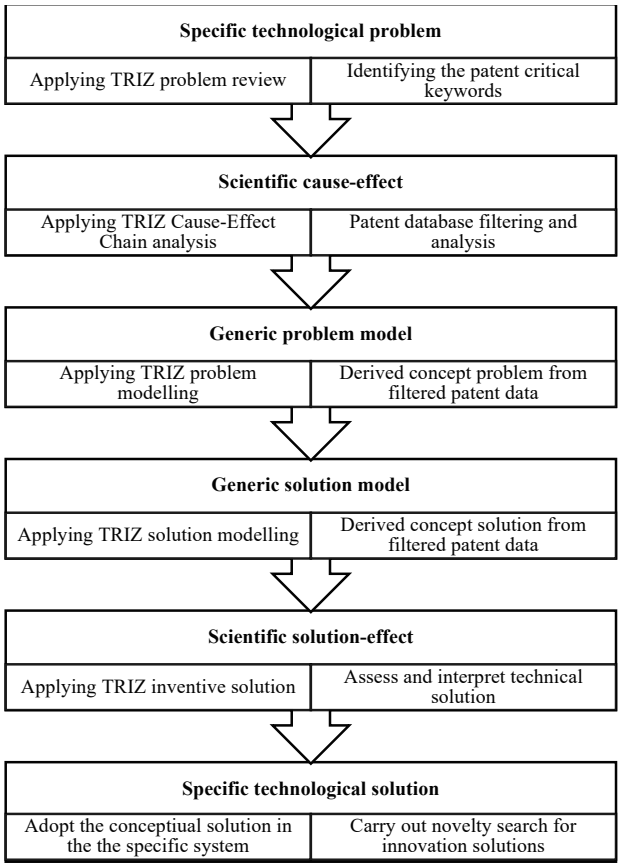


Fig. 2 The framework of TRIZ-PLR[8]

B. *The Sustainability of Green Network for 6G Technologies*

Since connection landscapes are constantly shifting, it is essential to exercise increased caution throughout the transition to 6G wireless networks. Because of this, the arrival of 6G will bring about disruptive developments that are a direct result of the technology of 5G. With machines and Internet of Things devices becoming the principal consumers for mobile data traffic, the scale of connectivity may only accelerate to a prediction of more than 55 billion devices linked to the Internet by the year 2025. This is a notion that has been compared to a turnaround to an industrial revolution. A quantity of data that is anticipated to be over 73 zettabytes per year will be produced by these devices because of their continuous sensing, processing, acting, and communication. This amount of data is unparalleled in the history of humanity. One of the most critical needs for 6G networks is the ability to accommodate diverse form factors, service demands, and hardware capabilities. This requirement is closely related to the requirement for specific technologies that enable the network to function [14].

In contrast to the 5G system, which is depicted in Figure 3, ultra-wideband contiguous bandwidth will be necessary for 6G networks for them to enable hyper-connectivity efficiently. If pervasive computing and ubiquitous connectivity usher in a new era of radical shifts, they also bring with them a great deal of vitally essential challenges. Among the most significant of these challenges are the presentation of terabits-per-second wireless connectivity and zero coverage-hole networks, which hinder computing for connected intelligence.

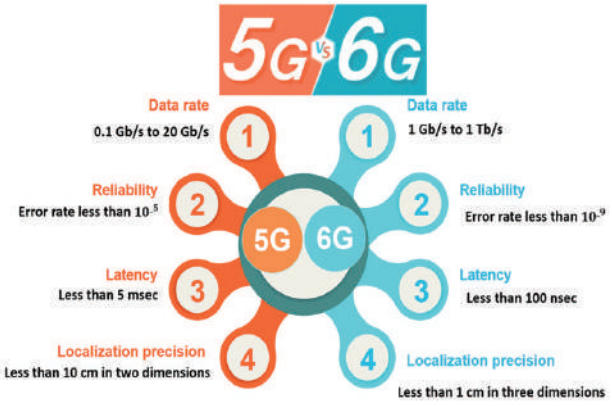


Fig. 3 The comparison features between 5G and 6G.

From a technological point of view, the fundamental objective of tackling these difficulties is to reinvent the hyper-connectivity of the sixth generation and beyond [15]. As a result of this innovation, green network technology has emerged as an essential component for the development of sustainable solutions that will improve the effectiveness and performance of 6G networks. By incorporating green network technologies, such as energy-efficient hardware, intelligent network management systems, and renewable energy sources, 6G networks can reduce their impact on the environment while simultaneously satisfying the requirements of hyper-connectivity [2,16].

III. Data Analysis of TRIZ-PLR on 6G Green Network Technology

Utilising TRIZ-PLR as its foundation, the technique that was deployed was able to identify several significant technological areas and trends in green network technologies. In addition to the incorporation of renewable energy sources, this data analysis highlights the importance of energy-efficient hardware solutions for 6G networks and sophisticated network management systems. It shows how the distribution of patent families according to scientific technology domains and CPC subclasses is responsible for the concentration of technological innovation and the actualisation of technologies in certain regions. Technology that has been patented and falls within the green network topic is centred on the development of ecologically friendly communication infrastructures. There are several critical technical domains, including the creation of hardware that is efficient in terms of energy consumption, the optimisation of traffic management, and the development of energy-saving algorithms for reducing power usage [17]. Several patents made use of materials that were either recyclable or biodegradable throughout the manufacturing process of equipment to reduce their influence on the environment. Other patents concentrate on network function virtualisation and software-defined networking, both of which offer flexibility and resource efficiency for 6G networks; this is the most significant area of development [18]. Several of the patents that pertain to hardware include energy-aware routing protocols, intelligent cooling solutions for data centres, and technology for managing electronic waste that cuts energy usage and minimise environmental damage. Continuing to strike a balance between the developments in 6G technology and the considerations of sustainability are demand response devices that change operations depending on real-time energy readings.

The implementation of the TRIZ-PLR framework resulted in the retrieval of several essential components that are associated with green network systems for 6G technologies. These components are already present in the patent database, as seen in Figure 3. Inside the first stage of the TRIZ-PLR process, a particular technical challenge inside the green network system that pertains to the reduction of energy consumption and the improvement of energy efficiency was identified. This will be accomplished by the development of technologies that generally lower the pervasive level of energy consumption that is inherent in network operations, as well as the optimisation of power management and the development of efficient hardware. The ideas include energy-saving algorithms, intelligent cooling systems, superior antenna designs, novel interference management, and low-power operating modes for infrastructure, cellular and broadcasting. These are all essential components of the inventions. The integration of communication networks and smart grid technologies would go hand in hand with this, with the goal of optimising energy distribution and real-time energy management.

In addition, thematic data on the same subject covers the use of renewable energy sources, such as solar or wind power, to power the infrastructure of the network. Additionally, the advancements that have been made in hybrid power systems, energy harvesting, and efficient methods for energy storage are of equal significance. The reduction of the environmental impact that may be achieved using cutting-edge eco-friendly materials and sustainable production techniques, in addition to

efficient management of electronic waste, is to be noted. Researchers can deduce significant trends related to 6G networks, emerging technologies, and opportunities for improving environmental performance by analysing the granted patents in these areas. This analysis can be used to assist in the formulation of future research and development towards more sustainable means of communication [19-20].

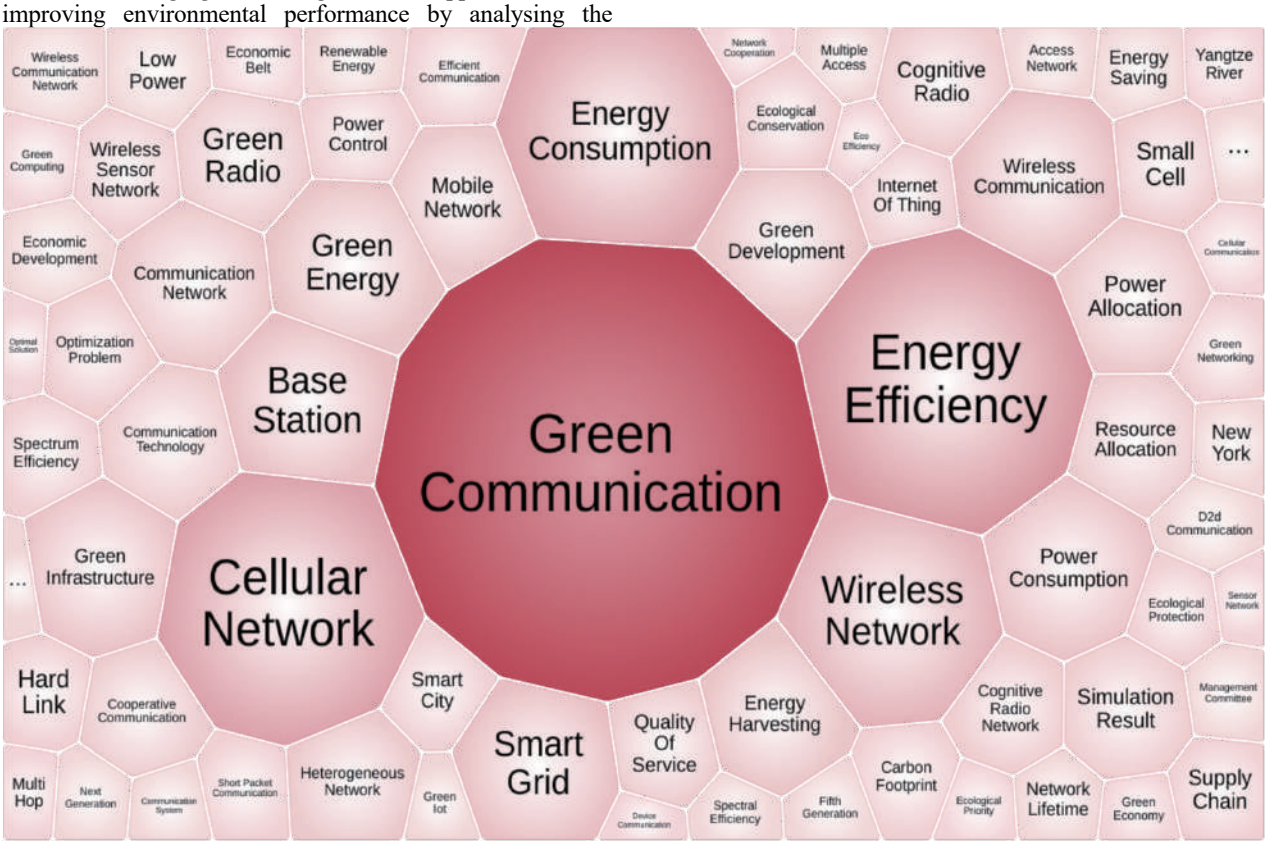


Fig. 3 Thematic analysis of TRIZ-PLR of green network patent analytics for 6G

For the second step of TRIZ-PLR, Fig. 4 shows the distribution of patent families by scientific technology domain in green networks for 6G—underlining areas of significant innovation. Connected with the management of energy consumption and optimisation of the operation of the network are the two most important segments: Control (21.51%) and Telecommunications (21.45%). This involves advanced power management systems, dynamic energy allocation, and energy-efficient communication protocols. The Digital Communication and Computer Technology segments account for 12.75% and 8.43%, respectively, of patents related to energy-aware routing, effective data transmission, and network virtualisation technologies like SDN and NFV that optimise resource utilisation with reduced hardware requirements. Other vital fields are IT Methods for Management at 6.78% and Electrical Machinery, Apparatus, and Energy at 6.05%, related to operational efficiency through better energy efficiency or next-generation data centre facilities, intelligent cooling systems, renewable integration, and superior battery technologies. It entails technologies for monitoring and optimising energy consumption. Other smaller segments, including Audio-visual Technology, 3.91%; Other Special Machines, 5.94%; and Civil Engineering, 1.17%, equally throw their efforts into developing innovations such as low-power multimedia processing, specialised equipment and construction of sustainable infrastructures. All these patents help to achieve a more sustainable, more power-efficient 6G network with reduced environmental impact.

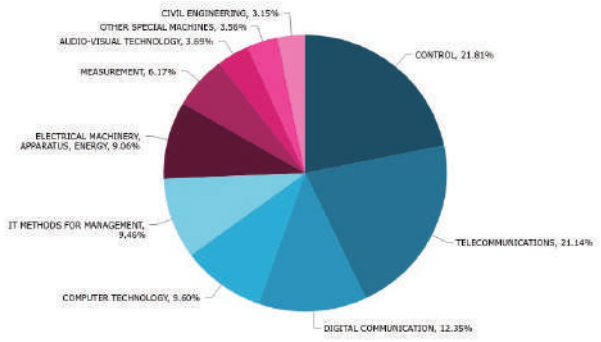


Fig. 4 The distribution analysis of TRIZ-PLR scientific technology domain of green network patent analytics for 6G

The third step in TRIZ-PLR will be to derive a conceptual problem from the patented database that can be analysed. This can be effectively done by identifying the cooperative patent classification subclasses of the patented technology for the green network in 6G. Fig. 5 illustrates the top 10 CPC subclasses that the patented technologies are classified into, indicating critical areas of innovation in the green network for 6G. The first subclass is H04W, covering wireless communication networks, including power control and energy-aware routing to enable meagre power. The next one is Y02D, which deals with methods and techniques for reducing the energy consumption of computing and telecommunication systems, including energy-efficient hardware and cooling systems. G06Q and G06F deal with

data-processing systems and electrical digital data processing, respectively, focusing on innovative grid technologies, energy management systems, and low-power computing.

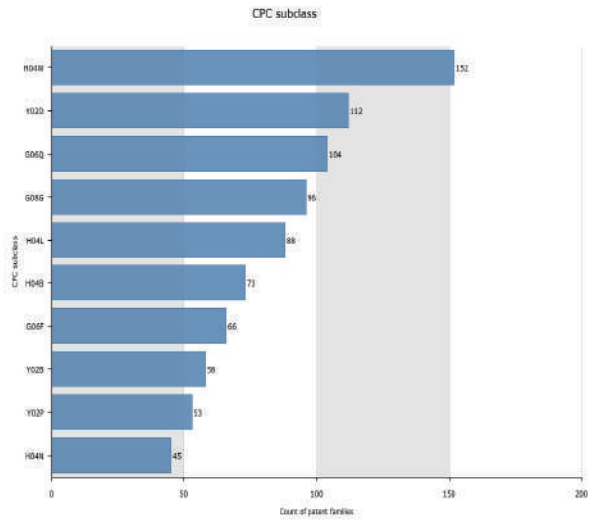


Fig. 5 The TRIZ-PLR CPC subclass analysis of green network patent analytics for 6G

Some other noteworthy subclasses are H04W, which is concerned with efficient methods of data transmission in terms of energy consumption, and H01B, which is centred on conductors or insulators for the purpose of improving energy efficiency in network architecture. Y02B and Y02E are technologies that are used to mitigate the effects of climate change. These technologies include energy-efficient infrastructure and the integration of renewable energy sources. In Y02P, sustainable manufacturing procedures for network components are discussed, and in H04N, approaches for energy-efficient video compression and transmission are included. Collectively, the purpose of these patents is to improve the energy efficiency and sustainability of 6G networks. They include a wide variety of technological breakthroughs in the areas of communication, data processing, and infrastructure.

IV. Case Study of Patented Technology Under CPC Subclass H04W of Green Network For 6G

The development of concept modelling for the technological solution of a green network for 6G is the next step that TRIZ-PLR will take. To accomplish the same thing, a case study is selected from the H04W subclass with the highest CPC. A comprehensive case study on patented technologies that fall under the CPC subclass H04W has been carried out, with the topic of wireless communication networks serving as the primary emphasis. Adaptive power control algorithms, base stations, mobile devices, and network management systems were some of the categories that were identified following TRIZ function analysis, which categorised the technologies, as shown in Table 1. In these evaluations, the crucial roles that power control and energy-efficient technologies play in the process of achieving sustainable 6G networks were outlined.

TABLE I. Function Analysis Power Control in 6G Green Network System

Tool	Action	Object	Result	Function
Adaptive power control algorithms	Adjust	Transmission power	Optimised energy consumption and maintained performance	Dynamic power optimisation
Base stations	Transmit	Data signals	High-speed, low-latency communication	Advanced connectivity
Mobile devices	Access	Network services	Seamless user experience and connectivity	User communication interface
Network management systems	Monitor	Network conditions	Efficient network operation	Network optimisation
Energy sources	Power	Network Infrastructure	Reduced carbon footprint	Sustainable energy supply
IoT devices	Monitor	Environmental Parameters	Real-time data collection and analysis	Environmental monitoring
Supporting infrastructure	Integrate	Network components	Cohesive operation and management	System Integration

In Figure 6, we see the technological landscape of the patent analysis framework that is based on the TRIZ-PLR approach. This framework displays the significant parts that are associated with H04W, which are energy efficiency, resource allocation, and power distribution. The focus of the case study: The provision of an optimisation function for energy consumption while simultaneously maintaining high performance across the network through the utilisation of adaptive power control algorithms is the essential beneficial function of Power Control in Wireless 6G Communication Networks that fall under the CPC Subclass H04W. When it comes to the execution of operations, the incorporation of sophisticated base stations, mobile devices, network management systems, and sustainable energy sources guarantees efficiency. It reduces the adverse effects on the environment. Using TRIZ function analysis under H04W, the innovations that have been granted patents may be classified into a few different categories: topic, function, object, and scientific impacts of the green network for 6G fields.



Fig. 6 The TRIZ-PLR technology cluster of green network patent analytics for 6G

The next step of TRIZ-PLR will be to develop concept modelling for the technological solution of the green network for 6G. A case study has been chosen from the highest CPC subclass of H04W. A detailed case study on patented technologies under CPC subclass H04W, focusing on wireless communication networks, was conducted. The technologies were then grouped into categories, such as adaptive power control algorithms, base stations, mobile devices, and network management systems, using TRIZ function analysis.

These analyses presented the critical roles of power control and energy-efficient technologies in attaining sustainable 6G networks. The primary helpful function of Power Control in Wireless 6G Communication Networks under CPC Subclass H04W is to provide an optimisation function for energy consumption while keeping high performance over the network through adaptive power control algorithms. Advanced base stations, mobile devices, network management systems, and renewable energy sources are integrated to ensure efficient and viable environmental impacts in executing operations. Some patented technologies using TRIZ function analysis fall into categories such as subject, function, object, and the scientific effects representing the green network for 6G fields under H04W.

To advance 6G green network technology innovation, forming research consortia and public-private partnerships among leading universities (like Nanjing University of Posts & Telecommunications and Xidian University) and industry leaders (such as Huawei and SGCC) is essential as shown in Figure 7. Collaborative efforts should focus on developing energy-efficient hardware, advanced power management systems, and integrating renewable energy sources. Establishing industry standards and gaining regulatory support will ensure scalability and compliance. Creating innovation hubs, deploying testbeds, and organising knowledge-sharing events will foster practical implementations and continuous improvement. Additionally, developing specialised educational programs will train the next generation of engineers and researchers, driving sustainable advancements in wireless communication.

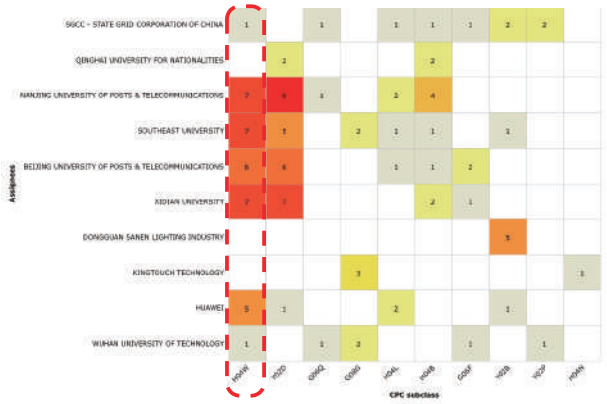


Fig. 7 The assignees of technology development for H04W subclass

V. Conclusion

The telecommunications industry is essential in global connectivity, with the innovation of 6G. It, however, forms one of the leading sources of environmental impacts. Green network communication practices are thus crucial in mitigating these impacts through optimisation of energy use,

reduction in carbon emission, and minimisation of electronic wastes. Applying TRIZ methodologies to patent literature identifies opportunities to enhance sustainability while transitioning from 5G to 6G networks. The findings of this study accentuate the building of research consortia, establishment of industry standards, and fostering of public-private linkages to propel sustainable developments in 6G network technologies.

ACKNOWLEDGEMENT

This author would like to acknowledge the financial support from the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2021/SS02/UTM/02/16) and Malaysia Communication and Multimedia Commission (Digital Society Research Grant MCMC(RED)700-8/2/11/JLD.3(64)

REFERENCES

[1] U.K. Chude-Okonkwo, B.S. Paul, and A.V. Vasilakos, "Enabling precision medicine via contemporary and future communication technologies: A survey". IEEE Access, 11, 2022, pp.21210-21240.

[2] J.R. Bhat, and S.A. Alqahtani, "6G ecosystem: Current status and future perspective". IEEE Access, 9, 2021, pp.43134-43167.

[3] Y. Lu, and X. Zheng, "6G: A survey on technologies, scenarios, challenges, and the related issues". Journal of Industrial Information Integration, 19, 2020, p.100158.

[4] R. Kumar, S.K. Gupta, H.C. Wang, C.S. Kumari, and S.S.V.P. Korlam, "From Efficiency to Sustainability: Exploring the Potential of 6G for a Greener Future". Sustainability, 15(23), 2023, p.16387.

[5] M. Matinmikko-Blue, S. Aalto, M.I. Asghar, H. Berndt, Y. Chen, S. Dixit, R. Jurva, P. Karppinen, M. Kekkonen, M. Kinnula, and P. Kostakos, White paper on 6G drivers and the UN SDGs. arXiv preprint arXiv:2004.14695, 2020.

[6] A.L. Imoize, O. Adedeji, N. Tandiya, and S. Shetty "6G enabled smart infrastructure for a sustainable society: Opportunities, challenges, and research roadmap"—sensors, 21(5), 2021, p.1709.

[7] Z. Abdul Rahim, and M.S. Iqbal, Hydrogen Fuel Cell Technology Revolution and Intervention Using TRIZ S‑curve Analysis for Automotive System Innovation. Chemical Engineering & Technology, 46(12), 2023, pp.2590-2599.

[8] Z. Abdul Rahim, and M.S. Iqbal, Patented Technology Innovation Portfolio in Producing Biofuel Using the Theory of Inventive Problem Solving (TRIZ) Methodology. Chemical Engineering & Technology, 46(12), 2023, pp.2654-2660.

[9] M.S. Iqbal, and Z.A. Rahim, The Rising Potential of Deep Eutectic Solvents: A Patent Analysis of Extraction and Energy Storage Applications. Authorea Preprints, 2024.

[10] Z.A. Rahim and M.S. Iqbal, The Review of Patent Literature and Analytics of Robo-Physic System Evolution Using Theory of Inventive Problem Solving (TRIZ). F. Saeed et al. (Eds.): IRICT 2023, LNDECT 210, pp. 1–13, 2024. https://doi.org/10.1007/978-3-031-59711-4_26

[11] M.S. Iqbal and Z.A. Rahim, The Patented Technology Innovation Portfolio on 4D Printer Using Theory of Inventive Problem Solving 2024.. F. Saeed et al. (Eds.): IRICT 2023, LNDECT 210, pp. 1–10, 2024. https://doi.org/10.1007/978-3-031-59711-4_16

[12] Z.A. Rahim, S.M. Yusof, N.A. Bakar, and W.M.S.W., Mohamad, "The application of computational thinking and TRIZ methodology in patent innovation analytics". In Recent Trends in Data Science and Soft Computing: Proceedings of the 3rd International Conference of Reliable Information and Communication Technology (IRICT 2018) Springer International Publishing. 2019. pp. 793-802.

[13] Z. Abdul Rahim, and M.S. Iqbal, "Innovation Process of Hydrogen Fuel Cell Storage Technology Using the Theory of Inventive Problem Solving (TRIZ)". Chemical Engineering & Technology, 46(12), 2023, pp.2635-2643.

[14] M., Banafaa, , I., Shayea J., Din, M.H., Azmi, A., Alashbi, Y.I. Daradkeh, and A., Alhammadi," 6G mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities. Alexandria Engineering Journal, 64, 2023. pp.245-274.

[15] A., Dogra, R.K. Jha, and S., Jain, "A survey on beyond 5G network with the advent of 6G: Architecture and emerging technologies". IEEE Access, 9, 2020, pp.67512-67547.

[16] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, "AI models for green communications towards 6G". IEEE Communications Surveys & Tutorials, 24(1), 2021, pp.210-247.

[17] D. Pradhan, and K.C., Priyanka, "A comprehensive study of renewable energy management for 5G green communications: Energy saving techniques and its optimisation". Journal of Seybold Report ISSN NO, 1533, 2020. p.9211.

[18] S. Moin, A. Karim, K. Safdar, I. Iqbal, Z. Safdar, V. Vijayakumar, K.T. Ahmed, and S.A. Abid, "GREEN SDN—An enhanced paradigm of SDN: Review, taxonomy, and future directions". Concurrency and Computation: Practice and Experience, 32(21), 2020, p.e5086.

[19] M.N. Mahdi, A.R. Ahmad, Q.S. Qassim, H. Natiq, M.A Subhi. and M. Mahmoud, From 5G to 6G technology: meets energy, internet-of-things and machine learning: a survey. Applied Sciences, 11(17), 2021. p.8117.

[20] O.L. López, O.M. Rosabal, D.E. Ruiz-Guirola, P., Raghuwanshi, K. Mikhaylov, L. Lovén, and S. Iyer, Energy- sustainable iot connectivity: Vision, technological enablers, challenges, and future directions. IEEE Open Journal of the Communications Society, 2023.

Enhancing the Dynamic Range of Refractive Index Sensors Using Neural Network Classification with MMI Coreless Fiber

Athi Shankhara Velusamy
Malaysian-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia Kuala Lumpur
54100 Kuala Lumpur, Malaysia
athishankhara@graduate.utm.my

Muhamad Yusof Mohd Noor
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru,Malaysia
yusofnor@utm.my

Fauzan Ahmad
Malaysian-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia Kuala Lumpur
54100 Kuala Lumpur, Malaysia
fauzan.kl@utm.my

Azizul Azizan
Razak Faculty of Technology and Informatics
Universiti Teknologi Malaysia Kuala Lumpur
54100 Kuala Lumpur, Malaysia
azizulazizan@utm.my

Abstract— The need for accurate RI measurement across a wide range is critical in various applications, yet traditional methods often face limitations in dynamic range and ambiguity. To overcome these limitations, machine learning (ML) techniques offer a powerful approach by identifying complex patterns in data to enhance measurement accuracy. To address these challenges, a supervised neural network machine learning (ML) algorithm is proposed to measure refractive index (RI) values both below and above the RI of the fiber material using a multimode interference (MMI) fiber sensor. The integration of ML techniques, specifically neural networks, enhances the ability to accurately predict RI values by leveraging complex patterns in the data. The sensor in this study is constructed by splicing a coreless multimode fiber (CMF) segment between two single-mode fiber (SMF) leads. The Neural Network classification algorithm is utilized to measure low and high RI regimes. This trained model demonstrates a wide dynamic range in RI measurement, covering ranges from 1.30 to 1.39 (low RI regime) and 1.46 to 1.55 (high RI regime), achieving an average accuracy of 58.1% for RI model prediction without significant RI ambiguity.

Keywords— Supervised Machine learning, classification learner, Decision Tree, refractive index fiber sensor, coreless multimode fiber, multimode interference

I. Introduction

Refractive index (RI) sensors are utilized in a variety of applications including energy production, chemical and biological analysis, and specimen detection. Optical fiber RI sensors, particularly popular for their compact size, portability, and immunity to electromagnetic interference, offer high sensitivity and have been developed in several forms such as surface-plasmon-resonance [1-2], interferometers [3-4], gratings [5-6], and multimode interference [7]. These sensors are generally optimized for detecting external mediums with a refractive index lower than that of the fiber itself, especially below 1.45 [8], to ensure minimal light loss and tight light confinement within the fiber. Conversely, some sensors are also designed for higher RI mediums above 1.45, useful in contexts like oil sensing [9].

The demand for RI fiber sensors that can accurately detect both low and high RI values is increasing. To address this, advanced signal processing techniques have been integrated into these sensors. For instance, Fourier Transform techniques have been applied in fiber Fabry-Perot Interferometers to produce a stable, clear interference spectrum across RI values from 1.3924 to 1.5882. Recently, an AI-based classification method has been applied into RI fiber sensing, using a support vector machine supervised algorithm in Fabry-Perot Interferometer designs to detect RIs ranging from 1.3158 to 1.5809 [10], demonstrating a significant advancement in fiber refractometry technology.

In this paper, we introduce the application of a classification machine learning algorithm to enable both low and high RI sensing regimes in a multimode interference (MMI) fiber sensor. The sensor comprises a short section of CMF between single-mode fiber (SMF) leads. The sensing mechanism of this sensor is based on guided and leaky multimode interference. This method employs a supervised Decision Tree (DT) algorithm. Unlike the conventional approach of examining a single or a few spectral features in the spectral response, this algorithm harnesses the rich information embedded in the entire spectrum. It considers each of the sampled wavelengths as a predictor variable derived from a set of sensing signals obtained with samples of different RI values. Subsequently, it utilizes these features to predict RI values based on the spectral signals obtained from each liquid sample. To the best of our knowledge, this is the first time DT classification learner has been applied in MMI fiber sensor type.

II. Methodology

The approach for predicting the refractive index (RI) signal in this proposed work is based on a Neural Network (NN), a supervised learning technique. NN are a popular machine learning technique used for categorizing data into predefined classes and making predictions. They work by using layers of interconnected nodes (neurons), where each connection has an associated weight. The network learns by adjusting these weights based on the error of its predictions.

In a NN for classification, the algorithm consists of an input layer, one or more hidden layers, and an output layer. Each

neuron in a layer receives input from the previous layer, processes it through an activation function, and passes the output to the next layer. The learning process involves forward propagation, where inputs pass through the network to generate an output, and backpropagation, where the error between the predicted and actual output is calculated and used to adjust the weights.

To construct the RI sensor, a coreless multimode fiber (CMF, FG125LA, Thorlabs) with a diameter of 125 µm was first spliced to an input single-mode fiber (SMF, SMF-28, Thorlabs) using a fusion splicer (FITEL S178A) set to auto mode. After splicing, the CMF was trimmed to a length of 3.5 cm. An output SMF with a flat, clean end face was then spliced to the previously combined input SMF and CMF, establishing the multimode interference (MMI) structure based on the CMF, as illustrated in Fig. 1. In the analysis of the MMI sensor's transmission spectra, shown in Fig. 2, the traditional method of detecting shifts through spectral dips or peaks faces difficulties with complex spectra that include overlapping features. This makes it challenging to track changes at a single wavelength. For more accurate sensing, it is advantageous to analyze the entire transmission spectrum, spanning multiple wavelengths. Therefore, the proposed DT algorithm takes the entire transmission spectrum of the sensor as input data. Each transmission at the sampled wavelengths is considered a feature variable in the input data array. In a tree-structured regression model, three types of nodes exist: the root node, representing the initial node for the entire sample, which may further split into additional nodes; branch nodes, representing input features of a data array, with branches symbolizing decision rules; and leaf nodes, denoting the output. The model predicts the output of the SRI based on the input spectrum of the sensor. For the experiment, a total of 200 spectral responses from the MMI sensor were recorded and used as data samples to train and test the DT model. These spectral responses corresponded to variations in RI spanning from 1.30 to 1.39 (low RI regime) and 1.46 to 1.55 (high RI regime), encompassing 10 cycles of measurements. Wavelength sampling points numbered 334, ranging from 1520 to 1580 nm, with a corresponding interval of 0.18 nm. Eighty per cent of the data samples were randomly chosen to train the model, and the remaining 20%, unseen by the model during training, were used to test the trained DT model. The DT training processes were implemented using MATLAB R2023b.

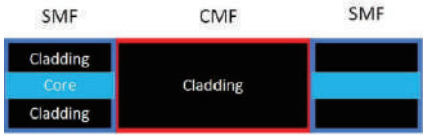


Fig. 1 The structure of CSF and SMF in cross-section geometry

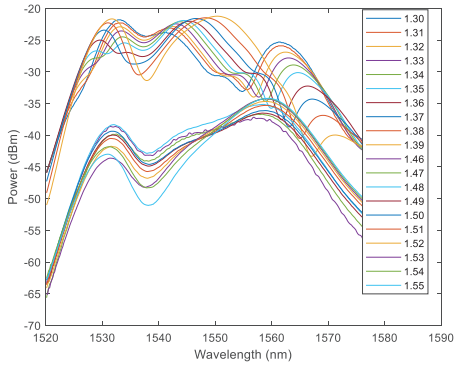


Fig. 2 Output transmission spectrum of the RI MMI fiber sensor

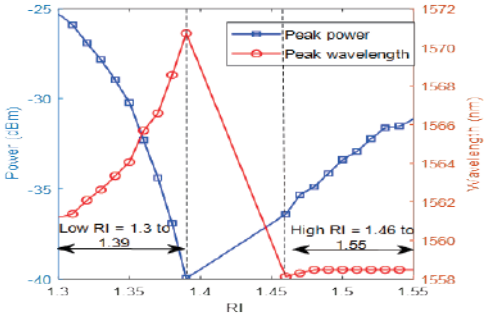


Fig. 3 Response of the MMI fiber sensor through the power change and wavelength shift.

In Fig. 3, the power change and wavelength shift were plotted for both low and high RI regimes in a transmission spectrum area in Fig. 2 from wavelength from 1550 nm to 1580 nm. For wavelength shift interrogation, the sensor exhibits a linear response only in the low RI regime, with minimal shift observed in the high RI regime. Conversely, power change interrogation demonstrates a distinct linear response across the low to high RI regimes, but the responses overlap. This results in ambiguity in the amplitude power readings of the MMI sensor, making traditional methods unsuitable for accurately interpreting the sensor's response across the full spectrum of refractive indices, from low to high RI regimes. Consequently, a new interrogation technique that incorporates machine learning is necessary to effectively address this issue.

III. Results and Discussion

Fig. 4 shows a confusion matrix for a predictive model of NN classification, where each row represents the actual class (True Class) and each column represents the predicted class (Predicted Class). Each cell within the matrix shows the percentage of predictions for that class pair, where the diagonal cells (highlighted by higher percentages and different shading) indicate correct predictions (true positives) where the predicted class matches the true class. The percentages off the diagonal represent misclassifications. The bottom rows of the matrix provide the Positive Predictive Value (PPV) and False Discovery Rate (FDR) for each predicted class. PPV (also known as precision) is the ratio of correct positive predictions to the total predicted positives, while FDR is the ratio of incorrect positive predictions to the total predicted positives. This matrix shows that most of the

RI are accurately predicted and are prone to error with an acceptable accuracy.

Fig. 5 displays the relationship between two variables out of 334 input wavelength variables, labelled as 1.5502 nm and 1.5500 nm, by plotting their values against each other. The wavelength around 1550 nm is chosen here as it is the standard wavelength for any power source in telco. The RI data points, represented in different colours, seem to form a linear trend, suggesting a strong correlation between these two variables. Each colour represents a different RI within the dataset from low to high RI regime. The presence of a few outlier points (the 'x' symbols in red) indicates instances where the model's predictions deviate significantly from the actual values.

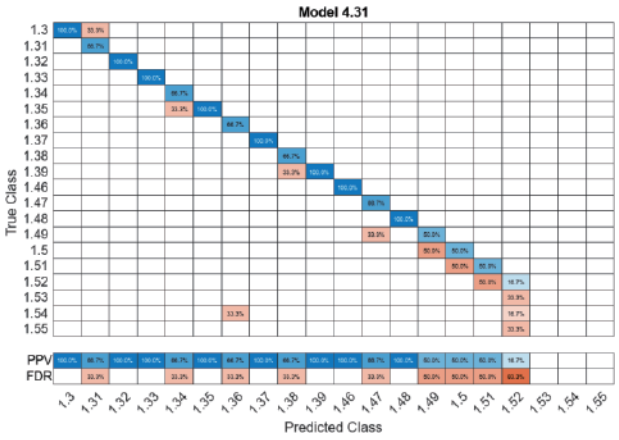


Fig. 4. Validation confusion matrix of the DT model

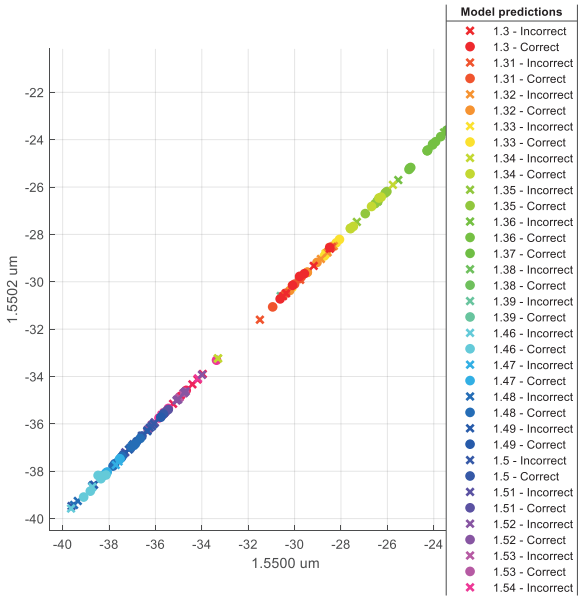


Fig. 5. Scatter plot of the NN model

Table I summarizes the performance of an NN model, specifically a "Trilayered NN" in two distinct phases: Validation and Test. In the Validation phase, the Trilayered achieved an accuracy of 73.8% with a total cost of 42. Conversely, during the Test phase, the model's performance dropped, exhibiting an accuracy of 67.5% with a considerably lower total cost of 13. This implies that while the model performed better in the validation phase, its effectiveness decreased in a real-world testing scenario, suggesting issues like overfitting or the need for further tuning to maintain consistency across different datasets, achieving a maximum classification accuracy of around 73.8%.

TABLE I
Training model accuracy NN classification algorithm

Model	Submodule	Validation		Test	
		Accuracy	Total cost	Accuracy	Total cost
NN	Trilayered NN	73.8 %	42	67.5 %	13

IV. Conclusion

In this paper, we have introduced a novel approach to address the challenges posed by overlapping and intricate signal patterns in low and high RI measurements using the MMI fiber sensor. This method leverages a classification machine learning algorithm, specifically the NN supervised classification algorithm, to predict RI values based on the spectral signals obtained from each liquid sample. The fabrication process for the sensor is straightforward, involving only fiber cleaving and splicing. The MMI sensor is created by splicing the CMF to the input and output SMFs. Our results demonstrate that the proposed method accurately identifies the RI values of liquid samples in both low and high RI regimes based on the MMI complex spectrum signal transformation, effectively mitigating ambiguity issues.

V. Acknowledgement

This work was supported by the Fundamental Research Grant Scheme (FRGS), grant number FRGS/1/2023/TK07/UTM/02/20 funded by the Ministry of Higher Education (MOHE), Malaysia.

References

[1] Z. Yin, X. Jing, G. Bai, B. Wu, Z. Gao, C. Liu, and Y. Li, "Experimental study of dual-parameter SPR sensor with integrated sensing channel," *IEEE Sensors Journal,* vol. 23, no. 8, pp. 8385-8390, 2023

[2] H. Wang, M. Wu, J. Zhou, S. Zheng, T. Xie, W. Dai, H. Fu, W. Lv, N. Chen, and Y. Bu, "SPR Sensor Based on Cascaded NCF and U-shaped Multimode Fibers for Simultaneous Detection of Refractive Index and Temperature," IEEE Sensors Journal, vol. 23, no. 15, pp. 16851-16858, 2023

[3] J. W. Gao, D. N. Wang, and B. Xu, " Highly Improved Refractive Index Sensitivity in Fiber Mach-Zehnder Interferometer Based on Two Tapers," *IEEE Photonics Technology Letters*, vol. 35, no. 17, pp. 935-938, 2023

[4] X. Li, R. Bai, L. Du, W. Zhang, and M. Li, " Femtosecond laser fabrication and refractive index sensing characteristics of highly sensitive fiber Michelson microcavity," IEEE Sensors Journal, 2023

[5] S. Gao, Y. Liu, J. Yang, Z. Duan, T. Yin, Z. Liu, J. Shi, L. Yuan, and C. Guan, "Ultrasensitive refractive sensor based on femtosecond laser inscribed long-period fiber grating with a few microcavities," *Journal of Lightwave Technology*, 2023

[6] Y. Zhao, T. -y. He, M. -q. Chen, Z. -b. Zhang, and R. -j. Tong, " Ultra-Short Fiber Bragg Grating Composed of Cascaded Microchannels in a Microprobe for Refractive Index Measurement," *Journal of Lightwave Technology,* vol. 41, no. 8, pp. 2555-2561, 2023

[7] K.Wang, X. Dong, M. H. Köhler, P. Kienle, Q. Bian, M. Jakobi, and A. W. Koch, " Advances in Optical Fiber Sensors Based on Multimode

Interference (MMI): A Review," *IEEE Sensors Journal,* vol. 21, no. 1, pp. 132-142, 2023

[8] J. Li, A. Razzaq, H. Zainuddin, F. Hanaffi, R. M. Chyad, H. Abdul Razak, and A. A. Latiff, " A review: Development of novel fiber-optic platforms for bulk and surface refractive index sensing applications," Sensors and Actuators Reports, vol. 2, no. 1, pp. 100018, 2020

[9] R. T. Ghahrizjani, H. Sadeghi, and A. Mazaheri, " A Novel Method for onLine Monitoring Engine Oil Quality Based on Tapered Optical Fiber Sensor," IEEE Sensors Journal, vol. 16, no. 10, pp. 3551-3555, 2016

[10] R. Martínez-Manuel, L. M. Valentín-Coronado, J. Esquivel-Hernández, K. J. J. Monga, and S. LaRochelle Lei, " Machine learning implementation for unambiguous refractive index measurement using a self-referenced fiber refractometer," IEEE Sensors Journal, vol. 22, no. 14, pp. 14134-14141, 2022

Deep Neural Network Towards URL Classification Using Effective URL Tokens

1st Zafar Ali
Faculty of Artificial Intelligence Malaysia, Kuala Lumpur.
https://orcid.org/0009-0007-5281-3176

2nd Siti Sophiayati Yuhaniz1
Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.
https://orcid.org/0000-0001-8309-5356

3rd Ghulam Mujtaba Shaikh2
Department of Computer Science, Sukkur IBA University, Sukkur 65200, Pakistan.
https://orcid.org/0000-0002-1563-1142

***Abstract*—**Methods for real-time web page classification are currently required due to the exponential growth of the Internet user base and the volume of web pages. To tackle these challenges, scholars have suggested URL-based methods in the literature, which offer classification speed and computational efficiency gains over content-based approaches. This study empirically evaluated deep neural networks using a large dataset of URL classification using simple URL tokens extracted from URI and Query String.

This study aims to diversify the range of URL tokens using the CNN network and enrich URL Tokens by extracting features from Query String and improving the quality of the training dataset. Such improvements have enabled the model to attain an F1 score of 0.8787, signifying a superior level of accuracy relative to prior models.

Keywords: URL Classification, URL Token, Feature Engineering, CNN, large Dataset

I INTRODUCTION

In the context of URL classification, the model's ability to accurately interpret and classify URLs based on simple tokenized features extracted from the URLs themselves is critical. These features include structural components of the URLs, such as the protocol, domain, path, and query parameters, which can provide significant insights into the content and purpose of a webpage. By employing models like CNNs, the study aims to capture spatial hierarchies in token sequences.

Investigating Convolutional Neural Networks (CNNs) for URL classification of the DMOZ dataset involves using CNNs to categorize URLs based on their content. The DMOZ dataset is a large, hierarchical directory of URLs. CNNs, known for their pattern recognition capabilities, are adapted to handle textual data by reprocessing URLs into numerical formats. The model, trained on labelled DMOZ data, uses convolutional layers to extract features and fully connected layers to classify URLs. Performance is evaluated using metrics like accuracy and F1-score, highlighting CNNs' potential in URL classification compared to traditional methods [1, 2].

In the realm of URL classification, the n-grams model has traditionally been dominant; however, studies based on the n-grams model have yet to attain outstanding performance. Thus, there's a compelling need to embrace alternative feature engineering methodologies within the URL training sets to secure superior results beyond those of previous feature engineering methodologies. Therefore, in recent years, a shift by some scholars towards the use of character-gram and word-gram models has led to better results compared to those obtained with traditional n-gram features [1, 2]. Another attempt [3-6] was made to improve the quality of features of the URL dataset by implementing character embedding and word embedding to capture all features at the most granular level, capturing patterns in character sequences that might signify malicious intent, specific content types, or domain categories. Moreover, it can establish a relationship between "irregular words" and regular words by comparing their alphabetical order.

In light of the above limitations of the studies, a new feature engineering technique that can produce good results with a large dataset needs to be introduced. Therefore, simple URL Word Net tokens generated based on the original URL that is more effective compared to conventional character and word features in terms of Convolutional Neural Networks [3, 4].

II. LITERATURE REVIEWS

In this section, we provide a comprehensive literature review on URL classification in terms of machine learning and deep learning.

In the literature, URL-based approaches applying both traditional machine learning and deep learning algorithms were recommended. Kan et al [5, 6] a complex segmentation/expansion technique, was utilized to extract features from URLs. Several methods with tokens and n-grams (n=4 to 8) features were recommended[7]. Rajalakshmi suggested a heuristic strategy for feature selection based on a dictionary[8]. This method leads to an increase in the size of the dictionary as the number of URLs increases, achieving

an F-measure value of 0.76 on the open directory project, also called DMOZ. A dictionary of 4-grams derived from Open Directory Project (ODP) URLs was used in the creation of the SVM binary classifier, which was used to classify web pages as either health pages or not. In another work, 7, a rejection framework was added for filtering URL features. A method utilizing Naive Bayes for URL classification was employed, incorporating supervised feature selection and a rejection framework to find significant properties of URLs through statistical methods[9]. Rajalakshmi et al.[1] proposed a deep learning method using a Recurrent Convolutional Neural Network to classify URLs specifically for children, and a CNN-based method with a Bidirectional Gated Recurrent Unit (BGRU) to extract comprehensive content features and retain the sequential information of the URL and Recurrent CNN achieved 82.04 F1-score of percent. Deep learning neural network (CNN and Bidirectional RNN, URLNet) on the DMOZ dataset achieved an F1 score of 81.17 per cent on Bi-LSTM while selecting a sample size of 10 percent[2]. On a dataset[10] that used Vietnamese words, a convolutional neural network (CNN) was implemented with a word embedding layer, and an F1 score of 0.9758 was produced.

III. METHODOLOGY

The integration of WordNet token embedding techniques with the DMOZ dataset offers a powerful approach to URL classification. This approach improves contextual comprehension of URL tokens and provides hierarchical feature representation, an important feature for embedding new unseen tokens to the trainset, but they are not part of the original dataset.

Dataset:

The DMOZ is the largest and the most comprehensive dataset for URL classification. This dataset includes 1.5 million URLs from 15 categories: Arts, Business, Computers, Games, Health, Home, News, Recreation, Reference, Science, Society, Shopping, Sports, Adult, and Kids. The details of the DMOZ dataset are shown in Figure 3.2. As seen here, the dataset is imbalanced, whereby each class contains a different number of URLs. Further, it can be noticed here that the 'Arts,' 'Society' and 'Business' classes contained 734K URLs; additionally,' 'Sports,' 'Science,' 'Computers,' 'Shopping,' 'Recreation,' 'Reference,' 'Health,' 'Games,' 'Kids,' 'Adult,' 'Home' and, 'News' hold 715K.

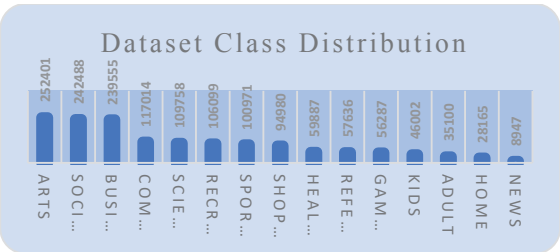


Fig. 1. DMOZ Dataset Class

The feature engineering technique is an integral part of this research because URLs have concise and limited information about the website, which poses a significant challenge in extracting meaningful and robust features for classification purposes. The URLs have concatenated strings that lead to words, non-meaningful words, abbreviations, and combinations of some special characters with characters and numbers. It is challenging to capture the full context or specific nuances inherent in a URL for a single word or token; the URLs are not only sources of tokens, but Query String is also part of tokens to increase URLs.

Pre-processing:

In natural language processing, data pre-processing is a critical part of the machine and deep learning models by ensuring clean, standardized, and meaningful data, as in Table 1. **Tokenization** splits the URL into smaller parts or tokens (like scheme, domain, path, and query parameters). **Delimiters** such as "/," "?", "&," =," and. " help separate URL components. Lowercasing leads to standardizing the URL by converting all characters to lowercase, ensuring uniformity. Removing Redundant 'www,' 'HTTP," https, ' etc, usually doesn't contribute to the semantic meaning. They are replacing encoded sequences (like "%20") with their actual characters ("cybermuse.gallery.ca"). Identifies the top-level domain to understand the domain category and parses the URL to extract TLDs like ".com," ".org," and ".edu." Query parameters are valuable data used to categorize URLs, often consisting of key-value pairs. Stopwords are famous words that are typically not effective for classification purposes. Stemming or Lemmatization is not applied to individual components of the URLs during text processing.

Token Embedding:

WordNet is a lexical database that organises English terms into sets of synonyms known as synsets, provides concise definitions, and records a variety of semantic relationships between these sets.

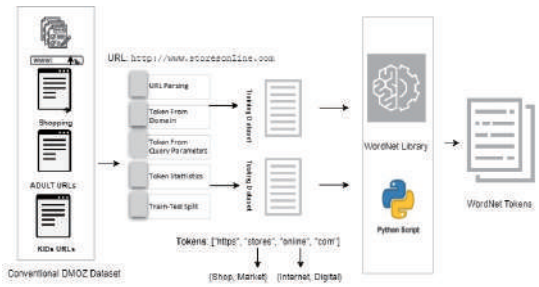


Fig 2. Word-Net Token Embedding Technique using DMOZ DS

Tokenization Split the URL and any associated text (**Tokens**: ["https", "stores", "online", "com"]) into individual words (tokens). Synsets provide synonyms and semantic relations that can enhance the understanding of the token in Fig. 2 token 'Shop', a synset includes retail, store, and market. If a URL is related to books, additional features might include

literature, novels, and publications. By adding synonyms for each token, we capture more contextual information. This reduces the sparsity of the data and improves the model's ability to generalize. It creates hierarchical features like a dog, a hypernym is an animal, and a hyponym is a poodle and semantic relations such as a car, a meronym is a wheel, and a holonym is a vehicle. Table 1 contains more WordNet Tokens examples to reflect semantic relationships among tokens.

TABLE 1.
URL WORD NET TOKEN

Man	Story	Profession	Sexual
Human	Tale	Profession	Sexual
Woman	Report	Professing	Intimate
Adult male	History	Business	None
Human	Fib	None	None
Store	**Clothes**	**Learning**	**Liberty**
Shop	Clothes	Teach	Impropriety
Market	Garment	Memorize	Autonomy
Retailer	Enclothe	Instruct	Familiarity
Sales	Dress	Acquire	Liberty

Let us suppose the 'Store' is part of the DMOZ Dataset but 'Shop' and 'Market' are not part of the dataset, but the model can see unseen features that are not part of the existing dataset. In future; if the model sees the 'Shop' or 'Market' token it can easily predict the "Shopping" category.

A. Convolutional Neural Network

CNNs, widely recognized for their effectiveness in image processing, have also been modified to handle text classification problems, such as URL classification. URLs can be effectively utilized to extract local and position-invariant information, as different URL sections typically provide clues about the nature of the connected material. Feature Extraction: CNNs are excellent at automatically detecting important features without needing specific instruction on what to look for, which can be crucial for identifying essential features.

An embedding layer converts input sequences into 100-dimensional vectors, allowing the model to represent words in high-dimensional space. It then creates a 1-dimensional convolutional layer (Conv1D) with X filters and a kernel size of 2 using the ReLU activation function to produce non-linearity. A dropout rate layer randomly deactivates a fraction of the input units during training to prevent overfitting. The MaxPooling1D layer selects the maximum value in a window, minimising dimensionality and computing effort. The multi-dimensional output of convolutional and pooling layers is flattened into a one-dimensional vector. Fully connected layers process this vector. These layers combine collected characteristics and use weighted sums and activation functions to predict. Final probability distributions for predefined categories are generated by the Softmax output layer.

TABLE 2.
CNN MODEL PARAMETER & CONFIGURATION

Layer Type	Output Shape	Param
Embedding	(None, 203, 100)	59140300
Conv1d	(None, 202, 512)	102912
MaxPooling1D	(None, 101, 512)	0
Dropout	(None, 101, 512)	0
Conv1d_1	(None, 100, 512)	524800
MaxPooling1	(None, 50, 512)	0
Dropout_1	(None, 50, 512)	0
Donv1d_2	(None, 49, 512)	524800
MaxPooling1	(None, 24, 512)	0
Dropout_2	(None, 24, 512)	0
Flatten	(None, 12288)	0
Dropout_3	(None, 12288)	0
Dense	64	786496
Dense	15	975
Total parameters	61,669,760	

In this experiment, various parameters were precisely optimized within the model configuration, utilizing simple URL token features extracted from the URI and Query String. This strategic adjustment led to an enhancement of model performance by an impressive 2%.

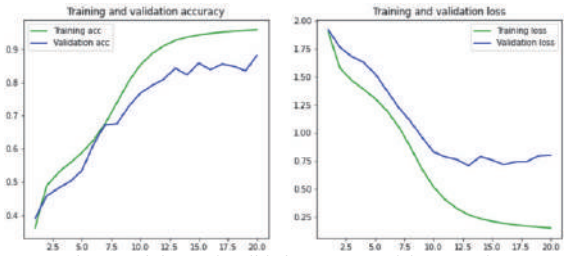


Fig 3. Training and Validation CNN Model 20 Epochs

The training output of the CNN module using the DMOZ dataset shows significant improvements over 20 epochs. Initially, the model starts with a high training loss of 1.8959 and an accuracy of 36.20%, with validation loss and accuracy at 1.9156 and 39.00%, respectively.
As training progresses, the loss consistently decreases, and accuracy increases. By epoch 20, the training loss drops to 0.1547, and the accuracy reaches 95.78%. The validation loss and accuracy improved to 0.7996 and 88.20%, respectively. This indicates effective learning and better generalization, showing CNN's ability to classify URLs accurately.

TABLE 3.

PRIOR EXPERIMENTS RESULTS OF MULTIPLE URL CLASSIFICATION

Features	All Grams [11]	n-Grams [12]	Char Word Embedding [2]	URL Word Net Tokens
	Machine Learning		Convolutional Neural Network	
Adult	87.60%	**32.03%**	**61%**	**0.5948**
Arts	81.90%	68.93%	64%	0.8872
Business	82.90%	86.24%	26%	0.8998
Computers	82.50%	95.34%	49.%	0.7109
Games	86.70%	96.28%	42%	0.7973
Health	82.40%	98.37%	35%	0.9806
Home	81.00%	95.02%	76%	0.8076
Kids	80.00%	81.18%	69%	0.6203
News	80.10%	91.26%	38%	0.9681
Recreation	79.70%	96.89%	36%	0.9560
Reference	84.40%	90.50%	68%	0.9486
Science	80.10%	94.83%	60%	0.9612
Shopping	83.10%	98.31%	**30%**	**0.9644**
Society	80.20%	93.12%	37%	0.9454
Sports	84.00%	96.21%	78%	0.9710
F1 Score	82.44%	87.63%	81.23%	**0.8787**

It can be easily noticed that Table.3 contained experiment outcomes of the prior investigations that could not achieve outstanding performance by applying several feature engineering techniques: all-grams, n-gram, char-grams, word-grams, and char & word embedding based on the large DMOZ dataset.

IV CONCLUSION

A token embedding approach using WordNet is the first attempt in term multiple URL classification using the DMOZ dataset.

WordNet, a comprehensive English lexical database, provided features that improved the CNN model. WordNet stores synsets of synonyms and their semantic links. WordNet lets us create URL token characteristics with richer semantic context. The model can comprehend the general and specific meanings of characters by utilising hypernyms and hyponyms to establish a hierarchical context. The token embedding technique seems very effective approach to comprehending new features

The machine learning model can see unseen features that are not part of the existing dataset. Similarly, the Store reflects 'Shop' & 'Market,' Man reflects 'Woman' & 'Human,' and Story reflects 'Tale' & 'Report.' The reflected features are not part of the existing dataset, but they are part of the trainset. Because of this methodology, the model can be capable of seeing new features that are not part of the DMOZ datasets. The expanded feature set led to improved robustness and generalization to unseen URLs in the model. It can be easily noticed that some categories in the DMOZ datasets are challenging such as 'Arts', 'Adult', 'Business', 'Health',' News' and 'Shopping' because the individual category has the lowest F1 score; but token embedding technique all categories have been performed well Table 3. This approach has a significant impact on the performance of the CNN model.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.rt

REFERENCES

[1] Rajalakshmi, R., et al., *Design of Kids-specific URL Classifier using Recurrent Convolutional Neural Network.* Procedia Computer Science, 2020. **167**: p. 2124-2131.

[2] Kexin, X., et al., *URL Classification with Deep Learning.* 2021.

[3] Feng, L., Y. Cui, and J. Hu. *Detection and classification of malware based on FastText*. in *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS)*. 2020. IEEE.

[4] Kale, A.S., et al., *Malware classification with word2vec, hmm2vec, bert, and elmo.* Journal of Computer Virology and Hacking Techniques, 2023. **19**(1): p. 1-16.

[5] Kan, M.-Y. and H.O.N. Thi. *Fast webpage classification using URL features*. in *Proceedings of the 14th ACM international conference on Information and knowledge management*. 2005.

[6] Kan, M.-Y. *Web page classification without the web page*. in *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*. 2004.

[7] Baykan, E., et al., *A comprehensive study of features and algorithms for URL-based topic classification.* ACM Transactions on the Web (TWEB), 2011. **5**(3): p. 1-29.

[8] Rajalakshmi, R. and C. Aravindan. *Naive bayes approach for website classification*. in *International Conference on Advances in Information Technology and Mobile Communication*. 2011. Springer.

[9] Rajalakshmi, R. and C. Aravindan, *A Naive Bayes approach for URL classification with supervised feature selection and rejection framework.* Computational Intelligence, 2018. **34**(1): p. 363-396.

10] Hung, P.D., N.D. Hung, and V.T. Diep. *URL Classification Using Convolutional Neural Network for a New Large Dataset*. in *International Conference on Cooperative Design, Visualization and Engineering*. 2022. Springer.

11. Baykan, E., et al. *Purely url-based topic classification*. in *Proceedings of the 18th international conference on World wide web*. 2009.

12. Shawon, A., et al. *Website classification using word based multiple n-gram models and random search oriented feature parameters*. in *2018 21st International Conference of Computer and Information Technology (ICCIT)*. 2018. IEEE.

On-Device Plastic Waste Classification for Waste Sorting Robots

Hyuntae Cho
Dept. of Digital Content
Tongmyong University
Busan, Republic of Korea
marine@tu.ac.kr

Young Jin Lee

RobOne Inc.
Busan, Republic of Korea
robert@robone.co.kr

***Abstract*— The ecological crisis, threats to biodiversity, threats to human health, and climate change caused by various types of plastic pollution are being seriously addressed globally. This paper proposes an on-device plastic waste classification system for waste sorting robots. The proposed system essentially operates on an embedded system, acquiring waste images and utilizing YOLOv8 to classify the types of plastic waste. We gathered a custom waste dataset from a waste recycling center in reality and trained the model to enhance classification accuracy. This paper also outlines the performance evaluation conducted to validate the effectiveness of the proposed approach.**

Keywords*— *Plastic waste, waste recycling, plastic reuse, YOLOv8, waste segregation, plastic recycling

I. Introduction

Plastic materials play an essential role in modern society due to their low cost, lightness, and excellent physical properties. However, their widespread use in various applications also contributes significantly to pollution and resource depletion when improperly managed. Plastic recycling is therefore crucial for safeguarding our environment, conserving valuable resources, minimizing waste, fostering economic prosperity, and advancing towards a more sustainable future [1]. At recycling centers, plastic waste undergoes a sorting process by size and material after being collected. This sorting can be performed manually by workers or with the assistance of machines such as conveyor belts, trommels, and optical sorters [2].

Several previous studies have utilized deep learning techniques to recognize waste materials. *Susanth* [3] introduced an image classifier that identifies objects and detects the type of waste material using convolutional neural networks (CNNs) such as ResNet50, DenseNet169, VGG16, and AlexNet. *Sudha* [4] proposed an automatic classification method for environmentally friendly waste segregation using deep learning, where objects are classified as biodegradable and non-biodegradable. *Sousa* [5] suggested automating waste sorting with deep learning. However, these studies are difficult to use in the actual field because they are trained using very clean waste like open datasets, and they do not sort in the way required by recycling centers.

This paper proposes a plastic waste classification system for waste sorting robots. In this study, we utilized YOLOv8, the latest version of YOLO, to classify different types of plastic. Additionally, we used the collected dataset consisting of 3,000 images of plastic waste from real environments. These images were then used to train the YOLOv8 model, aiming to enhance classification accuracy. Subsequently, we tested the trained model on the device in real-world environments. Furthermore, this paper presents experiments and performance evaluations of the proposed method.

This research was supported by Korea Institute for Robot Industry Advancement (Grant No. 2024-03015-3).

II. Plastic Waste Classification For Waste Sorting Robots

Delta robots are utilized to swiftly sort objects in dynamic environments and are employed in logistics, classification, and other fields. Recently, there has been a growing demand to employ delta robots for sorting recycling waste. As depicted in Figure 1, waste needs to be classified rapidly and accurately using cameras and AI technology, which serve as the eyes and brains of the system. Given that artificial intelligence technology must seamlessly integrate with robots, lightweight AI technology is imperative for efficient operation.

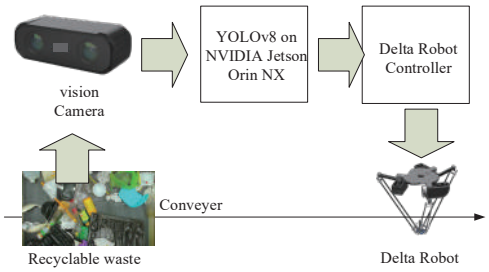


Fig. 1. Architecture of the automated recycling robot.

In actual waste recycling centers, plastics are typically classified into categories such as polyethylene (PE), polypropylene (PP), polystyrene (PS), and polyethylene terephthalate (PET) based on their material properties. In this paper, PET specifically refers to PET bottles among PET materials, while NPET encompasses products other than bottles made of PET materials, such as PET cups and packaging PET.

Object detection and classification technology plays a significant role in achieving this goal. YOLO (You Only Look Once) is a popular choice for real-time object detection and classification tasks, including those in video analytics and surveillance. To create an artificial intelligence model for plastic classification, we took 3,000 photos of plastic waste from a real recycling center (Busan Recycling Center) and labeled them. Figure 2 shows the collected plastic waste images and labeled examples used for training. After training with YOLOv8 [7] using this annotated data, it is used for plastic classification.

YOLOv8 stands out, ensuring real-time processing while maintaining high accuracy, even on small devices such as NVIDIA Jetson Orin. Its versatility and efficiency make it a valuable tool for waste recognition and location in waste classification systems. Based on the inferred results and bounding box information, Delta robot's picking tool determines the location where waste can be picked up, and the result is transmitted to the robot controller. Finally, delta robot performs actually pick plastic waste and sort it.

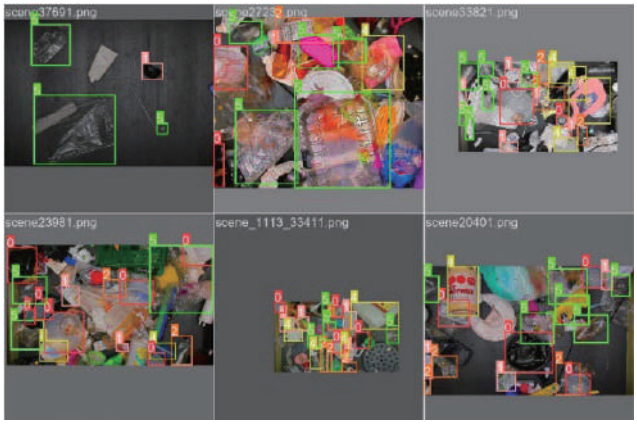


Fig. 2. Plastic waste dataset used for training.

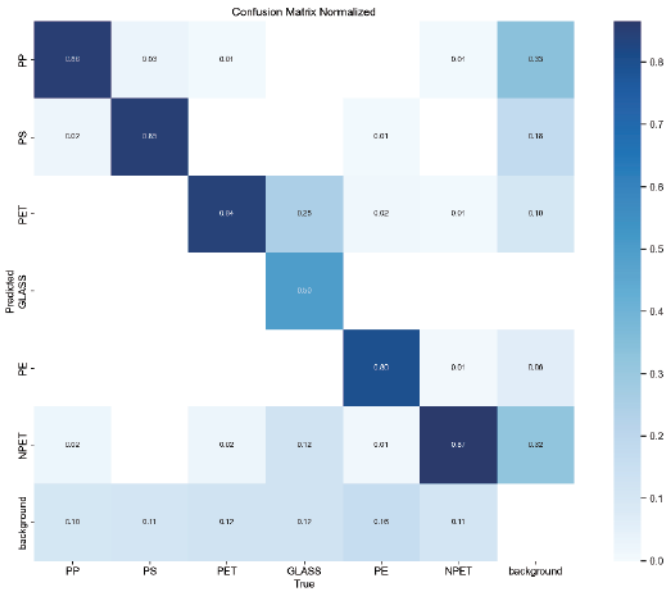


Fig. 3. Confusion matrix resulted in the collected custom dataset and YOLOv8.

III. Performance Evaluation

The dataset utilized for training encompassed diverse types of plastic waste, with many plastics exhibiting characteristics such as transparency, smooth edges, and lack of texture, rendering them exceedingly challenging to classify. Unfortunately, the reality is that highly contaminated and crumpled plastic waste, frequently encountered at actual recycling waste sorting sites, poses significant challenges for detection and classification owing to its irregular shapes and obscured features. Consequently, when employing a pretrained model or a general-purpose open dataset, the accuracy often falls short of 50%.

Upon applying our collected datasets to enhance the performance of the artificial intelligence model, Figure 3 illustrates a confusion matrix resulting from performance evaluation. Consequently, when glass waste was included, the accuracy reached 78.6%, whereas excluding glass raised the accuracy to approximately 84.4%. In actual recycling centers, materials like paper, glass, and plastic are sorted during the initial sorting process. Hence, glass and plastic should not be conveyed together. The Figure 4 depicts an example of inference running on the NVIDIA Jetson Orin board. The board performs inference at a rate of 10Hz or higher.



Fig. 4. Example of an inference result.

IV. Conclusion and Future Work

This paper introduces an on-device classification technique utilizing YOLOv8 to classify plastic waste by material within a delta robot system designed for sorting recycling waste. Training and inference were performed using data collected from a challenging real-world environment, encompassing five types of plastic and one type of glass. Additionally, tests on the Jetson Orin board validated its on-device operation. The method described in this paper holds promise as a viable solution for automating plastic waste sorting, potentially replacing human labor with high accuracy and swift sorting speed.

References

[1] J. Hopewell, R. Dvorak, E. Kosior, "Plastics recycling: challenges and opportunities," Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1526), pp.2115-2126, 2009.

[2] H. Cho, "Depth Extraction and Object Positioning System for Recyclable Waste Classification Using Delta Robots," IEEE International Conference on Consumer Electronics (ICCE), 2024.

[3] Susanth, G.S., Livingston, L.J. and Livingston, L.A., 2021. Garbage waste segregation using deep learning techniques. In IOP Conference Series: Materials Science and Engineering (Vol. 1012, No. 1, p. 012040). IOP Publishing.

[4] Sudha, S., Vidhyalakshmi, M., Pavithra, K., Sangeetha, K. and Swaathi, V., 2016, July. An automatic classification method for environment: Friendly waste segregation using deep learning. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 65-70). IEEE.

[5] Sousa, J., Rebelo, A. and Cardoso, J.S., 2019, September. Automation of waste sorting with deep learning. In 2019 XV Workshop de Visão Computacional (WVC) (pp. 43-48). IEEE.

[6] Glenn Jocher, Ayush Chaurasia, Jing Qiu, "Ultratics YOLOv8, https://github.com/ultralytics/ultralytics, 2023.

SmartRSD: An Intelligent Multimodal Approach to Real-Time Road Surface Detection for Safe Driving

Adnan Md Tayeb, Mst Ayesha Khatun, Mohtasin Golam, Md Facklasur Rahaman,
Ali Aouto, Oroceo Paul Angelo, Min-Seon Lee, Dong-Seong Kim, Jae-Min Lee, and Jung-Hyeon Kim
Networked Systems Laboratory, Department of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea.
(mdtayebadnan,20226138, golam248, facklasur, ali.aouto, oroceopaul, 20236117, dskim,ljmpaul, trizkim)@kumoh.ac.kr

***Abstract*—Precise and prompt identification of road surface conditions enables vehicles to adjust their actions, like changing speed or using specific traction control techniques, to lower the chance of accidents and potential danger to drivers and pedestrians. However, most of the existing methods for detecting road surfaces solely rely on visual data, which may be insufficient in certain situations, such as when the roads are covered by debris, in low light conditions, or in the presence of fog. Therefore, we introduce a multimodal approach for the automated detection of road surface conditions by integrating audio and images. The robustness of the proposed method is tested on a diverse dataset collected under various environmental conditions and road surface types. Through extensive evaluation, we demonstrate the effectiveness and reliability of our multimodal approach in accurately identifying road surface conditions in real-time scenarios. Our findings highlight the potential of integrating auditory and visual cues for enhancing road safety and minimizing accident risks.**

***Index Terms*—Road safety, image classification, audio classification, multimodal, and deep learning.**

I. Introduction

The real-time detection of road surface conditions is an essential measure to guarantee the safety of traffic, especially in the case of adverse weather conditions, including rain, snow, or ice. There is a direct link between tire-road friction, road surface condition, and the frequency of car crashes [1–3]. According to the recent statistics, accidents due to weather inflict almost 6,000 deaths and 445,000 injuries yearly. Wet weather stands out as the main cause in most of these incidents, with 73 percent of crashes happening on wet roads and 46 percent during rain. Additionally, over 200,000 accidents are associated with sleet and snow, with icy road conditions contributing to approximately 150,000 collisions [4].These statistics underscore the urgency of developing robust detection systems capable of identifying road surface condition in real-time. Furthermore, the impact of accurate surface condition detection extends beyond individual vehicles to encompass broader transportation networks. Timely and precise detection of road surface conditions enables transportation authorities to implement proactive measures, such as road maintenance and treatment strategies, to mitigate risks and ensure the smooth flow of traffic. The road condition detection is also important for the growing area of autonomous driving. Self-driving cars depend greatly on precise up-to-date information to quickly make decisions, particularly in challenging weather conditions when roads can quickly change. Therefore, the detection of road surface types has been a focal point of research for several decades, attracting considerable attention within academic and industry circles. Numerous ongoing projects in the field underscore the sustained interest and commitment to advancing our understanding and capabilities in road surface detection.

However, The majority of surface condition assessment techniques rely on visual data, which may be insufficient in certain situations, like low visibility, when the road is obstructed by debris, or in the presence of fog. Additionally, visual information can result in false information, particularly in varying lighting situations[9]. In low light, for example, a dry surface may appear wet, while surfaces reflecting light can give the impression of a dry surface being wet. Consequently, using only images may not suffice for accurate road condition detection. In response to these challenges, there has been growing interest in integrating multiple modalities of sensory information to enhance the accuracy and reliability of road surface condition detection systems. Detecting road conditions automatically by utilizing both visual and audio inputs may play a vital role in enhancing driver assistance systems (EDAS) and ultimately enhancing driver safety in the future. Furthermore, autonomous and semi-autonomous vehicles must be conscious of the state of the roads to adjust the vehicle's settings automatically to maintain a safe distance or reduce speed before entering the curve to the car ahead.

II. Methodology

This section will delve into the methodology employed in this study, providing an overview of multimodal fusion techniques, the approach to data collection, and the preprocessing steps undertaken. Additionally, it will discuss the weighted fusion technique utilized in this study.

A. Multimodal fusion techniques

Multimodal fusion techniques involve combining data or information from different sources or methods to enhance the performance or comprehension of a system or application. Multimodal fusion is especially important in affect recognition tasks, where analysis relies on data from various sources like audio, visual, physiology, and more, showing significant value. Its use in classification tasks is significant,

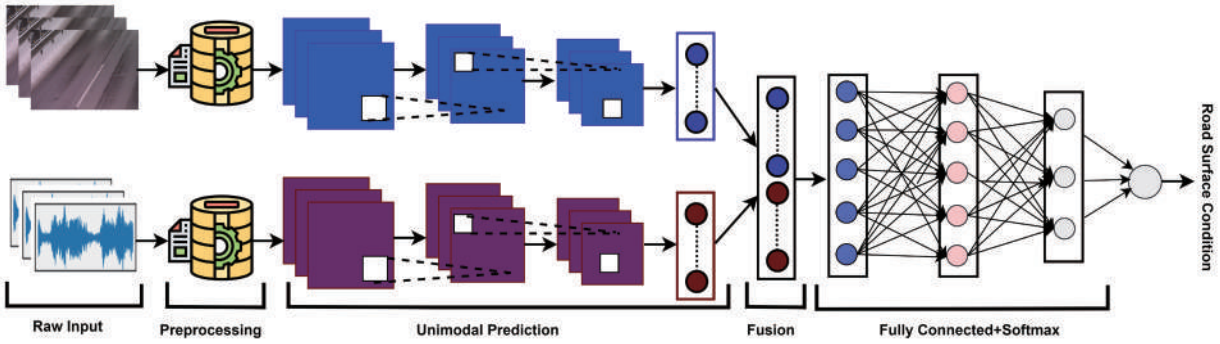


Fig. 1. Overview of the Proposed Multimodal Fusion

as combining data from different sources can enhance the feature space, strengthen the accuracy of models, and provide a more complete understanding of complex phenomena. The literature describes three main joint fusion strategies: feature-level, decision-level (or score-level), and model-level fusion. Every approach has unique benefits that are customized to meet different use cases.

- Feature-level fusion: Feature-level fusion is the process of combining features from various sources into one feature vector in order to incorporate information from all sources. This method reflects the way humans process information, combining various inputs like sound and sight before making forecasts. Feature-level fusion requires large datasets due to combining information from multiple modalities, but it offers the advantage of robustness to missing data from one modality, enabling continued predictions [7,8].
- Decision-level (Score-level) fusion: In decision-level fusion, each source is used separately to make predictions, and then the results from each source are combined. One downside is that if data from one source is missing, we can't make full use of it. Fusion methods can be straightforward, like taking a majority vote for classification tasks, or more complex, like learning weights. For regression tasks, we can train a linear regressor using predictions from each source and use its weights for fusion.
- Model-level (Hybrid-level) fusion: Model-level fusion combines the best parts of both feature-level and decision-level fusion methods. For example, it might mix features from some sources and then combine those predictions with scores from other sources that were looked at separately. This approach is adaptable and can adjust to what the task needs. An example is the method from , which mixes results from feature-level fusion with scores from separately processed sources. This hybrid method aims to get the advantages of mixing features from some sources while still considering the unique information from others. Using this fusion technique can make affect recognition tasks work better, especially when dealing with lots of different data from many sources.

B. Dataset

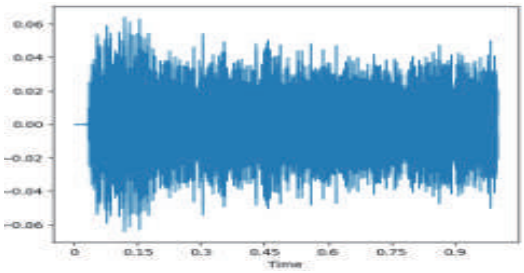


Fig. 2. Raw Audio

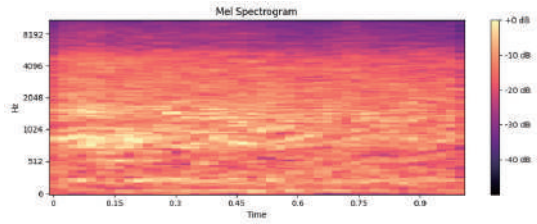


Fig. 3. Spectogram

Data collection for this study involved setting up a testbed equipped with both a camera for capturing images of roads and a microphone for recording audio. Over one year, data was collected in various environmental conditions (dry, rainy, and snowy) across different locations, capturing diverse road surfaces. This method enabled us to create a thorough collection of data that mirrors the situations vehicles face while driving. We meticulously chose combinations of images from the collected data to represent the best and worst cases, covering all possible scenarios and ensuring the model's ability to learn from a variety of situations. Altogether, we have 8,580 images in our dataset, with 2,868 images showing dry conditions, 2,548 images showing wet conditions, and 2,354 images showing snow conditions. Furthermore, a total of 75 minutes of audio recordings were gathered and grouped into three categories: dry, wet, and snow. The dataset is split into 80% training, 10% test, and 10% validation sets.

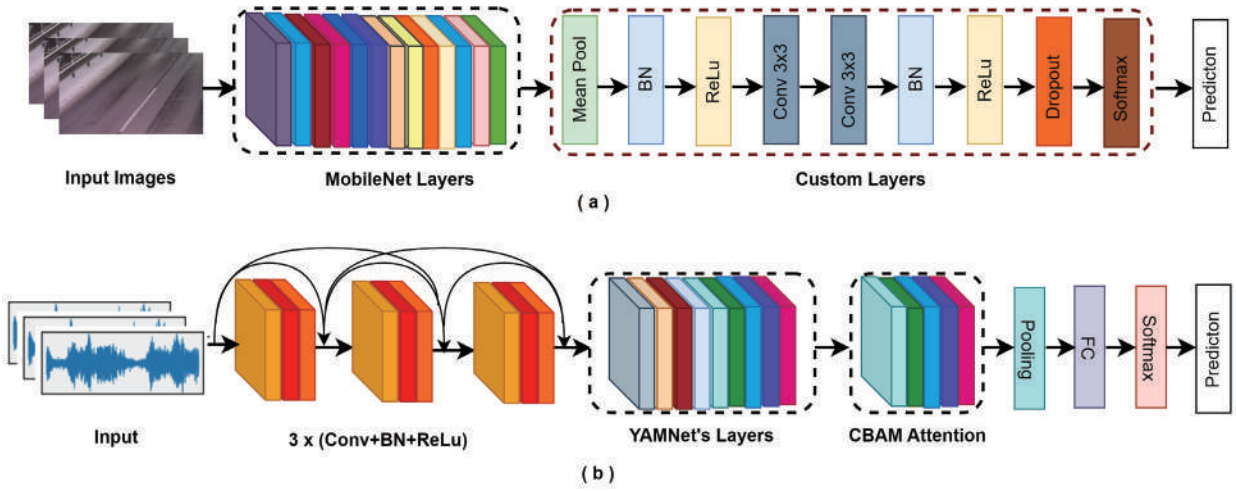


Fig. 4. Overview of the Proposed (a) Improved MobileNet, and (b) Improved YAMNet architecture for SmartRSD

C. Data Preprocessing

In our audio preprocessing pipeline, we standardized all audio clips by downsampling them to a uniform sampling rate of 16 kHz using the Sound library. A 1-second audio snippet is regarded as the input for all of our training and testing tasks, with each second of audio retaining the original recording level. To maintain consistency, we ensure that shorter audio clips are padded to match the 1-second duration. We generate Mel spectrograms (see Figure 3) from raw audio input (Figure 2), a transformation that provides a detailed representation of the frequency composition of the audio signal over time. By converting the audio signal into a visual representation, the Mel spectrogram serves as the input for our machine-learning models. This approach enables us to leverage well-established image classification techniques, as the spectrogram essentially becomes an image representation of the audio data. By treating the audio data as images, we can apply methodologies and algorithms commonly used in image processing and classification tasks, thus enhancing our ability to analyze and interpret the audio signals effectively.

D. Improved YAMNet architecture

YAMNet utilizes MobileNetV1 as its base network and comes pre-trained on the Google AudioSet dataset, encompassing 521 audio events. Consequently, in our research, we train YAMNet on unbalanced data and achieve notable performance gains. Before the feature extraction phase, we perform resampling into 16,000 Hz with one-channel audio to ensure uniformity. YAMNet is a deep learning-based model, which means it automatically extracts audio features through its feature extraction layer. This layer extracts audio features in the form of spectrograms, which are then passed to improved MobileNet layers for classification.

This process allows us to leverage the power of deep learning to automatically extract relevant audio features, which are then utilized for accurate classification using the improved MobileNet layers.

In our enhanced version of YAMNet (as shown in Figure 4, we've integrated three sets of deeply connected convolutional layers, batch normalization, and ReLU activations on top of the original YAMNet architecture. This addition has led to a notable improvement in performance. Moreover, the incorporation of deep connected skip pathways plays a crucial role in preserving information and ultimately enhancing classification accuracy. These skip pathways enable the model to retain important features while mitigating information loss, thereby contributing to more accurate and robust classification results. These enhancements elevate the capabilities of YAMNet by augmenting its capacity to capture intricate patterns and features within audio data, leading to improved classification accuracy and performance.

E. Improved MobileNet

MobileNet, developed by Google, is an open-source computer vision model designed for training classifiers. Depthwise convolutions employ one filter per channel, splitting the standard convolution into two separable layers: one for applying the filter and the other for concatenation. Additionally, pointwise convolutions of size 1×1 are used, merging their output with the depthwise convolution. This approach offers a significant reduction in the number of parameters compared to traditional networks, resulting in a lightweight deep neural network architecture [5,6]. Given its lightweight nature and efficiency, we have chosen to employ this approach in our model, especially considering that our model will run in a mini PC environment.

To further enhance classification performance, we've incorporated additional layers into the architecture. As illustrated in Figure 4(a), these include Convolutional layers, pooling operations, and ReLU activations. Additionally, a dropout layer has been integrated to mitigate the risk of overfitting.

Our experimentation has demonstrated a substantial improvement in the performance of MobileNet with the inclusion of these custom layers. By leveraging the efficiency and effec-

tiveness of the MobileNet architecture along with these additional layers, we've achieved enhanced classification accuracy while maintaining a lightweight and resource-efficient model suitable for deployment in resource-constrained environments such as mini PC setups.

F. Weighted Fusion

Given that the accuracy of image classification typically surpasses that of audio classification, combining results from both modalities may introduce a mismatch. To address this potential issue, we've implemented a weighted fusion technique, assigning slightly more weight to image predictions and less to audio predictions. Specifically, we've assigned a 60% weight to image predictions and 40% to audio predictions initially and updated according to prediction accuracy.

Algorithm 1 Weights assignment algorithm for Multimodal Classifier

Require: Initial weights for α^0 audio, and I^0 image, N Number of epochs

procedure WEIGHTS(α^0, I^0, N)
  Train α^0 with $\{W_{\alpha_i}\}_{i=1}^{k}$
for $i = 1, 2, ..., N$ **do**
    process batch
    update weights α_i
end for
  Train I^0 with $\{W_{I_i}\}_{i=1}^{k}$
for $i = 1, 2, ..., N$ **do**
    process batch
    update weights I_i
end for
  Initialize $\omega^0 \leftarrow w1 \cdot \alpha^N + w2 \cdot I^N$
  Train Multimodal classifier with ω^0
for $i = 1, 2, ..., N$ **do**
    process batch
    compute fusion output
    compute loss
    update weights ω^0
end for
  return ω^0 (final weights for the fusion model)
end procedure

This weighted fusion approach ensures a balanced integration of information from both image and audio modalities, enhancing overall prediction accuracy while minimizing the risk of mismatch errors.

III. RESULTS AND DISCUSSION

In this study, we trained various combinations of unimodal models to evaluate prediction accuracy. We used MobileNet for image classification and YAMNet for audio classification as our base models. To enhance performance, we modified both models and tested the best possible combinations with our multimodal classifier to determine the highest accuracy.

As Table I illustrates, the best accuracy is achieved by the combination of the proposed Improved MobileNet and Improved YAMNet, reaching 94.91%. This surpasses other combinations by approximately 2%.

TABLE I
CLASSIFICATION RESULTS OF THE PROPOSED IMPROVED(I) MOBILENET AND YAMNET WITH MULTIMODAL CLASSIFIER

Model	Accuracy	Precision	Recall	F1-score
MobileNet+YAMNet	0.9118	0.9111	0.9118	0.9114
IMobileNet+YAMNet	0.9379	0.9241	0.9314	0.9254
MobileNet+IYAMNet	0.9275	0.9341	0.9336	0.9354
IMobileNet+IYAMNet	0.9491	0.93.87	0.9356	0.9384

IV. CONCLUSION

This research presents a multimodal approach that combines audio and images to detect road surface conditions in real-time. we aim to address the limitations of existing methods that solely rely on visual data, particularly in challenging environmental conditions such as low light, fog, or debris covered roads.

ACKNOWLEDGMENT

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 33.3%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33.3%) and by the MSIT, Korea, under the ICAN support program(IITP-2024-00156394, 33.3%) supervised by the IITP

REFERENCES

[1] M. Golam, A. Riztia, P. Daely, A. Khatun, J.-H. Kim, J.-H. Lee, Y.-R. Cho, D.-S. Kim, and J. M. Lee, "Development of the System Architecture for Black Ice Detection to Prevent Transportation Calamities," in Proceedings of Symposium of the Korean Institute of Communications and Information Sciences, 2022, pp. 692-693.

[2] M. A. Khatun, D.-S. Kim, J. M. Lee, and J.-H. Kim, "Comparison Between Vision Transformer and CNN for Ice Image Classification," in Proceedings of Symposium of the Korean Institute of Communications and Information Sciences, 2023, pp. 1822-1823.

[3] G. O. Anyanwu, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Optimization of RBF-SVM Kernel Using Grid Search Algorithm for DDoS Attack Detection in SDN-Based VANET," IEEE Internet of Things Journal, vol. 10, no. 10, pp. 8477-8490, May 15, 2023.

[4] G. C. Amaizu, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Investigating Network Intrusion Detection Datasets Using Machine Learning," in 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2020, pp. 1325-1328.

[5] T. Zhao, J. He, J. Lv, D. Min, and Y. Wei, "A comprehensive implementation of road surface classification for vehicle driving assistance: Dataset, models, and deployment," IEEE Transactions on Intelligent Transportation Systems, 2023.

[6] A. M. Tayeb, K.-H. Lee, J.-Y. Seo, G.-W. Kim, H.-B. Park, H.-J. Lee, and T.-H. Kim, "Facial skin condition analysis based on pore and wrinkle segmentation," in Proceedings of the Symposium of the Korean Institute of Communications and Information Sciences, 2023, pp. 1648-1649.

[7] N. Saeed, M. Alam, and R. G. Nyberg, "A Multimodal Deep Learning Approach for Gravel Road Condition Evaluation through Image and Audio Integration," Transportation Engineering, vol. 100228, 2024.

[8] T. Rateke and A. Von Wangenheim, "Road surface detection and differentiation considering surface damages," Autonomous Robots, vol. 45, no. 2, pp. 299-312, 2021.

Vision-Based Black Ice Detection using Convolutional Neural Networks for Road Safety

Ali Aouto*, Jae-Min Lee* and Jung-Hyeon Kim†
*Networked Systems Lab., Dept. of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea.
†Industry-Academic Cooperation Foundation, Kumoh National Institute of Technology, Gumi, South Korea.
Email: {ali.aouto,ljmpaul,trizkim}@kumoh.ac.kr

***Abstract*—Black ice poses a significant hazard to road safety, especially during winter months when its transparent nature makes it virtually indistinguishable from the road surface. Traditional methods of black ice detection rely on environmental sensors and pavement temperature monitoring, which often lack precision and real-time capabilities. In this paper, we propose a novel approach utilizing Convolutional Neural Networks (CNNs) for vision-based black ice detection. Our method leverages high-resolution images captured by onboard cameras mounted on vehicles to identify regions of potential black ice formation. We present a custom CNN architecture designed to process these images efficiently and accurately classify road segments as either safe or potentially hazardous due to black ice. Training data for the CNN model is collected from diverse driving scenarios, encompassing varying weather conditions, road surfaces, and lighting conditions. Extensive experiments conducted on real-world datasets demonstrate the effectiveness of our approach, achieving high detection accuracy and robustness against environmental factors. Additionally, we discuss the integration of our black ice detection system with existing vehicle safety systems, enabling proactive alerts to drivers and stakeholders for mitigating the risk of black ice-related accidents. Overall, our vision-based approach offers a promising solution for enhancing road safety by providing timely and reliable detection of black ice hazards.**

***Index Terms*—Black Ice, Convolutional Neural Network, Vehicle-to-vehicle**

I. Introduction

Road safety is a critical concern, particularly during winter months when adverse weather conditions, such as snow and ice, significantly increase the risk of accidents. Among these hazards, black ice is particularly dangerous due to its transparent nature, making it almost indistinguishable from the road surface. This thin layer of ice can form without any visible cues, catching drivers off guard and leading to loss of vehicle control. Therefore, effective detection of black ice is paramount for preventing accidents and enhancing road safety [1]. Traditional methods for black ice detection primarily rely on environmental sensors, such as road temperature sensors, humidity sensors, and salinity sensors, which are installed on roads or vehicles. While these methods can provide useful data, they often suffer from limitations such as delayed response times, high installation and maintenance costs, and the inability to provide precise, localized detection. Additionally, these sensors can be affected by dirt, wear, and environmental factors, which can degrade their performance over time.

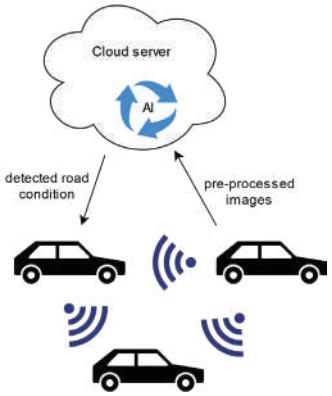


Fig. 1. Proposed model for Black ice detection. Vehicles have a mounted camera that takes images of the street pre-process it then send it to the cloud server to train the model and if black ice is detected the cloud shares with the vehicles a warning

More advanced approaches have explored the use of infrared thermography and laser-induced breakdown spectroscopy to detect ice on roads. Infrared thermography can identify temperature anomalies on road surfaces, potentially indicating the presence of ice. However, this technique requires expensive equipment and may struggle with varying environmental conditions such as sunlight and shadow. Laser-induced breakdown spectroscopy, while highly accurate, is also costly and complex, requiring specialized equipment and safety precautions due to the use of high-energy lasers.

In recent years, advancements in computer vision and machine learning, particularly Convolutional Neural Networks (CNNs), have opened new possibilities for road condition monitoring. CNNs have shown remarkable performance in various image recognition tasks due to their ability to automatically learn and extract relevant features from raw image data. This makes them well-suited for detecting subtle visual cues associated with black ice.

Furthermore, we enhance our detection system by integrating vehicle-to-vehicle (V2V) communication with 5G connectivity to the cloud, where the processing and AI models reside. This architecture allows for the rapid collection and sharing of road condition data between vehicles, ensuring that

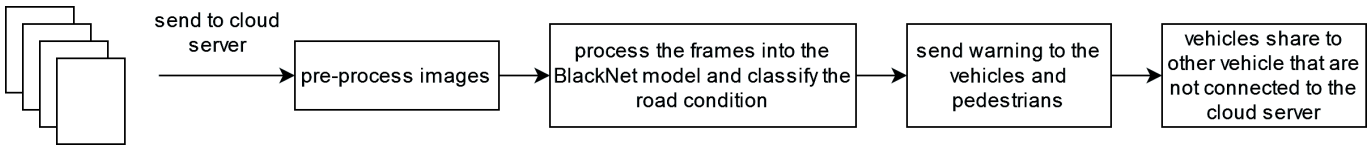


Fig. 2. Proposed system work flow. the cars capture the images then sends them to the cloud where the training and classification is happening. after detecting black ice the cloud send the warning message to vehicles that are connected to the cloud. If a car does not have the system installed but still have some connection to the system through V2V will receive warning

drivers are alerted to hazardous conditions ahead [2]. The use of 5G ensures low-latency communication, enabling real-time data processing and updates. Our custom CNN architecture is designed to efficiently process these images in the cloud, accurately classifying road segments as safe or potentially hazardous. Fig. 1 shows the proposed system where the AI model is in the cloud server.

This paper presents the development and evaluation of our vision-based black ice detection system. We describe the design of our CNN model, the collection and preprocessing of training data, and the extensive experiments conducted to validate our approach. The results demonstrate the robustness and accuracy of our system in diverse driving scenarios, highlighting its potential to significantly enhance road safety during winter months.

II. Related Works

Lee *et al.* presents a method to detect black ice on roads using Convolutional Neural Networks (CNN) to prevent accidents in automated vehicles. The authors collected image data from various environments and augmented it to create a comprehensive dataset. They used grayscale images of road environments and trained a CNN model, achieving 96% accuracy in black ice detection [3]. Abdalla *et al.*proposed a black ice detection system using a depth imaging sensor based on the Kinect device. The system classifies different types of ice (soft ice, wet snow, hard ice, black ice) and measures ice thickness and volume. It operates within a range of 82 cm to 1.52 m from the camera. The system was tested on various backgrounds such as wood, glass, ceramic, plastic, and concrete, demonstrating high proficiency in detecting black ice by utilizing Kinect's depth-sensing capabilities [4]. Ma*et al.* focuses on using a tri-wavelength non-contact optical technology to detect black ice. The method distinguishes dry, wet, black ice, ice, and snowy conditions by measuring the reflectance at three different wavelengths (1310 nm, 1430 nm, 1550 nm). The study confirmed the effectiveness of this approach in detecting black ice through wavelength-specific reflectance patterns, suggesting its potential application in developing advanced equipment for road condition monitoring [5].

III. Proposed Model

Our proposed system, BlackNet, is designed to enhance road safety by providing real-time detection of black ice and other hazardous road conditions using a vision-based approach. BlackNet is a modified version of MobileNet v2, chosen for its lightweight architecture and high performance in image classification tasks, making it suitable for deployment in real-time scenarios on vehicles with limited computational resources. The modifications to MobileNet v2 include enhanced feature extraction layers to capture fine-grained features relevant to road surface conditions and a custom classification head tailored.

The modified MobileNetV2 architecture incorporates ResNet-style long-range residual connections, enhancing its capability for training deeper networks while maintaining the lightweight efficiency characteristic of MobileNetV2. This hybrid model retains the original inverted residual blocks, which include expansion layers, depthwise convolutions, and projection layers, but introduces additional residual connections that span multiple blocks. These connections facilitate improved gradient flow and feature reuse across the network, addressing the vanishing gradient problem and enabling more effective training of deeper layers. By combining the efficiency of MobileNetV2 with the robust gradient propagation of ResNet, this modified architecture offers a balanced solution that enhances performance in image classification tasks without significantly increasing computational complexity. This approach leverages the strengths of both architectures, potentially resulting in a more powerful and versatile model suitable for a wide range of applications. Fig. 3 shows the traditional inverted residual block architecture, traditional MobileNetv2 and the proposed model architectures.

Intensity-based preprocessing, specifically through histogram equalization, is a powerful technique for enhancing the contrast in road surface images, making it easier to distinguish between icy and non-icy regions. This method involves converting the color images to grayscale, as histogram equalization is most effective on single-channel images. By redistributing the pixel intensity values, histogram equalization stretches the distribution of these values across the entire range, thereby highlighting the contrast between different surface conditions. For even more refined results, Contrast Limited Adaptive Histogram Equalization (CLAHE) can be used. CLAHE applies histogram equalization locally, which adjusts the contrast adaptively in different regions of the image, thereby handling varying lighting conditions more effectively. Normalizing the images after equalization ensures they are ready for further processing or classification tasks. This preprocessing step is crucial in improving the accuracy and robustness of machine learning models tasked with identifying road surface conditions.

To train BlackNet, we initially pre-trained the model on the ImageNet dataset to leverage a broad base of visual features, ensuring robust initial weights. Following this, we fine-tuned the model using a specialized, open-source dataset containing

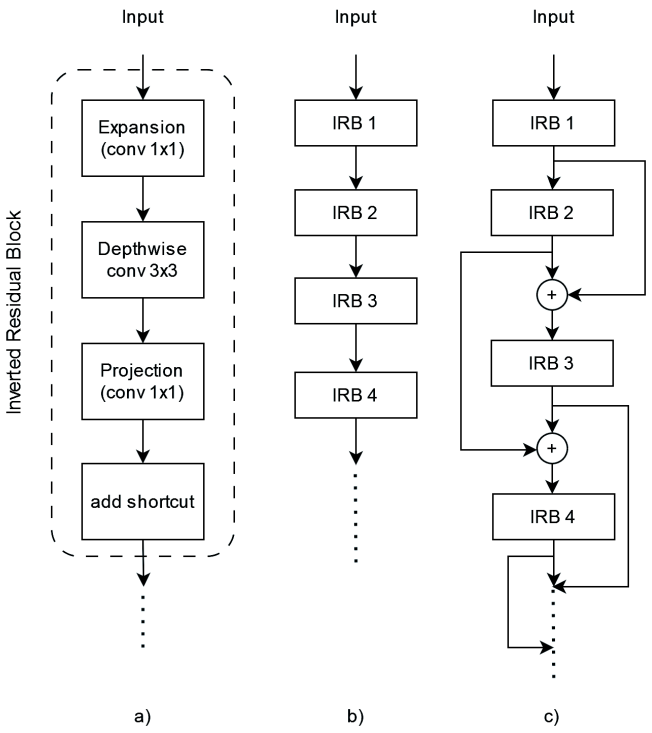


Fig. 3. a) Inverted residual block architecture (IRB). b) Traditional MobileNetv2 IRB connection architecture. c) Proposed BlackNet architecture with ResNet-style long-range residual connections.

high-resolution images of road surfaces labeled into the four classes. This dataset includes diverse images captured under varying weather conditions and lighting scenarios, ensuring the model's robustness. Data augmentation techniques such as rotation, flipping, contrast adjustment, and noise addition were applied to improve the model's generalization capabilities. The dataset was also balanced to ensure equal representation of each road condition class, addressing potential biases during training.

The system comprises several key components, starting with onboard cameras and sensors mounted on vehicles to capture real-time images of the road surface. These images are preprocessed by a lightweight edge computing device onboard the vehicle, which prepares the images by resizing and normalizing them before transmission. Using high-speed 5G connectivity, the preprocessed images are transmitted to the cloud, ensuring low-latency communication. The cloud infrastructure hosts the BlackNet model, which processes the images to classify the road conditions.

In the real-time detection workflow, the onboard camera continuously captures images of the road surface. These images are preprocessed on the edge computing device to prepare them for analysis, including tasks like resizing and normalization. Once preprocessed, the images are transmitted to the cloud via 5G connectivity. In the cloud, BlackNet processes the images and classifies them into one of the four classes: dry, wet, snow, or black ice. Based on the classification result, the system generates an alert if black ice or other hazardous conditions are detected. These alerts are communicated to nearby vehicles using vehicle-to-vehicle (V2V) technology, enabling drivers to receive timely warnings.

The proposed system is designed to integrate seamlessly with existing Advanced Driver Assistance Systems (ADAS) and Vehicle-to-Everything (V2X) frameworks. This integration enhances the overall safety features of modern vehicles by providing real-time road condition updates and proactive hazard warnings. To validate the effectiveness of BlackNet, extensive experiments are conducted using real-world datasets. Performance metrics include accuracy, latency, and robustness across diverse environmental conditions. Preliminary results indicate that BlackNet achieves high accuracy in detecting black ice, with low latency and robust performance, making it a promising solution for enhancing road safety during winter months.

IV. Simulation Results

For training our model, we utilized a comprehensive dataset comprising over 370,000 images of asphalt and concrete roads under various conditions, including dry, wet, snowy, and icy surfaces. This diverse dataset provides a robust foundation for developing a machine learning model capable of accurately identifying and distinguishing between different road conditions. By including images of roads in multiple states, the model can learn to recognize and adapt to a wide range of environmental factors, enhancing its reliability and performance in real-world scenarios. The extensive nature of the dataset ensures that the model is exposed to a vast array of visual features and patterns associated with different road conditions, ultimately contributing to more precise and effective classification. This meticulous preparation of the dataset is crucial for building a robust and versatile model, pivotal for applications in autonomous driving and road safety systems. Fig. 4 show some samples of the dataset used in to train the proposal.



Fig. 4. samples of the dataset Black ice, dry, wet, and snow, respectively.

A. State-of-the-art comparision

The results of our simulation demonstrate that BlackNet outperforms existing state-of-the-art models in the detection of black ice and other road conditions. Table I presents the accuracy of various models, including GoogLeNet, VGG-19, ResNet-18, DarkNet-19, ResNet-50, MobileNetV2, and our proposed BlackNet. The accuracy of BlackNet, at 93.79%, is higher than all other compared models, indicating its superior performance in accurately classifying road conditions under diverse scenarios. In detail, BlackNet's performance surpasses that of MobileNetV2 by approximately 1.49%, showcasing the

effectiveness of our modifications. The inclusion of ResNet-style long-range residual connections has significantly contributed to this improvement by enhancing gradient flow and feature reuse across the network. This improvement is particularly evident in challenging conditions such as wet or icy roads, where subtle visual differences are critical for accurate classification.

TABLE I
SIMULATION RESULTS USING DIFFERENT EXISTING STATE-OF-THE-ART

Model	Accuracy
GoogLeNet [6]	88.52
VGG-19 [7]	89.81
ResNet-18 [8]	90.66
DarkNet-19 [9]	91.25
ResNet-50 [8]	91.4
MobileNetv2 [10]	92.3
BlackNet	93.79

In our comparison we have tried to use other models that are considered as small models to be fair in the comparison with the proposal and Mobilenetv2. the challenge in detecting black ice is that it has 2 forms either looks like water or snow and this shows in the confusion matrix where most of the wrong classifications has appeared when the model cant distinguish between the wet and black ice or snow and black ice.

B. Pre-processing effectiveness

To demonstrate the effectiveness of intensity-based preprocessing through histogram equalization, we conducted a comparative analysis of our BlackNet model with and without the use of (CLAHE). Additionally, we performed the same comparison on the original MobileNetV2 model to further highlight the impact of CLAHE on enhancing image contrast and classification accuracy. This allowed us to assess whether the improvements observed with BlackNet could be generalized to other models and to validate the overall effectiveness of CLAHE in enhancing image classification tasks. The results of these experiments are summarized in Table II. The inclusion of CLAHE preprocessing significantly improved the accuracy of both models. For BlackNet, the accuracy increased from 91.5% to 93.79%, demonstrating a substantial gain in performance due to the enhanced contrast in preprocessed images. Similarly, the original MobileNetV2 model showed an improvement in accuracy from 90.1% to 92.3% when CLAHE was applied.

TABLE II
IMPACT OF CLAHE PREPROCESSING ON MODEL ACCURACY

Model	Without CLAHE	With CLAHE
MobileNetV2	90.1%	92.3%
BlackNet	91.5%	93.79%

V. CONCLUSION AND FUTURE WORKS

In this paper, we presented BlackNet, a modified MobileNetV2 architecture enhanced with ResNet-style long-range residual connections for vision-based black ice detection. Our approach leverages high-resolution images captured by onboard cameras and utilizes histogram equalization techniques to enhance image contrast, facilitating accurate classification of road conditions. Extensive experiments demonstrate that BlackNet outperforms existing state-of-the-art models in accuracy, achieving 93.79%.The integration of our system with 5G connectivity and V2X communication ensures real-time detection and timely alerts to drivers, contributing to improved road safety.

Future work will focus on addressing current limitations by exploring lower-resolution image processing, incorporating additional sensor data, optimizing edge computing for reduced latency, and investigating advanced model architectures. With these enhancements, BlackNet has the potential to significantly advance the field of road safety and autonomous driving technology, providing a reliable solution for detecting hazardous road conditions. Also we would like to address the huge challenge in the cases where it is super dark that the images are not visible.

ACKNOWLEDGMENT

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 50%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 50%)

REFERENCES

[1] H. Lee, M. Kang, J. Song, and K. Hwang, "The detection of black ice accidents for preventative automated vehicles using convolutional neural networks," *Electronics*, vol. 9, no. 12, 2020.

[2] S. Bhardwaj and D.-S. Kim, "Federated learning-based joint radar-communication mmwave beamtracking with imperfect CSI for V2X communications," pp. 201–206, 2023.

[3] H. Lee, K. Hwang, M. Kang, and J. Song, "Black ice detection using cnn for the prevention of accidents in automated vehicle," pp. 1189–1192, 2020.

[4] Y. E. Abdalla, M. T. Iqbal, and M. Shehata, "Black ice detection system using kinect," pp. 1–4, 2017.

[5] X. Ma and C. Ruan, "Method for black ice detection on roads using tri-wavelength backscattering measurements," *Appl. Opt.*, vol. 59, no. 24, pp. 7242–7246, 2020.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, vol. arXiv:1409.1556, 2014.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[9] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271, 2017.

[10] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

Input feature analysis for LSTM-based maximum wave height prediction

Geonwoo Hong
Kongju National University
Cheonan 31080, Korea
redgil77@smail.kongju.ac.kr

Yonggang Kim*
Kongju National University
Cheonan 31080, Korea
ygkim@kongju.ac.kr

***Abstract*—This study investigates the effect of varying input features on the prediction of maximum wave height and evaluates the prediction accuracy and computational efficiency of a long short-term memory (LSTM) model under varying numbers of input features. Using oceanographic data from the National Weather Service, the research selects combinations of 2 to 7 features out of a possible 13 to predict maximum wave height. It identifies an optimal balance between model accuracy and computational time, highlighting the trade-offs inherent in feature selection. The results show that including up to five input features significantly improves model performance without significantly increasing training time or computational requirements. This research enhances marine forecasting by highlighting the strategic selection of input features, which improves the efficiency of predictive models within feasible computational limits.**

I. Introduction

Accurate prediction of wave heights is critical for a wide range of maritime activities, including navigation, coastal engineering, and safety management. With advances in computational methods and the availability of extensive meteorological and oceanographic data, machine learning models, particularly those based on long short-term memory (LSTM) networks, have emerged as powerful tools for predicting wave dynamics. These models use historical data to capture the temporal and spatial variability inherent in the marine environment, offering the potential for highly accurate predictions.

Recent research on wave height prediction has increasingly focused on optimizing input features to enhance model performance. The selection of relevant features is a critical step in the development of predictive models, as it directly affects their accuracy, complexity, and computational efficiency. However, determining the most effective combination of input features remains a challenging task given the dynamic and multifaceted nature of oceanic systems.

Building on previous studies, this research aims to systematically evaluate the impact of varying the number of input features on the performance of LSTM-based wave height prediction models. By analyzing combinations of features derived from meteorological and oceanographic data, the study aims to identify the optimal number of input features that provide a balance between prediction accuracy and computational efficiency.

II. Related Work

In the field of wave height prediction, the application of advanced neural network architectures has shown promising results. Osawa *et al.* used a recurrent neural network (RNN), specifically LSTM, to predict wave height and peak wave period using only historical wave data [1]. Their approach highlighted the strength of LSTM in capturing temporal dependencies. However, the focus on wave data alone highlighted a limitation in predicting short-term variability, suggesting an opportunity for improvement through the integration of a broader range of meteorological features. Our study seeks to address this gap by augmenting LSTM networks with additional input parameters to improve prediction accuracy. Guan *et al.*'s exploration of combining convolutional neural networks (CNN) with LSTM (CNN-LSTM) for short-term wave height prediction presents a robust method for feature extraction and time series analysis [2]. Their model, which uses CNN for feature mapping followed by LSTM for temporal relationship analysis, shows superior prediction performance compared to traditional machine learning methods. This dual-structured approach inspires our research to further investigate the synergistic potential of CNN and LSTM in improving wave height prediction, with a particular focus on enriching the model input with diverse meteorological data to refine prediction accuracy.

While both studies have pioneered the use of neural networks in wave height prediction, they also highlight areas for further exploration. In particular, there is potential for optimizing model structures and expanding the set of input features to more accurately represent the dynamic and complex nature of marine environments. Our research builds on these findings and aims to advance the field by exploring refined LSTM architectures and extending the range of input features. This approach aims to capture the intricate interplay of atmospheric and oceanic factors influencing wave dynamics, thereby improving the accuracy and reliability of wave height predictions.

III. LSTM-Based Prediction

A. Data Utilization

This study used weather observation data from lighthouses provided by the Korea Meteorological Administration. Specifically, we selected data from Jiguido covering the period

from 1 January 2004 to 31 December 2015. In the pre-processing phase, entries with missing values (NaN values) were removed, reducing the data from 112,348 entries to 100,921 entries. Subsequently, min-max normalization was applied to standardize the data.

TABLE I: Feature List

Index	Input Feature
0	Wind Speed
1	Wind Direction
2	Maximum Gust Direction
3	Maximum Gust Speed
4	Sea Level Pressure
5	Temperature
6	Minimum Daily Temperature
7	Maximum Daily Temperature
8	Water Temperature
9	Maximum Wave Height
10	Significant Wave Height
11	Wave Period
12	Water Level

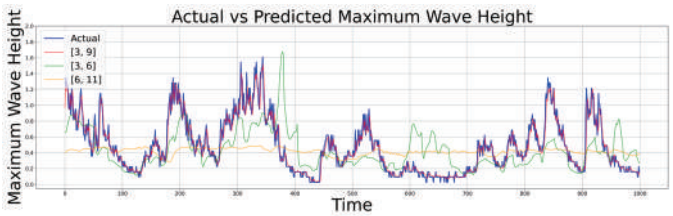
B. Learning with LSTM

Our LSTM model was designed to accommodate inputs with 2 to 7 input features, with a single hidden layer of 64 units and an output layer to predict maximum wave height. The dataset, comprising hourly records from 2004 to 2015, was divided into training and test sets with an 80% to 20% split, respectively.

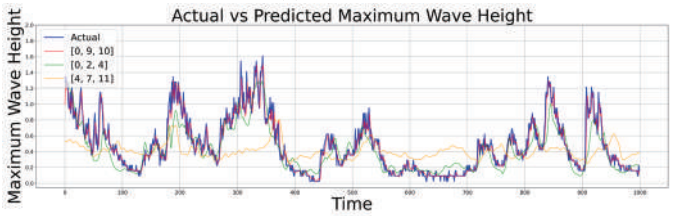
The choice of optimizer is crucial in training deep learning models as it directly influences the convergence speed and the quality of the final model. After evaluating various optimizers, we selected the Adam optimization algorithm due to its adaptive learning rate capability, which helps in converging faster and more effectively even with minimal hyperparameter tuning. Adam combines the advantages of two other extensions of stochastic gradient descent, specifically AdaGrad, which works well with sparse gradients, and RMSProp, which works well in online and non-stationary settings [3]. During the training phase, characterized by a batch size of 10,192 and a cap of 500 epochs, we used Adam. An early stopping mechanism was used to mitigate overfitting, stopping training if no improvement in loss was detected for 5 consecutive epochs.

C. Prediction of Maximum Wave Height

The input to the LSTM model consisted of unique combinations of 2 to 7 features selected from the 13 available features provided by the Met Office. The total number of possible input combinations for each feature group was 78, 286, 715, 1287, 1716, and 1716 respectively. Based on the performance results of these combinations, we identified three representative combinations for further analysis: the best-performing, the median-performing, and the worst-performing. The performance of these combinations is illustrated in the following graphs, which show the predictive capabilities of the LSTM model across different input feature configurations.



(a) 2 feature



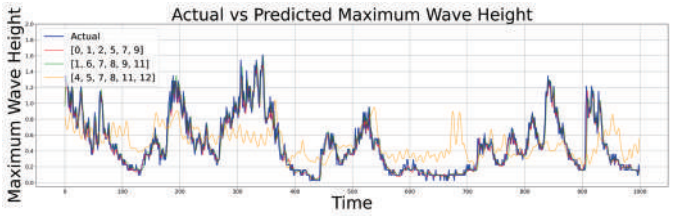
(b) 3 feature



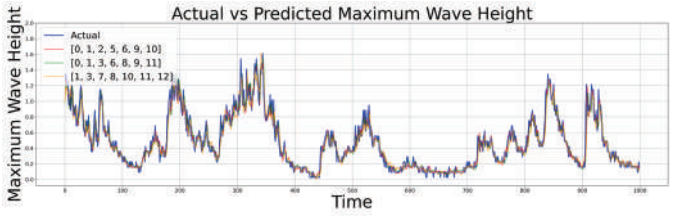
(c) 4 feature



(d) 5 feature



(e) 6 feature



(f) 7 feature

Fig. 1: Prediction of maximum wave height over 1000 hours using combinations of input features.

In Figure 1, the most effective combination is shown in red, the median in green, and the least effective in gray. The accompanying graphs show that the most effective combination

consistently achieves remarkable prediction accuracy, which closely mirrors the actual results. Conversely, the median combination shows some predictive ability but falls short when it comes to accurately predicting sudden changes. The least effective combination faces significant challenges in making accurate predictions. We evaluated the effectiveness of these combinations in all scenarios using three critical evaluation metrics: root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2). This analysis showed that their performance levels were comparatively stable across varying numbers of input features.

IV. Influence Evaluation of the Input Features

Our study examines the influence of a selected set of 13 features on model performance, considering combinations of 2 to 7 input features. Superior model performance is identified by combinations that fall within the top 10% according to the evaluation metrics of RMSE, MAE, and R^2. The methodology used quantifies the influence of each feature by counting its occurrence in the best-performing combinations across these metrics, selected from subsets of features without repetition. This approach highlights the relative importance of each feature, as indicated by its frequency of occurrence in the most effective 10% of combinations. This analysis serves to highlight the contribution of each feature and provides insight into its proportional impact on model performance.



Fig. 2: The Influence of Input Features

The heatmap in Figure 4 provides invaluable insight into the contribution of each meteorological feature to our forecast model. It shows that the percentage impact of each feature changes significantly as more features are included. Initially, with two features, 'Wind Speed' has a significant impact of 33.33%, confirming its crucial role in the model's performance when minimal features are used. Close behind, 'Significant Wave Height' has a significant influence of 22.92%, further emphasizing its importance. When the model is expanded to include three features, there is a noticeable shift in the distribution of influence. Wind Speed' remains a key factor, albeit with a reduced impact of 25.67% compared to the two-feature scenario, suggesting that the influence is spread across a greater number of features. Similarly, the influence of 'Significant Wave Height' decreases to 12.64%, highlighting the dynamic interplay between feature combinations and their collective impact on model accuracy. The introduction of four features adds further complexity: 'Wind Speed' remains the leading influence, although its impact decreases to 19.21%. This pattern illustrates the complex relationship between the increasing number of input features and their individual contributions to model performance. In addition, features such as 'Maximum Wind Gust Speed' and 'Atmospheric Pressure at Sea Level' begin to show significant influence, with contributions exceeding 8%, indicating their increased importance in more complex configurations. With the integration of five features, the trend evolves such that no single feature dominates; instead, a balanced mix of influences drives model performance. In this scenario, 'wind speed' remains influential but becomes part of a broader group of features that collectively improve prediction accuracy. When six features are included, it is evident that features that previously had little influence begin to increase their influence, suggesting that these combinations can provide new insights that are not apparent when considered in isolation. At seven features, the heatmap shows a flattening out of the influence of individual features, suggesting that adding more features beyond this point does not necessarily lead to significant improvements in prediction accuracy, but rather adds complexity that needs to be managed carefully.

In summary, the heatmap highlights the need for a strategic approach to feature selection. The optimal set of features may vary depending on the specific characteristics of the model and the nature of the prediction task. This in-depth analysis serves as a critical guide for refining and optimizing model design, with the goal of improving predictive performance while maintaining computational efficiency.

Figure 4 provides a detailed examination of how increasing the number of input features from two to seven affects the model's performance metrics. It shows a consistent decrease in RMSE and MAE, while R^2 value shows an increase, indicating improved prediction accuracy and a better fit of the model with the addition of more features. With an initial setup of two features, the RMSE is recorded at 0.7205, the MAE at 0.4589 and the R^2 at 0.4527. As the number of features is increased, both the RMSE and the MAE show a gradual decrease, while the R^2 value shows an upward movement, particularly noticeable when moving from three to four features. At the same time, the time required to train the model increases from 00h 06m 02.37s to 04h 16m 27.06s for a seven-feature scenario. This raises an important consideration regarding the efficiency of performance improvement relative

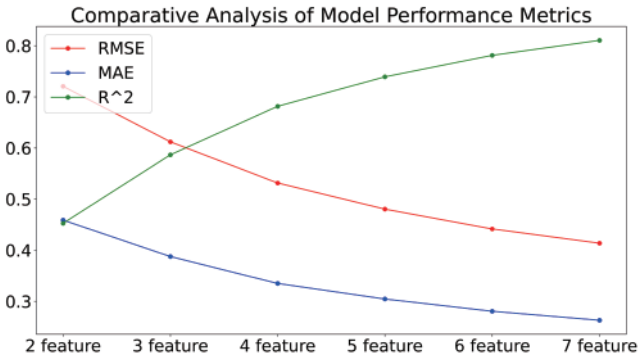


Fig. 3: Performance metrics RMSE, MAE, and R^2 across feature sets ranging from 2 to 7 features.

to the time invested. The analysis suggests that including up to five input features optimizes the generalization performance of the model in an efficient manner relative to the increase in training time.

This finding has important implications for the development of predictive models. While increasing the number of input features tends to improve prediction accuracy, it also increases computational complexity. Identifying the optimal number of features becomes essential at the point where additional performance improvements do not proportionally outweigh the additional time required.

V. Discussion

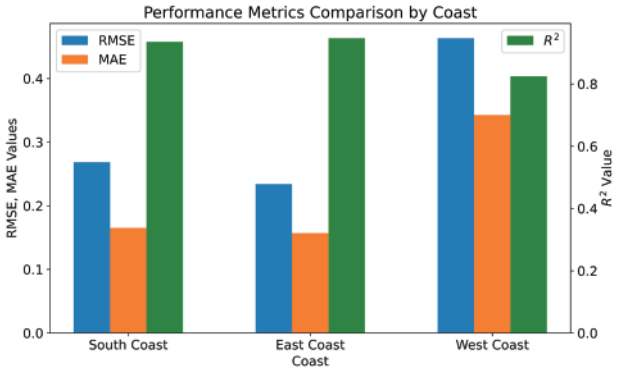


Fig. 4: Performance metrics RMSE, MAE, R^2 of the model applied to South, East, and West Coasts.

Figure 4 presents the performance metrics RMSE, MAE, R^2 of our proposed model applied to datasets from different coastal regions: south coast (Jiguido), east coast (Ideokseo), and west coast (Seosudo). Initially, our research used data from Jiguido. To further evaluate the robustness of our model's best combination of 5 input features, and to increase the model's generalizability and evaluate its performance under diverse conditions, we conducted additional experiments using data from Ideokseo and Seosudo.

The results show that the model performed similarly on the East Coast data compared to the South Coast, indicating the proposed feature combination's effectiveness across different regions. However, the West Coast data resulted in higher errors. This discrepancy may be attributed to the geographical differences in Korea; the East Coast features steep cliffs and rough waves, while the West Coast has gentler terrain and calmer seas, leading to different wave dynamics and prediction challenges.

VI. Conclusion

This study underscores the complex balance between incorporating additional input features and the associated computational costs in wave height prediction models. The systematic improvement in performance metrics, observed through a decrease in RMSE and MAE and an increase in R^2 as the number of features increases from two to seven, demonstrates that a richer feature set can significantly improve the model's ability to understand the intricate dynamics of wave behavior. Importantly, the most favorable balance between model accuracy and computational efficiency is achieved when the number of input features is limited to five. Beyond this threshold, the marginal increase in accuracy does not justify the additional computational resources and complexity. At this critical point, the model achieves a practical balance, exploiting the predictive power of a large set of features while keeping training time and resource consumption within reasonable limits.

This finding serves as a strategic guide for future model development, advocating careful selection of input features to maintain model effectiveness without imposing undue computational demands. Future studies could investigate ways to improve algorithmic efficiency and simplify models, potentially allowing for the inclusion of an expanded set of features without significantly increasing computational load. This direction could further refine the performance of the model and its applicability to real-world forecasting scenarios.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2022-00166739).

References

[1] K. Osawa, H. Yamaguchi, M. Umair, M. A. Hashmani, and K. Horio, "Wave Height and Peak Wave Period Prediction Using Recurrent Neural Networks," in *Proceedings of the International Conference on Computational Intelligence (ICCI)*, Universiti Teknologi PETRONAS, Malaysia, October 2020.

[2] X. Guan, "Wave height prediction based on CNN-LSTM," in *Proceedings of the International Conference on Machine Learning Big Data and Business Intelligence (MLBDBI)*, Southwest University, Chongqing, China, 2020.

[3] D. P. Kingma and J. Lei Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, San Diego, 2015.

Derivation of controllable region of towfish according to change in towing point, towing speed, and center of gravity

line 1: 1st Min-Kyu Kim
line 2: *Maritime ICT & Mobility Research Department*
line 3: *Korea Institute of Ocean Science & Technology*
line 4: Busan, Korea
line 5: kmk7059@kiost.ac.kr

line 1: 2nd Jihyeong Lee
line 2: *Maritime ICT & Mobility Research Department*
line 3: *Korea Institute of Ocean Science & Technology*
line 4: Busan, Korea
line 5: wlgud0818@kiost.ac.kr

line 1: 3rd Sang-Ki Jeong
line 2: *Maritime ICT & Mobility Research Department*
line 3: *Korea Institute of Ocean Science & Technology*
line 4: Busan, Korea
line 5: jeongsk313@kiost.ac.kr

***Abstract*—This paper derives a controllable region to improve the image quality of a sound navigation ranging (sonar) system for a towfish (underwater towed vehicle) equipped with two elevators and a single rudder. The towfish connected to the mother ship though a towing cable changes its movement characteristics not only according to the movement of the cable but also according to the towing point, towing speed, and center of gravity, and can be able to respond to fluctuations caused by the unpredictable underwater environment and various operating conditions. Specifically, in order to properly control the attitude of the towfish, we propose a method to derive the feasible region of the towing point according to changes in the center of gravity and towing speed. A water tank experiment was conducted to verify the proposed method, and it was confirmed that the feasible region can be used as an indicator to ensure attitude control of the towfish.**

Keywords—towfish, towing point, towing speed, center of gravity, attitude control

I. Introduction

Un manned underwater vehicles (UUVs) are widely used for marine environmental data collection and military purposes [1], and among UUVs, autonomous underwater vehicles (AUVs) and towfish are mainly used. AUVs have the advantage of being able to operate autonomously without operator intervention, but there are limitations in performing long-term missions due to limited power supply [2]. In contrast, the towfish is connected to the mother ship via a towing cable, so its movement is limited, but long-term and wide-area observations are possible because it is continuously supplied with power from the mother ship [3].

Although towfish is generally designed to move stability, they are difficult to perfectly respond to the unpredictable underwater environment and various operating conditions. In particular, when interferometric synthetic aperture sonar InSAS which has 10 times higher resolution than side scan sonar (SSS), is mounted on the towfish, the irregular movement of the towfish causes the InSAS image to be distorted or out of focus [4]. Therefore, precise attitude control is required to improve image quality.

Related studies on towfish for various purposes area mainly focused on aspects of control methods [5], but there is little research on finding region where the attitude control of towfish can be more easily controlled by complexly analyzing the effects of towing point, towing speed, and center of gravity. Therefore, this paper analyzes the relationship between elements to make it easier to control the attitude of the towfish. In this study, a dynamic analysis is conducted between the towing point, towing speed, center of gravity, and elevator capacity for the purpose of controlling the attitude of an towfish with elevators attached to the left and right horizontal tail wings. And based on this mechanical analysis, we examine how to select the appropriate towing point and center of gravity.

II. Target Towfish

Fig. 1 shows the target towfish used in this study, To control the attitude of the towfish, elevators are installed on the left and right horizontal tail wings, and there are towing cable and towing points for connection to the mother ship.

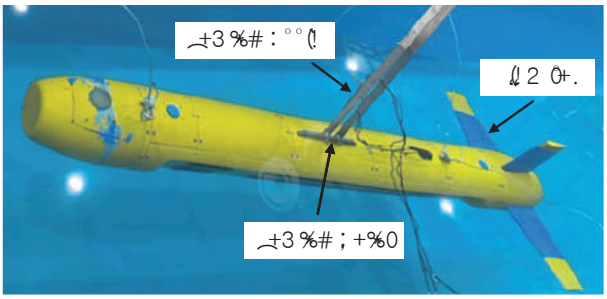


Fig. 1. Towfish used in this study

TABLE I. Specifications of Target Towfish

Towfish			
Length	$3.5m$	***Diameter***	$0.4m$
Single elevator area	$0.025m^2$	***Rudder area***	$0.03m^2$
Weight in air	$2940N$	***Buoyancy***	$490N$
Operating depth	$\leq 200m$	***Towing speed***	$\leq 2m/s$

III. Dynamic Model of Towfish

The towfish can be expressed as 6 degrees of freedom (DOF) equations of motion consisting of the linear and angular velocities on the three-axis direction [6]. For attitude control, the towfish's equations of motion considering the drag forces generated by the elevators attached to the horizontal tail wings and the towing force by the cable are as follows:

$$M\dot{\boldsymbol{\nu}} + C(\boldsymbol{\nu})\boldsymbol{\nu} + D(\boldsymbol{\nu})\boldsymbol{\nu} = \begin{pmatrix} f_c \\ r_c \times f_c \end{pmatrix} + \begin{pmatrix} f_a \\ r_a \times f_a \end{pmatrix} + \boldsymbol{f_b} \quad (1)$$

where, $\boldsymbol{\nu} = (u, v, w, p, q, r)^T \in R^T$, which represents the linear and angular velocities in the three-axis direction of the

body frame. Linear and angular velocities are (u, v, w) and (p, q, r), respectively. M, $C(\boldsymbol{v})$ and $D(\boldsymbol{v})$ are inertia matrix, Coriolis and centripetal matrix, and damping matrix, respectively. The forces acting on the towfish is expressed in Fig. 2.

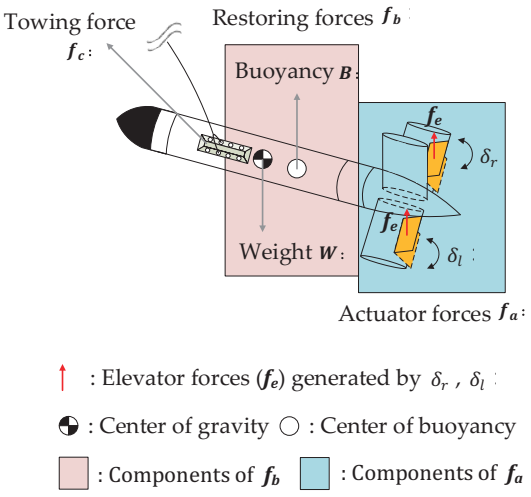


Fig. 2. Towfish used in this study

where, $\boldsymbol{f_c}$ is the towed force and $\boldsymbol{f_a}$ is the actuator force generated by the left and right elevator angles δ_l, δ_r. $\boldsymbol{f_b}$ is the weight and buoyant forces and $\boldsymbol{r_c}$, $\boldsymbol{r_a}$ are the position vectors from the center of gravity to the towing point and the center of actuators, respectively.

Meanwhile, in order to measure the position of the towfish, first the position of the mother ship must be obtained from GPS. Then the position of the towfish can be obtained by using communication with the USBL (Ultra-short baseline) mounted on each of the mother ship and the towfish. The linear and angular velocities in the body frame of the towfish are measured by IMU (Inertial Measurement Unit). However, in this study, the linear and angular velocities in the body frame of the towfish were converted into the positions and Euler angles in the Earth-fixed frame through the following equation [7].

$$\dot{\eta} = J(\eta_2)\nu \tag{2}$$

We define $\eta = (\eta_1^T, \eta_2^T)^T$, $\eta_1 = (x, y, z)^T$, $\eta_2 = (\phi, \theta, \psi)^T$, where x, y and z are the positions of towfish in the Earth-fixed frame and ϕ, θ and ψ are the Euler angles for roll, pitch and yaw of the towfish in the Earth-fixed frame. $J(\eta_2)$ is the rotation matrix from the body frame to the Earth-fixed frame.

IV. Selection of Feasible Region That Can Realize Attitude Control

In general, the center of gravity of the towfish is designed slightly below the center of body, so the roll can be restored naturally. For this reason, the roll control was omitted in this study. However the pitch is difficult to restore naturally. Therefore it is necessary to control it in order to maintain an accurate attitude.

A. Feasible towing point region for pitch control

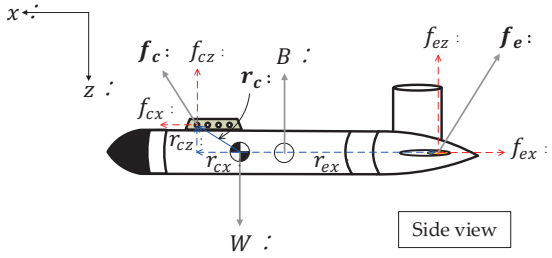


Fig. 3. The forces acting on the towfish in the x − z plane (side view)

Fig. 3 show the forces acting on the towfish in the $x - z$ plane. Where, $\boldsymbol{f_c} = (f_{cx}, f_{cz})^T$, $\boldsymbol{f_e} = (f_{ex}, f_{ez})^T$ are the towing force and drag force of elevators, $\boldsymbol{r_c} = (r_{cx}, r_{cz})^T$, $\boldsymbol{r_e} = (r_{cx}, 0)^T$ are the position vector from the center of gravity to the towing point and center of elevators. Using these forces, the pitching moment of the towfish generated by towing forces, restoring force, and drag force of elevators can be calculated as follows:

$$M_y = \boldsymbol{r_c} \otimes \boldsymbol{f_C} + \boldsymbol{f_e} \otimes \boldsymbol{f_e} + f_{b_pitch} \tag{3}$$

where, f_{b_pitch} is the restoring force response to pitch and $\otimes$ is the outer product of vector in the plane. $\otimes$ satisfies the following relation equation.

$$\boldsymbol{a} \otimes \boldsymbol{b} = (E\boldsymbol{a})^T \boldsymbol{b} = (E^T \boldsymbol{a}^T)\boldsymbol{b}, \quad \boldsymbol{a}, \boldsymbol{b} \in R^2, \quad E = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{4}$$

where, E is the rotation matrix that rotates 90° counterclockwise on a plane. Thowfish needs to be interpreted in various cases because the direction and force of the pitching moment vary depending on the towing point, center of gravity and buoyancy. In this study, five cases given in Fig. 4 are selected and interpreted. Basically, if the center of gravity is behind the center of buoyancy, it is difficult to secure the stability of the towfish. So this case was ignored.

The velocity of the towfish cannot be neglected because the drag force of elevator changes to square according to velocity. Therefore, considering not only the center of gravity but also the change in velocity we will find out how to calculate the range of the towing point that can control the attitude of the towfish. Since the center of buoyancy is located in the center of the towfish's body, it is assumed that this position does not change.

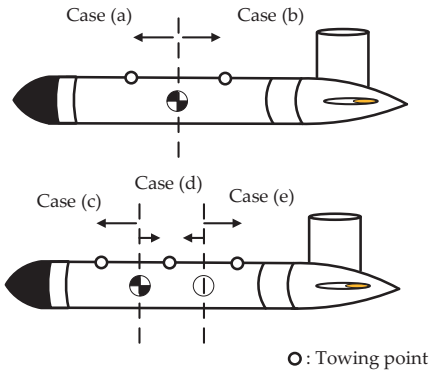


Fig. 4. Five cases ((a)-(e)) to interpret the effect of the towing point and center of gravity

B. *Cases where only the towing point changes (cases (a) and case(b))*

In cases (a) and (b), where the towing point is in front of and behind the center of gravity, the center of gravity and center of gravity and center of buoyancy are fixed at the same position. The effect of the restoring force term, f_{b_pitch}, can be ignored because the moments caused by the weight and buoyancy are the same position. r_{cx} which is in equilibrium with the maximum elevator forces in the positive and negative directions in (3), is calculated follows:

$$r^{+}_{cx\,max} = \frac{r_{cz}f_{cx}+r_{ex}f^{-}_{ez\,max}}{f_{cz}} \tag{5}$$

$$r^{-}_{cx\,max} = -\left(\frac{-r_{cz}f_{cx}+r_{ex}f^{+}_{ez\,max}}{f_{cz}}\right)$$

The $r^{+}_{cx\,max}$ and $r^{-}_{cx\,max}$ are the maximum towing points where the pitching moment occurs in the positive and negative directions and $f^{+}_{ez\,max}$ and $f^{-}_{ez\,max}$ are the maximum drag force of the elevators in the positive and negative directions.

C. *Cases where the center of gravity and towing point changes (case(c)-case(e))*

In case (c)-(e), the center of gravity is in front of the center of buoyancy. As the position of the center of gravity changes, we defined that the case in front of the center of gravity is as (c), the case between the center of gravity and the center of buoyancy is as (d), and the case behind the center of buoyancy is as (e). In these cases, the restoring force must be considered because the center of gravity and center of buoyancy are at different position. The restoring force f_{b_pitch} is can be simplified to $-x_g W$ when the center of buoyancy is fixed at the origin and the towfish is stabilized. Therefore the maximum towing points are obtained as follows:

$$r^{+}_{cx\,max} = \frac{r_{cz}f_{cx}+r_{ex}f^{-}_{ez\,max}+x_g W}{f_{cz}} \tag{6}$$

$$r^{-}_{cx\,max} = \frac{r_{cz}f_{cx}+r_{ex}f^{+}_{ez\,max}+x_g W}{f_{cz}}$$

The method of obtaining the region of towing point that can control the attitude of the towfish was examined. Based on this we implement a range that can perform attitude control by grasping it to the towfish. Based on Table 1, which is the specification of the towfish, to find the range of the towing point that can maintain the pitch angle, the simulation was conducted while changing the speed of the towfish to 1 to 2 m/s and position of center of gravity to 0 to 0.5 m. It is assumed that the force is balanced by generating a towing force f_{cz} -2450 N in the direction opposite to the underwater weight of the towfish. In the specification of the towfish r_{cz} is 0.2 m and when x_g is at the center of body, r_{ex} is 1.7 m.

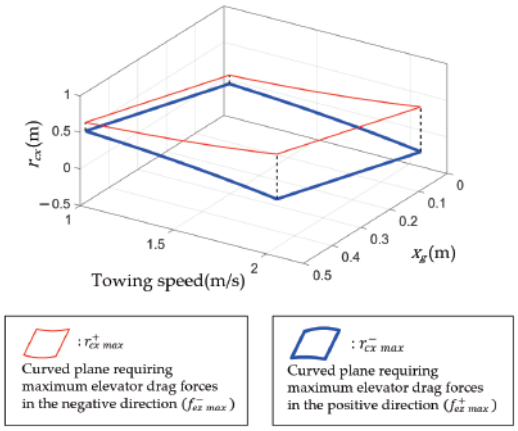


Fig. 5. Feasible region of r_{cx} when towing speed and x_g change

Fig. 5 shows the feasible region of the towing point r_{cx} when the towing speed varies from 1 to 2 m/s and x_g varies from 0 to 0.5 m. The region surrounded by the red line ($r^{+}_{cx\,max}$) is the position requiring maximum elevator drag forces in negative direction ($f^{-}_{ez\,max}$) and the region surrounded by the blue line ($r^{-}_{cx\,max}$) is the position requiring maximum elevator drag forces in the positive direction ($f^{+}_{ez\,max}$). Therefore, in order to control the towfish more easily, it would be beneficial to select the towing point inside the region surround by red and blue line.

V. Water Tank Experiment

A. *Experiment conditions*

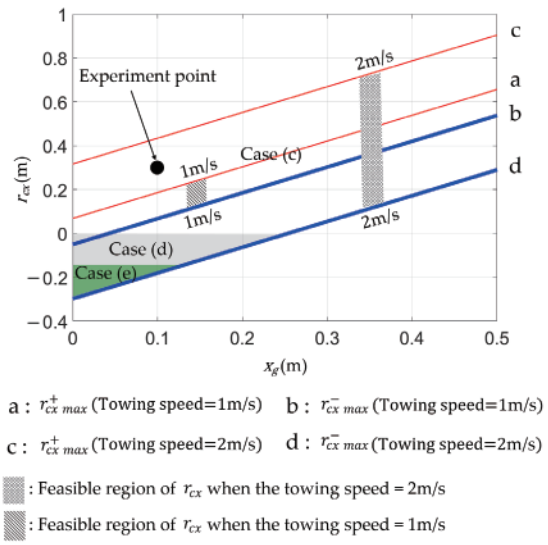


Fig. 6. Feasible regions of r_{cx} when expressed along the $r_{cx} - x_g$ plane.

Fig.6 shows the feasible regions corresponding to three cases in Fig. 4. Case (a) and case (b) are the special cases where x_g is 0, and the range of towing point that can control the pitch is similarly distributed in the positive and negative directions. So these two cases were excluded from this study. In case (c) where the towing point is in front of x_g, it can be seen that the feasible region is wide. However, in case (d) and case (e) where the towing point is behind the x_g, it can be seen that the feasible region is narrow. These two cases

belong to the feasible region is difficult to control the pitch because $f_{ez\,max}^{+}$ is concentrated.

The x_g of the towfish was $0.1\,m$, and the towing point r_{cx} for the water tank experiment was selected as $0.3\,m$. Under these conditions, the flow speed of water tank was selected as $1\ m/s$ and $2\ m/s$, respectively, and the experiment was conducted. The pitch controller used for the experiment was a PID controller, and its purpose was to maintain a pitch angle of 0 degree.

1) Towing speed is 1m/s

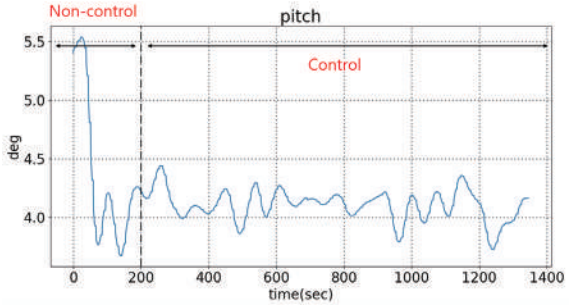


Fig. 7. Pitch control when towing speed is $1\ m/s$

Fig. 7 shows the pitch control when towing speed is $1\,m/s$. When towing speed is $1\,m/s$, because it is outside the region where the attitude of the towfish can be controlled, it can be confirmed that the reference pitch input value does not maintain 0 degree and comes out to about 4 degrees even if the elevator angle is applied to the maximum.

2) Towing speed is 2m/s

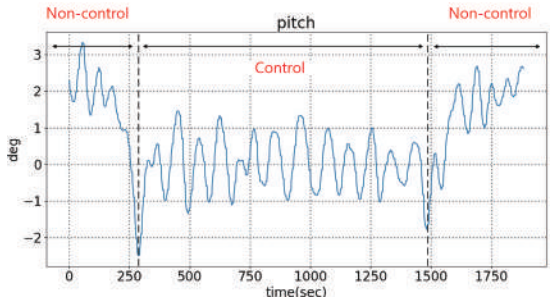


Fig. 8. Pitch control when towing speed is $2\ m/s$

Fig. 8 shows the pitch control when towing speed is $2\,m/s$. When towing speed is $2\,m/s$, it is feasible region where the attitude of the towfish can be controlled, so it can be confirmed that the designed elevator capacity can sufficiently control the attitude of the towfish.

From these result, is it judged that if towing speed is $1\ m/s$, it will be difficult to respond to unexpected environmental disturbances, and if the towing speed is $2\ m/s$, it will be possible to respond with elevators even if environmental disturbances exist.

VI. Conclusion

In this study, for the purpose of more easily controlling the attitude of the towfish equipped with high-resolution InSAS, the method of obtaining the feasible region of the towing point that can control the attitude of the towfish was examined. In order to select the feasible region of the towing point that can control the pitch angle, the dynamic equations of the towfish were used, and dynamic analysis was performed by comprehensively considering the center of gravity, center of buoyancy, towing point, and towing force.

As a result of simulation, it was confirmed that there are regions that are easy to control and regions that are not, and through the water tank experiment, it was confirmed that pitch control was possible in the feasible region and that pitch control was impossible out of feasible region. Based on these results, the method proposed in this study can be used as indicator to select the optimal towing point in the design stage of the towfish or in the field operation.

Acknowledgment

This research was a part of the project titled "Development of smart maintenance monitoring techniques to prepare for disaster and deterioration of port infra structures", funded by the Ministry of Oceans and Fisheries, Korea.

References

[1] S. Pang, Y. Li, and H. Yi, "Joint formation control with obstacle avoidance of towfish and multiple autonomous underwater vehicle based on graph theory and the null-space-based method," Sensors, vol. 19, 2591, 2019..

[2] T. Murashima, T.Aoki, S. Tsukioka, S. Hyakudome, H. Yoshida, S. Nakajoh, S. Ishibashi, and R. Sasamoto, "Thin cable systems for ROV and AUV in JAMSTEC," Oceans, vol. 5, pp. 2695-2700, 2003.

[3] B. Buckham, M. Nahon, M. Seto, X. Zhao, and C. Lambert, "Dynamics and control of a towed underwater vehicle system, part 1: Model development," Ocean engineering, vol. 30, pp. 453-470, 2003.

[4] J. Lee, Y. Oh, S. Park, and H. Kim, "Development of towed synthetic aperture sonar system,: Korea Soc. Nav. Sci. Technol, vol. 2, pp. 28-31, 2019.

[5] G. Antonelli, S. Chiaverini, R. Finotello, and R. Schiavon, " Real-time path planning and obstacle avodiance for RAIS: An autonomous underwater vechicle," Ocean engineering, vol. 26, pp. 216-227, 2001.

[6] T.I. Fossen, Handbook of marine craft hydrodynamics and motion control. John Wiley & Sons Ltd: Sussex, UK, 2011.

[7] T.I. Fossen, Guidance and control of ocean vehicles. John Wiley & Sons Ltd: Sissex, UK, 1994.

A Study on Composite Platform and Field Test

1st Jihyeong Leei
Maritime ICT·Mobility
Research Department
KIOST
Busan, Republic of Korea
wlgud0818@kiost.ac.kr

2nd Sang Ki Jeong
Maritime ICT·Mobility
Research Department
KIOST
Busan, Republic of Korea
jeongsk313@kiost.ac.kr

3rd Haeyong Park
Maritime ICT·Mobility
Research Department
KIOST
Busan, Republic of Korea
hypark@kiost.ac.kr

Abstract*— This paper presents the development and sea area testing results of a composite marine unmanned platform consisting of Unmanned Surface Vehicle (USV) and Remotely Operated Vehicle (ROV). The platform's structure, control system, sensor system, and control algorithms are developed and introduced. The composite platform collects information on underwater structures such as walls, floors, bridges, and harbors. The ROV can be launched and recovered using a winch system installed on the USV. Data collected from both vehicles can be transmitted in real-time to the operations center via Radio Frequency (RF) communication. *A Cross Track Error (CTE) tracking algorithm and dynamic positioning algorithm are proposed. The proposed algorithms are implemented through sea area testing, and performance test results are presented. Particularly, for the hovering control of the ROV, a PID controller and robust Sliding Mode Control (SMC) are applied, demonstrating the excellent performance of SMC.

Keywords—Unmanned Surface Vehicle, Unmanned Underwater Vehicle, Composite Marine Unmanned Platform, Cross-Track Error

I. INTRODUCTION

The increasing importance of various resources such as petroleum, gas, and rare earth elements underscores the growing emphasis on the development of unmanned exploration vehicles, particularly due to safety and cost concerns[1]. Currently, most construction or maintenance operations of marine structures involve divers working directly in the sea[2]. However, it is challenging for individuals to dive deeper than 30 meters or to continue working underwater for more than an hour. Therefore, several studies are being conducted on the utilization of Unmanned Surface Vehicles (USVs) and Unmanned Underwater Vehicles (UUVs) in relation to USVs[3,4]. Remotely Operated Vehicles (ROVs) among underwater robots are operated from the mother ship, allowing stable acquisition of location data through GPS, unrestricted operating time due to energy limitations, and real-time reception of underwater data. However the disadvantage lies in the tether cable connecting the mother ship and the ROV, which restricts the maneuverability and requires operation by trained operators. In 2018, Lachaud et al. proposed a combined USV-ROV and conducted communication analysis between the USV and ROV howvere, they did not propose a coordination system between the USV and ROV[5,6]. Furthermore, the mathematical modeling of the composite platform and the analysis of underwater cable drag were presented without actual platform or field test results[7]. Therefore, in this study, the advantages of combining USV and ROV were utilized to construct a platform capable of perfoming tasks on coastlines inaccessible to large work vessels, and further investigations were conducted into the control of USV and the relative control of ROV. Typically, composite platforms are affected by the tether cable connecting the USV and ROV. However, the composite platform developed in this study was deisghed and fabricated for long-term operation, real-time underwater data collection, and relative position measurement using GPS and Ultra-Shore BaseLine (USBL).

II. DYNAMICS AND CONTROL SYSTEM

A. Dynamics of Composite Platform

Fig. 1 and Table 1 shows the frame of the composite platform and outlines the coordinate system, respectively.

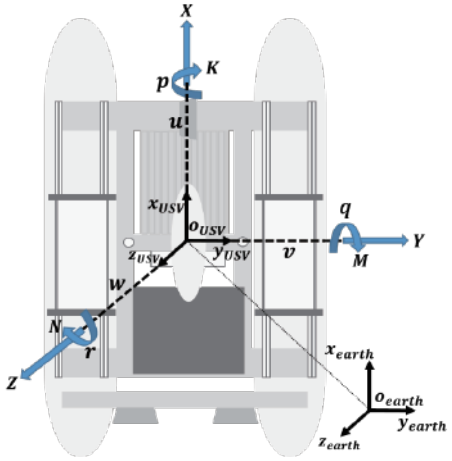


Fig. 1. Coordinates of Compostie Platform

TABLE I. NOTATION OF SNAME FOR COMPOSITE PLATFORM

	Force and moments	***Linear and angular Velocity***	***Position and Euler angle***
Motion in the X direction	X	u	X
Motion in the Y direction	Y	v	Y
Motion the Z direction	Z	w	Z
Rotation about the X- axis	K	p	φ
Rotation about the Y-axis	M	q	θ
Rotation about the Z-axis	N	r	ψ

The basic formulas of dynamcs are based on the Newton-Euler equations and vector-matrix dynamics. As the USV moves along the x-axix and y-axis, it possesses three degrees of freedom and is capable of surge, sway, and yaw motions. The dynamic equations of the USV are expressed in equatios (1a), (1b), and (1c).

$$(m - X_{(\dot{u})})\dot{v}(u) - r + (Y_{(\dot{v})}vr + Y_{(\dot{r})}r - mx_G\, r)r - (X_u + X_{u|u|}|u|)u = \tau_x \quad (1a)$$

$$(m - Y_{(\dot{v})})(\dot{v}) + (mx_G - Y_{(\dot{r})})(\dot{r}) + \mu r - X_{(\dot{u})}u_r r - (Y_v + Y_{v|v|}|v|)v - (Y_r + Y_{r|r|}|r|)r = \tau_Y \quad (1b)$$

$$(mx_G - N_{(\dot{v})})(\dot{v}) + (I_Z - N_{(\dot{r})})(r - (Y_{(\dot{v})}v_r r - \dot{Y}_{(\dot{r})}r + mx_G\, r)u) + X_{(\dot{u})}u_r v - (N_v + N_{v|v|}|v|) - (N_r + N_{r|r|}|r|)r = \tau_N \quad (1c)$$

here, τ represents the thruste force. Detailed definitions of related parameters and conditions are provided in the related studies[8,9]. The motion of the ROV has six degrees of freedom: surge, sway, heave, roll, pitch, yaw. The heave motion of the ROV is achieved throuth winch control; therefore, the dynamic equation of the ROV is expressed as in equations (2a), and (2b)

$$M_{RB}\dot{\nu} + M_A\nu_w + C_{RB}(\nu)\nu + C_A(\nu)\nu + D(\nu_w)\nu_w + g(\eta) = \tau \quad (2a)$$

$$\tau = [\tau_1\ \tau_2 \dots \tau_5\ \tau_6]^T \text{ where, } \tau_3 = \tau'_3 + \tau_w \quad (2b)$$

here τ'_3 represents the thrust force for heaving motion, and τ_w denotes the winch force. The conditions and detailed definitions of related parameters can be found in related studies[8].

B. Algorithm of Cross-Track Error

The cross track error (CTE) refers to a linear distance error caused by disturbances such as waves, wind, and tides in the target and actual moving paths of the platform. In this study, an algorithm to overcome CTE was proposed, and its simplified schematic diagram is shown in Fig. 2.

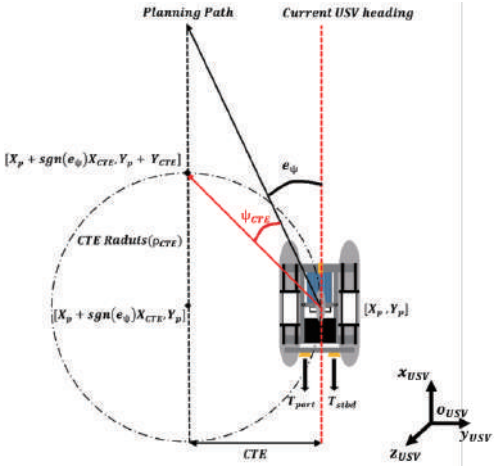


Fig. 2. Schematic of CTE algorithm

In this study, to reduce the CTE during the USV surging control, the center of the platform $[X_p, Y_p]$ is shifted by Y_{CTE} in the y-axis direction to obtain new coordinates $[X_p, Y_p + sgn(e_\psi)Y_{CTE}]$. Next, a circle ρ_{CTE} is drawn with the new point as the center, and the new target direction angle ψ_{CTE} is calculated using the intersection point $[X_p + X_{CTE}, Y_p + sgn(e_\psi)X_{CTE}]$, between ρ_{CTE} and the target path. Then, the calculated ψ_{CTE} is subtracted from the target direction angle obtained using the LOS, e_ψ.The difference between ψ_{CTE} and e_ψ denoted as E_ψ is used as the error in the control. Therefore, the CTE algorithm is expressed as follows in equation (3).

$$T_{CTE} = K_{P\psi}(E_\psi) + K_{I\psi}\int E_\psi + K_{D\psi}(\dot{E_\psi}) \quad (3)$$

where $E_\psi = e_\psi - \psi_{CTE}$

In the CTE algorithm, the output of the controller is the thruster force of either the port-side thruster or the starboard-side thruster, depending on tih sign of e_ψ. The output of each thruster is expressed in equation (4) and Fig. 8 shows the block diagram of the PID controller considering the CTE algorithm.

$$\begin{aligned} T_{Port} &= T_{CTE} \quad \text{if } e_\psi < 0 \\ T_{Stbd} &= T_{CTE} \quad \text{if } e_\psi > 0 \end{aligned} \quad (4)$$

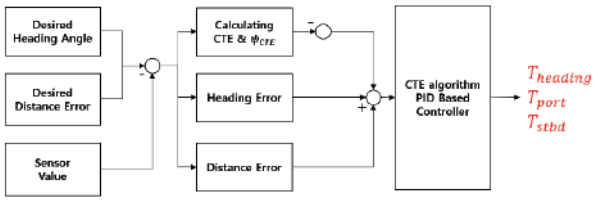


Fig. 3. Block-Diagram of CTE Algorithm

C. Dynamic Positoning Algorithm of USV

Dynamic positioning is an algorithm used to maintain a specific angle or position of the platform. The DP algorithm develod in this study is programmed to allgn the bow of the USV, which receives less drag, in the direction of tidal currents to minimize the impact of drag in areas with high currents. Subsequently, GPS detects the USV being pushed by the currents, and the algorithm dynamically adjusts the bow direction to maintain its position. The DP algorithm ensures that both the USV and ROV remain at the task position. Fig. 4 depicts the simplified schematic diagram of the DP algorithm used in this study.

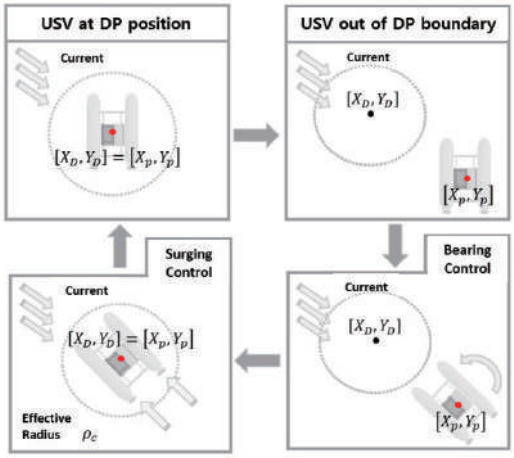


Fig. 4. Schematic of DP algorithm

The objective of the dynamic positioning is to ensure that the distance error of the USV remains within the effective radius ρ_c. This is achieved through the utilization of the thruster control system and the aforementioned CTE algorithm. Equation (5) expresses this goal as follows:

$$[X_D - X_p]^2 + [Y_D - Y_p]^2 < {\rho_c}^2 \quad (5)$$

Where, X_D and Y_D represent the position of the target point, and X_p and Y_p represent the current position of the USV.

D. SlidingMode Control of ROV

In this section, we describe the attitude control necessary for the hovering motion of the ROV. The motion of the ROV related to yaw, roll, and pitch angles is given by equation (6).

$$\mathrm{M}\ddot{\mathrm{x}} + \mathrm{C}(\mathrm{x}, \dot{\mathrm{x}}) + g(\mathrm{x}) = \mathrm{u} \tag{6}$$

where,

$$\dot{\mathrm{x}} = [\mathrm{p} \quad \mathrm{q} \quad \mathrm{r}]^T, \mathrm{u} = [\mathrm{u_p} \quad \mathrm{u_q} \quad \mathrm{u_r}]^T.$$

$$\mathrm{M} = \mathrm{M}_{RB} + \mathrm{M}_A = \begin{bmatrix} I_x - K_{\dot{p}} & 0 & 0 \\ 0 & I_y - M_{\dot{q}} & 0 \\ 0 & 0 & I_z - N_{\dot{r}} \end{bmatrix},$$

$$\mathrm{C} = \begin{bmatrix} (I_z - I_y)qr - (\dot{r} + pq)I_{xz} + (r^2 - q^2)I_{yz} + (pr - \dot{q})I_{xy} \\ (I_x - I_z)rp - (\dot{p} + qr)I_{xy} + (p^2 - r^2)I_{zx} + (qp - \dot{r})I_{yz} \\ (I_y - I_x)pq - (\dot{q} + rp)I_{yz} + (q^2 - p^2)I_{xy} + (rq - \dot{p})I_{zx} \end{bmatrix},$$

$$g(x) = \begin{bmatrix} W_{z_g} \cos(\theta) \sin(\theta) \\ W_{z_g} \sin(\theta) \\ 0 \end{bmatrix}$$

Here, the coriolis added mass term and hydrodynamic damping are omitted due to their small values under the situation of slow hovering motion. Furthermore, to overcome disturbances such as uncertain hydrodynamic forces and achieve stable attitude control of the ROV, a controller using Sliding Mode Control (SMC) was designed and validated through actual experiments. Parameters can be found in related studies [8].

To control these 3 axes robustly, SMC was applied. When defining error as $\tilde{\mathrm{x}} = \mathrm{x} - \mathrm{x}_d$, s and $\dot{\mathrm{s}}$ are expressed in equation (7) and (8), respectively.

$$\mathrm{s} = \mathrm{M}\left(\frac{d}{dt} + \lambda\right)\tilde{\mathrm{x}} = \mathrm{M}\left(\dot{\tilde{\mathrm{x}}} + \lambda\tilde{\mathrm{x}}, \ \lambda\right) > 0 \tag{7}$$

$$\dot{\mathrm{s}} = \mathrm{M}\left(\ddot{\tilde{\mathrm{x}}} + \lambda\dot{\tilde{\mathrm{x}}}\right) = u - \mathrm{C}\dot{\mathrm{x}}|\dot{\mathrm{x}}| - g(\mathrm{x}) - M(\ddot{\mathrm{x}}_d - \lambda\dot{\tilde{\mathrm{x}}}) \tag{8}$$

At this time, the Lyapunov function must be established. Therefore, λ is set to $\frac{1}{2}\mathrm{M}\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{s}^2 = \mathrm{M}(\dot{\mathrm{s}})\mathrm{s} \leq 0$.

Thus, the control inputs u and gain k are expressed in (9).

$$\mathrm{u} = \widehat{\mathrm{M}}\left(\ddot{\mathrm{x}}_d - \lambda\dot{\tilde{\mathrm{x}}}\right) + \hat{C}\dot{\mathrm{x}}|\dot{\mathrm{x}}| + g(\mathrm{x}) - \mathrm{ksgn}(\mathrm{s}) \tag{9}$$

where,

$$\mathrm{k} > \left\|(\widehat{\mathrm{M}} - \mathrm{M})\left(\ddot{\mathrm{x}}_d - \lambda\dot{\tilde{\mathrm{x}}}\right) + (\widehat{\mathrm{C}} - \mathrm{C})\dot{\mathrm{x}}|\dot{\mathrm{x}}|\right\|$$

Fig. 5 shows the block diagram of a PD controller and the SMC. In SMC controller, robust action ksgn(s) is added to PD control.

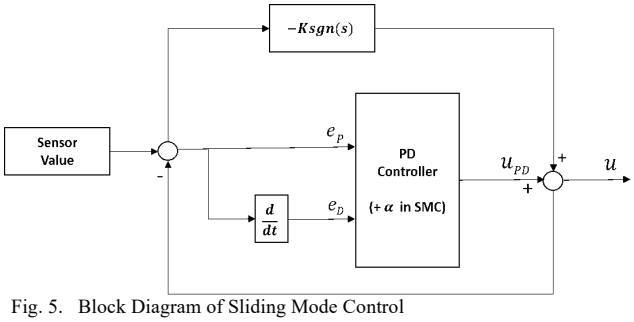


Fig. 5. Block Diagram of Sliding Mode Control

III. Performance Test of Composite Platform

A. DP Performance Test of Composite Platform at Sea

The DP performance test of the composite platform involved conducting dynamic positioning of the Usv at coordinates [35.0750216, 129.0849246] for 10 minutes and setting the DP point of ROV to the centroid position of the USV [0, 0]. Since the ROV and tether cable may affect the USV during DP, the effective radius ρ_c of the USV DP was changed from 3m to 4m. From 433 collected data points, it was confirmed that the USV's DP point remained within a 4m effective radius. The ROV exhibited an error of approximately 40cm in the x-axis direction and about 50cm in the y-axis direction due to the influence of currents. The average values between the DP points of the ROVm were [-0.4095m, -0.5348m]. Fig. 6, and Fig. 7 depict the composite platform during the DP performance test and results of the DP performance test of the composite platform.

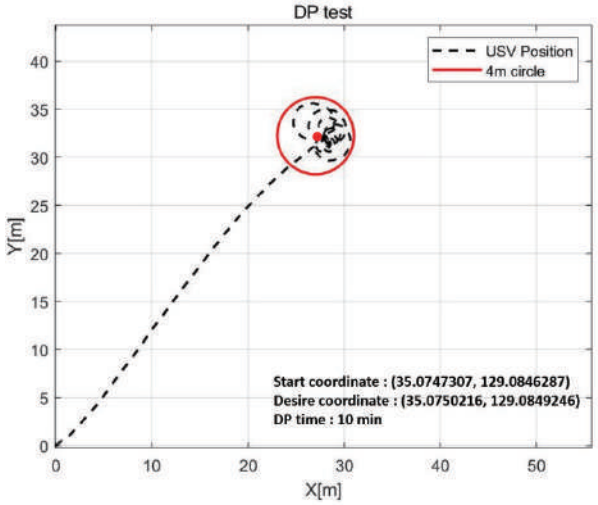


Fig. 6. Result of Composite Platform(USV) DP Performace test

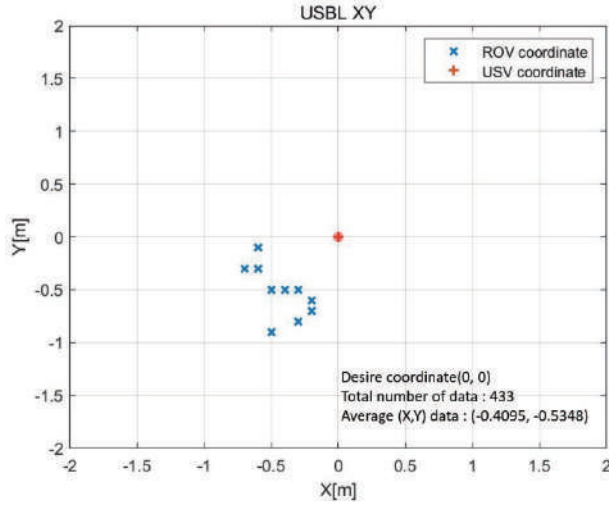


Fig. 7. Result of Composite Platform(ROV) DP Performace test

B. *Performance Test of ROV Hovering Control at Sea*

In this section, the USV was used to maneuver the ROV to the task position. At the task position, the USV engaged in DP while the ROV hovered. This study compared the performance of the PD controller, widely used for underwater ROV attitude control, with the SMC, designed to robustly compensate for disturbances in water. Gyro sensors were employed for attitude sensing. After establishing ROV control used the PD controller, the SMC, including robust control elements to eliminate disturbances at the same PD gain, was tested. Control experiments were conducted under nearly identical conditions for approximately 1 minute. Data were measured at intervals of 0.1 seconds, with target angles of 0°, 0°, and 50°for roll, pitch, and yaw respectively. Fig. 8-10 illustrate the results of stability tests for roll, pitch, and yaw. The results demonstrate superior performance and lower error of the SMC compared to PD control. Table 2. Presents the Root Mean Squar (RMS) control error results for PD and SMC. According to the test results in Table 2. The robust SMC showed better performance than PD control for roll, pitch, and yaw angle control of the ROV by 21%, 33%, and 30% respectively.

TABLE II. COMPARISON BETWEEN THE PD AND SMC RESULT

	PD error	***SMC error***
Roll(°)	-0.77940	-0.61474
Pitch(°)	0.59657	0.39623
Yaw(°)	-0.42525	-0.29591

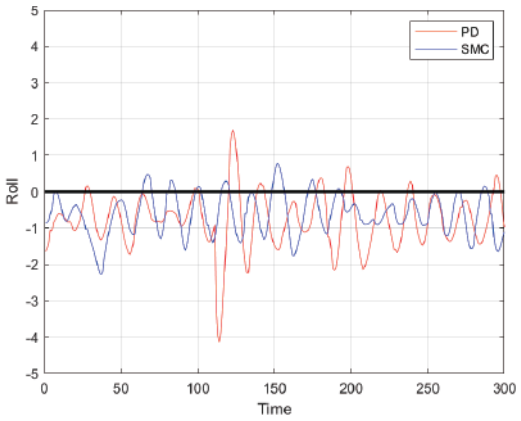


Fig. 8. Roll Test of ROV Applying PD and SMC

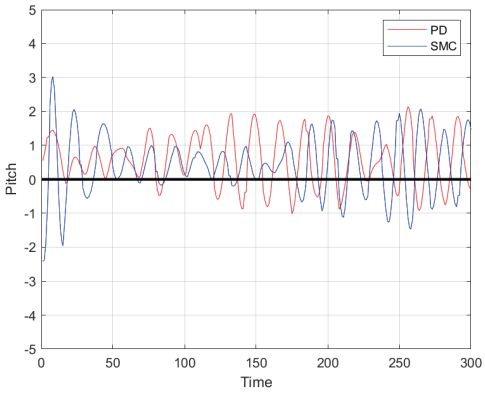


Fig. 9. Pitch Test of ROV Applying PD and SMC

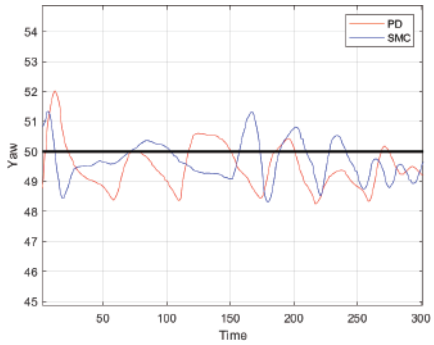


Fig. 10. Yaw Test of ROV Applying PD and SMC

IV. CONCLUSION

In this study, a highly operable USV and an ROV capable of performing underwater tasks were developed for the research and development of a composite platform that provides stable power supply and real-time acquisition of underwater data. Additionally, the structure and control system of the composite platform, consisting of USV, ROV, and winch systems, were designed, manufactured, and integrated with control algorithms. The performance of the developed composite platform was validated through real sea area tests. Particularly, a CTE algorithm was proposed for surging control, and in the DP test of the composite platform, the effective DP radius condition was set to 4m, with the DP control satisfying this condition. In sea trials for ROV hovering control, PD and robust SMC were applied, with robust SMC showing 21%, 33%, and 30% better performance than PD control in roll, pitch, and yaw angle control of the ROV, respectively.

ACKNOWLEDGMENT

This research was a part of the project titled 'Development of smart maintenance monitoring techniques to prepare for disaster and deterioration of port infra structures', funded by the Ministry of Oceans and Fisheries, Korea.

REFERENCES

[1] J.W. Necholson, A.J. Healey, "The present state of autonomous underwater vehicle (AUV) applications and technologies", Proceeding of Marine Technology Society Journal, 2008, vol.42(1), pp 44-51.

[2] T.S. Kim, C.H. Kim, M.K. Lee, "Study on the design and the control of an underwater construction robot for port construction", Proceeding of Journal of Navigation and Port Research, 2015, vol.39(3), pp. 253-260.

[3] Y. Chen, Y. Liu, Y. Meng, S. Yu, Y. Zhuang, "System modeling and simulation of an unmanned aerial underwater vehicle", Proceeding of J. Mar. Sci. Eng, 2019, vol.7, pp. 444.

[4] J. Rodriguez, H. Castaneda, L. Gordillo, "Design of an adaptive sliding mode control for a micro-AUV subject to water currents and parametric uncertainties", Proceeding of J. Mar. Sci. Eng, 2019, vol.7, pp. 445.

[5] E. Lachaud, Y. Monbeig, P. Nolleau, A. Hardy, M. Thompson, M. Lardeux, "Opportunities and challenges of remote operating a ROV embarked on a USV", Proceeding of Offshore Technology Confence, 2018, Houston, TX, USA, 30

[6] B. Lyu, Z. Zeng, D. Lu, L. Lian, J. Tu, L. Xiao. Y. Liu, "Combined small sized USV and ROV observation system for long-term, large-scale, spatially explicitaquatic monitoring", Proceeding of the 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO), 2018

[7] M. Vu, M. Van, D.H.P. Bui, Q.T. Do, T.T. Lee, S.-D., H.S. Choi, "Study on dynamic behavior of unmanned surface vehicle-linked unmanned underwater vehicle system for underwater exploration", Proceeding of Sensor, 2020, vol. 20, pp. 1329.

[8] T.I. Fossen, "Handbook of marine craft hydrodynamics and motion control", Proceeding of John Wiley & Sons: Hoboken, Nj, USA, 2011.

[9] C.J. Wu, B. Eng, "6-DoF Modelling and Control of a Remotely Operated Vehicle", Proceeding of Flinders University, Adelaide, South Australia, 2018

Enhancing Maritime Security via Blockchain Integration for Cross-Border Vehicle Authentication

Mohamed A. Dini, Esmot Ara Tuli, Dong-Seong Kim and Taesoo Jun
Networked Systems Laboratory, Department of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea.
mohamedabubakar407@gmail.com, (esmot, dskim, taesoo.jun)@kumoh.ac.kr

***Abstract*—Maritime borders hold significant importance for a nation's economic and political stability. Ensuring the security of these extensive maritime frontiers is a task of priority importance for naval forces, particularly in managing and mitigating unauthorized entries by foreign vessels into international maritime zones. To simplify and automate this task, this paper proposes a blockchain-enabled maritime vehicle authentication and verification system. The system employs octopus robots, designed for monitoring and performing authentication checks. A smart contract is utilized to implement the demonstrative authentication process, ensuring that operations are secure, decentralized, and tamper-proof. This integration enhances the efficiency and reliability of maritime security measures.**

***Index Terms*—Blockchain, smart contract, maritime security, vehicle authentication.**

I. Introduction

Maritime activities, crucial for global trade, defense, and environmental monitoring, have grown significantly, increasing the need for security in these vast, often ungoverned waters [1]. Tracking unauthorized vehicles is vital for protecting maritime borders against smuggling, unauthorized fishing, and national security threats. Therefore, securely and autonomously monitoring and authenticating cross-border vehicle movements is essential [2], [3]. The complex and dynamic maritime environment challenges current security frameworks, necessitating innovative solutions for real-time tracking, data integrity, and operational resilience against unauthorized access, ensuring the safety and security of maritime activities and borders [4].

Traditional approaches struggle to ensure data integrity, face limitations in autonomous navigation in complex underwater environments, and lack efficient, secure data transmission capabilities [3]. This scenario necessitates the development of an innovative system that leverages the maneuverability and adaptability of biomimetic robotics, combined with the immutable, decentralized data management afforded by blockchain technology. Such integration aims to enhance operational efficiency, ensure secure and tamper-proof data handling, and set new standards in maritime security operations [5]. The lack of a system addressing both advanced navigation and secure data management in the maritime domain underscores a significant research gap, highlighting the need for an inventive approach to maritime security and surveillance.

To tackle the issues mentioned in the comprehensive problem outline, this study presents a blockchain-based octopus robot designed for marine cross-border vehicle identification. While the integration of advanced biomimetic robotics provides the necessary physical infrastructure for complex and efficient underwater navigation and data gathering, the primary focus of this effort is on the blockchain component. The importance of blockchain technology cannot be overstated, as it has the potential to revolutionize maritime security operations by enhancing data security, integrity, and operational efficiency, particularly in vehicle authentication. This approach aims to ensure dependable and secure maritime operations by leveraging the strong, immutable nature of blockchain. The contribution of this work, centered on the blockchain dimension, is as follows:

- To enhance security and trust in maritime operations, this work utilizes blockchain technology to secure and authenticate data collected and transmitted by an octopus robot.
- To innovate at the intersection of robotics and blockchain, this project optimizes blockchain to support the operational needs of an octopus-inspired robot in maritime environments.

The rest of the paper is organized as Section II for related works, Section III for system design and experimental setup, Section IV for result discussion and system evaluation, and Section V to conclude the study.

II. Related works

Recent advancements in blockchain technology have led to significant enhancements in the field of marine transportation and intelligent transportation systems. For instance, Zhang et al. (2023) developed a blockchain-based IoT collaborative processing system that notably improves the security and efficiency of marine transportation flow scheduling and management. Their proposed system not only resists replay and camouflage attacks but also enhances the security and transaction processing speed by 8% and 6% respectively, without incurring additional communication costs [6]. Additionally, Ying and Wang (2023) addressed cross-border data privacy concerns through a blockchain model that facilitates secure identity authentication without the need for sensitive data to traverse borders. By integrating SM2 and SM9 cryptographic algorithms, their model ensures robust data security and supports efficient cross-border identity authentication [7].

Furthermore, blockchain technology is making strides in improving communication security, consensus efficiency, and

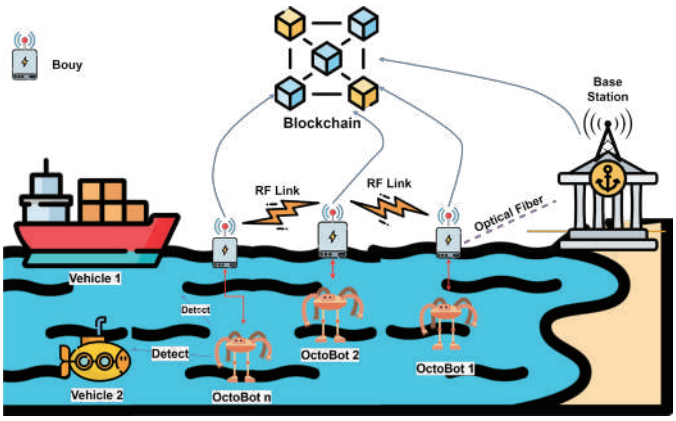


Fig. 1. An illustration of blockchain-based secure maritime communication

fault tolerance within the Internet of Vehicles (IoV) [8]. Hu et al. (2019) implemented a blockchain-based Byzantine consensus algorithm that significantly bolsters the IoV's capability to handle authentication challenges in intelligent transport [9]. Another innovative approach by Si et al. (2023) introduced a heterogeneous cross-chain interaction mechanism that utilizes relay nodes and identity encryption to enhance data communication and interoperability across the Intelligent Transportation System (ITS), ensuring secure and efficient cross-domain interactions [10]. These blockchain-based systems not only tackle the traditional challenges of centralization and inefficiency in authentication protocols but also shift the computing load to Roadside Units (RSUs), streamlining the authentication process and communication within cross-trusted Authority (TA) networks. These developments underscore the transformative impact of blockchain technology across various aspects of transportation and security management. However, our paper presents a novel approach that encompasses a blockchain-based authentication system, complemented by an automated mechanism for the detection and identification of unauthorized vehicles, facilitated by the deployment of octopus robots.

III. Proposed System

Fig. 1 illustrates the proposed system. A network of miniature octopus robots is strategically deployed across maritime borders to leverage their exceptional maneuverability for underwater navigation and monitoring tasks. Drawing inspiration from prior successful implementations as noted in [11], these robots are uniquely suited for operating in complex underwater environments where precise and agile movements are essential. Due to the inherent challenges associated with underwater communication, such as signal instability and noise, the robots are connected via cables. These cables serve multiple purposes: they facilitate reliable data and power transmission and provide a means for physical retrieval of the robots if necessary. To enable data transfer from the different octopus robots and guarantee real-time connectivity, RF link transmission will be used. Optical fiber will be used to establish all connections to the base station in order to maximize transfer speeds and maintain high reliability. A decentralized and impenetrable record of all operations will be established since every piece of data that is received or communicated among all nodes will be delivered straight to and securely kept on the blockchain network. This system architecture supports the overall security and integrity of the marine surveillance system, in addition to improving data handling efficiency.

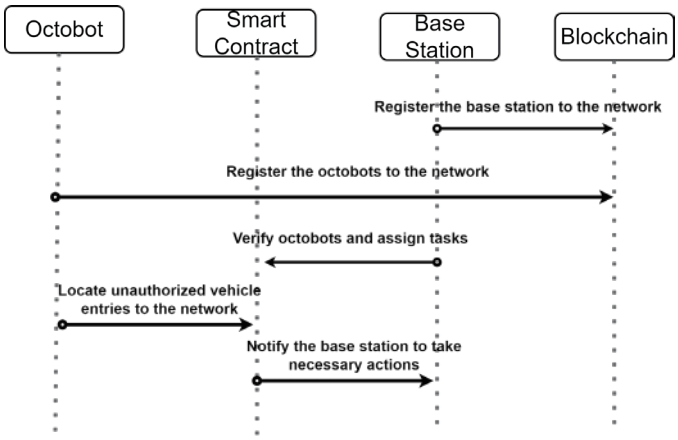


Fig. 2. Workflow of the Blockchain-Based Octopus Robot System for Maritime Cross-Border Vehicle Authentication

Fig. 2 showcases the proposed system pseudocode, which begins with the registration of both the base station and octobots on the blockchain network, ensuring secure integration. The base station then registers authorized vehicles using a smart contract. After verifying the octobots, the base station assigns them specific tasks and deployment positions in the sea. When an octobot detects an unauthorized vehicle, it sends the vehicle's location to the base station, which then takes appropriate action. All actions and data are securely recorded by the smart contract, ensuring data integrity and operational efficiency.

The core focus of this work is on the implementation of blockchain technology for the registration and authentication of vehicles. Each octopus robot is equipped to identify and authenticate maritime vehicles via the "MaritimeSecurity" smart contract [12]. This blockchain-based approach ensures that all data regarding vehicle movements are securely logged and immutable. In instances of unauthorized vehicle detections, the system is designed to instantly report such events to the blockchain network, which in turn notifies border security agencies promptly. The MaritimeSecurity smart contract employs several key constructs:

- V: A vehicle represented by a struct containing its ID (id), authorization status ($auth$), last access timestamp ($lastAccess$), registrant address ($regAddr$), and access history count ($histCount$).
- M_v: A mapping of vehicle IDs to their corresponding V structs.
- M_h: A mapping of vehicle IDs to their access history timestamps.
- $owner$: The contract deployer with exclusive administrative privileges.

A. Contract Functionalities

The blockchain-based smart contract for maritime security encompasses key operations essential for managing vehicle access and authentication, all encapsulated within a comprehensive mathematical framework:

1. Vehicle Registration: Algorithm 1 introduces additional variables and checks to ensure a more robust registration process. Initially, it sets up collections for registered vehicles M, valid vehicle types T, and valid states for registration S, as well as an empty set for errors. The algorithm retrieves the vehicle type and state using helper functions. If the vehicle is already registered, it emits an error and reverts. It then checks if the vehicle type and state are valid, adding any errors to the 'errors' set. If errors are detected, it emits each error and reverts the process. If no errors are found, the vehicle ID is added to the registered vehicles set, and a 'VehicleRegistered' event is emitted. This approach ensures that only valid and unregistered vehicles are processed, making the system more reliable and efficient.

Algorithm 1 Vehicle Registration Algorithm

procedure REGV(id)
  $M \leftarrow$ Set of registered vehicles
  $T \leftarrow$ Set of valid vehicle types
  $S \leftarrow$ Set of valid states for registration
  $errors \leftarrow \emptyset$
  $type \leftarrow$ `getType`(id)
  $state \leftarrow$ `getState`(id)
  if $id \in M$ **then**
    emit `Error`("Vehicle already registered")
    revert
  end if
  if $type \notin T$ **then**
    $errors \leftarrow errors \cup \{"Invalidtype"\}$
  end if
  if $state \notin S$ **then**
    $errors \leftarrow errors \cup \{"Invalidstate"\}$
  end if
  if $errors \neq \emptyset$ **then**
    for $err \in errors$ **do**
      emit `Error`(err)
    end for
    revert
  end if
  $M \leftarrow M \cup \{id\}$
  emit `VehicleRegistered`(id)
end procedure

2. Vehicle Authentication: A function evaluates whether a vehicle, identified by its unique ID, is authorized to access restricted maritime zones. This function checks the vehicle's registration status within the blockchain registry to ensure it has the necessary permissions for entry.

$$\text{authenticateVehicle}(id) \rightarrow auth_{id} \tag{1}$$

This process in Equation. 1 determines if the vehicle represented by id is listed as authorized in the mapping M_v. The output, $auth_{id}$, is a boolean value that indicates whether the vehicle is permitted access. A return value of true signifies that the vehicle is authorized, while false indicates that it is either not registered or not authorized, thereby enforcing secure and controlled access to designated maritime areas.

3. Access Logging: This function systematically records each instance a vehicle enters or exits a restricted maritime zone. This is crucial for maintaining a chronological log that supports both security measures and operational analysis.

$$\text{logAccess}(id, t) \rightarrow M_h[id].append(t),\ lastAccess_{id} = t \tag{2}$$

In Equation 2, t represents the timestamp when the access occurred. This function updates the access history of a vehicle in the mapping M_h by appending the new timestamp t to the list of timestamps associated with the vehicle ID id. Simultaneously, it updates the 'lastAccess' time for the vehicle to t. These updates ensure that a comprehensive and traceable log of vehicle movements is maintained, which is essential for subsequent reviews and verifications of security and operational activities.

4. Vehicle Status Update: Equation 3 showcases the main variables needed to update the status of vehicles in the system, which is crucial to dynamically managing the authorization status of registered vehicles. It allows administrative adjustments to a vehicle's access permissions based on ongoing security assessments or compliance changes.

$$\text{updateVehicleStatus}(id, auth) \rightarrow auth_{id} = auth \tag{3}$$

In this definition, $auth$ is the new authorization status, which could be either true (authorized) or false (unauthorized). This function modifies the authorization status $auth_{id}$ for the vehicle identified by id in the registry M_v. The update directly impacts whether the vehicle will be allowed access to secure maritime zones in the future, ensuring that the vehicle access permissions remain aligned with the latest security protocols and regulatory requirements. This capability is essential for maintaining the integrity and security of maritime operations.

B. Experimental Setup

Using Remix IDE and smart contracts deployed on the Ethereum-based Goerli test network, the proposed system effectively captures and records vehicle data detected by the octopus robot at maritime borders. This process is managed in a decentralized manner, leveraging the Solidity programming language for smart contract development. The testing environment included an MSI PC running Windows 10 with specifications that include an Intel(R) Core(TM) i5-8500 CPU at 3.00GHz, 6 cores, an NVIDIA GeForce GT 1030 GPU with CUDA capabilities, and 36GB of RAM. The development framework Goerli was also employed to facilitate testing. The system's efficiency and operational feasibility were affirmed

by analyzing gas cost metrics, confirming the proposed solution's viability.

IV. Results

TABLE I
GAS COST METRICS FOR BLOCKCHAIN-ENABLED MARITIME VEHICLE AUTHENTICATION FUNCTIONS

Function	Gas	Function	Gas
registerVehicle	114458	logAccess	134823
revokeVehicleAccess	42931	updateVehicleAccess	60256
owner	0	authenticateVehicle	0
vehicles	0	checkVehicleHistory	0
vehicleAccessHistory	0		

Gas costs for the MaritimeSecurity smart contract are analyzed in Table. I. Gas costs for registerVehicle and logAccess, for example, are significantly higher—114,458 and 134,823 gas units, respectively. The main reason for this increased expense is because these services require a lot of storage. Due of the processing resources needed for data persistence, the registerVehicle function requires writing to the blockchain state in order to create and save a new vehicle entry. This process is inherently gas-intensive. Similar to this, the logAccess function adds access logs to the blockchain, updating its state and raising the amount of work required to process the transactions due to the larger ledger.

On the other hand, since revokeVehicleAccess and updateVehicleAccess mostly entail editing already-existing entries rather than adding new ones, these actions have lower gas costs of 42,931 and 60,256 gas units, respectively. Notably, read-only operations such as authenticateVehicle, checkVehicleHistory, and associated query operations are free of gas costs, indicating that they have no effect on the blockchain state. This disparity in gas prices highlights how crucial it is to optimize smart contract functions—which alter the blockchain state—in order to reduce operating expenses and improve system performance, particularly for essential features like access logging and vehicle registration.

V. Conclusion

Blockchain technology is widely used in many different fields, and its combination with robotics offers a promising path for further development. Through the development of the MaritimeSecurity smart contract, we have investigated this integration in this study with the goal of improving maritime border security via Cross-Border Vehicle Authentication. This work expands on earlier research on octopus robots intended for underwater navigation by applying blockchain technology to secure, blockchain-enabled activities. The functionality for car registration, access logging, termination of access, and status updates were carefully constructed into the smart contract.

The Goerli test network was used to deploy the system, and there, its functionality was extensively assessed, with a focus on gas cost efficiency. Looking ahead, field testing in real marine conditions and integrating the blockchain system with real-time sensor data will be the main priorities of future work. The goal of these endeavors is to evaluate the reliability of the system and the stability of data transmission in operational scenarios. This represents a noteworthy advancement towards the pragmatic implementation and actualization of resilient, blockchain-based marine security remedies.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 34%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33%) and by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology, under Grant (2022R1I1A1A01069334, 33%)

References

[1] S. Tsiulin, K. H. Reinau, and N. Goryaev, "Conceptual comparison of port commmunity system and blockchain scenario for maritime document handling," in *2020 Global Smart Industry Conference (GloSIC)*, 2020, pp. 66–71.

[2] S. Bauk and L. H. Ntshangase, "Maritime blockchain constraints' analysis by ism and micmac techniques," in *2023 12th Mediterranean Conference on Embedded Computing (MECO)*, 2023, pp. 1–6.

[3] S. Marenković, E. Tijan, and S. Aksentijević, "Blockchain technology perspectives in maritime industry," in *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2021, pp. 1414–1419.

[4] P. Rahimi, N. D. Khan, C. Chrysostomou, V. Vassiliou, and B. Nazir, "A secure communication for maritime iot applications using blockchain technology," in *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2020, pp. 244–251.

[5] B.-M. Batalden, P. Leikanger, and P. Wide, "Towards autonomous maritime operations," in *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2017, pp. 1–6.

[6] P. Zhang, Y. Wang, G. Aujla, A. Jindal, and Y. D. Al-Otaibi, "A blockchain-based authentication scheme and secure architecture for iot-enabled maritime transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, pp. 2322–2331, 2023.

[7] Z. Ying and K. Wang, "Blockchain distributed identity management model for cross-border data privacy protection," *Journal of Surveillance, Security and Safety*, 2023.

[8] E. A. Tuli, J.-M. Lee, and D.-S. Kim, "Integration of quantum technologies into metaverse: Applications, potentials, and challenges," *IEEE Access*, vol. 12, pp. 29 995–30 019, 2024.

[9] W. Hu, Y. Hu, W. Yao, and H. Li, "A blockchain-based byzantine consensus algorithm for information authentication of the internet of vehicles," *IEEE Access*, vol. 7, pp. 139 703–139 711, 2019.

[10] H. Si, W. Li, Q. Wang, H. Cao, F. Bacao, and C. Sun, "A secure cross-domain interaction scheme for blockchain-based intelligent transportation systems," *PeerJ Computer Science*, vol. 9, 2023.

[11] M. A. Dini, S. O. Ajakwe, D.-S. Kim, J. M. Lee, and T. Jun, "Octopus-inspired robot for underwater exploration in military internet of battle operations and surveillance," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 1900–1904.

[12] [Online]. Available: https://github.com/Dini-bot/MaritimeSecurity

Infrastructure-Linked CAV Overtaking Decision Algorithm

Jihun Kim*, Youngjin Yoon, Seoksan Shin
Vehicle Safety R&D Center
Korea Automotive Technology Institute
Daegu, Republic of Korea
*jhkim6@katech.re.kr

Dong Seog Han
School of Electronics Engineering
Kyungpook National University
Daegu, Republic of Korea
dshan@knu.ac.kr

***Abstract*—Autonomous vehicles have been the focus of extensive research, through which they have been developed beyond those relying on standalone functions to connected autonomous vehicles (CAVs). Connected autonomous vehicles are connected to various objects based on Vehicle-to-Everything (V2X) communication technology, thus enabling high-level autonomous driving functions. In this paper, we propose an overtaking decision algorithm that links CAVs with road infrastructure. The proposed algorithm is implemented and evaluated using road infrastructure and autonomous vehicle data collected on real roads. We aim to overcome the limitations of the perception area and driving strategy by leveraging road infrastructure. Ultimately, this work is expected to contribute to the realization of high-level autonomous vehicles.**

***Index Terms*—Autonomous Vehicle, Connected Autonomous Vehicles(CAVs), V2X Communication, Driving Decision, Overtaking, Road Infrastructure**

I. Introduction

Autonomous vehicles (AV) are considered to be the future of the automobile industry and the foundation for next-generation intelligent vehicles [1]. Over the past decade, autonomous driving has attracted both substantial research interest and associated market expectations. Autonomous driving is expected to lead to an era in which safe and convenient travel can replace the need for human driving. It is also expected that self-driving services will improve quality of life by reducing traffic accidents. In terms of the global market, it is predicted that the economic impact of autonomous driving will amount to 3.5 trillion dollars. By eliminating the human element of road hazards, autonomous driving technology is expected to make roads safer and lower the accident rate while driving. It will provide convenience. This possibility has attracted great interest not only in academia but also in industry, and the size of the research market is increasing daily. For the systematic development of autonomous driving technology, the Society of Automotive Engineers (SAE) divides autonomous driving into six levels ranging from levels 0 through 5 [2]. At level 0, the driver is responsible for looking ahead, steering, and accelerating/decelerating. Only in emergencies does the autonomous system assist the driver by warning the driver or temporarily intervening in performance, such as in rear-end collision warning or forward collision-avoidance assistance. At level 1, the system provides steering or acceleration/deceleration in certain situations. Examples include lane-keeping assist and adaptive cruise control. In level 2, both steering and acceleration/deceleration are performed. In each of levels 0, 1, and 2, the autonomous driving system assists the driver in certain situations, but in levels 3, 4, and 5, the system takes over driving such as by ensuring that the vehicle stays in its intended lane as well as by performing changes, overtaking, obstacle avoidance, etc. For these autonomous vehicles to operate on complex roads, it is important to consider not only the stability of the autonomous vehicle but also the time it takes to drive to the destination.

Overtaking algorithms are essential in reducing driving time, and they have therefore attracted research interest with related studies actively underway. Existing overtaking research is represented by methods for overtaking by calculating the optimal path [3] as well as methods that implement lane changes and overtaking without calculating the path [4]. The representative overtaking algorithm changed lanes by considering the presence of other vehicles in the next lane based on the distance to the vehicle in front, which allowed for the implementation of the overtaking algorithm. However, this algorithm does not apply to the existing front vehicle after changing lanes. The situation of overtaking a vehicle was not carefully considered, which could lead to a risk of collision in real-world environments where the vehicle in front moves dynamically. To solve this problem, the current paper proposes an overtaking decision algorithm that utilizes sensor infrastructure on the road [5]. The proposed algorithm transmits and receives information from sensors such as cameras, lidar, and radar on the road through Vehicle-to-Everything (V2X) to enable autonomous driving. Overtaking is determined by assessing the environment around the vehicle. This is because, when using V2X communication, it is easier to obtain traffic information on the distance ahead, as is required for mid-to long-term overtaking, compared to the difficulty of making decisions using only the sensors possessed by existing drivers and autonomous vehicles.

The rest of this paper is organized as follows. Section II

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE), Korea, under the "Regional Industry Infrastructure and RD Support Program" (P0018599) supervised by the Korea Institute for Advance- ment of Technology (KIAT).

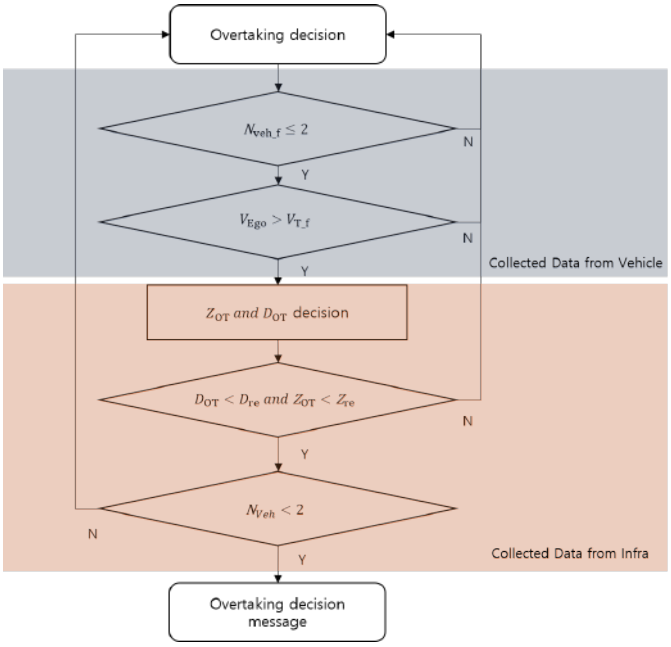


Fig. 1. Flowchart of the proposed overtaking decision algorithm.



Fig. 2. Radar road infrastructure and camera in autonomous driving testbed.

introduces the current status of V2X communication for sharing road infrastructure information with autonomous vehicles. The proposed road infrastructure-linked overtaking decision algorithm is formalized in Section III. The proposed algorithm is verified using results collected through experiments conducted within an autonomous driving testbed. Finally, the paper concludes in Section IV.

II. Road Infrastructure and V2X Communication

Autonomous driving methods can be divided into the Standalone type [6], [7], wherein the system embedded within the vehicle itself solely performs recognition, judgment, and control, and the connected type [8], [9], wherein the driving system performs perception, decision, and control in cooperation with surrounding vehicles and road infrastructure. To cooperate with surrounding vehicles or road infrastructure in this way, a V2X communication technology must be used. There are two representative methods of V2X communication for CAVs: wireless access in vehicular environments (WAVE) of the Institute of Electrical and Electronics Engineers (IEEE) [10] and cellular-V2X of the 3rd generation partnership project (3GPP) [11].

Recent applications of V2X communication for autonomous vehicles have aimed to implement advanced autonomous driving functions such as extended sensing, platooning, and remote control beyond the capabilities of cooperative-intelligent transport Systems (C-ITS) services [12]. Advanced autonomous driving ultimately means fully autonomous driving. To implement and commercialize fully autonomous driving functions, it is essential to have a cooperative structure between V2X and road infrastructure. Connected road infrastructure services began with the purpose of collecting traffic information, but their functions are expanding due to the development of autonomous vehicles and V2X technologies [13], [14]. Road infrastructure that supports autonomous driving is used to expand the perception area through the use of sensors such as cameras, radar, and lidar, or to provide traffic signal information using V2X communication. Representative infrastructure of this type includes camera or radar-based traffic information collection devices and lidar-based intersection information collection devices. The composition of road infrastructure and autonomous vehicles expands to smart cities or urban demonstration complexes.

III. Proposed Overtaking Decision Algorithm

The overtaking decision algorithm for autonomous vehicles proposed in this paper is depicted in Fig. 1. The proposed algorithm provides a safe and accurate overtaking function by allowing autonomous vehicles to determine overtaking in coordination with road infrastructure. There are basic prerequisites for the proposed algorithm to work: First, there must be a section in which applicable road infrastructure (such as radar and cameras) exists, and second, the autonomous vehicle must satisfy the conditions of driving at a lower speed than the driving speed set by obstacles or low-speed vehicles.

The overtaking driving decision algorithm is operated in an autonomous driving control center that applies the V2X method introduced in Section II. The autonomous driving control center collects and manages information or data on road infrastructure and autonomous vehicles and distributes messages that are relevant to autonomous driving as necessary. The preceding operation of the proposed algorithm determines whether or not to overtake based on data collected from the autonomous vehicle. The overtaking condition is established

when the number of obstacles or vehicles ($N_{\text{Veh_f}}$) detected from the front perception sensor is two or fewer. The speed of the autonomous vehicle (V_{Ego}) is lower than the driving target speed ($V_{\text{T_f}}$). Next, the overtaking distance (D_{OT}) and area for overtaking (Z_{OT}) are decided based on the current speed of the autonomous vehicle, which is obtained from road infrastructure data. The overtaking distance is calculated as follows:

$$D_{\text{OT}} = V_{\text{Ego}} \times T_{\text{OT}} \tag{1}$$

where T_{OT} is the time required to overtake driving. This is determined to be inversely proportional to the current speed of the vehicle [15]. By comparing the distance and area from the current location of the autonomous vehicle to the maximum detection point of road infrastructure, overtaking is selected only when the remaining distance (D_{re}) and area (Z_{re}) for overtaking are large. Finally, if the number of obstacles or driving vehicles in the detection area is less than two, then a final overtaking decision is made. If it is determined that overtaking is possible, the relevant information is transmitted to the autonomous vehicle to determine whether or not overtaking is possible in real time.

Fig. 2 shows the radar road infrastructure within the autonomous driving testbed [16]. Radar road infrastructure consists of radar sensors and controllers that are capable of detecting objects up to 1,000 meters away using the frequency of the Ka-Band (34 GHz) band. A tracking camera is also configured to monitor real-time road conditions at the point where the radar road infrastructure is located. Fig. 3 shows the image data that collects information on the direction of traffic flow and driving vehicles detected in the area in real-time by the radar road infrastructure through tracking cameras. In this paper, the image data from the tracking camera was used to verify the detection data of the radar road infrastructure. The detection data of the radar road infrastructure is expressed in the form of a two-dimensional radar coordinate system value comprising latitude and longitude values, as well as the driving direction, driving area, speed, and type of obstacle or vehicle based on the system time and the ID value of the detected obstacle or vehicle. Detection data from the radar road infrastructure is collected at 100 ms and transmitted to the autonomous driving integrated control center. The collected radar road infrastructure data is utilized in the latter part of the proposed overtaking decision algorithm in Fig. 1. The first half of the proposed overtaking decision algorithm utilizes obstacle or vehicle information detected from cognitive sensors such as the cameras and lidar of autonomous vehicles. The autonomous driving control center collects the information through a V2X communication device, and the collected messages use the SAE J2735 standard [17] or a customized message form [18].

Fig. 4 shows the data and results for operating the overtaking decision algorithm of an autonomous vehicle linked to the radar road infrastructure. V_{Ego} is the velocity of the autonomous vehicle obtained from the vehicle, $V_{\text{Obs IDS}}$ is



Fig. 3. Radar road infrastructure detection area in the autonomous driving test bed.

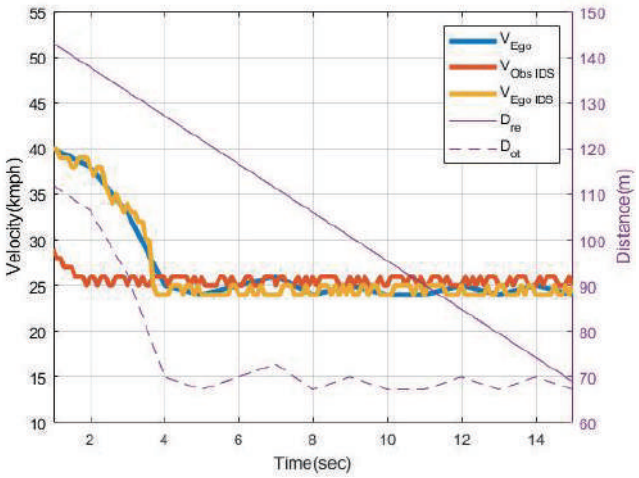


Fig. 4. Data and results for overtaking decision algorithm.

the velocity of the vehicle classified as a forward obstacle obtained by radar infrastructure, $V_{\text{Ego IDS}}$ is the velocity of the autonomous vehicle obtained by radar infrastructure, D_{re} is the remaining distance within the detection area, D_{ot} is the minimum distance required for overtaking. The area in which the data was collected has a standard speed of 50 km/h. However, due to the presence of a low-speed vehicle ahead, the autonomous vehicle is driven at a low speed of about 25 km/h. Since the speed of the autonomous vehicle is lower than the set standard speed, it corresponds to an algorithm operation, and after the condition is established, the minimum overtaking distance which is calculated based on radar road infrastructure data, and the infrastructure detection distance are compared to finally determine the behavior of the autonomous vehicle. An overtaking action is chosen. As can be seen in Fig. 4, overtaking was determined to be possible for up to 15 seconds. After 15 seconds, the autonomous vehicle continued to drive. but overtaking could no longer be chosen because the detection of the radar road infrastructure was limited. Advancements in autonomous driving functions linked to road infrastructure clearly contribute to increasing safety. However, it is difficult

to guarantee complete reliability, because unexpected variables exist on the road. Therefore, the safety of autonomous vehicles must be improved as much as possible, and functional research on fully autonomous vehicles must continue.

IV. Conclusion

This paper has proposed an overtaking decision algorithm for autonomous vehicles linked to radar road infrastructure. Our proposed overtaking decision algorithm provided performance verification results using autonomous vehicle and radar road infrastructure data that was collected through autonomous driving testbed testing. We expect that technology for fully autonomous vehicles can be realized by continuing to advance autonomous driving functions linked to road infrastructure.

Acknowledgment

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE), Korea, under the "Regional Industry Infrastructure and R&D Support Program" (P0018599) supervised by the Korea Institute for Advancement of Technology (KIAT).

References

[1] C. Badue, R. Guidolini, R. V. Carneiro, *et al.*, "Self-driving cars: A survey," *Expert systems with applications*, vol. 165, p. 113 816, 2021.

[2] I. Barabás, A. Todoruţ, N Cordoş, and A. Molea, "Current challenges in autonomous driving," in *IOP conference series: materials science and engineering*, IOP Publishing, vol. 252, 2017, p. 012 096.

[3] M. Reda, A. Onsy, A. Y. Haikal, and A. Ghanbari, "Path planning algorithms in the autonomous driving system: A comprehensive review," *Robotics and Autonomous Systems*, vol. 174, p. 104 630, 2024.

[4] S. Dixit, S. Fallah, U. Montanaro, *et al.*, "Trajectory planning and tracking for autonomous overtaking: State-of-the-art and future prospects," *Annual Reviews in Control*, vol. 45, pp. 76–86, 2018.

[5] S. S. Lodh, N. Kumar, and P. K. Pandey, "Autonomous vehicular overtaking maneuver: A survey and taxonomy," *Vehicular Communications*, p. 100 623, 2023.

[6] J. Kim, D. S. Han, and B. Senouci, "Radar and vision sensor fusion for object detection in autonomous vehicle surroundings," in *2018 tenth international conference on ubiquitous and future networks (ICUFN)*, IEEE, 2018, pp. 76–78.

[7] C. Badue, R. Guidolini, R. V. Carneiro, *et al.*, "Self-driving cars: A survey," *Expert systems with applications*, vol. 165, p. 113 816, 2021.

[8] H. U. Ahmed, Y. Huang, P. Lu, and R. Bridgelall, "Technology developments and impacts of connected and autonomous vehicles: An overview," *Smart Cities*, vol. 5, no. 1, pp. 382–404, 2022.

[9] P. Kopelias, E. Demiridi, K. Vogiatzis, A. Skabardonis, and V. Zafiropoulou, "Connected & autonomous vehicles–environmental impacts–a review," *Science of the total environment*, vol. 712, p. 135 237, 2020.

[10] A. M. Abdelgader and W. Lenan, "The physical layer of the ieee 802.11 p wave communication standard: The specifications and challenges," in *Proceedings of the world congress on engineering and computer science*, vol. 2, 2014, pp. 22–24.

[11] S. Chen, J. Hu, Y. Shi, *et al.*, "Vehicle-to-everything (v2x) services supported by lte-based systems and 5g," *IEEE Communications Standards Magazine*, vol. 1, no. 2, pp. 70–76, 2017.

[12] M. H. C. Garcia, A. Molina-Galan, M. Boban, *et al.*, "A tutorial on 5g nr v2x communications," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1972–2026, 2021.

[13] Y. Liu, M. Tight, Q. Sun, and R. Kang, "A systematic review: Road infrastructure requirement for connected and autonomous vehicles (cavs)," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1187, 2019, p. 042 073.

[14] H. Farah, S. M. Erkens, T. Alkim, and B. van Arem, "Infrastructure for automated and connected driving: State of the art and future research directions," *Road vehicle automation 4*, pp. 187–197, 2018.

[15] G. Asaithambi and G. Shravani, "Overtaking behaviour of vehicles on undivided roads in non-lane based mixed traffic conditions," *Journal of traffic and transportation engineering (English edition)*, vol. 4, no. 3, pp. 252–261, 2017.

[16] S.-S. Shin, H.-J. Kang, and S.-J. Kwon, "A study on data analysis for improving driving safety in field operational test (fot) of autonomous vehicles," *Machines*, vol. 10, no. 9, p. 784, 2022.

[17] C. Hedges and F. Perry, "Overview and use of sae j2735 message sets for commercial vehicles," SAE Technical Paper, Tech. Rep., 2008.

[18] J. Kim, T. Kim, W. Kim, D.-W. Lee, T.-L. Kim, and S.-J. Kwon, "A study on the integrated v2x communication data collection for data storage system of automated driving," *Transactions of the Korean Society of Automotive Engineers*, vol. 31, no. 12, pp. 1003–1010, 2023.

Analysis of Safety Regulation Violations by Electric Kick Scooter Users

Chang-Ryul Park
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
blueskyaoa7@gmail.com

Gangmin Lee
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
ganmin1003@gmail.com

Yukang Kim
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
yukang7376@gmail.com

JaeYeong Lee
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
jaeyeong221@gmail.com

Dae-Won Jeong
CEO
401SPAG Co., Ltd.
Seongnam, South Korea
dw@401spag.com

Kyungyong Chung
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
dragonhci@gmail.com

***Abstract*— In recent years, the urban traffic environment has been rapidly evolving. The widespread adoption of personal modes of transportation, such as electric kick scooters, has significantly influenced our mobility patterns. However, these advancements have also introduced new traffic safety challenges, prompting increased calls for stricter safety regulations. This study investigates the violations of safety regulations by electric kick scooter users through image analysis using YOLOv5 and OpenCV. By analyzing these violations, this research aims to enhance the monitoring and enforcement of safety regulations, contributing to improved urban traffic safety.**

Keywords—YOLO, OpenCV, electric kick scooter, identification

I. Introduction

In recent years, the urban traffic environment has been rapidly changing. In particular, with the spread of shared means of transportation such as electric kick scooters, our way of transportation has been innovatively changing. However, these advancements are accompanied by new safety issues. In particular, the rapid spread of electric kick scooters poses new challenges to road safety. According to recent studies, the accident rate of electric kick scooters is increasing rapidly, and the number of violations of safety regulations is also rising[1].

For safety, electric kick scooter users must comply with regulations such as wearing a helmet, driving on roadways, and boarding a single person. However, in reality, there are many instances of violating these safety regulations. Although efforts are being made to enforce these regulations, judicial authorities also face challenges. Unlike vehicles equipped with surveillance cameras, monitoring kick scooters is difficult because it requires direct enforcement. Additionally, there is a website for reporting safety regulation violations, but it handles various types of violations. This can cause delays in response, and it is often unclear where to report specific issues, such as safety reports, crime prevention, and public safety proposals. The reporting process can also be cumbersome due to the amount of required information[2].

This paper proposes a program that analyzes images reported as safety regulation violations using YOLOv5, and OpenCV and analyzes and informs committed violations of electric kick scooter users. Through this, a new technical solution will be presented to improve road safety and manage the traffic environment more safely and efficiently.

The proposed program contributes to surveilling users in violation of safety regulations at the same time as providing violation information by analyzing violations with a few images uploaded for reporting. This means that a witness who observes a safety regulation violation takes a picture and uploads an image through the web, and the server analyzes the image through YOLOv5 and OpenCV. This program analyzes whether safety regulations, such as helmet usage, are being followed by electric kick scooter users. In addition, it reads the Exif (Exchangeable image file format) data included in the image file and extracts image shooting time and location information. This has the advantage of being able to easily analyze and supervise violations of safety regulations with a few images of electric kick scooter users, who have been directly enforced by judicial authorities. In addition, when uploading an image to the web, the accuracy of the analysis can be further improved through simple options.

The structure of this paper is as follows: Chapter 2 discusses how to analyze images. Chapter 3 describes the technology of analyzing images uploaded through web pages from servers through YOLO and OpenCV, and the technology of extracting Exif data from image files. Finally, Chapter 4 concludes this paper.

II. Related Work

Recent research is advancing in the fields of object detection and image analysis, aiming for innovative developments. In particular, techniques such as YOLO (YOLOv5) and OpenCV have outstanding performance in object detection and classification in many studies, such as H. Deshpande [3]. YOLO is a deep learning algorithm that detects and classifies objects in real-time, and OpenCV is a library widely used for computer vision and image processing.

Many studies have shown that OpenCV alone can be used for object detection. However, this study aims to combine the YOLO algorithm, which is optimized for real-time object detection and can detect objects at very high speeds. With OpenCV, which supports various color space conversions and can transform images into multiple color spaces such as RGB, HSV, and LAB, this can enhance performance and speed in object detection while improving accuracy and efficiency in both object and color detection.

A. Image Analysis with YOLOv5

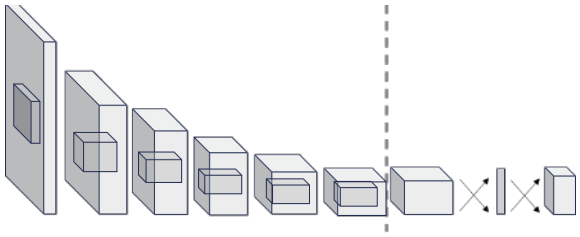


Fig. 1. YOLO algorithm structure

Fig. 1 shows the overall structure of YOLO. It is a structure that receives an image of 448x448 and finds out the location and identity of an object in the image through multiple layers. One model achieved object detection. The model that performs object detection is largely divided into two structures: the left part of Figure 1 is the Backbone and the right part is the Head. Backbone plays the role of extracting the characteristics of the input image and the head plays the role of object detection by receiving a characteristic map from which the characteristics are extracted. As a result of examining recent research trends in drone detection technology, B. Aydin[4] confirmed that the YOLO algorithm outperforms other object detection technologies. YOLO is used for many object detection tasks with fast detection and excellent accuracy, and various versions of YOLO have been used for drone detection. Through the research, YOLOv5 has led to higher accuracy and performance improvement on drone datasets compared to previous versions of YOLO.

B. Image Analysis with OpenCV

The Open Source Computer Vision Library (OpenCV) is an open-source library used to perform computer vision and image processing tasks. OpenCV is widely used in the field of computer vision by providing various functions such as image processing, object detection, tracking, machine learning, and video processing. For example, OpenCV may be used in the study of developing autonomous vehicle technology. It may be used to support autonomous driving by detecting and analyzing the surrounding environment through a camera. B. Bavitha[5] implements the function of recognizing colors through Red, Green, and Blue values of the image by utilizing the RGB color model of the OpenCV and detecting, and tracking objects recognized by colors in real-time through the HSV color model. This function has been implemented due to OpenCV supports various images/vision and real-time processing, and at the same time is an open source in a form that may be easily utilized. As another possible field application example, there is color-based object classification in industrial robots. Fig. 2 shows one of the methods of expressing color detection through OpenCV.

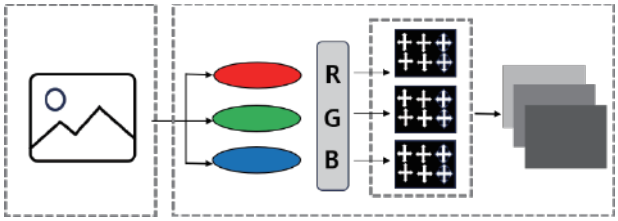


Fig. 2. The RGB model Color processing in OpenCV

Fig. 2 shows how to process the colors of an image using the RGB model. The RGB model expresses the color to be represented by mixing the three primary colors of light, red, green, and blue, in various ways. In OpenCV, the image is displayed as a matrix of PxQx3 pixels, and then the colors in the BGR order are represented by values between 0 and 255 for each pixel value. As the value increases, the light of the corresponding color becomes brighter.

III. Analysis of Safety Regulatory Violations and Extraction of Analysis through Algorithms

Fig. 3 shows the process of analyzing safety regulation violations in the images uploaded using YOLO and OpenCV and providing what the administrator needs.

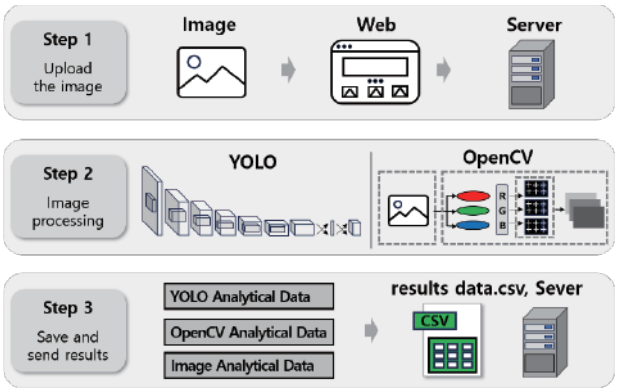


Fig. 3. Image analysis and analysis extraction process

Witnesses who observe a situation that violates safety regulations take a picture and upload it to a web page. The uploaded image is moved to a server and analyzed through YOLO and OpenCV. In addition, Exif data contained in an image file is extracted, and all information that can be obtained in the situation at the time of the safety regulation violation is analyzed, extracted, and provided to the administrator. Fig. 4 shows the appearance of the implemented web page.

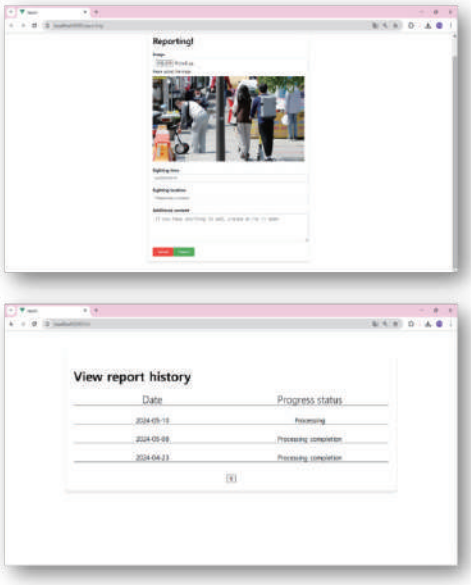


Fig. 4. The top web page (a) receives a report, and the bottom web page (b) inquires about a report

On the upper (a) web page, there is a place where you can upload a picture of the electric kick scooter user who violated safety regulations at the time of the incident. In addition, it consists of a place where you can write down the date, time, and location information at the moment of the violation in case no Exif data is included within the image file. On the web page below (b), you can check the progress of the contents reported by the reporter through the web page as being processed and completed.



Fig 5. Objects detected with YOLO v5

A. Detection of objects in an image using the YOLO algorithm

P. k. Yadav [6] used the confidence score to evaluate the performance of YOLOv5 that detects cotton growing voluntarily in a corn field. Based on this idea, we intend to divide the confidence score, with 0.8 as the threshold, depending on whether or not an electric kick scooter user wears a helmet. Figure 5 shows the process of analyzing the image through YOLO. Figure 6 illustrates the process of image analysis using YOLO conducted in this study.

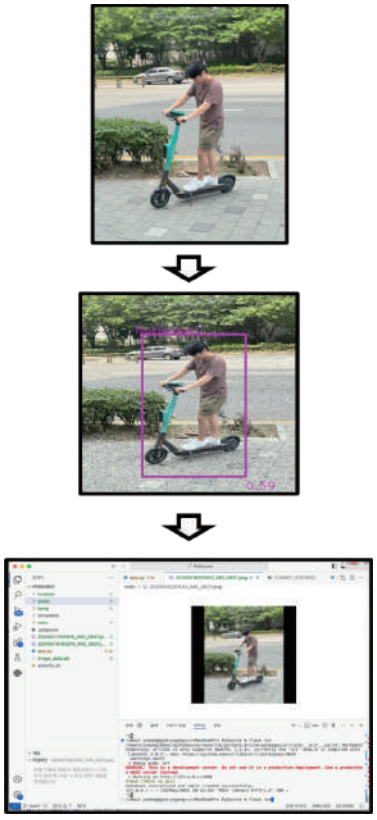


Fig. 6. YOLOv5 based analysis process.

The uploaded image is transmitted to the server through a web and analyzed using YOLOv5. First, check if the electric kick scooter and the user are detected as objects in the image. After that, it is determined whether the electric kick scooter user is wearing a helmet. This analysis result is provided to the administrator, and as shown in Figure 6, the analyzed content of the image, along with location and timestamp information from the image file's Exif data, is stored in the administrator's database. This allows the administrator to verify the analyzed image, helmet usage evaluation, the location and time when it was captured, with an accuracy rate of 0.83.

B. How to detect the color of the electric kick scooter with OpenCV

N. Wang [7] conducted object recognition through YOLO and color recognition through OpenCV to recognize greenhouse cucumber fruits based on color space characteristics. In this way, we proceed with object recognition on the kick scooter with YOLOv5 and match the RGB color of the kick scooter with the company's representative color through OpenCV to identify the company of the reported electric kick scooter. OpenCV's approach to infer the electric kick scooter color consists of several steps. First, the boundary box of the electric kick scooter object is extracted and the corresponding area is separated from the image. Then, K-means clustering is used to extract the color information of pixels.

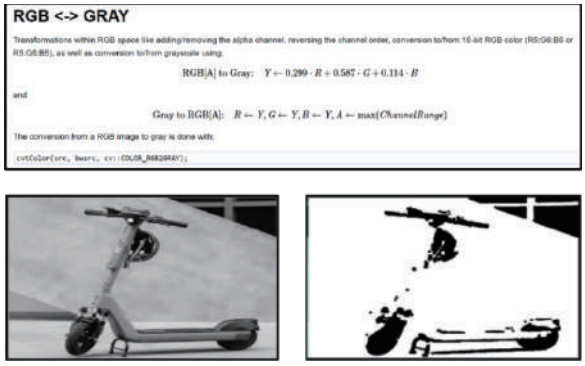


Fig. 7. OpenCV's grayscale process

First, there is the pre-processing stage. As depicted in Fig. 7, the image being analyzed is converted to grayscale, and the outline of the object is extracted to narrow down the color detection range. Next, selecting the number of clusters. The number of major colors to be extracted from the image is determined, and the color of the object is set as the initial central value, taking into account the color of each company. Next, each pixel is assigned to the nearest cluster using the Euclidean distance measurement formula. The center of the cluster is adjusted by updating the representative value of each cluster with the average value of the pixels belonging to the cluster. The allocation and update steps are repeated to continuously update the cluster representative values until there is no change in the representative values. Finally, the clustered representative color value is generated, and the color is determined by comparing it with the color of the electric kick scooter for each company.

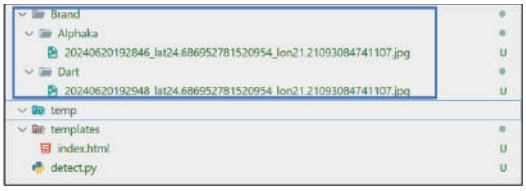


Fig. 8. Color-based company classification

Figure 8 shows the results of the color detection in the research system using OpenCV, where the detected colors are matched with companies and classified accordingly. If the color detection is performed in low resolution, the image of the electric scooter is enlarged, and noise is removed using Gaussian Blur. After that, the range of color detection is narrowed down using contouring, and the color value is found using k-means clustering, mapping it to the company's color within an acceptable error range. However, in such cases, the accuracy drops below 0.3, indicating that further research is needed.

C. *Extracting Exif data from an image file*

In order to extract Exif data from an image file, Python's PILLOW library should be used. First, an image file is opened using the image module of the PIL, and the _getexif() method is called to get the EXIF data of the image. First, it is a method of extracting GPS information. In the EXIF data, GPS information is stored in the key '34853' ('GPSInfo'). This key can be used to get GPS information. In the GPS information dictionary, latitude and longitude are stored as keys 2 and 4, respectively, with their references (north/south and east/west) are stored as keys 1 and 3, respectively. Using this information, it is possible to calculate the actual latitude and longitude values. Latitude and longitude are stored in the form of degrees, minutes, and seconds, and they are processed by using the function converter_to_degrees that converts them into decimal. By checking the latitude and longitude references, the final value is converted into negative numbers in the case of the south latitude or the west longitude. Next, there is a method of extracting date and time information. The date and time of the original photographing may be extracted using the 'DateTimeOriginal' key (key number 36867) from the Exif data. If the key does not exist, a 'DateTime' key (key number 306) may be identified. The extracted date and time information will be stored in the form of a character *string*, and be converted into a Python data time object if necessary. This will make more useful use of date and time information.

IV. Conclusion

This paper proposes a method to enforce regulations on electric kick scooter users who violate safety regulations through image analysis. The proposed approach utilizes YOLO and OpenCV to analyze images, specifically focusing on detecting whether helmets are worn and matching information such as the company of the electric kick scooter used by violators. The analyzed results are communicated to administrators, providing them with details such as helmet usage status, the company of the electric kick scooter used, date, and location. Through this study, it becomes possible to report electric kick scooter users based on a few photos and analyze regulatory violations. Police can effectively enforce regulations on electric kick scooter users using data provided by the server, contributing to maintaining road safety.

In future studies, the following plans are scheduled to be implemented. First, additional analysis technology will be added to detect passengers riding together. Analyzing whether vehicles adhere to the safety regulation of single occupancy can strengthen enforcement efforts. Second, technology for analyzing whether electric kick scooters are driven on sidewalks will be incorporated. This technology aims to analyze the regulation of electric kick scooter usage based on safety regulations that vary depending on whether they are on sidewalks or bike lanes. Lastly, using Raspberry Pi to detect and analyze violators of safety regulations automatically, substituting for informants. Raspberry Pi, a small and affordable computer, will establish an environment capable of running camera modules, YOLO, and OpenCV, enabling real-time monitoring of road conditions and detection of violators of safety regulations. This system aims to address issues of accuracy in object detection due to resolution problems in photos taken by eyewitnesses. It is expected to contribute to improving road safety and encouraging compliance with traffic regulations.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation)"(2021-0-01393).

References

[1] S. G. Ko, (2023, June 18), Electric kick scooter accidents surge, 526 cases in northern Gyeonggi Province alone... cardiac arrest transfer. THE FACT. https://news.tf.co.kr/read/national/2024825.htm

[2] Safety e-Report, www.safetyreport.go.kr/.

[3] H. Deshpande, A. Singh, and H. Herunde. "Comparative analysis on YOLO object detection with OpenCV." International journal of research in industrial engineering, Vol.9, Issue.1, pp.46-64, March 2020

[4] B. Aydin, "Drone Detection Using YOLOv5", MDPI Eng 4, no.1, pp. 416-433. February 2023.

[5] B. Bavitha, R.Navya, M.Haritha and K.H.Grace, "COLOR IDENTIFICATION", ANITS., Visakhapatnam, AP., India, ICSMEC21-0103, 2021

[6] P. k.Yadav, "Assessing the performance of YOLOv5 algorithm for detecting volunteer cotton plants in corn fields at three different growth stages", Artificial Intelligence in Agriculture, vol. 6, no. 25, pp. 292-303, Nov. 2022.

[7] N. Wang, "An Enhanced YOLOv5 Model for Greenhouse Cucumber Fruit Recognition Based on Color Space Features", Agriculture, vol. 12, no. 10, p.1556 , Sep. 2022.

Beam Propagation Method Simulation of Uncoated Coreless Silica Fiber for Humidity Sensing

Athi Shankhara Velusamy
Malaysian-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia Kuala Lumpur
54100 Kuala Lumpur, Malaysia
athishankhara@graduate.utm.my

Muhamad Yusof Mohd Noor
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru,Malaysia
yusofnor@utm.my

Siti Mahfuza Saimon
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru,Malaysia
sitimahfuza@utm.my

Fauzan Ahmad
Malaysian-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia Kuala Lumpur
54100 Kuala Lumpur, Malaysia
fauzan.kl@utm.my

Abstract— The use of coated Coreless Silica Fiber (CSF) humidity sensors presents significant challenges, including increased costs and complications during the fabrication process due to the inherent limits and nature of foreign materials. In response to these issues, this paper proposes a CSF humidity sensor design without any coating. We introduce a simulation of a bare CSF design for sensing humidity and evaluate the sensor's performance by simulating changes in the analyte refractive index, ranging from 1.00 RIU (0% RH) to 1.33 RIU (100 %RH). Our findings demonstrate that the uncoated CSF design effectively detects humidity, with a sensitivity of 313.1 x 10^{-9} RIU per unit of RI shift and a strong linear relationship, evidenced by an R^2 value of 0.987. These results establish a foundation for potential future design modifications to further enhance the performance of the uncoated CSF humidity sensor. Furthermore, the proposed sensor design holds significant potential for maritime IoT applications, such as monitoring humidity levels within ship cargo holds to prevent damage to sensitive goods.

Keywords—CSF, SMF, MFD, RI, Fiber, Rsoft, BeamPROP, Humidity Sensing, Analyte medium

I. INTRODUCTION

CSF is a crucial component of optical fiber sensors, utilized in various applications such as sensing and material conversion [1]. Unlike multimode fiber (MMF), which has a core with a higher refractive index than its cladding, CSF lacks a distinct core, allowing light to interact more freely with its surroundings and external materials [2]. This unique characteristic enables CSF to serve as a transition material between fibers with different properties, as a splice material in fiber lasers and amplifiers to reduce splice losses, and in sensing applications [3, 4]. Originally developed to overcome the limitations of MMF, CSF's ability to modify light propagation parameters makes it highly effective for these diverse applications [5,6].

Optical humidity fiber sensors offer several advantages over traditional electronic sensors, including reduced weight and size, resistance to electromagnetic interference, multi-channel and remote operation, and corrosion resistance [2,3]. Most fiber humidity sensors, including those based on CSF, use hygroscopic thin films or layers for humidity sensing in various configurations. To enable humidity sensing, a sensitive material such as hydrogel or polymer [7,8,9] is coated on the CSF. CSF-based humidity sensors are compact, low-cost, highly sensitive, and have a simple structure.

However, the main problem is that the coated sensitive material makes the fabrication of coated CSF more complex and costly, and introduces high temperature cross-sensitivity. To address these issues, a CSF humidity sensor without any coating could provide an effective solution. Bare CSF, without any coating, makes the sensor insensitive to temperature fluctuations, enhancing its reliability and accuracy. This characteristic is particularly beneficial in maritime environments, where temperatures can vary widely. The uncoated CSF design ensures accurate and consistent data collection for applications such as cargo hold monitoring, structural health monitoring of ships, and environmental monitoring in maritime operations, making it an excellent choice for critical applications in the maritime industry.

In this paper, we simulate a CSF configured as a Multimode Interference (MMI) structure, with the input and output spliced to Single Mode Fiber (SMF) to sense humidity. To detect humidity without the use of any sensitive humidity material, this study examines the sensitivity performance of the bare CSF fiber. We evaluate its humidity sensing abilities by analyzing the electric field distribution throughout the sensor and the interaction between evanescent waves and effective refractive index (RI) changes in various analyte RI mediums, ranging from 1.00 RIU (0 %RH) to 1.33 RIU (100 %RH). To the best of our knowledge, this is the first publication that simulates bare CSF as a fiber humidity sensor. Furthermore, this design offers significant potential for maritime IoT applications, such as monitoring humidity levels within ship cargo holds to prevent damage to sensitive goods.

II. METHODOLOGY

CSF which is employ in this experiment is illustrated in Fig. 1. This CSF consist of outer diameter of 125µm and length of 58000µm. Then, analyte diameter or known as background index included. This background index plays an important role as a sensing medium. The essential notion in this CSF is that the fiber doesn't have a core in the inner diameter, hence this inner diameter shouldn't be confused with a core. For this reason, the 125µm smaller diameter is referred to as the

cladding diameter. In addition, as we added a SMF design, we cut a certain depth and width in the middle part of the CSF because to understand and view the spectrum of the CSF. Based on Fig. 1, there is a cross-section view of SMF in the CSF background index (analyte) in the cut-off region.

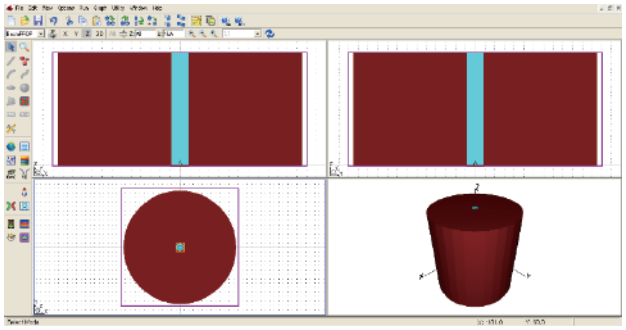


Fig. 1 The structure of CSF and SMF in cross-section geometry

The simulation for this study, a software called "BeamPROP" which is a device tool based on the Beam Propagation Method (BPM) under the software Synopsys Rsoft CAD is used.[10] In this software, we can design (2D) and (3D) but the reason why two-dimensional(2D) space was implemented as to simulate a sensing device because it saves time by giving the simulation output in a short period of time. Therefore, it is less complex to mesh and uses less computer CPU and memory when compared to three-dimensional (3D) simulation. Wave optics module and beam module was chosen for the electrical field study because its wavelength is comparable to and significantly smaller than that of the device under examination and for a forward and backward propagation. Cross section method (CSM) is used to examine the mode shape of the waveguide cross-section region and offers virtual information in the mode field diameter (MFD) [11]. It is possible to see the multiple electric field distribution of waveguides (EWFD) or power and assess the strength and penetration depth of the evanescent wave because of the interaction between the sensor and the various analyte medium values. Hence, the geometry settings of CSM of CSF are shown in Table 1. The CF diameter is set within 125µm, the length CF was set in 58000µm, and diameter of SMF core was set in 9µm. There were also some other parameter data added for future use, but these doesn't affect the simulation until is declared. In (BeamPROP), we add the symbol and values in the "Symbol Table Editor" and the "Name" is known as "Domain" and "Expression" is known as "Dimension". Hence, we must put the values and accept the symbol values to start the simulation. Table I shows the geometry settings of simulated model which were used.

TABLE I. GEOMETRY SETTINGS

Domain	Dimension
CF diameter	125µm
SMF diameter	9µm
CF length	58000µm
MMF core 1 diameter	62.5µm
MMF core 2 diameter	105µm
Dimension	3
Grid size	1
Grid size Y	1
Alpha	0
Step size	1nm

The RI was assigned to each domain of the drawn parameter by applying material settings. To express the RI of CSF in the near-infrared spectrum, a three term Sellmeier equation was applied as expressed in Eq. (1) [12].

$$\frac{0.6961663\lambda^2}{\lambda^2-(0.0684043)^2}+\frac{0.4079426\lambda^2}{\lambda^2-(0.1162414)^2}+\frac{0.8974794\lambda^2}{\lambda^2-(9.896161)^2}=n^2-1 \quad (1)$$

Based on the equation Eq. (1), n represents the CSF and λ represents the operating wavelength. Hence, the RI mediums are set to isotropic medium. Technically, the CSF and the analyte are mostly made up of glass and liquid which possess one RI in all directions. The λ known as operating wavelength is set to 1500 nm to 1600nm in a practical fiber. The applied parameter values are tabled in Table II [13].

TABLE II. MATERIAL SETTINGS

Parameter	Value
CSF RI	1.449 RIU
Analyte RI	1.00 RIU – 1.20 RIU
Medium RI	Isotropic
Wavelength λ	1500nm – 1600nm

The length of the self-image in a fiber sensor can be customized to fit the desired self-image index. Previous studies reported four self-image indices for CF lengths. The CF self-image lengths were calculated using a fixed CF with a diameter of 125µm and an RI of 1.4440 RIU. The calculated values were rounded to the closest tenth to facilitate measurements. The output intensity of the sensor was optimized prior to exposure to a high-RI environment. Optimal sensor performance was attained with an even self-image index, specifically the second index showing high sensitivity. This research aims to conduct simulations and experiments focusing on the length of fiber with the second self-image index to further investigate and validate sensor performance with respect to high RIs. The optimal sensor length selected was 5.9 cm for easy handling fiber fabrication and higher output power.

TABLE III. CF SELF-IMAGE LENGTH CALCULATED AT VARIOUS SELF-IMAGE INDICES

Index of self-image	Length of CF self-image (cm)	Rounded-off Length (cm)
1	1.4556	1.5
2	2.9112	2.9
3	4.3669	4.4
4	5.8226	5.9

Sensing ability, often referred to as detection ability, was applied to the evanescent wave interaction between the sensor and analyte medium. The variations in the RI distribution are sensed or detected by the evanescent field. This is caused by the analyte medium, which effects the effective mode index, n_{eff} of the sensor. The "Penetration Depth" or known as dp represents the exponential decay from the evanescent wave with the distance in the analyte medium, which is expressed in Eq. (2).

$$\frac{\lambda}{2\pi\sqrt{n_{eff}^2\sin^2\theta-n_a^2}}=d_p \quad (2)$$

Based on the equation Eq. (2), n_a represents the analyte RI and λ represents the operating wavelength. The d_p values change with the change in na which makes and alloys random value of analyte medium detection. The sensor sensitivity, S can be expressed as in Eq. (3), which can be obtained from the slope of the graph.

$$\frac{\Delta n_{eff}}{\Delta n_a} = S \tag{3}$$

In equation Eq. (3), p represents the self-image number, L represents the CF self-image length, n_{NCF} represents the RI_{NCF} , D_{NCF} represents the CF diameter, and $\lambda_{dip}/_{peak}$ represents the dip or peak wavelength. The length of the self-image can be customized to fit the desired self- image index. A calculation was performed to estimate the optimal self-image length for the CF using an equation based on the zeroth-order approximation (14).

$$L = \frac{p.n_{NCF}.D_{NCF}^2}{\lambda_{dip}/_{peak}} \tag{4}$$

III. Results and Discussion

In Fig. 2, it shows electric field intensity thru in the SMF in the CF. Hence, we get to know that CF allows multiple modes of light to propagate because of its large size in diameter. In addition, there were total of 200 steps of wavelengths because from the 1500nm to 1600nm which makes a larger data taken and more understanding in spectrum waves. But in this study, we just focus on the fundamental mode for the ease of evaluation. In a selective study, the CSF with different analyte medium values with 1.00 RIU to 1.20 RIU with its respective spectrum are illustrated through red to blue colour and from bottom to top, which represents the maximum and minimum electric field intensity. In general, a CSF structure, the electric field focuses on the centre and equally dispersed which shows the light is traveling or constrained in the fiber. In Fig. 2, we can see the light travel from bottom to top part. This explains that the spectrum travelled in the SMF and to the output of CSF with acceptable power output.

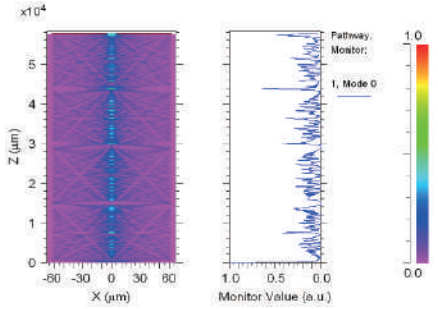


Fig. 2 Shows a travel of spectrum from bottom to top of the SMF in the CF

The spectrum also known as power distribution of the CSF can be seen and analyse by studying MFD from 1.00 RIU to 1.20 RIU as shown from the Fig 3. It displays the MFD which stands for power towards various analyte RIs ranging between 1.00 RIU to 1.20 RIU. The RI difference between the CSF and the analyte medium determines the penetration depth. The RI difference between the CSF and the analyte medium determines the penetration depth.

This shows that the bare or uncoated CSF can detect the refractive index (RI) of an analyte medium within the range of 1.00 to 1.20 RIU. Hence, the light intensity at 1.20 RIU is higher than at lower RI values light intensity, which can attribute to the absorption characteristics or the nature of the RI at 1.20 RIU, which is more towards liquid states with higher moisture content rather than those below 1.20 RIU, which are more like air which has lesser moisture content.

This states that the sensor can effectively detect humidity since the relative humidity (RH) increases, the air particles surrounding the sensing region, which acts as cladding layer, become more saturated with water molecules [15,16]. Consequently, this saturation leads to an elevation in the RI of the surrounding air from 1.00 towards the water's RI of 1.20, essentially the signalling changes in different humidity levels.

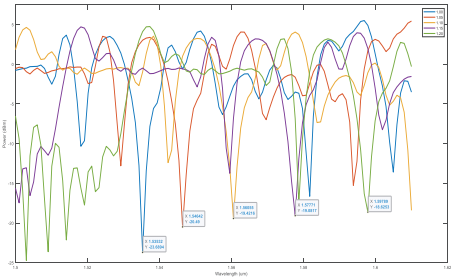


Fig. 3 Shows a spectrum from the waveform between 1500nm till 1600nm.

Based on Fig. 4, the RIU sensor's computed sensitivity of CSF is 313.1 x 10^{-9} which has a high degree of linearity with R^2 value of CSF at 0.987 which is an optimal linearity value. [17,18] This demonstrates that the penetration depth will rise in proportion to the height of the evanescent wave. Hence, the overall sensitivity of the sensor will also increase. As a result, the sensor's overall sensitivity will rise. In Fig. 3, the shift signal has already entered the band over 1600nm, which is not predicted and has crossed the desired limit. Initially, we should test up to RI 1.35 RIU, but as it entered RI 1.25 RIU. Thus, we began at RI 1.00 RIU and continued to RI 1.20 RIU. However, some overlapping or evanescent waves continue to exist at 1620nm.

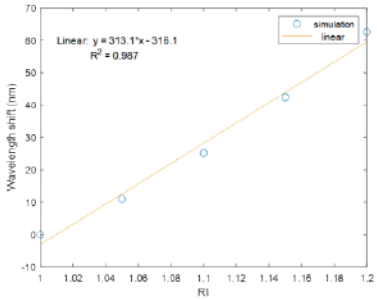


Fig. 4 Shows difference linear of wavelength shift (nm) between RI

The figure 5 shows the mode patterns in a 125 μm sensing fiber with different lengths of coreless fiber (CF) as 15 cm, 29 cm, 44 cm, and 59 cm. Within the 1500-1600 nm waveform, the 59 cm CF exhibits distinct characteristics, suggesting better humidity sensing. Its power spectrum fits

the waveform bandwidth, unlike the other CSF lengths which have broader spectra exceeding 1600 nm. This indicates the 59 cm CSF may be more effective for humidity sensing in the specified range.

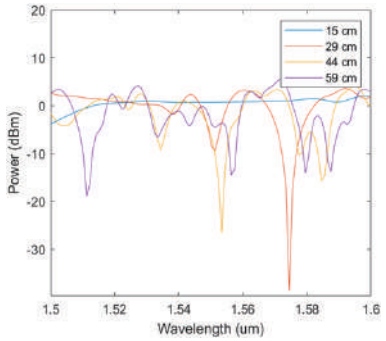


Fig. 5 Shows difference linear of wavelength shift (um) between power

IV. Conclusion

In this study, we addressed the challenges associated with coated CSF humidity sensors, such as increased costs and fabrication complexities. We proposed an innovative design for a CSF humidity sensor without any coating and conducted simulations to evaluate its performance. The simulation results indicated that the bare CSF design can effectively detect humidity, demonstrating a sensitivity of 313.1 x 10^{-9} RIU per unit of refractive index shift and a strong linear relationship with an R^2 value of 0.987. These promising results lay the groundwork for future design modifications to further enhance the sensor's performance. Additionally, the uncoated CSF humidity sensor has a significant potential for maritime IoT applications, particularly in monitoring humidity levels within ship cargo holds to prevent damage to sensitive goods. The insensitivity to temperature fluctuations further enhances its reliability and accuracy in maritime environments.

V. Acknowledgement

This work was supported by the Fundamental Research Grant Scheme (FRGS), grant number FRGS/1/2023/TK07/UTM/02/20 funded by the Ministry of Higher Education (MOHE), Malaysia.

References

[1] Zhang, Ya-nan, Lebin Zhang, Bo Han, Peng Gao, Qilu Wu, and Aozhuo Zhang. "Reflective mercury ion and temperature sensor based on a functionalized no-core fiber combined with a fiber Bragg grating." Sensors and Actuators B: Chemical 272 (2018): 331-339.

[2] Xu, W.; Shi, J.; Yang, X.; Xu, D.; Rong, F.; Zhao, J.; Yao, J. Relative Humidity Sensor Based on No-Core Fiber Coated by Agarose-Gel Film. *Sensors* 2017, *17*, 2353.

[3] Zhao, Yong, Jian Zhao, and Qiang Zhao. "Review of no-core optical fiber sensor and applications." Sensors and Actuators A: Physical 313 (2020): 112160.

[4] O. Laccourreye, H. Maisonneuve,French scientific medical journals confronted by developments in medical writing and the transformation of the medical press,European Annals of Otorhinolaryngology, Head and Neck Diseases,Volume 136, Issue 6,2019,Pages 475-480,ISSN 1879-7296.

[5] Baharin, N. F., Azmi, A. I., Abdullah, A. S., & Mohd Noor, M. Y. (2018). Refractive index sensor based on lateral offset of coreless silica interferometer. Optics & Laser Technology, 99, 396–401.

[6] Sun, Bing, Fei Li, Kandi Xu, Kaiming Zhou, and Zuxing Zhang. "Temperature-insensitive fiber-optic refractometer based on immobilization of polydimethilsiloxane film." IEEE Photonics Technology Letters 34, no. 3 (2022): 165-168. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[7] [11] Younus, Shahad I., Anwaar A. Al-Dergazly, and A. K. Abass. "Characterization of Multimode Interference Based Optical Fiber." In IOP Conference Series: Materials Science and Engineering, vol. 1076, no. 1, p. 012060. IOP Publishing, 2021.Shimura, Y.; Suzuki, Y.; Morisawa, M. Improving Measurement Range of a Swellable Polymer-Clad Plastic Fiber Optic Humidity Sensor by Dye Addition. Sensors 2022, 22, 6315.

[8] Y.L. Hoo, W. Jin, C. Shi, H.L. Ho, D.N. Wang, S.C. Ruan, Design and Modeling of a Photonic Crystal Fiber Gas Sensor, Appl. Opt. 42 (2003) 3509.

[9] Zhao, Yong, Jian Zhao, and Qiang Zhao. "Review of no-core optical fiber sensor and applications." Sensors and Actuators A: Physical 313 (2020): 112160.

[10] Stanislava Serecunova, Dana Seyringer, Heinz Seyringer, Frantisek Uherek, Tomas Mizera, Dusan Pudis; Design and simulation of polymer based 1x4 multimode interference splitter.AIP Conf. Proc. 2 November 2021; 2411 (1): 060004.

[11] aofei, Yuchan Hu, Hongda Cheng, Feng Yan, Qianyu Lin, Yu Chen, Pengjun Wu et al. "Side-polished singlemode-multimode-single-mode fiber structure for the vector magnetic field sensing." Journal of Lightwave Technology 38, no. 20 (2020): 5837-5843. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[12] [11] Younus, Shahad I., Anwaar A. Al-Dergazly, and A. K. Abass. "Characterization of Multimode Interference Based Optical Fiber." In IOP Conference Series: Materials Science and Engineering, vol. 1076, no. 1, p. 012060. IOP Publishing, 2021.Shimura, Y.; Suzuki, Y.; Morisawa, M. Improving Measurement Range of a Swellable Polymer-Clad Plastic Fiber Optic Humidity Sensor by Dye Addition. Sensors 2022, 22, 6315

[13] Malitson, Ian H. "Interspecimen comparison of the refractive index of fused silica." Journal of the Optical Society of America 55, no. 10 (1965): 1205-1209

[14] Xiuxin Liu, Xinpu Zhang, Yun Liu et al., "Multi-point fiber-optic refractive index sensor by using coreless fibers", Optics Communications 365, pg. 168, (2016); doi: 10.1016/j.optcom.2015.10.07

[15] A. Thakallapelli, S. Ghosh and S. Kamalasadan, "Real-time frequency based reduced order modeling of large power grid," 2016 IEEE Power and Energy Society General Meeting (PESGM), *Boston, MA, USA, 2016, pp. 1-5.*

[16] Zhao, Yong, Jian Zhao, and Qiang Zhao. "Review of no-core optical fiber sensor and applications." Sensors and Actuators A: Physical 313 (2020): 112160.

[17] Chai, Quan, Hyeonwoo Lee, Seongjin Hong, Yongsoo Lee, Junbum Park, Jianzhong Zhang, and Kyunghwan Oh. "Nanoliter liquid refractive index sensing using a silica V-groove fiber interferometer." Photonics Research7, no. 7 (2019): 792-797. I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[18] K. Suzuki, H. Kubota, S. Kawanishi, M. Tanaka, M. Fujita, Optical properties of a low-loss polarization-maintaining photonic crystal fiber., Opt. Express. 9 (2001) 676–680.

Applicability of block pavement in a harbor based on deflection and settlement characteristics

Yong-Soo Ha
Maritime ICT & Mobility Research
KIOST
Busan, South Korea
ysha@kiost.ac.kr

Yun-Tae Kim
Department of Ocean Engineering
Pukyong national university
Busan, South Korea
yuntkim@pknu.ac.kr

Minh-Vuong Pham
Department of Ocean Engineering
Pukyong national university
Busan, South Korea
vuongpm@dau.edu.vn

Haeyong Park
Maritime ICT & Mobility Research
KIOST
Busan, South Korea
hypark@kiost.ac.kr

Sang Ki Jung
Maritime ICT & Mobility Research
KIOST
Busan, South Korea
jeongsk313@kiost.ac.kr

Myoung Hak Oh
Maritime ICT & Mobility Research
KIOST
Busan, South Korea
omyhak@kiost.ac.kr

***Abstract*— In ports, the pavement often experiences differential settlement and cracks due to soft ground and heavy equipment. This study explores two configurations of port blocks applied in a harbor to understand deflection and settlement characteristics. Falling weight deflectometer (FWD) tests were conducted on asphalt pavement and block pavements to assess deflection. A cement-treated base showed better bearing capacity and settlement management during harbor operations compared to asphalt pavement. LiDAR measurements detected multiple settlements in the crushed-stone base post-construction due to surface loads. The relative deflection ratio correlated with settlements. Settlements were minimal (less than 5 cm), indicating that block pavement, with its versatile application and the strength of the cement-treated base, can be effectively used in various harbor settings.**

Keywords*— *Concrete Block pavement, Port pavement, Deflection, Settlement

I. Introduction

Heavy equipment is frequently used in port areas situated on claimed soft ground, exerting significant axial loads that make the pavement susceptible to plastic deformation. Therefore, cases of pavement damage such as differential settlement have frequently occurred. Generally, asphalt pavement and concrete pavement are widely used as road pavement types. The block pavement method has the characteristics of concrete materials such as strength, durability, and lifespan, and has the advantage of overcoming the maintenance difficulties of existing concrete pavement. Various studies have employed falling weight deflectometer (FWD) tests to analyze the performance of existing pavements. However, a lack of analysis was conducted comparing new paving methods with existing ones and their long-term deflection and settlement characteristics. While LiDAR measurements, such as mobile mapping systems (MMS), were useful for generating spatial data, they typically focus on surface coordinates over time, necessitating geotechnical analyses to understand deflection and settlement patterns thoroughly.

II. Material and Method

A. Site and design

A port block (length of 250 mm, width of 120 mm, thickness of 160 mm), which are longer, wider, and thicker than a sidewalk block, was manufactured. The performance of the port block was verified through tests for the bending strength, absorbtion, freeze and thaw durablity, and surface skid resistance, and the manufactured block satisfied a compressive strength with 70 MPa and a bending strength with 7 MPa. Figure 1(b) shows block pavement at the port with a herringbone arrangement pattern using manufactured blocks and joints.

B. FWD test

In this study, to evaluate a quantitative performance considering the deflection characteristics on block pavement, the central surface deflection (D_0) and the deflection ratio (DR), which compares the relative amount of deflection with different types and times of the pavement, were used. The equation to calculate DR using the deflection at the same location is as follows.

$$\mathrm{DR}_{A,\ B} = \frac{D_A}{D_B} \tag{1}$$

where, $\mathrm{DR}_{A,\ B}$ is the deflection ratio at A compared to B, D_A is target deflection, D_B is comparable deflection.

C. LiDAR measurement

In this study, the settlement characteristics, that occur over time after construction, were measured for the entire block pavement including CSB and CTB. The elevation distribution of the point cloud according to time changes was compared, and the overall trend was evaluated. Multiscale model-to-model cloud comparison (M3C2), that performs direct comparison between points, was used to analyze the settlement characteristics that occur owing to the heavy equipment operation in the port [1]. DEM of difference (DoD) which creates a DEM by gridding and comparing two point clouds [2], simple and fast cloud-to-cloud comparison (C2C) that was not required gridding and meshing of the point clouds [3], and cloud-to-mesh distance (C2M) which uses distance between a point cloud and a 3D mesh [4] have individual errors, such as errors occurring in roughness, between adjacent point clouds, and due to comparison with the average value [2].

III. Results and Discussion

A. Deflection charactieristic

Figure 1 shows the DR in the horizontal direction for the CSB and CTB. The DR was calculated based on the deflection amount for the block pavement conducted at the same location with different times. The average DR of CSB was 1.65, which was relatively high at immediately after construction. After 7 months of construction, it was decreased to 1.21 and converged to 1, securing a bearing capacity equivalent to that of asphalt pavement. In CTB, a slightly larger deflection occurred immediately after construction compared to asphalt pavement, however, after 7 months of construction, the bearing capacity improved according to the compaction of the bedding sand. The DR measured 7 months after construction showed an average of 1.21 for CSB and 0.72 for CTB, and the bearing capacity of CSB and CTB was 0.74 and 1.68 times that of asphalt pavement. After 7 months of construction, the DR of CTB was 0.53 times lower than that of CSB.

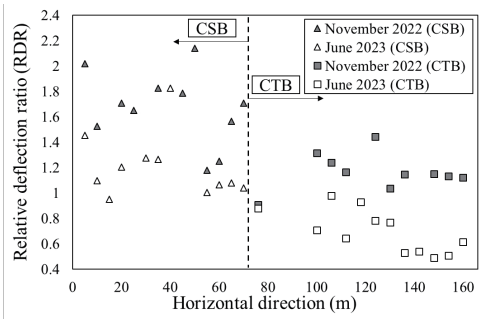


Fig. 1. Relative deflection ratio for the longitudinal direction

B. Settlement characterlistic

Figure 2 shows the settlements that occurred in CSB and CTB for each period using the M3C2 technique. In M3C2, a normal scale (D) and projection scale (d) of each core point and normal were set as 0.3677. The calculation mode was set to the vertical direction to calculate settlement and prevents errors in selecting the normal vector that may occur since the normal scale is small compared to roughness.

Figures 2(a) and (b) compare the settlement amount in CSB and CTB for the entire period of operation of heavy equipment. The maximum settlement -5.78 cm measured in CSB, and the maximum settlement -2.21 cm was detected in CTB. In the area around the west entrance within CSB that the maximum settlement occurred, the same settlement was occurred on the asphalt pavement situated parallel to the block pavement (Fig. 3(a)), which occurred due to the low bearing capacity of the basement across north-south direction. Additionally, a subsidence of -4.42 cm was detected at a location approximately 48 m from the western entrance within the CSB. In this section, a ground survey was performed to check the state of replacement for the initial crushed stone layer. Even though compaction was performed afterward, the bearing capacities were not sufficiently revealed (Fig. 3(b)) Overall settlement were distributed within -3.1cm except the locations which had the geotechnical problem as shown in Fig. 3.

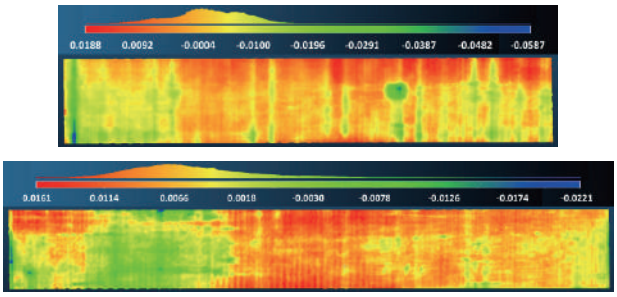


Fig. 2. Relative deflection ratio for the longitudinal direction

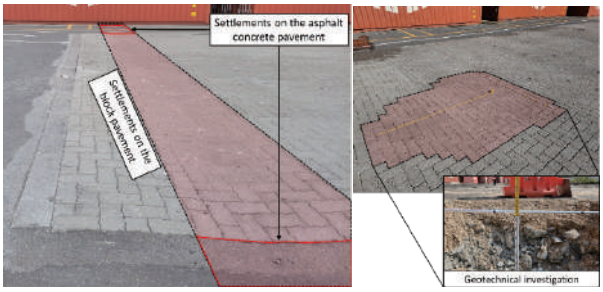


Fig. 3. Relative deflection ratio for the longitudinal direction

IV. Conclusion

In this study, the blocks for ports were developed to solve and supplement problems occurring in asphalt and concrete pavement. Block pavement was constructed in the container storage located at the domestic OO port, divided into CSB and CTB. The FWD test-based deflection and bearing capacity characteristics and the LiDAR measurement-based settlement characteristics that occurred due to the operation of the heavy equipment after construction were analyzed.

In ports where heavy equipment operates, sufficient ground bearing capacity can be secured through CTB rather than CSB, and it is believed that the block pavement can be widely used in the port area based on the various advantages of the block pavement technique.

Acknowledgment

This research was supported by Research service “ICT-based smart block development and application” funded by the Ministry of Oceans and Fisheries, Korea and Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries, Korea (20210659).

References

[1] Lague, D., Brodu, N., and Leroux, J., “Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z)” ISPRS Journal of Photogrammetry and Remote Sensing, 2013, pp. 10-26.

[2] Lane, S.N., Westaway, R.M., Hicks, D.M., “Estimation of erosion and deposition volumes in a large, gravel-bed, braided river using synoptic remote sensing” Earth Surface Processes and Landforms, 2003, 249–271.

[3] Girardeau-Montaut, D., Roux, M., Marc, R., Thibault, G., “Change detection on points cloud data acquired with a ground laser scanner” International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 2005, 30–35.

[4] Cignoni, P. and Rocchini, C., . “Metro: measuring error on simplified surfaces” Computer Graphics Forum, 1998, 167–174.

Simulation of Bare Trench Silica Fiber Sensors for Humidity Sensing between Length of Depths of 50µm and 110µm

Athi Shankhara Velusamy
Malaysian-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia Kuala Lumpur
54100 Kuala Lumpur, Malaysia
athishankhara@graduate.utm.my

Muhamad Yusof Mohd Noor
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru,Malaysia
yusofnor@utm.my

Siti Mahfuza Saimon
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
81310 Johor Bahru,Malaysia
sitimahfuza@utm.my

Fauzan Ahmad
Malaysian-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia Kuala Lumpur
54100 Kuala Lumpur, Malaysia
fauzan.kl@utm.my

Abstract— One kind of coreless silica fibre (CSF) that is highly well-known for its use in humidity-sensitive materials is trench coreless silica fibre, or TCSF. However, because of inherent limits and the nature of foreign materials or substances, there is a significant rise in cost and a complicating factor during the fabrication process of TCSF or CSF-based humidity sensors. Relevant to this paper introduces a simulation of a two TCSF which the width and depth of 50µm design and another width of 50µm and depth of 110µm design for measuring humidity as opposed to mimicking a CSF to determine the TCSF's sensitivity. To test the sensor's ability to sense humidity, we simulated variations in the analyte refractive index, which ranges from 1.000 RIU (representing air) to 1.350 RIU (representing water). These, results indicate that the two bare or uncoated TCSF designs successfully detect humidity (1.00 to 1.35 RI), with one achieving a sensitivity of 375.6 x 10^{-9} RIU per unit of RI change and another achieving a sensitivity of 593.7 x 10^{-9} RIU per unit of RI and demonstrating strong linearity with two different R^2 values of 0.9412 and 0.9299. As a result, the findings set the stage for potential design changes that could be made in the future to improve the performance of the uncoated TCSF humidity sensor.

Keywords—TCSF, Trench, MFD, RI, Fiber, Comsol Multiphysics, Humidity Sensing, Analyte medium

I. INTRODUCTION

Coreless silica fiber (CSF) sometimes referred to as coreless fiber or NCF (No-core fiber), is a crucial component of optical fiber sensors that are useful in a wide range of application.[1] CSF does, in contrast to other fibers, have a distinct core region with a greater refractive index than the cladding around it. Because it is not steered like a standard fiber, light can be used for additional purposes, such as sensing applications and material conversion between various characteristics and fibers. This is because, it is coreless, the CSF modifies the parameters of light propagation, enabling light to interact with surroundings and external materials. [2,3,4] Originally, CSF was developed to overcome the constraints that the multimode fiber (MMF) was subject too. Furthermore, the CSF functions as a sensor. [5,6]

With CSF, scientists have recently started experimenting with novel structural changes, such as V-shaped, bend-shaped, tapered and trench-shaped designs. [7] These developments signify a substantial progression in CSF design. The presence of trenches is essential. When a CSF has trenches added to it, the refractive index changes and the way light flows through the fiber is impacted. [8,9] The light travels in a different shape here than it does in a typical circle. The electric field and refractive index values will vary with depth and width. Trenched CSF, also known as TCSF, is technically more sensitive because of its larger trench diameter and cladding design, which interacts with the surrounding sensing medium more well. [7] This TCSF sensor based has been applied to detect liquid refractive index ranging from 1.3000 RIU to 1.4000 RIU.

In this paper, we simulate two TCSF designs with the two trenches depth of 50µm and 110µm and both same widths of 50µm via wave optic module COMSOL Multiphysics® software to sense humidity. Hence, without the use of any sensitive humidity material, this study examines the sensitivity performance of the naked TCSF fiber to fell humidity. To evaluate its humidity-sensing abilities, we have also talked about the electric field distribution throughout the sensor and the interaction between evanescence waves and effective RI changes at various analyte RI mediums between the range of 1.000 RIU (representing air) and 1.350 RIU (representing water). To the best of our knowledge, the simulation of bare TCSF as fiber humidity sensors was reported here for the first time.

II. METHODOLOGY

In this experiment, we use a trench coreless silica fiber (TCSF) design which is illustrated in Fig. 1. This TCSF consist of outer diameter or PML (Perfectly Meshed Layer) within 4µm. The analyte diameter of 200µm is included. The diameter of the analyte is a crucial sensing medium. The primary notion of this TCSF is that the fiber lacks a core in the inner diameter, hence this inner diameter shouldn't be confused with a core. Hence, this smaller diameter is known as a cladding diameter which is 125µm. Additionally, we cut

a specific depth and width into the top portion of the cladding because we are utilising a trench design. Based on Fig. 1, shows a cross-sectional view and PML in the cut-off zone and the TCSF sensing medium (analyte).

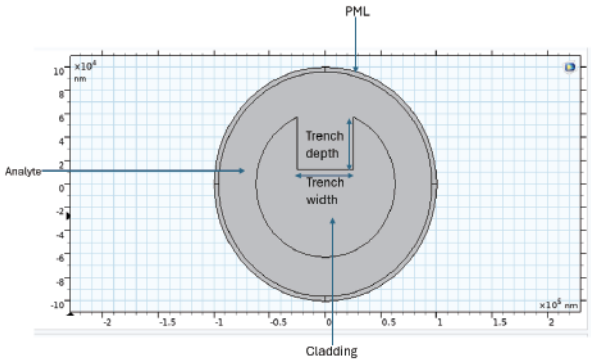


Fig. 1 The structure of TCSF in cross-sectional shape

The software used to create the study's simulation is called "COMSOL Multiphysics" version 6.2® or known as (COMSOL) in shortform was used. The implementation of two-dimensional (2D) space was motivated by its ability to simulate a sensing device in a time-efficient manner by providing simulation output quickly. As a result, it reduces difficulty during the meshing process and less core usage in computer CPU and memory rather than doing simulation in three-dimensional (3D). Wave optics module was chosen for the electrical field study because its wavelength is comparable to and significantly smaller than that of the device under examination. Cross section method (CSM) is used to examine the mode shape of waveguide cross-section region and offers virtual information in the mode field diameter (MFD).[10] It is possible to see the multiple electric field distribution of waveguides, or "EWFD" and assess the strength and penetration depth of the evanescent wave because of the interaction between the sensor and the various analyte medium values. Hence, the geometry settings of CSM of TCSF are shown in Table 1. The TCSF outer diameter or known as PML is set within 4µm, the analyte domain diameter was set in 200µm, the cladding diameter in 125µm, the trenches depths of (50µm and 110µm) and both trenches width were set in 50µm. In (COMSOL), the "Domain" is known as "Name" and "Dimension" is known as "Expression". Table I shows the geometry settings of simulated model which were used.

TABLE I. GEOMETRY SETTINGS

Domain	Dimension
Cladding diameter	125µm
Analyte diameter	200µm
PML diameter	4µm
Trench depth	50 µm & 110µm
Trench width	50 µm

The RI was assigned to each domain of the drawn parameter by applying material settings. In, order to express the RI of TCSF in the near-infrared spectrum, a three term Sellmeier equation was applied as expressed in Eq. (1) [11].

$$\frac{0.6961663\lambda^2}{\lambda^2-(0.0684043)^2}+\frac{0.4079426\lambda^2}{\lambda^2-(0.1162414)^2}+\frac{0.8974794\lambda^2}{\lambda^2-(9.896161)^2}=n^2-1 \quad (1)$$

Based on the equation Eq. (1), n represents the TCSF and λ represents the operating wavelength. Hence, the RI mediums are set to isotropic medium. In theory, the analyte and the TCSF are mostly composed of liquid and glass, each of which has a single RI in every direction. The λ known as operating wavelength is set to 1550 nm which is the minimum attenuation window for a practical fiber. The applied parameter values are tabled in Table II [12].

TABLE II. MATERIAL SETTINGS

Parameter	Value
TCSF RI	1.449 RIU
Analyte RI	1.000 RIU – 1.350 RIU
Medium RI	Isotropic

Sensing ability, often referred to as detection ability, was applied to the evanescent wave interaction between the sensor and analyte medium. The variations in the RI distribution are sensed or detected by the evanescent field. This is caused by the analyte medium, which effects the effective mode index, n_{eff} of the sensor. The "Penetration Depth" or known as dp represents the exponential decline from the evanescent wave with the distance in the analyte medium, which is expressed in Eq. (2).

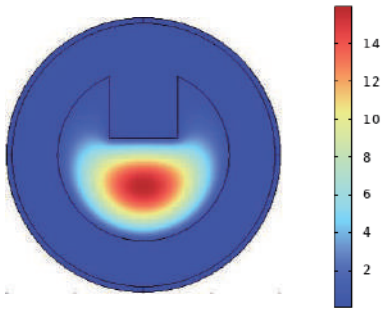
$$\frac{\lambda}{2\pi\sqrt{n_{eff}^2\sin^2\theta-n_a^2}}=d_p \quad (2)$$

Based on the equation Eq. (2), n_a represents the analyte RI and λ represents the operating wavelength. The d_p values change with the change in na which makes and alloys random value of analyte medium detection. The sensor sensitivity, S can be expressed as in Eq. (3), which can be determined by the slope of the graph.

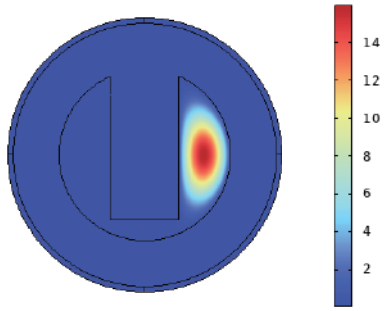
$$\frac{\Delta n_{eff}}{\Delta n_a}=S \quad (3)$$

III. RESULTS AND DISCUSSION

In Fig. 2 (a & b), it shows electric field intensity around in the 2 TCSF designs. Thus, we learn that the huge diameter of TCSF permits the propagation of various modes of light. Out of the ten modes, the fundamental mode and the other four are more likely to be researched because they exhibit the highest power and are highly excited for a Gaussian beam. [13] But in this study, we just focus on the fundamental mode for the ease of evaluation. In a selective study, both TCSFs with different analyte medium values with 1.000 RIU to 1.350 RIU with its respective electric field distributions are illustrated through red to blue colour, which represents the maximum and minimum electric field intensity. The electric field is evenly distributed and focused on the centre of a TCSF construction, indicating that light is either travelling through the fibre or being restricted by it. In both Fig. 2 (a & b), we can see the middle part which is redder than the outer part which is from yellow to blue. This clarifies why the centre has greater power than the peripheral part. The colour fades out from red to blue indicates that the electric field gradually decreases when reaching the analyte due to the attenuation of the light energy as it reaches the boundary with different RI mediums.



(a)

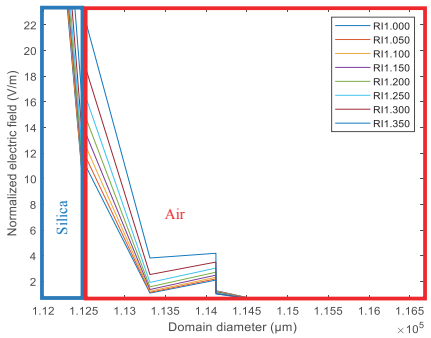


(b)

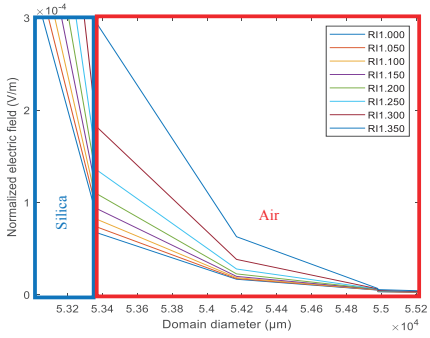
Fig. 2 Coloured illustration of normalized electric field intensity around the TCSF of depth and width of 50µm (a) and TCSF of depth of 110µm and width of 50µm (b).

The power distribution of the TCSF, often referred to as the electric wave field distribution (EWFD), can be observed and examined by examining the MFD from 1.000 RIU to 1.350 RIU, as demonstrated by the Figure 3(a) and (b). It shows both MFDs that represents the power towards the different analyte RI from 1.000 RIU to 1.350 RIU. The RI difference between the TCSF and the analyte medium determines the penetration depth. Based on both Fig. 3(a & b), the evanescent wave can be seen as tiny light waves that extend from the cladding medium into the analyte medium, taking on an oval form as opposed to a circular one. It is observed in Fig. 3(a) that the 50µm trench marks the end of the silica border at 112.5µm, while the wave seems to leak up to 114.2 µm in the direction of the air. While Fig. 3(b), the boundary of silica ceases at 5.33µm because of the larger depth of 100µm and the in addition, a larger wave leakage extends up to 5.5µm which indicate towards the air. Hence, when the light overlaps with the surrounding material, it signifies the existence of the evanescent wave. This proofs than the sensitivity of the TCSF with the depth of 110µm is better than the TCSF with the depth of 50µm because as the depth is deeper, the wave can sense the surrounding air faster and in a lesser domain diameter than the TCSF with the depth of 50µm. The optimisation of the meshing procedure is the cause of the evanescent wave's lack of smoothness.

This demonstrates that the refractive index (RI) of an analyte medium in the range of 1.00 RIU to 1.35 RIU may be detected by the bare or uncoated TCSF. The nature of the RI at 1.350 RIU, which is more towards liquid states with higher moisture content than those below 1.350 RIU, which are more like air with lower moisture content, can be attributed to the absorption characteristics or the higher light intensity at 1.350 RIU than at lower RI values. This states that the sensor can effectively detect humidity since the relative humidity (RH) increases, the air particles surrounding the sensing region, which acts as cladding layer, become more saturated with water molecules [14,15]. As a result, this saturation causes the surrounding air's relative humidity (RI) to rise from 1.000 to 1.350 near the water, effectively communicating changes in humidity levels.



(a)

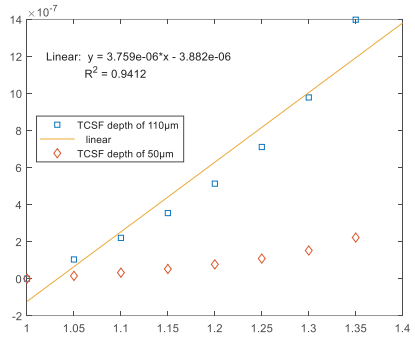


(b)

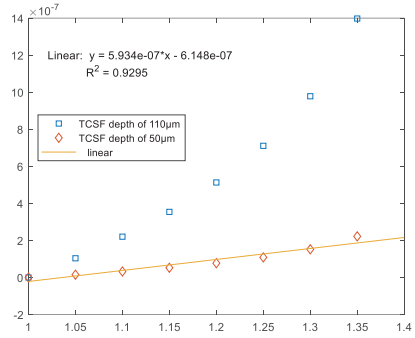
Fig. 3 MFD of simulated design and the evanescent wave of (a)TCSF (depth 50µm and width 50µm) and (b)TCSF (depth 110µm and width 50µm).

Based on Fig. 4, the calculated sensitivity of sensor of RIU per unit RI change of TCSF with the depth of 110µm is 375.6 x 10^{-9} which exhibits a stronger linearity with R^2 value of trench at 0.9412 rather than the TCSF with the depth of 50µm is 593.7 x 10^{-9} with R^2 value of trench at 0.9295 which are an ideal value for linearity. Both trenches had achieved that ideal value of R^2 which are above 0.5.[16,17] This demonstrates that the penetration depth will rise in proportion to the height of the evanescent wave. As a result, the sensor's overall sensitivity will rise.

A maximum wavelength sensitivity of 108 pm/RH% was attained throughout a relative humidity range of 30% to 100%, based on the work of Al-Swefee et al. who developed a trenchless CSF humidity sensor.[18] Based on the data shown in Fig. 4, we anticipate that our sensor, which uses two trench CSF designs, the one TCSF with the depth of 110µm will show significantly higher sensitivity than the one TCSF with the depth of 50µm. This shows the difference between them. Based on this paper, the simulation results of this TCSF will provide a foundation for future experiments. These experiments will help validate and enhance the sensor's performance through controlled environment testing and offer valuable guidance.



(a)



(b)

Fig. 4 (a & b) Shows difference linear of effective index change between TCSF (depth 50µm and width 50µm) and TCSF (depth 110µm and width 50µm).

IV. Conclusion

In this paper, the design and simulation of the two Trench Coreless Silica Fiber (TCSF) sensors are explained in this work with two different depths (50µm and 110µm) and with both same width of 50 µm with various analyte RI values falling between 1.000 RIU and 1.350 RIU. Throughout the study, we can conclude that the two TCSF sensors can detect RI, indicating that the sensor's sensitivity are good. It's excellent that the optimised TCSF sensors detected the difference in analyte RI's electric field distributions and evanescent fields without overlapping throughout the simulation assessment. To sum up, this work demonstrates the sensitivity of the TCSF sensors for humidity detection, which will be a strong suit for subsequent applications.

V. Acknowledgement

This work was supported by the Fundamental Research Grant Scheme (FRGS), grant number FRGS/1/2023/TK07/UTM/02/20 funded by the Ministry of Higher Education (MOHE), Malaysia.

References

[1] Zhang, Ya-nan, Lebin Zhang, Bo Han, Peng Gao, Qilu Wu, and Aozhuo Zhang. "Reflective mercury ion and temperature sensor based on a functionalized no-core fiber combined with a fiber Bragg grating." Sensors and Actuators B: Chemical 272 (2018): 331-339.

[2] Xu, W.; Shi, J.; Yang, X.; Xu, D.; Rong, F.; Zhao, J.; Yao, J. Relative Humidity Sensor Based on No-Core Fiber Coated by Agarose-Gel Film. *Sensors* 2017, *17*, 2353.

[3] Zhao, Yong, Jian Zhao, and Qiang Zhao. "Review of no-core optical fiber sensor and applications." Sensors and Actuators A: Physical 313 (2020): 112160.

[4] O. Laccourreye, H. Maisonneuve,French scientific medical journals confronted by developments in medical writing and the transformation of the medical press,European Annals of Otorhinolaryngology, Head and Neck Diseases,Volume 136, Issue 6,2019,Pages 475-480,ISSN 1879-7296.

[5] Baharin, N. F., Azmi, A. I., Abdullah, A. S., & Mohd Noor, M. Y. (2018). Refractive index sensor based on lateral offset of coreless silica interferometer. Optics & Laser Technology, 99, 396–401.

[6] Sun, Bing, Fei Li, Kandi Xu, Kaiming Zhou, and Zuxing Zhang. "Temperature-insensitive fiber-optic refractometer based on immobilization of polydimethilsiloxane film." IEEE Photonics Technology Letters 34, no. 3 (2022): 165-168. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[7] Nazirah Mohd Razali, Hafizal Yahaya, Muhammad Quisar Lokman, Siti Nur Fatin Zuikafly, & Fauzan Ahmad. (2023). Simulation of Trenched Coreless Silica Fiber Sensor for Refractive Index Measurement. Journal of Advanced Research in Applied Sciences and Engineering Technology, 32(1), 87–94.

[8] Shimura, Y.; Suzuki, Y.; Morisawa, M. Improving Measurement Range of a Swellable Polymer-Clad Plastic Fiber Optic Humidity Sensor by Dye Addition. *Sensors* 2022, *22*, 6315.

[9] P. Hu, X. Dong, K. Ni, L.H. Chen, W.C. Wong, C.C. Chan, Sensitivity-enhanced Michelson interferometric humidity sensor with waist-enlarged fiber bitaper, Sensors Actuators B Chem. 194 (2014) 180–184.2013.12.081.

[10] aofei, Yuchan Hu, Hongda Cheng, Feng Yan, Qianyu Lin, Yu Chen, Pengjun Wu et al. "Side-polished singlemode-multimode-single-mode fiber structure for the vector magnetic field sensing." Journal of Lightwave Technology 38, no. 20 (2020): 5837-5843. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[11] Younus, Shahad I., Anwaar A. Al-Dergazly, and A. K. Abass. "Characterization of Multimode Interference Based Optical Fiber." In IOP Conference Series: Materials Science and Engineering, vol. 1076, no. 1, p. 012060. IOP Publishing, 2021.Shimura, Y.; Suzuki, Y.; Morisawa, M. Improving Measurement Range of a Swellable Polymer-Clad Plastic Fiber Optic Humidity Sensor by Dye Addition. *Sensors* 2022, *22*, 6315.

[12] Y.L. Hoo, W. Jin, C. Shi, H.L. Ho, D.N. Wang, S.C. Ruan, Design and Modeling of a Photonic Crystal Fiber Gas Sensor, Appl. Opt. 42 (2003) 3509.

[13] Malitson, Ian H. "Interspecimen comparison of the refractive index of fused silica." Journal of the Optical Society of America 55, no. 10 (1965): 1205-1209

[14] A. Thakallapelli, S. Ghosh and S. Kamalasadan, "Real-time frequency based reduced order modeling of large power grid," 2016 IEEE Power and Energy Society General Meeting (PESGM), *Boston, MA, USA, 2016, pp. 1-5.*

[15] Zhao, Yong, Jian Zhao, and Qiang Zhao. "Review of no-core optical fiber sensor and applications." Sensors and Actuators A: Physical 313 (2020): 112160.

[16] Chai, Quan, Hyeonwoo Lee, Seongjin Hong, Yongsoo Lee, Junbum Park, Jianzhong Zhang, and Kyunghwan Oh. "Nanoliter liquid refractive index sensing using a silica V-groove fiber interferometer." Photonics Research7, no. 7 (2019): 792-797. I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[17] K. Suzuki, H. Kubota, S. Kawanishi, M. Tanaka, M. Fujita, Optical properties of a low-loss polarization-maintaining photonic crystal fiber., Opt. Express. 9 (2001) 676–680.

[18] Huda T. Al-Swefee, Sarah K. Al-Hayali, Abdulhadi A. Al-Janabi, "Enhanced Relative Humidity Sensor via Diameter of No-Core Fiber Structure" (2019) Iraqi Journal of Laser, 18(1), pp. 7–11.

Construction of a simulation environment for federated learning-based LEO satellite communications systems

Gwangun Yu
Kongju National University
Cheonan 31080, Korea
yugwangun@smail.kongju.ac.kr

Yonggang Kim*
Kongju National University
Cheonan 31080, Korea
ygkim@kongju.ac.kr

***Abstract*—Low-earth orbit (LEO) satellite communications have attracted substantial research attention because they are feasible and applicable in next-generation communications systems compared to medium-Earth orbit (MEO) and geostationary (GEO) satellites. However, the low altitude of an LEO satellite results in a limited coverage area per satellite, unlike MEO or GEO satellites, which provide wireless connectivity services to broader regions. The limited network coverage of LEO satellites can be enhanced by deploying an increased number of LEO satellites. However, as the diversity of satellites increases, the constellations of LEO satellites and constellation-aware resource management become important to efficiently enhance the satellite services. For efficient resource management, a federated learning approach could be adopted between a ground station and LEO satellites. Information from multiple satellites could be aggregated to produce a global model, which can be utilized by satellites that contact the ground station. In this paper, we construct simulation environments for efficient federated learning in LEO satellite communications systems. With the constructed environments with actual parameters of federated learning updates, we show the feasibility of federated learning with an insight into learning delays impacts on LEO satellite systems.**

I. Introduction and Motivation

Low Earth Orbit (LEO) satellite networks are extensively utilized across various environments due to their ability to provide universal coverage, which is unattainable by terrestrial networks. However, the significantly higher velocities of LEO satellites compared to terrestrial networks necessitate the development of robust satellite constellations to maintain effective communication links. Therefore, the strategic design and deployment of these constellations are crucial for ensuring sustained and reliable communication capabilities within LEO networks. Given the presence of thousands of satellites, traditional centralized processing approaches might be rendered inefficient due to the substantial volumes of data generated by satellites. Furthermore, the computational capabilities of satellites are continually advancing [1]. Consequently, satellite federated learning emerges as a pivotal solution to address the limitations of conventional learning [2].

In satellite federated learning, rather than a single node processing and learning from all satellite data, each satellite engages in learning using its own data. During this process, the satellites receive the necessary gradients for machine learning from a parameter server, which is assumed to be located at a ground station in this scenario. The satellites then integrate these received gradients with their own specific data, which could include their velocity, position, and other relevant information, to conduct machine learning. Subsequently, the information derived from this learning process is transmitted back to the parameter server. The parameter server continuously updates the global model based on the incoming data, thereby refining the learning process across the network. Additionally, each satellite node can only communicate for a limited period because signals from GS are no longer reachable once the node passes over the Earth's horizon. It is conceivable to consider scenarios where the duration of the learning process exceeds the available communication time. In such cases, the residual learning tasks could be transferred to the subsequent satellite in the network to avoid delays in updating the global model as depicted in Figure 1. This strategy is particularly pertinent given that the typical orbital period for LEO satellites ranges from 90 to 120 minutes, necessitating efficient management of both learning and communication tasks within these time constraints to maintain continuous and timely model updates

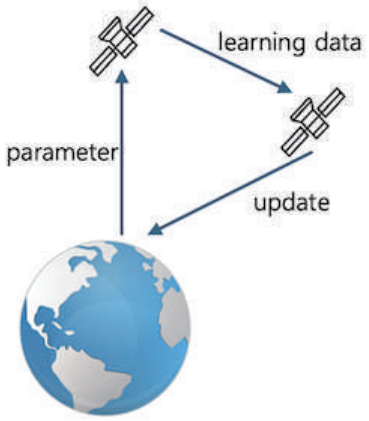


Fig. 1. Transfer of residual learning data from one satellite to the next satellite.

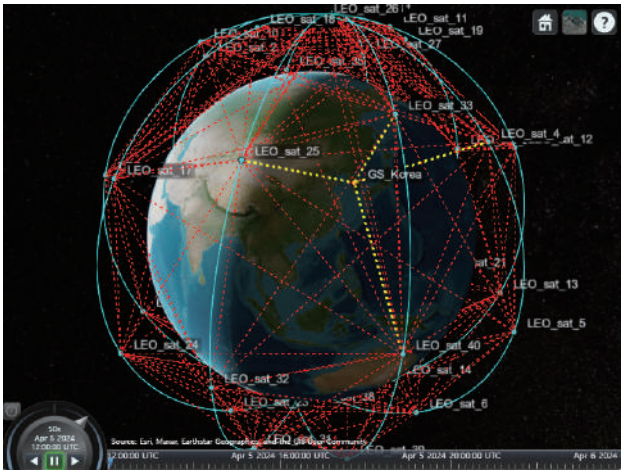


Fig. 2. visualization of simulation

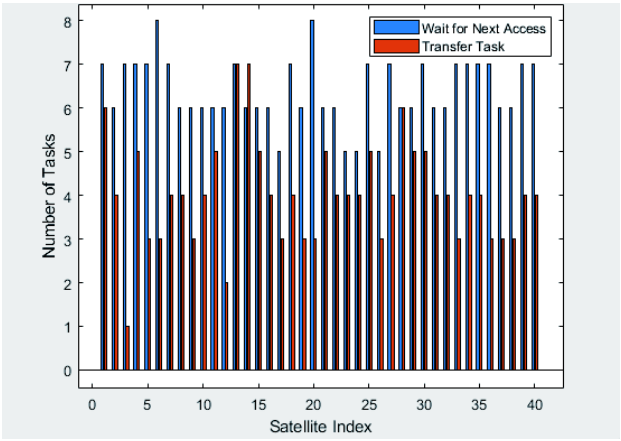


Fig. 3. transfer task is way less then waiting task

[3]. In this paper, we construct simulation environments for federated learning-based LEO satellite systems by adapting existing methodologies on federated learning and integrating them with LEO satellite communication techniques. The constructed simulation environments demonstrate the feasibility of federated learning in LEO satellite communications and provide insights into learning delay for future research.

II. Simulation Environments for Satellite Federated Learning

In the considered system model of satellite networks, Walkerstar constellation 85°: 40/5/1 is applied. The Walkerstar constellation is represented as $\{ i : n/p/s \}$ where i represents the inclination of the orbital planes in degrees, which determines the latitudinal coverage of the satellites. n indicates the total number of satellites in the constellation, reflecting its capacity for global coverage. p denotes the number of orbital planes, which affects the spatial distribution and redundancy of the network. Lastly, s signifies the relative shift in right ascension of the ascending node between adjacent orbital planes, crucial for optimizing the overlap and timing of satellite passes over specific regions of Earth.

Figure 2 illustrates that the red line represents connectivity between satellites, the yellow line depicts connections from satellites to the GS, and the blue line indicates the orbital plane. Assuming the ground station is located in South Korea, the maximum communication duration with each satellite is limited to 20 minutes per pass, dictated by the satellite's communication capabilities. Additionally, the learning time for each satellite ranges from 10 to 30 minutes, varying according to satellite availability. To simulate federated learning-based LEO satellite communication, the maximum time interval between learning sessions is set to be greater than the communication time. Evaluating the effectiveness of transmitting remaining learning tasks to the next satellite when the learning duration exceeds the maximum communication time. This method is investigated for its efficiency since a satellite that has exceeded its communication window must wait for the next opportunity to transmit data, potentially delaying the integration of its data into the global model and impacting the timeliness and relevance of the model updates.

Figure 3 presents a color-coded analysis where the blue marker signifies tasks that were not completed within the communication window, and the red marker denotes tasks that were finished within the allotted communication time.

III. Conclusion

Simulation results demonstrated that when the duration of learning surpasses the allocated communication time, delays in updating the global model are inevitable. This observation underscores the necessity for timely updates to ensure the efficiency of satellite communications. In instances where a satellite is available and not actively engaged, delegating the task to an incoming satellite may serve as an effective strategy to circumvent these delays. Further research should explore the transfer of residual data tasks across various orbital planes and consider the integration of multiple-parameter servers, including additional ground stations not previously considered in the simulation. Implementing these measures has the potential to substantially improve the operational robustness and adaptability of satellite network systems. Such advancements could lead to more efficient data handling and improved continuity of service across the satellite constellation.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2022-00166739).

References

[1] M. Y. Abdelsadek, A. U. Chaudhry, T. Darwish, E. Erdogan, G. Karabulut-Kurt, P. G. Madoery, O. B. Yahia, and H. Yanikomeroglu, "Future space networks: Toward the next giant leap for humankind," *IEEE Transactions on Communications*, vol. 71, no. 2, pp. 949–1007, 2022.

[2] B. Matthiesen, N. Razmi, I. Leyva-Mayorga, A. Dekorsy, and P. Popovski, "Federated learning in satellite constellations," *IEEE Network*, 2023.

[3] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, "Ground-assisted federated learning in LEO satellite constellations," *IEEE Wireless Communications Letters*, vol. 11, no. 4, pp. 717–721, 2022.

Proof of Concept (POC) for Queue Management in Sustainable Smart Community Automated Convocation

Jing Huey Khor
Communications, Signal Processing, and Control Group
University of Southampton (Malaysia Branch)
Johor Bahru, Malaysia
j.khor@soton.ac.uk

W. Ismail
Auto-ID Laboratory (AIDL), School of Electrical and Electronic Engineering, Universiti Sains Malaysia
Penang, Malaysia
eewidad@usm.my

***Abstract*—Before 2010, Universiti Sains Malaysia (USM) employed a proprietary convocation system utilizing barcode technology. However, this system presented limitations such as vendor lock-in and insufficient customization options, leading to congestion and long queues during ceremonies due to the slow detection rate of barcode technology. In response, this paper introduces the Sustainable Smart Community Automated Convocation (SSCAC) system, integrating USM's Roll Call System (USM CRC) with UHF RFID technology. Implemented since 2010, SSCAC has streamlined the management of approximately 5700 graduates per year at USM's convocation ceremonies. The SSCAC system optimizes flow management by significantly reducing human intervention and improving detection rates compared to the previous proprietary system. Through a proof of concept, SSCAC has demonstrated its efficacy in developing a simple, reliable, and systematic flow, effectively addressing the shortcomings of the previous system.**

Keywords*—Queue *management, Convocation, Roll call system, Proof of concept, UHF RFID

I. Introduction

The convocation ceremony, a hallmark event in universities worldwide, celebrates the academic achievements of thousands of graduates through the conferral of degrees and recognition of honorary recipients. Managing such a large cohort efficiently is essential, necessitating robust roll call and data management systems. The roll call process is pivotal in orchestrating the orderly progression of graduates onto the stage to receive their scrolls. A streamlined system minimizes queuing times and optimizes the flow of graduates from their seats to the center stage, mitigating congestion within the venue. To address these challenges, universities have increasingly turned to advanced technologies such as Ultra-High Frequency Radio Frequency Identification (UHF RFID).

By leveraging UHF RFID technology alongside efficient data management systems, universities can streamline convocation ceremonies, ensuring a seamless experience for graduates and attendees alike. This integration exemplifies the ongoing commitment to innovation in academia, harnessing technology to elevate the ceremonial process and honor academic accomplishments effectively.

This approach aligns with broader trends in technology adoption within academic institutions, where research [1] indicates a growing reliance on RFID systems for various administrative functions. The scalability and versatility of RFID technology make it well-suited for large-scale events like convocation ceremonies [2], where the efficient management of participants is paramount to the success of the event. Furthermore, the integration of RFID systems with data management platforms enhances the overall efficiency of academic operations [3], demonstrating the multifaceted benefits of technological innovation in higher education.

The rest of this paper is structured as follows. The background of USM convocation ceremony is described in Section 2. In Section 3, the system architecture of the sustainable smart community automated convocation (SSCAC) is presented. Section 4 shows the configuration of the SSCAC and Section 5 analyses the issues faced during the implementation of SSCAC. Finally, Section 6 provides the conclusion of this paper.

II. Sustainable Smart Community Automated Convocation

USM hosts two convocation ceremonies each year to award certification scrolls to a large number of graduates. The April ceremony consists of two sessions per day, with approximately 1,000 graduates receiving their scrolls. In contrast, the September ceremony spans five days, with nine sessions in total, accommodating around 4,700 graduates. This brings the total number of graduates attending each year to about 5,700. Given the scale of attendees, there is a crucial need for a systematic system to manage the flow of graduates and ease the graduation process. Before 2010, USM utilized a barcode system for convocation, where graduates carried cards with barcode labels that were scanned upon passing a terminal. However, this system had limitations, including slow reading rates and congestion issues due to line-of-sight requirements. To address these challenges, USM implemented a more efficient automated convocation system in 2010, integrating a roll call system (USM CRC) with Ultra-High Frequency Radio Frequency Identification (UHF RFID) technology.

The SSCAC system, as it's called, offered several advantages: a. Simplified the convocation process and reduced manpower requirements. b. Utilized reusable UHF RFID tags, significantly lowering tag costs compared to barcode systems. c. Allowed for customization and enhancements by USM staff, unlike previous vendor-dependent systems. d. In-house development of RFID readers and systems facilitated easy maintenance, support, and troubleshooting. e. Served as a showcase of local technological expertise, offering guidance to other institutions with similar needs or customization requirements.

The successful implementation of the SSCAC system not only streamlined convocation ceremonies at USM but also demonstrated the potential for local technological solutions

to address complex operational challenges in educational institutions.

III. SSAC System Architecture

The SSCAC system consisted of six terminals with different functionality as shown in Figure 1. It was designed to enable a real time-sharing graduate's information at different terminals using a single server. The system was designed to obtain graduates' information from the back-end database in which this database was developed using Microsoft SQL Server Management Studio and it was installed on the database terminal. The application and database servers must be loaded in balance at the application level to reach maximum stability and availability. The system required five notebooks to run the application and they were used as the terminals to display the data output. The system required a local area network (LAN) environment to enable large data transferring and sharing among computers. A switch was utilized as a backbone of the LAN connection.

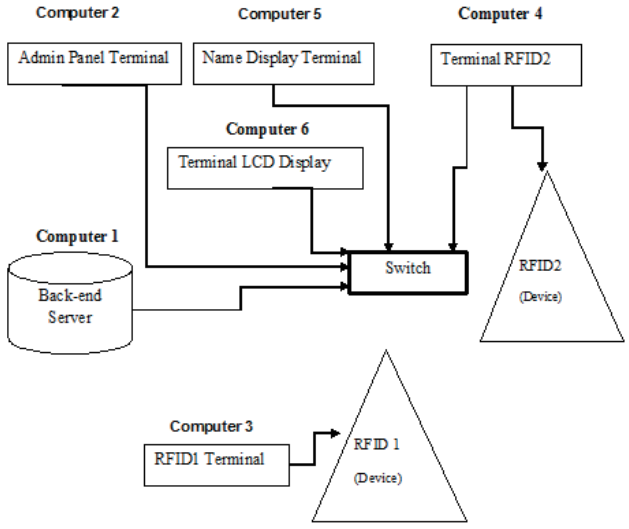
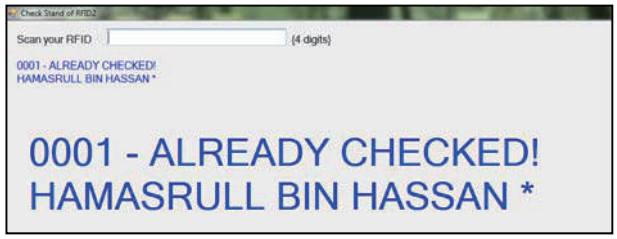


Fig. 1. Architecture of SSAC System

Firstly, a list of graduates was selected by an administrative to store in the back-end database at the Admin Panel terminal. It was stored in Excel file and later imported to the SQL server database. In addition, a full list can be generated and displayed at the end of each session at this terminal. The functionality of RFID tag was examined at RFID 1 terminal in which it was a stand-alone system. The RFID 1 terminal acted as a checking station. The tag attached at the academic robe of graduates was checked to ensure it was able to send ID properly to the reader. The graduates were required to make sure that their data including their names and academic program were displayed correctly on the LCD screen. Next, the detection of the UHF RFID tag which was attached to the graduate's academic robe was done at the RFID 2 terminal. The graduate's information was stored in the queue list at the Name Display Terminal after a successful detection of RFID tag was obtained at the RFID 2 terminal. The graduate needed to ensure that the reader can successfully detect the tag attached at the academic robe as indicated in Figure 2(a) while the Figure 2(b) showed an error message at the RFID 2 terminal when an invalid tag ID was detected. Successful tag detection enabled the graduate's information to be displayed simultaneously at the Name Display and LCD Display terminals. The system was developed to display a queue list at the left side of graphical user interface (GUI) as shown in Figure 3(a) at the Name Display terminal. The queue list can display the graduates' names up to three names. The first graduate's name in the queue list was displayed on the right side of GUI after the fourth graduate's data was verified by the system. The lists of the graduates were arranged in a proper order and a maximum number of three graduates were remained in the queue list. The graduate's information including the graduate's name, programme, and degree classification obtained were displayed on the right side of GUI. The LCD Display terminal showed the name and programme of the graduates to the audiences in the hall. It automatically displayed them once it received an input data from the RFID 2 terminal. The data displayed on the LCD Display terminal was similar to the data displayed on the right side of the Name Display terminal as shown in Figure 3(b).

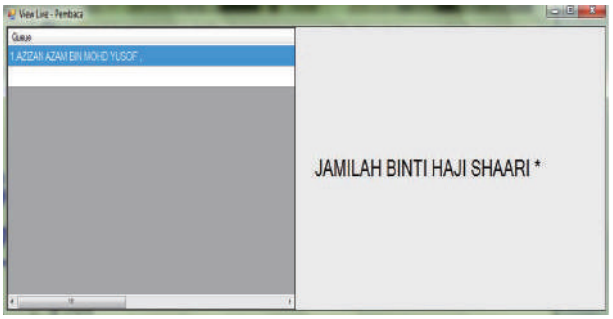


(a)

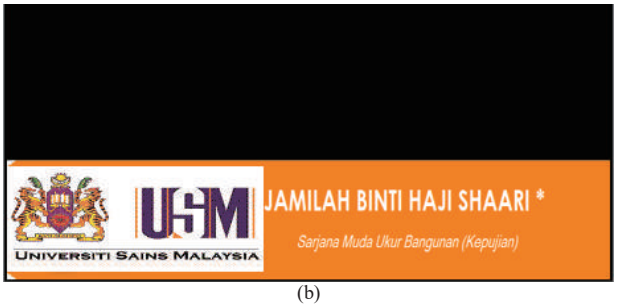


(b)

Fig. 2. (a) The graduate's name was shown once the tag ID was found in the database (b) An error message was shown if tag ID was not in database



(a)



(b)

Fig. 3. (a) Name display terminal (b) LCD display terminal

IV. Movement Flow of SSCAC System

The SSCAC system was designed to create a systematic movement flow of graduates from their seat at the hall until they received the scroll on the stage. The initial movement flow of graduates was shown at Figure 4(a). The first, second, and third graduates would queue alongside the stage and they would walk passing through the RFID 1 and 2 terminals once their names were announced. Once the fourth graduate's tag was detected at the RFID 2 terminal, the data of the fourth graduate will enable the name of the first graduate to be displayed on the screens. Hence, the Dean can immediately announce the name of the first graduate at the Name Display terminal. The first graduate walked to the middle of the stage to receive the scroll from the Chancellor and it was followed by the second, third, and fourth graduates as shown as in Figure 4(b). The data of the second graduate was displayed on the screens simultaneously after the fifth graduate walked passing the RFID 2 terminal. Hence, the Dean announced the name of the second graduate once it was displayed at the Name Display terminal. The process was repeated until the last three graduates for each session. A staff was needed to manually key in a specific number, which was 9999, as a signal to the system. The system automatically showed the name of the last three graduates at the Name Display and LCD Display terminals.

V. Challenges of SSCAC System Implementation

A month before the convocation ceremony, the information of graduates including their ID tags, names and programs were stored in the database. In addition, RFID tags were examined properly to make sure that the graduates' data were stored in the database in which each ID tag encompassed the correct graduates' information. Next, the data stored in the database was checked to eliminate any duplicate data. In addition, the components utilized in the SSCAC system were tested and verified to ascertain the functionality and performance of the components. These processes were done before the convocation ceremony to ensure a smooth and systematic movement flow during the convocation ceremony.

During the convocation ceremony, RFID tags were distributed to all graduates and affixed to their academic robes. Graduates queued up after receiving instructions from staff. They first checked the functionality of their RFID tags at RFID 1 terminals, then proceeded to RFID 2 terminals, where the system authenticated their tag IDs against the SSCAC database. If verified, their names and programs were displayed on Name and LCD terminals; otherwise, an error message appeared. Graduates then moved to the stage to receive their scrolls, with audience screens displaying their movement and information. Several issues arose during SSCAC system implementation, all of which were addressed:

a. Integrating UHF RFID technology with roll-call and audio visual systems was challenging but resolved through pilot tests.
b. Designing a systematic movement flow with minimal queuing required optimizing waiting areas and reducing walking distances.
c. Ensuring real-time data detection and sharing necessitated employing UHF RFID technology for fast detection rates and immediate data display.
d. Misreading of RFID tags occurred due to body interference, which was mitigated by ensuring tags were positioned away from direct contact with the body.

These improvements enhanced the reliability and efficiency of the SSCAC system, ensuring a smooth convocation experience for graduates and attendees.

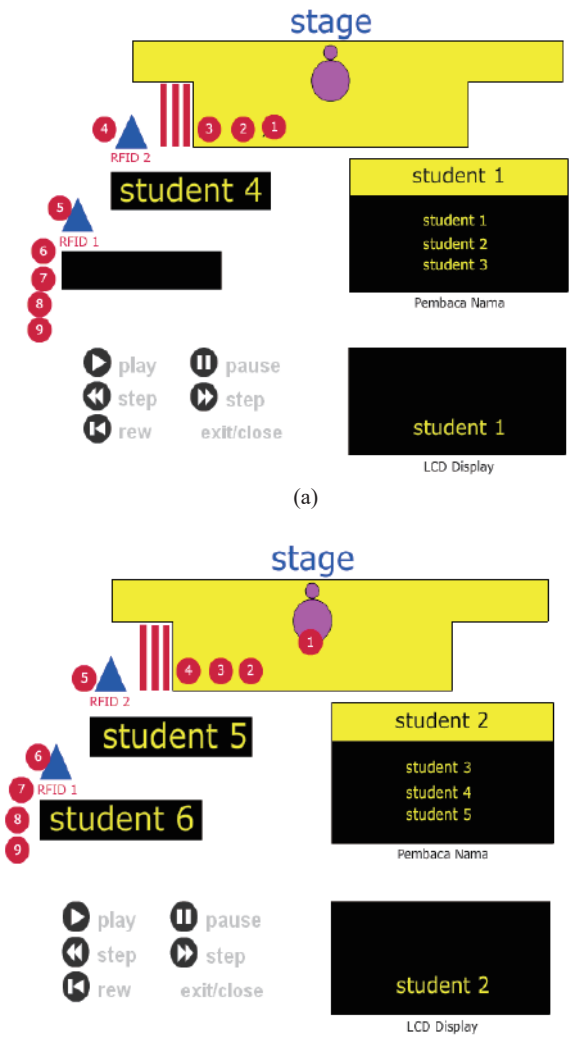


Fig. 4. (a) Initial movement flow of graduates – left wing (b) Movement flow of second, third, fourth and fifth graduates – right wing

VI. Proof of Concept

The Proof of concept (POC) is a realization of a certain method to establish its validity and feasibility based on testing results [4][5]. Hence, POC was conducted to verify the performance of the proposed SSCAC for implementation in the USM convocation ceremony. Five tests were carried out and each of them had its own objective, as presented in Table 1.

There were three types of passive tags used in this testing, namely Passive Gen 2 Thin PVC Card, Metalcraft RFID Windshield Tag, and SMARTRAC Web RFID Wet Inlay (NXP G2iL). These three tags will be named as card, windshield, and sticker web tags in the following session. The summary for the specifications for each passive tag is shown in Table 2.

TABLE I. PILOT TESTS OBJECTIVES

Test	Description
Test 1	To test performance of tags on single and double antennae
Test 2	To test the best position to attach the tag on the graduation robe during the convocation
Test 3	To test which types of tag that performed the best when communicating with the RFID reader
Test 4	To test the best distance and placement of reader antennae from the ground
Test 5	To test the effect of using T-connector and power divider on the double antennae to the successful detection rate
Test 6	To test the overall estimated time taken for a convocation event of 1000 graduates

TABLE II. SUMMARY OF TESTED TAGS

Type of passive tag / Specification	Card	Windshield	Sticker web
Operating frequency (MHz)	860-960	860-960	860-960
Reading range (m)	1.5	5.0	5.0
Operating distance (m)	6.6	5.5	4.0 - 6.0
EIRP (W)	2.0	0.25	2.0

In each pilot test, the passive RFID tag was attached on seven parts of the body, namely upper arm (cloth), arm (cloth), upper arm (inner robe), arm (inner robe), upper arm (outer robe), arm (outer robe), and loose (the tag was held by student themselves without attaching the tag onto the other places). The measurements of each pilot test were collected from a distance of 10 cm to 100 cm, with increment of 10 cm and the detection of tag that was attached on 7 parts of body was measured. There were ten laps of tests that were conducted for each detection range for detection rate of single or double antennae test and the performance of passive tag on different parts on academic robe. Therefore, a trial consisted of 700 readings comprised of 10 detection range x 7 parts x 10 laps for each pilot test was run.

As an example, Test 1 was conducted to test the performance of passive tags on single and double array antennae. Test 2 was done to locate the best part to attach the tag on the convocation robe during the convocation ceremony.

The results from the Figure 6(h) shows that the best part to attach card tag was the upper arm (inner robe) when using the two antennae array. There was a smaller loss in signal transmission from the reader to the tag because the upper arm part was in-line with the antennae.

Conclusion

This paper introduces a Sustainable Smart Community Automated Convocation (SSCAC) system, which integrates the Roll Call System (USM CRC) with UHF RFID technology. The SSCAC system features six terminals: admin panel, RFID 1, RFID 2, name display, LCD display, and database terminals. USM CRC streamlines roll call processes, while UHF RFID technology enhances graduate identification speed. A redesigned movement flow reduces congestion and queuing during convocation ceremonies. The implemented SSCAC, resolving implementation issues to enhance reliability and efficiency by recommending guideline of usage to automate queuing system by profiling the RFID to suit body movements. Proof of concept (POC) trials confirmed SSCAC's feasibility for USM convocations. To address issues such as congestion and long queues during convocation ceremonies, a novel systematic movement flow for graduates was devised. This sustainable approach minimizes the physical and environmental impact of large gatherings, aligning with sustainability objectives. Challenges encountered during its deployment have been successfully resolved, resulting in a more dependable and efficient system. SSCAC system's feasibility validation underscores its potential as a sustainable ICT solution. Successful implementation positions SSCAC as a model for local technological expertise, offering guidance for other institutions with similar needs.

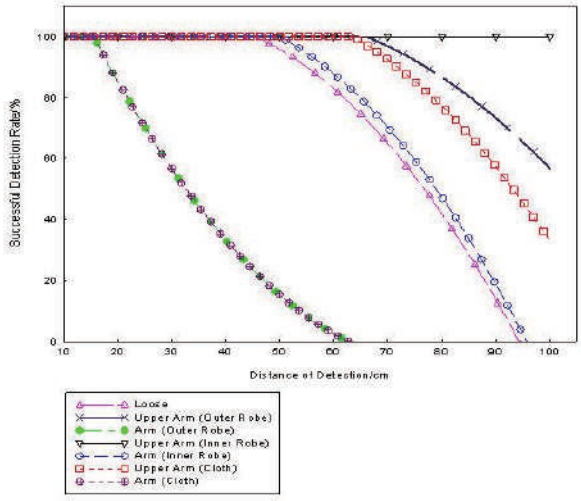


Figure 6(h): The graphs of distance of detection versus percentage of successful detection of RFID card tag at different location using double antennae

ACKNOWLEDGMENT

The authors would like to thank the USM RU (Research University - 814241) grant secretariat and Malaysia Ministry of Higher Education (FRGS Fund - 6071306) for sponsoring the development of the in house built in system of RFID devices.

REFERENCES

[1] Swan, M. (2017). "The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery." In Big Data, Vol. 1, No. 2, pp. 85-99.

[2] Yang, L., Xie, R., & Li, X. (2018). "Design and Implementation of RFID-based University Campus Intelligent Attendance Management System." In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics, pp. 356-365.

[3] Yeo, S., Kim, Y., & Kim, S. (2020). "The Development of an RFID-based Tracking System for Academic Resources in Higher Education." In Journal of Educational Technology & Society, Vol. 23, No. 2, pp. 23-36.

[4] Hasan, A. (2007). Proof of Concept. In C. Phillips (Ed.), Project Management for Information Systems (5th ed., pp. 149-150). Pearson Education Limited.

[5] Khademi, M., Arabi, M., & Nabatchian, M. (2009). The role of proof of concept in software development projects. International Journal of Computer Science and Network Security, 9(10), 100-106

Design of a Scheme to Enhance Real-Time Performance for DDS-Based Encrypted Communication in ROS2

Hee-Jae Shin*, Jae-Min Lee*, Tae-Soo Jun†, Dong-Seong Kim*
IT Convergence Engineering*, Computer Software Engineering†
Kumoh National Institute of Technology, Gumi, Korea
{shinheejae*, ljmpaul*, taesoo.jun†, dskim*}@kumoh.ac.kr*

***Abstract*— ROS/ROS2 is a robot software development framework used across multiple robots, enabling developers to create advanced robotic applications. Particularly, ROS2 is based on DDS (Data Distribution Service) to enhance security and real-time performance. ROS2 follows the DDS-Security standard for encrypted communication, designed with system performance and scalability. The use of asymmetric keys is restricted to the discovery, authentication, session, and shared secret setup phases, while data distribution is encrypted using symmetric keys. Authentication can be performed using signatures, but for real-time performance, MAC through individual keys between nodes is also used. Unlike plaintext communication, multicast in encrypted communication requires a group key. For real-time performance, a common key method is used among specific nodes, ensuring multicast authentication and encryption are conducted together. This paper proposes a method to improve the existing multicast authentication process, ensuring maximum real-time performance, and enhancing the overall efficiency of the system.**

Keywords— ROS2, DDS, SROS2, DDS-Security

I. INTRODUCTION

ROS2 (Robot Operating System 2) has established itself as an essential platform for developing various automated mobility systems such as autonomous vehicles, electric cars, drone-based delivery, and customized public transportation solutions. The fields where ROS2 is used are predominantly safety-critical systems, where the precise timing of task execution is crucial. These applications demand a high level of reliability and performance to ensure safe and effective operation. [1] Unlike its predecessor ROS1, ROS2 incorporates DDS (Data Distribution Service) as a standardized middleware, which provides real-time performance, scalability, and flexibility using the publish and subscribe concept. [2] This architecture allows for efficient data distribution and management, crucial for complex robotic systems. Moreover, ROS2's secure communication follows DDS security standards, allowing encrypted communication using DDS-Security options. The DDS security standard supports the overall encryption of messages, implying that RTPS (Real-Time Publish-Subscribe) messages can be encrypted. DDS encrypted communication typically recommends using public keys during the registration process and symmetric keys during communication. This is because DDS is generally used in systems requiring high performance and extensive scalability with numerous processes and computers, and unnecessary performance degradation and scalability issues must be avoided.[3] Additionally, DDS can be extended to edge computing platforms through DDS-XRCE, further emphasizing the need for efficient and scalable communication protocols.[4] The use of asymmetric key encryption is recommended to be limited to discovery, authentication, session, and shared secret setting phases. Asymmetric key usage in the data distribution process is not favorable for real-time performance, hence symmetric keys are used. In plaintext communication, publishers can multicast data by default, but encrypted communication requires the implementation of a group key method. This promotes efficient key management and optimization of network resources, allowing multiple receivers to share a single encrypted key, enabling specific nodes within the group to share and decrypt the encrypted data using the same key. Additionally, for group security, multicast authentication and multicast encryption processes are necessary. During initial connection, the reliability of each node is verified using asymmetric key (public and private key) methods, and symmetric keys are securely exchanged. Once symmetric keys are exchanged, subsequent communication is encrypted using these symmetric keys. Symmetric key encryption methods such as AES-256 are faster and more efficient than asymmetric key methods like RSA, making them suitable for real-time data transmission.

II. EXISTING SYSTEM

In DDS, there are attacks that generate false information by impersonating an information distributor. To prevent this, the DDS security authentication process is divided into asymmetric key signature verification and MAC (Message Authentication Code) verification methods. Signature-based encryption is a public key-based verification process that allows all subscribers to verify whether the data is from the correct publisher with about a 1024-bit overhead, making it a multicast-authenticatable method by default. However, due to the computationally intensive nature of public key-based encryption, real-time performance is compromised, which is undesirable in time-sensitive applications. In contrast, MAC-based authentication is a symmetric key-based process with significantly faster verification speeds compared to public key-based authentication. Specifically, HMAC (Hash-based Message Authentication Code) used in the DDS Security standard can perform calculations more than 1000 times faster than public key-based authentication.[3] However, for general verification, since the publisher uses different symmetric keys for each subscriber, MAC authentication must be included in proportion to the number of subscribers. This proportionality can increase complexity but offers a much-needed speed advantage, making it more suitable for real-time applications.

The DDS security standard document introduces HMAC and AES-GCM/GMAC-based GMAC (Galois Message Authentication Code) for MAC authentication. HMAC uses general hash functions such as SHA-256, while GMAC generates authentication through GHASH. The authentication information hashing process uses part of the data section as input. The actual data encryption process involves encrypting plaintext data sections with symmetric keys along with hash fragments, ensuring that the data remains secure and tamper-

proof. When a distributor node communicates with a single subscriber node, individual symmetric keys are used, whereas communication with a group via multicast uses a group key for encryption. This method ensures that data remains secure and only accessible to authorized nodes, thereby maintaining data integrity and confidentiality. In the decryption and hash verification process, the actual data decryption is carried out using symmetric keys in the same manner as the encryption process. Hash verification is conducted by comparing the hash values computed separately by the subscriber with the transmitted hash values. This method allows for quick verification through hash computation, unlike public/private key-based signature authentication which requires the entire packet to be received for verification. This reduces latency and improves efficiency, making it more viable for real-time applications.

To verify packet integrity using MAC values, the sender generates a MAC through the message and sends it as a packet along with the data. The receiver then separates the message part, computes the MAC independently, and verifies it against the MAC sent by the sender. This process ensures that the data has not been tampered with during transmission, providing a reliable means of data verification.

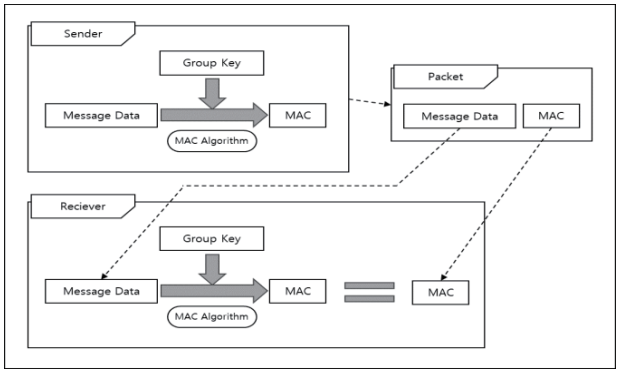


Fig. 1. MAC Verification

There is extensive research on authentication processes using signatures and MAC. For signature-based authentication, methods such as HashTree and HashChains have been proposed.[5] These methods involve verifying an appropriate block size instead of every transmitted packet, buffering the received packets until the block transmission is complete, and then performing overall verification with the transmitted verification packet. In the case of HashTree, the frequency of including the signature is not specified and can be arbitrary, using the most secure stream authentication method that includes a signature on each packet followed by a hash value. This increases the overhead size per packet but maintains adequate real-time performance by calculating the signature only for the block. HashChains involve including the hash value and signature of all packets in the last packet for verification.[6] This method, while efficient, requires careful management of packet sequences to ensure accuracy and integrity throughout the transmission process.

For symmetric key-based authentication, TESLA (Timed Efficient Stream Loss-tolerant Authentication) method has been proposed.[7] This method, similar to signature-based methods, uses a common symmetric key generated arbitrarily instead of MAC values proportional to the number of subscribed nodes. The sender hashes the key sequentially and includes it in packets in reverse order. As a result, subscribers cannot predict the MAC value for the next key and must buffer the packets until the key is revealed, finally verifying whether the sequential hash values form a legitimate series. This method provides robust security while maintaining efficiency and is particularly useful in environments where real-time performance is critical.

Most security implementations default to encrypting all aspects of communication to ensure confidentiality and integrity. However, in real-world applications, performance issues such as network load and processing speed can arise. Therefore, a strategy to control the number of encryptions is necessary. This control ensures that while maintaining security, the system's performance does not degrade, which is critical in real-time applications such as autonomous vehicles and drones. Balancing security and performance is essential to ensure the seamless operation of these advanced systems.

III. Proposed System

This paper primarily employs AES-GCM to suit ROS2. The system introduces encryption and authentication using AES-GCM for a predefined number of nodes and divides the multicast group for transmission if this number is exceeded. This approach ensures that the system remains efficient even as the number of nodes increases by dynamically adjusting the group size and managing the encryption workload effectively. Moreover, if the number of subscribers becomes excessive, preventing multiple MAC-based methods from maintaining real-time capability, the system enables switching to other methods such as TESLA. This flexibility is crucial for maintaining system efficiency and ensuring that performance does not degrade under heavy load conditions.

Among MAC-based authentications usable in the DDS Security standard, GMAC based on AES-GCM is preferred over HMAC. The AES-GCM process integrates encryption and authentication value generation, offering faster verification compared to HMAC. Furthermore, GMAC inherently prevents duplicate values through Nonce. Using HMAC would require specifying an arbitrary AES encryption process known to subscribers and adding a Nonce item to prevent duplicate value-based authentication. Additionally, AES-GCM is convenient as a plug-in embedded in DDS security, simplifying the implementation and ensuring compatibility with existing systems. The verification hash value can be delivered with different sequences, reference data, and hash calculation counts for each node in the group due to the fast hash computation speed. This ensures that nodes in the group cannot transmit false information to other nodes with different hash sequences while enabling instant verification of packets without waiting until the end. This instant verification is critical for real-time applications where delays can have significant negative impacts, such as in autonomous driving and industrial automation.

When the number of subscriber nodes is small, AES-GCM/GMAC is applied as a DDS-Security plugin. The AES-GCM/GMAC method is efficient as the authentication part is generated along with the encryption process. The MAC value for the Common MAC message is followed by the MAC generated by the symmetric key corresponding to the Publisher and Subscriber nodes. Subscriber nodes can verify if the packet sender is the Publisher, preventing other Subscriber nodes from impersonating the Publisher and transmitting false information. Although HMAC is also available as a MAC method, considering factors such as parallelism and the inclusion of Nonce values, AES-GCM/GMAC is more appropriate.

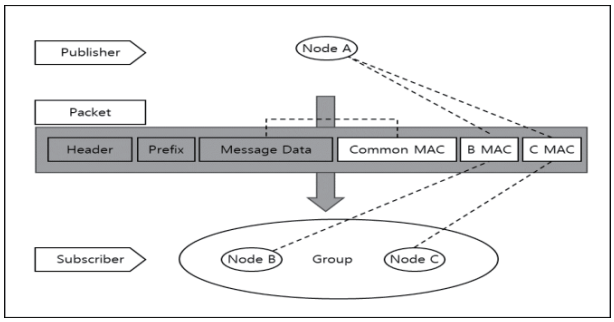


Fig. 2. MAC Values in Packet Postfix

When the number of subscribers increases to the point where it is impractical to use separate MAC values for each subscriber, a scheme like TESLA, designed to handle a large number of subscribers, is used. TESLA can replace multiple MACs with a single MAC, although it may affect real-time performance due to its delayed authentication nature. For signatures, the computational performance of the nodes should be considered. Any other new methods can be applied if available. This centralized approach simplifies key management and reduces the risk of key compromise. Additionally, periodic key rotation can be implemented to further enhance security, ensuring that encryption keys are regularly updated to prevent long-term exposure to potential threats. Overall, the proposed system aims to balance security and performance, providing a robust solution for real-time applications in ROS2. By leveraging advanced encryption methods and dynamic group management, the system can maintain high performance and security even under challenging conditions.

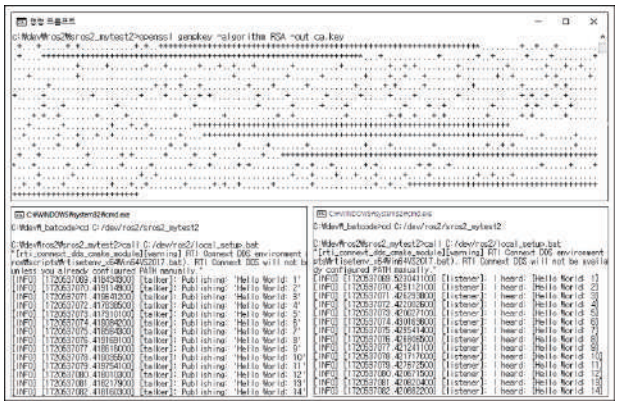


Fig. 3. CA key generation and Pub/Sub encrypted communication

Figure 3 illustrates the preliminary experiments conducted, where a basic CA certificate for the entire node is generated, followed by generating and testing certificates for Publisher and Subscriber. These initial tests, which include the creation of CA keys and the execution of communication tests between Publisher and Subscriber, serve to validate the basic functionality of the encryption and communication setup. The comprehensive testing phase will follow, involving the application of AES-GCM-GMAC and extensive experiments with multiple Publishers and Subscribers. Performance comparisons will be made with existing methods, including signature-based approaches. Subsequently, an analysis incorporating the proposed variable structure will be conducted to thoroughly evaluate its impact on real-time performance and security. These further experiments will be conducted in various environments and under diverse conditions to ensure a comprehensive evaluation of the system's efficiency and robustness.[8]

IV. Conclusion

This paper proposes a system to enhance real-time performance by improving the multicast authentication process in group key-based encrypted communication. The system suggests employing AES-GCM for encryption and authentication for a predefined number of nodes set by the publisher, and dividing the multicast group for transmission when this number is exceeded. This approach ensures that encryption and authentication do not become bottlenecks as the number of nodes increases. If the number of subscribers grows excessively, making it challenging for multiple MAC-based methods to maintain real-time capability, the system supports switching to the TESLA method. It also proposes a variable structure to switch to signature-based authentication when MAC-based authentication cannot meet real-time requirements due to the excessive number of subscribers. This flexibility maximizes real-time efficiency by minimizing the use of RSA-based asymmetric encryption, except in essential situations for node-to-node connections.

By addressing the limitations of existing methods and proposing a flexible, scalable solution, this work aims to enhance the real-time performance and security of ROS2 in various applications. Future research will focus on refining the proposed system, implementing it in real-world environments, and optimizing its performance and security features. Further experiments aim to provide a deeper understanding of the system's efficiency and robustness under various conditions. These efforts are expected to advance secure and efficient communication in robotic and automated systems.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 50%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 50%)

References

[1] S. H. Kim, H. S. Chwa, “Prioritized ROS2 Communication Framework for Future Mobility: Minimizing Latency and Optimizing Network Throughput,” Communications of the Korean Institute of Information Scientists and Engineers, Vol. 41, pp. 23-32, 2024

[2] OMG Data Distribution Service (DDS) Version 1.4. http://www.omg.org/spec/DDS/1.4

[3] DDS Security Version 1.1. http://www.omg.org/spec/DDS-Security/1.2

[4] J. H. Cha, J. W. Lee, J. M. Lee, and D. S. Kim, “Network Load Reduction Scheme of Secure DDS based Edge Computing Platform in Industrial IoT Environment”, KICS Fall Conference 2021, pp. 181-182, 2021

[5] B. H. Chung, J. N. Kim, H. S. Cho, “Security Technology for DDS,” Electronics and Telecommunications Research Institute, vol.26, pp.112-122, 2011

[6] A. Pannetrat, and R. Molva, “Efficient Multicast Packet Authentication,” Proc. Netw. Distributed Syst. Security Symp., San Diego, CA, 2003.

[7] A. Perrig, J. D. Tygar, and D. Song, “The TESLA Broadcast Authentication Protocol,” RSA CryptoBytes, Vol. 5, pp.2-13 2002.

[8] M. H. Jang, H. J. Kim, J. H. Lee, J. H. Lyu, J. M. Lee, and D. S. Kim, “Design and Implementation of Performance Evaluation Scheme for DDS based Naval Ship Combat Network”, KICS Winter Conference 2024, pp. 398-399, 2024

Hybrid Energy Harvesting Development for Internet of Things (IoT)

Shum Kar Yen
School of Electrical and Electronic Engineering
Universiti Sains Malaysia
14300 Pulau Pinang, Malaysia
karyenkarmei0612@gmail.com

Widad Ismail
School of Electrical and Electronic Engineering
Universiti Sains Malaysia
14300 Pulau Pinang, Malaysia
eewidad@usm.my

Abstract— **Energy harvesting (EH) is an attractive and green solution to the problem in powering internet of things (IoT) in a rural environment which do not support power supply. Energy harvesting harvest energy from the ambient environment and replenishes its rechargeable battery had overcome the pitfalls associated with defining lifetime, which arise because EH never stops but can occasionally run out of energy. Due to the small dimension of a sensor node and exterior environment, the power supply attached to the sensor node has to be very limited in size and not able to produce enough power continuously. The proposed solar powered energy-harvesting system is the key technology for solving finite power problem. This scheme works on the principle of photo-voltaic effect. In order to enhance the efficiency and robustness of the system, the solar energy-harvesting system is designed by using a switching circuit and managing the charging of lithium battery. Lithium battery-charging is also used to afford a stable power supply with 5-V output voltage through a standard Universal Series Bus (USB) interface. The battery is charged once the voltage of lithium battery drops below the predetermined level with sufficient illumination of sun. The circuit was built and tested by a several series of experiment to investigate its performance and characteristic. It is demonstrated that the proposed energy harvesting mechanism could operate perpetually, given that the system is exposed to voltage more than 5V from solar panel for at least 5 to 6 hours in a day. It is tested that the solar panel implemented operates best in between 8V to 11V, in which providing a power of more than 1.465W. When the sun shines perpendicularly to the solar panel, the solar panel will also provide a relatively high output power which is at 4.219W.**

***Index Terms*—IoT, energy harvesting, solar energy**

I. Introduction

Energy harvesting is the process of capturing and storing energy from the environment. In the Internet of Things (IoT), powering devices can be challenging for engineers due to the need for numerous wireless sensor nodes. EH offers a solution by powering these nodes, ensuring continuous operation without relying solely on environmental power [1]. Solar energy, particularly from photovoltaic cells, is a highly effective EH technique for IoT devices due to its power density and efficiency [2]. This project focises specifically on studying solar energy harvesting for IoT systems.

Other common EH methods include RF, piezoelectric, thermal and kinetic energy harvesting.

The state-of-the-art energy storage techniques for energy-harvesting system in sustainable wireless sensor nodes can be classifies into two technologies, i.e., supercapacitors and rechargeable batteries [3]. Since the supercapacitors have significantly lower power density and higher leakage overhead than rechargeable batteries [4], which makes them impractical for small-package WSN nodes, we used a lithium battery as the storage in an energy-harvesting system employment. The most popular source of environmental energy is the sun. One reason why solar energy is becoming more widely used is that it has a higher power density (about 15 mW/cm3) than other renewable energy sources, which enables wireless sensor nodes to be completely self-sustaining. This project will be focused on powering the IoT – on Adruino Board using solar energy.

This paper focuses on an solar energy-harvesting system based which prefers to use the solar energy as power and battery as supplementary under inadequate illumination of sun. The main aim of this project is to develop an energy harvesting system for supplying long term and stable power using photovoltaic (PV) technologies for IoT based system developed in Adruino. The specific objectives are as follow:

1. To design a solar charging controller based on the power needed by the IoT based system on Adruino.
2. To run the IoT based system by implementing a solar energy-harvesting system as its' power source
3. To analyse and characterize the proposed prototype system at real location as validation to its eligibility.

This paper focusses on the development of solar energy-harvesting system to supply long term and stable power to IoT based system on Adruino. The designed circuitry can be divided into two main functional parts, i.e., the control circuit and the charging circuit. The whole system is designed, built and tested on breadboard. Experimental results show that the system can work under the requirement of pre-designing. They will be explained in detail in Section III.

II. RELATED WORK

Apart from solar energy-harvesting, there are a few types of energy harvesting which were done by researchers around the globe. Solar or wind power are mainly to be used as renewable energy sources for wireless sensor network. Ramasur et.al. [5] took some efforts in a wind energy harvester (WEH) model in order to increase the lifespan of WSN with powering-up methods. The WEH model consisted a wind generator and a power management unit to store and condition the generated energy [5]. The results showed that their aero-elastic flutter generator could produce more power compared with the other small-scaled generators. However, there are some disadvantages for this model as well. The circuit proposed was not equipped with maximum power point tracking (MPPT) resulting in poor efficiency of the WEH. [6] compares the power generation potential of different energy harvesting techniques. Among all the energy harvesting technologies, solar energy harvesting through photovoltaic conversion and vibration energy harvesting through piezoelectric elements provide the highest power densities [7]. A complete energy harvester must include a transducer, an energy control unit, and an energy storage unit.

III. METHODOLOGY

A. *Design Requirement*

a. *Current Consumption Analysis*

In the developed IoT based system on modified Adruino Uno Board, wireless sensor nodes are commonly connected to the board which has to be powered with 3.3V. In this system, rain sensor which is powered by 5V [8] is used as the sensor and XBee-Pro S3B [9] is used for transmitting the collected data from sensor to the computer in a long range. Since the XBee-Pro S3B 900HP is embedded with the modified Adruino Uno Board which is shown in Figure 1, current consumed by the entire board is analysed and calculated in prior of designing the solar energy-harvesting system. The measured current consumption in the IoT based system in the modified Adruino Uno Board in one hour usage is 0.032A as shown in Table 1. Hence, the energy-harvesting system should be designed with current that is more than 0.032A to support the system.

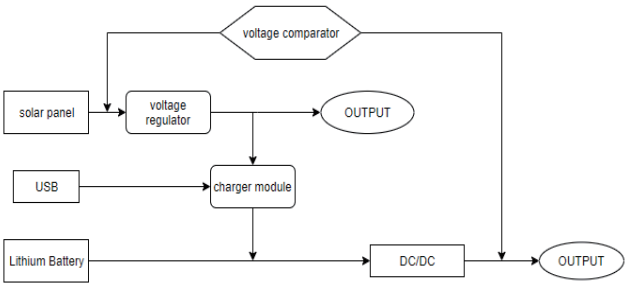


Figure 1: Block Diagram for EH System

TABLE I. CALCULATED CURRENT CONSUMPTION IN ONE HOUR OF USAGE [10]

MODE (TAG)	AVERAGE CURRENT (MA)	% OF TIME	CURRENT CONSUMPTION
TRANSMIT	230	1.5556×10^{-4}	0.3578MA
RECEIVE	229	1.8333×10^{-4}	0.4198MA
IDLE	45	0.0132	5.94MA
SLEEP	32	99.9865	0.032A
		TOTAL	0.032A

b. *Power Consumption Analysis*

In this paper, the sensor node includes the following components: Adruino Uno Board, XBee-Pro S3B 900HP Module and Sensors (Rain Sensor, dissolved oxygen, PH etc)

The power consumption of all the components is calculated as shown in Table II. *Vin* represent the input voltage which is the output voltage of the solar energy-harvesting system. *I* stands for the average current. Sensor's current consumption is set to the average operating current. *T* stands for the working hour in one day. In this paper, 1h is used means that a data monitoring period lasts for 2.5 min every hour. Hence the power consumption of one sensor node is about 0.31*Wh* every day.

B. *System Construction*

a. *Scheme of Design*

The proposed solar energy harvesting system is physically composed of a solar panel, a lithium battery and a control circuit. Control circuit consist of an automatic switching circuit, charging circuit, and a boost DC/DC converted module for the lithium battery. The system schematic diagram is shown in Figure 1.

TABLE II. POWER CONSUMPTION OF ONE SENSOR NODE

	Vin/V	*I/mA*	*P/mW*	*T/h*	*W/Wh*
Transmit & Receive Node	5	32	160	1	0.160
Modified Adruino Uno Board	5	20	100	1	0.100
Rain Sensor	5	10	50	1	0.05
Sum		62	310		0.310

The scheme of solar energy harvesting design in this research is illustrated in Figure 1. The 3 inputs which are solar panel, USB port and lithium ion battery are designed to feed power to the entire circuit. Voltages provided by solar panel will go through voltage regulator to buck down the voltage to the desired 5V. Input solar voltage will be compared with the

output of DC/DC Boost Converter which is from battery in order to determine the type of power to be used to support the design. Battery charger module is connected post voltage regulator, which signify that the battery will only be charged under when the voltage from the solar panel is more than 5V signifying there is sufficient light from the sun. The control charging chip of the charging circuit is TP4056 IC Chip. In order to provide desired 5V from lithium ion battery which had a nominal voltage of only 3.7V, DC/DC Boost Converter is used to boost up the voltage from 3.7V to 5V. Hence, the design consists a total of 2 outputs which can be connected to the different inputs in the IoT based system design.

b. Circuit Construction Design and Principle

LM311N IC Chip [11] is a voltage comparator, and has two inputs, the inverted input as a reference which is provided by the output of DC/DC Boosting Module and the positive-going input is provided by the solar panel's voltage. When the positive-going input voltage is higher than the reference 5V, LM311N outcome a high-level output, which makes the gate of the N Channel Power MOSFET transistor (IRF540N) high. In this design, N Channel Power MOSFET transistor will function as a switch and turn on when the gate is in HIGH state. Current flow from source to drain, the system selects the solar panel for power support. At the same time, the gate of P Channel Power MOSFET (IRF9530N) is high, which makes the P Channel MOSFET to go off and the output of the lithium battery is closed.

On the other hand, when the positive-going input voltage is beneath 5V reference voltage, the comparator will output low voltage, and N Channel Power MOSFET will be turned off whereas the P Channel Power MOSFET will be turned on. As a matter of fact, the lithium ion battery will provide the power to the system now instead of solar panel.

IV. RESULT AND DISCUSSION

A. Solar Energy-Harvesting System Experiment

After the circuit of solar energy-harvesting system is constructed and built on breadboard, a real-time data is collected and analysed. Observations were carried out and recorded when there is sufficient sunlight and insufficient of sunlight. The solar harvesting circuit during sufficient sunlight provides voltage more than 5V from the solar panel. The LED light up, signified that the system used solar energy as power to support the system. Comparator output produces high value and turn on the N-Channel Power MOSFET. Since the source of the MOSFET is connected to the output of regulator circuit, hence the MOSFET is drawing current from Drain to Source, providing voltage and current flow into the IoT based system on the modified Adruino Uno Board and turning the LED ON at the same time. The measured current was at 0.051A which is enough for the solar energy to supply power to the board and the system. Figure 2 shows the successful transmission of the data collected by the sensors through XBee-Pro S3B 900HP to the Laptop to be displayed. This had proved that the power by solar to the IoT system is enough for the transmission to be made.

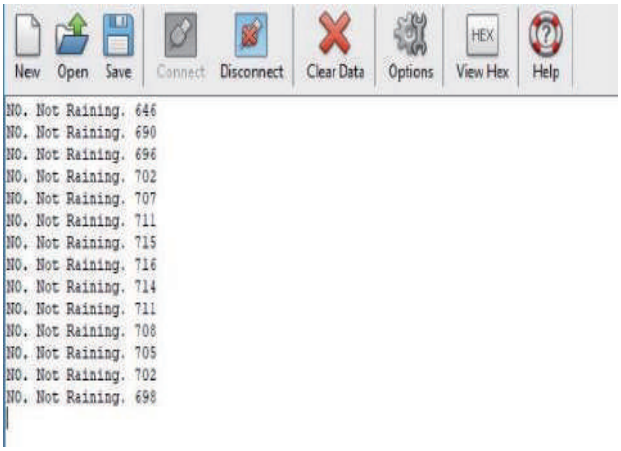


Figure 2: Transmission of data from the system to the Laptop using solar power

When insufficient light going through the solar harvesting circuit, the circuit selected battery as power supply to support the system. The LED at the solar side turned off, comparator output LOW value of voltage and turn off the N-Channel Power MOSFET. At the same time, P-Channel Power MOSFET will turn on, allowing the current flow from Source to Drain of the MOSFET. The measured current is at 0.052A which is enough to support the system under 5V boosted voltage from the DC/DC boost converter module.

B. Performance Analysis

The solar energy harvesting system was tested and run from day to night. Regulator output voltage, battery voltage, solar operating voltage, and output voltage to the developed IoT based system were measured. The values were plotted in a graph as shown in Figure 3 that shows the graph of raw data collected in a day. The operating voltage (blue) is increasing from morning to the afternoon signifying that the voltage provided by the solar panel had increased.

The sudden drop of voltage from 3pm to 3.30pm is due to the heavy rain. The voltage raised back at 4pm after the rain and gradually decreases when night falls. It is clear that the output voltage of the regulator (yellow) increased and decreased just as the operating voltage provided by the sun.

Comparator will give output HIGH (1) or LOW (0) in this design. When LOW, the system selects lithium ion battery as power whereas when HIGH, the system will select the solar power instead. Comparator output in the graph shows that the system used battery as power support in the early morning, during rain and at night when sunlight is still

insufficient. Besides, the graph also shows that the overall output voltage provided by the sun or battery into the IoT based system was always approximately at 5V. The system functioned stably with low-power loss and high reliability.

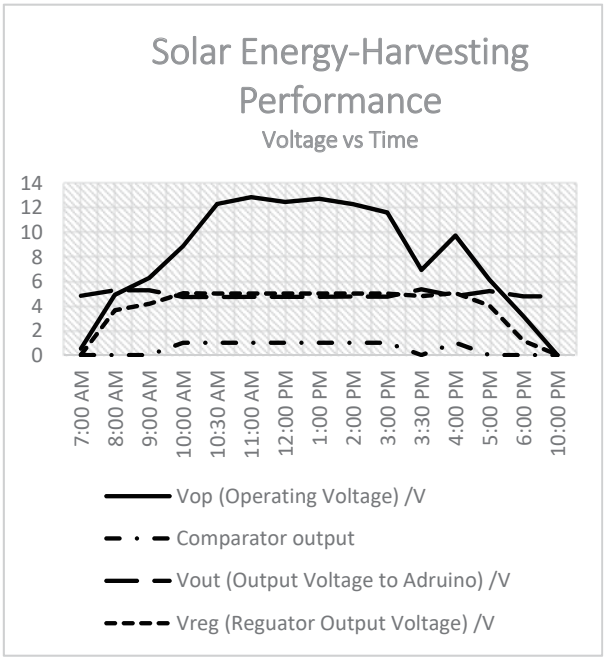


Figure 3: Raw Data collected in real time from morning to night

V. CONCLUSION

In this paper, a solar energy-harvesting system is presented and designed for IoT applications using a modified Arduino Uno Board. The system uses hardware to control solar charging and manage a lithium-ion battery, enhancing robustness. The power requirements for transmission and WSN nodes in IoT are analysed, ensuring a stable 5V output from both solar and battery power sources. The lithium battery charging system includes a protection circuit to avoid charge-discharge cycles, extending battery life. Experimental results show that the system can automatically switch power supply branches when solar panel voltage drops below a predetermined value. When the battery voltage falls below 2.9V, it is properly charged if the solar panel provides more than 5V. The system demonstrated stable performance with low power loss and high reliability, maintaining a consistent 5V output.

VI. FUTURE WORKS

The solar energy harvesting system can be improved in terms of providing various of output voltages so that the system is able to support different designs of IoT based system. The system can also use the Maximum Power Point Tracking (MPPT) Algorithm in order to increase the efficiency of the system. Stability of the power from solar panel in providing transmission by XBee should also can be improved to ensure the data received were accurate and fast.

ACKNOWLEDGEMENT

Appreciation and gratitude to School of Electrical and Electronic Engineering, Universiti Sains Malaysia lecturers, staffs and Dr Samer Ahmed Al-Gailani for the support and ideas. Special appreciation to Nur Afifah Ahmad Nizam and Kamarul Hafiz Bin Kamaludin for the assistance in this project.

REFERENCES

[1] C. P.Souza, "On Harvesting Energy from Tree Trunks for Environmental Monitoring," *International Journal of Distributed Sensor Networks.,* vol. 12, no. 6, p. 9, 2016.

[2] I. J. Minchul Shin, "Energy management algorithm for solarpowered energy harvesting wireless," *IET Communications,* p. 14, 2016.

[3] Fayaz Akhtar, Mubashir Husain, Rehmani, "Energy replenishment using renewable and traditional energy resources for sustainable wireless sensor networks: A review," *COMSATS Institute of Information Technology, Wah Cantt, Pakistan,* vol. 45, pp. 769-784, 2015.

[4] Zhanwei Xu, Zhi Li,Chris M. B. Holt,Xuehai Tan,Huanlei Wang,Babak Shalchi Amirkhiz,Tyler Stephenson,and David Mitlin, "Electrochemical Supercapacitor Electrodes from Sponge-likeGraphene Nanoarchitectures with Ultrahigh Power Density," *The Journal of Physical Chemistry Letters,* p. 2928–2933, 2012.

[5] Hancke, D. Ramasur and G. P., "A wind energy harvester for low power wireless sensor networks," *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings,* pp. 2623-2627, 2012.

[6] R. V. P. P.T.V.Bhuvaneswari, "Solar Energy Harvesting For Wireless Sensor Networks," *2009 First International Conference on Computational Intelligence, Communication Systems and Networks,* 2009.

[7] Wenqi GUO, and William M. HEALY, "Power Supply Issues in Battery Reliant Wireless Sensor Networks: A Review," *International Journal of Intelligent Control and Systems,* 2014.

[8] "Openhacks," 13 February 2018. [Online]. Available: https://www.openhacks.com/uploadsproductos/rain_sensor_module.pdf. [Accessed 13 May 2019].

[9] T. Jimblom, "Sparkfun," 5 June 2014. [Online]. Available: https://cdn.sparkfun.com/datasheets/Wireless/Zigbee/90002173_N.pdf. [Accessed 23 April 2019].

[10] K. H. B. Kamaludin, *Internet of Things Based Wireless Sensor Network System for Water Quality Monitoring,* Penang, Malaysia, 2017.

[11] Li, Y. and Shi, R, "An intelligent solar energy-harvesting system for wireless sensor networks.," *EURASIP Journal on Wireless Communications and Networking,* 2015.

Model Plant Mismatch in FCS-MPC Implementation: A Tuning Guideline for Robust Performance

Ahmad Takiyuddin Abdullah
Mechatronics Engineering Technology Section, UniKL MFI, Universiti Lumpur Lumpur, Bandar Baru Bangi, Selangor,Malaysia
takiyuddin@unikl.edu.my

Sevia Mahdaliza Idrus*
Faculty of Electrical Engineering
Universiti Teknolgi Malaysia
Johor Bahru, Malaysia
*sevia@utm.my

Shahrin Md Ayob
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Johor Bahru Malaysia
e-shahrin@utm.my

***Abstract*—Finite Control Set Model Predictive Control (FCS-MPC) is essential for managing multilevel inverters in mobile, military, and maritime communication systems. However, discrepancies between the model and actual plant dynamics, known as Model Plant Mismatch (MPM), can degrade system performance and communication reliability. This paper addresses MPM in FCS-MPC by proposing a comprehensive tuning guideline to enhance robustness and control accuracy. The guideline includes systematic procedures for identifying critical parameters, adaptive tuning techniques to counteract parameter variations and strategies for real-time model adaptation. Simulation confirms the effectiveness of this approach in improving performance and stability of multilevel inverters, ensuring reliable communication across diverse and challenging environments. This study provides a valuable framework for optimizing FCS-MPC systems, supporting the advancement of ICT convergence and resilient communication infrastructures.**

Keywords— SBBMLI, FCS-MPC, MPM, no model plant mismatch

I. INTRODUCTION

The convergence of Information and Communication Technologies (ICT) has revolutionized the landscape of modern communication systems, driving significant advancements across diverse sectors including mobile, military [1] and maritime. These sectors, each with their unique operational demands and challenges, have greatly benefited from the integration of cutting-edge ICT solutions that enhance connectivity, improve performance, and ensure robust operations in dynamic and often harsh environments [2][3][4].

One of the key technological innovations that has emerged from this convergence is the deployment of advanced control strategies such as Model Predictive Control (MPC). Among these, Finite Control Set Model Predictive Control (FCS-MPC) has gained traction, particularly in power electronics applications like multilevel inverters [5][6]. These inverters are critical components in various communication infrastructures, supporting everything from mobile base stations to naval communication systems, where reliable and efficient power conversion is paramount.

However, a significant challenge in the effective implementation of FCS-MPC is model plant mismatch (MPM), which occurs when there is a discrepancy between the predictive model and the actual system [7]. This mismatch can arise due to parameter uncertainties, unmodeled dynamics, or external disturbances, leading to suboptimal performance and reduced robustness. Addressing MPM is essential for ensuring the reliability and efficiency of multilevel inverters, which in turn supports robust communication systems across mobile, military, and maritime domains.

This paper proposes a comprehensive tuning guideline aimed at mitigating the effects of model plant mismatch in FCS-MPC implementations for multilevel inverters. By systematically analysing the impacts of MPM and exploring effective tuning strategies, we aim to enhance the robustness and performance of these critical systems. This work not only contributes to the advancement of control strategies in power electronics but also supports the broader goal of improving communication infrastructures through ICT convergence.

The structure of this paper is as follows: Section II provides an overview of FCS-MPC and identifies common sources of model plant mismatch in multilevel inverters. Section III introduces the proposed tuning methodology, outlining the steps and considerations necessary for robust FCS-MPC implementation. Section IV presents simulation and experimental results, demonstrating the effectiveness of the proposed guidelines in improving control performance. Finally, Section V concludes with a discussion on the implications of our findings and potential future research directions.

By addressing the critical issue of model plant mismatch and offering a systematic approach to tuning, this paper aims to enhance the reliability and efficiency of multilevel inverters, thereby supporting the continued advancement and convergence of ICT in mobile, military, and maritime communication systems.

II. OVERVIEW OF FCS-MPC AND SOURCES OF MODEL PLANT MISMATCH IN MULTILEVEL INVERTERS

Finite Control Set Model Predictive Control (FCS-MPC) is a powerful control strategy employed in various applications, especially in power electronics systems such as multilevel inverters [8][9][10]. Unlike traditional MPC methods, which typically operate on continuous control inputs, FCS-MPC directly manipulates a finite set of control actions, making it well-suited for systems with discrete control variables like multilevel inverters. By optimizing control actions over a finite set, FCS-MPC offers advantages in terms of computational efficiency and real-time implementation [11][12][13].

In the context of multilevel inverters, model plant mismatch (MPM) arises from several factors. Firstly, parameter uncertainties play a significant role, stemming from variations in component values such as inductors, capacitors, and switching devices, as well as fluctuations due to aging effects or temperature changes [14]. These uncertainties can lead to deviations between the modelled and actual behaviour of the system. Additionally, multilevel inverters may exhibit unmodeled dynamics that are not fully captured in the predictive model. These dynamics could include nonlinearities in power electronic components, dynamic interactions with connected loads, or disturbances from the environment that are not accounted for in the model.

Moreover, external disturbances represent another source of model plant mismatch in multilevel inverters [9]. Deployed in mobile, military, or maritime environments, these inverters are subject to various disturbances such as electromagnetic interference, voltage fluctuations, or rapid load changes. These disturbances can perturb the system and introduce discrepancies between the predicted and actual behaviour. Furthermore, practical implementations of multilevel inverters involve inherent switching delays and losses associated with power semiconductor devices [15]. These delays, influenced by operating conditions and device characteristics, may not be accurately represented in the control model, leading to disparities between predicted and actual switching behaviour.

Understanding these sources of model plant mismatch is crucial for developing effective tuning strategies to mitigate their impact on the performance and robustness of FCS-MPC in multilevel inverter applications. In the following section, we will discuss a comprehensive tuning guideline aimed at addressing these challenges and enhancing the reliability of multilevel inverter control systems.

III. Introducing the Proposed Tuning Methodology for Robust FCS-MPC Implementation

To mitigate the effects of model plant mismatch and enhance the robustness of FCS-MPC in multilevel inverter applications, we propose a comprehensive tuning methodology. This methodology is designed to systematically address the sources of mismatch identified earlier and optimize the controller's performance under varying operating conditions. Fig. 1 shows the structure of comprehensive approach for FCS-MPC controller design in switched-battery boost-multilevel inverter (SBBMLI) voltage control. The approach includes problem formulation, open-loop configuration, FCS-MPC controller design, performance evaluation and assessment of implementation efficiency. This methodology aims to demonstrate the effectiveness, robustness, and practicality of the proposed FCS-MPC controller for voltage control in the SBBMLI.

The process begins by validating the accuracy of the predictive model against the actual plant behavior to ensure proper model validation and identification. This involves collecting data from the physical system and comparing it with the model predictions. Any discrepancies observed should be carefully analysed to identify areas where the model may need refinement or adjustment.

The parameter uncertainties and variations by parameter tuning and adaptation employing adaptive tuning strategies. This may involve online parameter estimation techniques or adaptive control algorithms that continuously adjust controller parameters based on real-time system feedback. By adapting to changes in the system dynamics, the controller can maintain optimal performance despite variations in operating conditions.

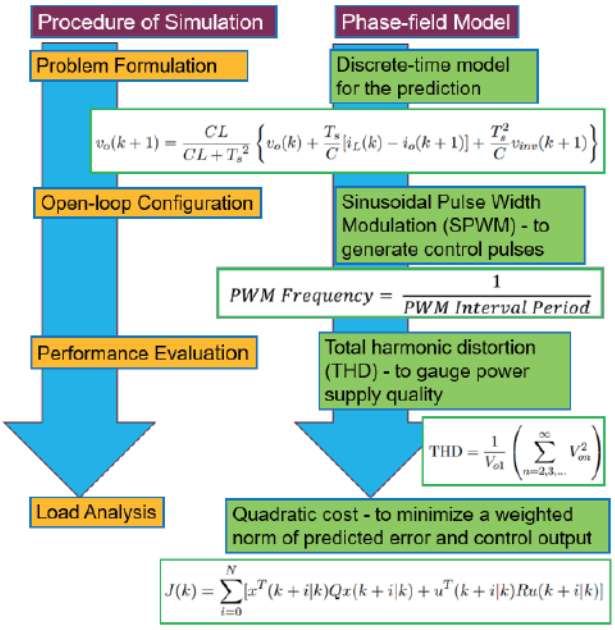


FIGURE 1: The structure of comprehensive approach for FCS-MPC controller design in SBBMLI voltage control

The strategies to mitigate the impact of external disturbances on the control system is developed. This includes incorporating disturbance observers or feedforward compensation techniques to pre-emptively counteract disturbances before they affect the system response. By actively rejecting disturbances, the controller can maintain stability and performance in the presence of external perturbations.

Switching control optimization is applied to minimize switching losses and delays while ensuring accurate tracking of reference signals. This involves advanced control algorithms that optimize the switching sequence based on predicted system behaviour and constraints. By carefully managing switching transitions, the controller can improve efficiency and reduce the impact of switching-related mismatches.

Robustness analysis was conducted to evaluate the performance of the tuned controller under various operating conditions and disturbances. This involves simulation studies as in real-world environments. By systematically assessing the controller's robustness, any remaining deficiencies can be identified and addressed before deployment.

By following these steps and considerations, the proposed tuning methodology aims to enhance the robustness and performance of FCS-MPC in multilevel inverter applications. In the following section, simulation and experimental results are presented to demonstrate the effectiveness of this methodology in mitigating the effects of model plant mismatch and improving control performance.

The section provides a structured outline of the proposed tuning methodology, highlighting its key components and

considerations for robust FCS-MPC implementation in SBBMLI. The results show FCS-MPC using the generalized SBBMLI problem formulation has given rise to less THD compared to FCS-MPC using the Euler method in all three load configurations; resistive load where THD using the generalized SBBMLI problem formulation is 0.29% and THD using the Euler method is 0.68%, resistive and inductive load where THD using the generalized SBBMLI problem formulation is 0.25%. THD using the Euler method is 1.50%, and resistive, capacitive load where THD using the generalized SBBMLI problem formulation is 0.29% and THD using the Euler method is 0.77% [5].

IV. Simulation Validation of Proposed Guidelines

Simulations were conducted to evaluate the performance of the proposed tuning methodology in improving the control performance of multilevel inverters under FCS-MPC. The simulations were performed using a detailed model of the SBBMLI system, incorporating the proposed tuning strategies and considering various operating conditions and disturbances of 3 load configurations as shown in Fig. 2. Key performance metrics such as tracking accuracy, settling time, and robustness to disturbances were evaluated to assess the effectiveness of the tuned controller. The performance of the tuned FCS-MPC controller was compared against a baseline controller with default tuning parameters to quantify the improvements achieved through the proposed methodology.

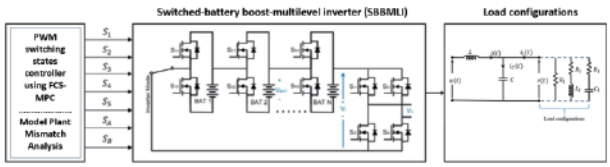


FIGURE 2: The 3 load configurations for the simulation

The execution began with the model validation and identification of the predictive model against the actual plant behavior. Fig. 3 shows the load analysis segment evaluates the FCS-MPC controller's performance, robustness, and stability across various load conditions. By subjecting the controller to diverse load scenarios, we assess its ability to maintain stable operation and satisfactory performance. This analysis provides insights into the controller's adaptability to fluctuations in load magnitude, frequency, and type, enabling optimization of its performance for reliable operation in real-world applications.

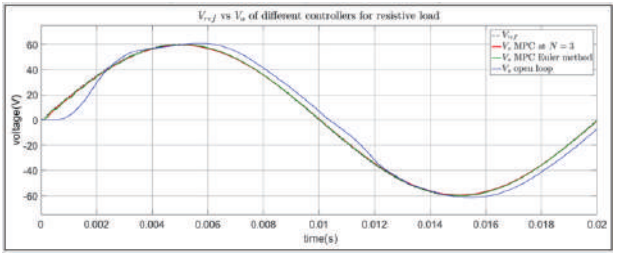


FIGURE 3: V_{ref} *vs* V_o of different controllers for resistive load

Simulation validation was conducted using SBBMLI no model mismatch and model mismatch setup to validate the findings from the simulation studies and assess the practical feasibility of the proposed tuning methodology. Table 1 shows the details of the no model mismatch setup, whereas Table 2 shows sample for model mismatch setup for resistive load with various true plant. The performance of the tuned FCS-MPC controller was evaluated under the 3 different load configurations i.e. resistive load, resistive and inductive load, and resistive and capacitive load. Measurements of key performance metrics were recorded and analysed to validate the effectiveness of the proposed tuning methodology.

Table 1: Parameters for the no model mismatch setup

Parameter	Value
Filter inductance, L	$5.7mH$
Filter capacitance, C	$60\mu F$
Sampling time, T_s	$1/40000s = 25\mu s = 40kHz$
Resistive load, R_l	20Ω
Inductive load, L_l	$10mH$
Capacitive load, C_l	$80\mu F$

Table 2: Parameters for the model mismatch setup for resistive load with various true plant

Parameter	True Plant Value (+5%/-5%)	True Plant Value (+3%/-3%)	True Plant Value (+7%/-7%)
Filter inductance, L	$6mH/5.4mH$	$5.9mH/5.5mH$	$6.1mH/5.3mH$
Filter capacitance, C	$63\mu F/57\mu F$	$62\mu F/58\mu F$	$64\mu F/56\mu F$
Load resistance, R_l	$21\Omega/19\Omega$	$20.6\Omega/19.4\Omega$	$21.4\Omega/18.6\Omega$

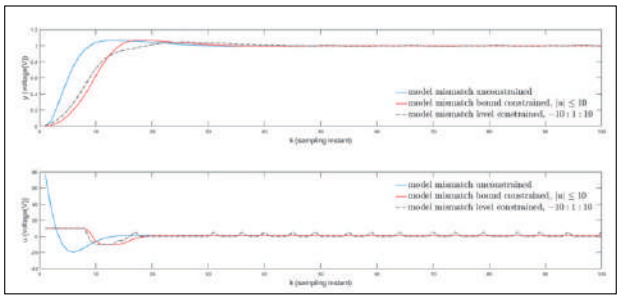


FIGURE 4: Plot of output and input of model mismatch for unconstrained, bound constrained and level constrained

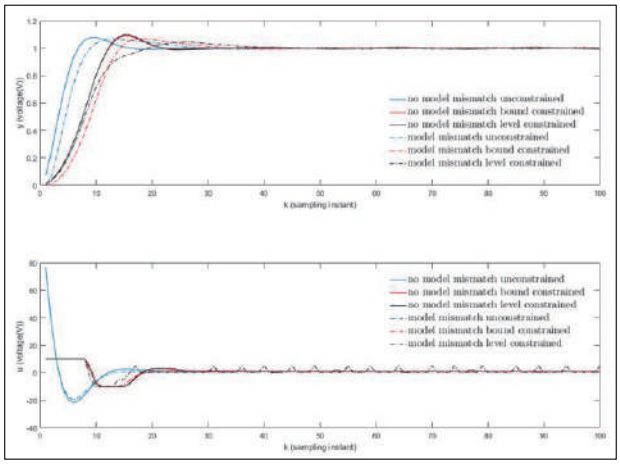


FIGURE 5: Plot of output and input of performance comparison between model mismatch and no model mismatch.

The simulation and experimental results demonstrated significant improvements in control performance achieved through the proposed tuning methodology. Fig. 4 shows the sample plot of output and input of model mismatch for unconstrained, bound constrained and level constrained and Fig. 5 shows the sample plot of output and input of performance comparison between model mismatch and no

model mismatch. Enhanced tracking accuracy, reduced settling time, and improved robustness to disturbances were observed compared to the baseline controller. The tuned controller exhibited robust performance across a range of operating conditions and disturbances, highlighting the effectiveness of the proposed strategies in mitigating the effects of model plant mismatch. The experimental validation confirmed the practical feasibility of implementing the proposed tuning methodology in real-world multilevel inverter systems, providing confidence in its applicability to industrial settings.

V. Conclusion

The sample simulation results presented in this section, along with variations in configurations showcased throughout, provide compelling evidence of the effectiveness of the proposed tuning methodology in improving control performance and robustness of multilevel inverters under FCS-MPC. These findings underscore the importance of addressing model plant mismatch and the potential benefits of adopting advanced tuning strategies in practical applications.

Based on the analysis, the significant effects of parameter setup throughout the simulation lead to the following tuning parameter guidelines:

- The sample rate should ideally exceed 25 times the bandwidth to ensure optimal digital control performance.
- Implementing disengaged constraints yields similar output responses under both unconstrained and constrained conditions (e.g., bound constrained and level constrained).
- Decreasing the MPC weight, Q, increases steady-state error while reducing control effort, thereby minimizing the impact on input bounds, u.
- Increasing the predictive horizon, N, results in smaller overshoot and reduced oscillation. However, higher predictive horizons entail significantly longer execution times.
- The effects of Kalman estimator parameters, QN and RN, are considerably less significant compared to other MPC parameters in cases of model mismatch.
- Model mismatch leads to degradation in steady state, overshoot, and settling time compared to scenarios with no model mismatch.
- Employing a large step interval for level constraints introduces a discrepancy in steady-state error that does not align with the unconstrained case.

The tuned FCS-MPC controller showed significant improvements in tracking accuracy, settling time, and robustness to disturbances compared to baseline controllers. This highlights the importance of advanced tuning strategies in achieving superior control performance in multilevel inverter applications.

The proposed tuning methodology effectively mitigated the effects of model plant mismatch, ensuring robust control performance across a range of operating conditions and disturbances. This emphasizes the importance of addressing uncertainties and variations in system parameters for reliable control implementation.

Acknowledgment

The authors would like to thank Universiti Teknologi Malaysia (UTM) for the support of the project under UTM Institutional Grant vote 08G49.

References

[1] A. L. Imoize, O. Adedeji, N. Tandiya, and S. Shetty, '6G enabled smart infrastructure for sustainable society: Opportunities, challenges, and research roadmap', *Sensors*, vol. 21, no. 5, p. 1709, 2021.

[2] G. Gardašević, K. Katzis, D. Bajić, and L. Berbakov, 'Emerging wireless sensor networks and Internet of Things technologies - Foundations of smart healthcare', *Sensors*, vol. 20, no. 13, p. 3619, 2020.

[3] Y. Gui *et al.*, 'Voltage Support With PV Inverters in Low-Voltage Distribution Networks: An Overview', *IEEE J. Emerg. Sel. Top. Power Electron.*, 2023.

[4] E. A. Bhuiyan, M. Z. Hossain, S. M. Muyeen, S. R. Fahim, S. K. Sarker, and S. K. Das, 'Towards next generation virtual power plant: Technology review and frameworks', *Renew. Sustain. Energy Rev.*, vol. 150, p. 111358, 2021.

[5] A. T. Abdullah, S. M. Idrus, and S. M. Ayob, 'Performance Analysis of FCS-MPC Using the Generalized Formulation and Euler Method on SBBMLI', *5th IEEE Conf. Energy Conversion, CENCON 2021*, pp. 85–90, 2021, doi: 10.1109/CENCON51869.2021.9627244.

[6] A. T. Abdullah, S. M. Idrus, and S. M. Ayob, 'Model predictive control using Euler method for switched-battery boost-multilevel inverter', *Int. J. Power Electron. Drive Syst.*, vol. 14, no. 3, pp. 1497–1508, 2023.

[7] A. Dahlmann, 'Predictive Control of Grid-tied Photovoltaic Energy Systems with Constant Switching Frequency'. Northern Arizona University, 2020.

[8] W. Wu, L. Qiu, J. Rodriguez, X. Liu, J. Ma, and Y. Fang, 'Data-driven finite control-set model predictive control for modular multilevel converter', *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 11, no. 1, pp. 523–531, 2022.

[9] I. Harbi *et al.*, 'Model predictive control of multilevel inverters: Challenges, recent advances, and trends', *IEEE Trans. Power Electron.*, 2023.

[10] C.-H. Lin, S. A. Farooqui, H.-D. Liu, J.-J. Huang, and M. Fahad, 'Finite Control Set Model Predictive Control (FCS-MPC) for Enhancing the Performance of a Single-Phase Inverter in a Renewable Energy System (RES)', *Mathematics*, vol. 11, no. 21, p. 4553, 2023.

[11] S. Vazquez *et al.*, 'An artificial intelligence approach for real-time tuning of weighting factors in FCS-MPC for power converters', *IEEE Trans. Ind. Electron.*, vol. 69, no. 12, pp. 11987–11998, 2021.

[12] T. Chen, O. Abdel-Rahim, F. Peng, and H. Wang, 'An improved finite control set-MPC-based power sharing control strategy for islanded AC microgrids', *IEEE Access*, vol. 8, pp. 52676–52686, 2020.

[13] M. Azab, 'A finite control set model predictive control scheme for single-phase grid-connected inverters', *Renew. Sustain. Energy Rev.*, vol. 135, p. 110131, 2021.

[14] O. M. H. Anssari, M. Badamchizadeh, and S. Ghaemi, 'Designing of a PSO-Based Adaptive SMC With a Multilevel Inverter for MPPT of PV Systems Under Rapidly Changing Weather Conditions', *IEEE Access*, vol. 12, pp. 41421–41435, 2024.

[15] M. R. Chowdhury, M. A. Rahman, M. R. Islam, and A. M. Mahfuz-Ur-Rahman, 'A new modulation technique to improve the power loss division performance of the multilevel inverters', *IEEE Trans. Ind. Electron.*, vol. 68, no. 8, pp. 6828–6839, 2020.

Recent Studies on Serverless Computing: Addressing the Cold Start Problem

Jaehwan Lee
Department of Computer Science and Engineering
Kongju National University
Cheonan, South Korea
jhnlee@kongju.ac.kr

Yeunwoong Kyung
Division of Information and Communication Engineering
Kongju National University
Cheonan, South Korea
ywkyung@kongju.ac.kr

Seungmin Oh
Department of Computer Science and Engineering
Kongju National University
Cheonan, South Korea
smoh@kongju.ac.kr

***Abstract*— Serverless computing continues to be a promising and growing area of study in the field of cloud computing. In this paper, we investigate current research trends in serverless computing, focusing on addressing the cold start problem. Future research directions to enhance resource management efficiency are also discussed. In conclusion, this study highlights the importance of addressing the cold start problem and implementing more intelligent workload predictions to improve the efficiency of serverless computing systems.**

Keywords—serverless computing, cloud computing, lambda function, cold start problem

I. INTRODUCTION

Cloud computing represents a technological paradigm that delivers computing resources and services remotely over the internet [1]. It emerged as a solution to overcome the challenges associated with high initial setup costs, maintenance, and scalability issues that plagued traditional computing infrastructures. Its representative advantage is providing users with the ability to utilize resources and services on demand, by their unique requirements. The performance of services in cloud computing depends heavily on the developer's knowledge and experience with the cloud. In cloud computing, resources frequently remain occupied even during idle periods, leading to unnecessary wastage of resources.

In the pre-era of serverless computing, the efficiency of services within cloud computing relied heavily upon the cloud-related knowledge and experience of developers. This scenario often led to the continuous occupation of resources, even in the absence of tasks, thereby causing a wasteful expenditure of resources. Serverless computing is a next-generation cloud computing paradigm that allows resource utilization in units of functions that are executed without the need to consider the existence of a server [2]. The functions are often referred to as lambda, allowing for the efficient utilization of resources on a function-by-function basis. The cloud providers now can offer a system where the independent execution unit with its resources is freed upon the completion of a function's execution, significantly improving resource efficiency.

The concept of serverless services began to gain attention with the launch of AWS Lambda [3], leading to various cloud providers including Google [4] and Microsoft [5] now offering serverless solutions. Concurrently, a plethora of open-source projects such as Kubeless [6], OpenFaaS [7], OpenWhisk [8], and OpenLambda [9] are being actively developed. In the field of serverless computing, a variety of research is currently underway [10]-[15]. Although not limited to, these studies primarily focus on objectives such as:

- In the serverless model, each client request triggers the creation of a new lambda instance. To handle sudden surges in demand, rapid replication of lambda instances is required. Research is being conducted on various techniques to enhance the fast scaling of instances, including advanced caching methods, lightweight instances, and scheduling strategies.
- Research is being conducted on technologies that can monitor resources and the execution environment in real-time, while maintaining low system overhead across various levels, including functions and instances.

This paper investigates current research trends in serverless computing and analyzes the requirements for its efficient architecture. Specifically, it explores the monitoring and scheduling techniques within existing research in the field. The rest of the paper is structured as follows. In section 2, we investigate and analyze recent research results and related studies, and in section 3, we provide the conclusion.

II. RECENT TRENDS AND ANALYSIS

Serverless computing involves the implementation of microservices within a cloud environment. A service developer can simply leverage resources provided directly by cloud service providers, eliminating the need for developers to prepare infrastructure or configure service runtime environment. It is facilitated by the decomposition of services into smaller, independently operable units, commonly organized at the function level. The functions are also called lambda, or "Function as a Service" (FaaS) because it executes functions on-demand in the cloud, in response to each request. Serverless computing leverages virtualization technologies such as Virtual Machines (VMs) and containers to execute lambda functions providing an abstraction layer that automates the allocation of computation resources.

The functions in serverless computing are designed to be stateless, meaning they do not retain any previous state between executions. The execution process begins with the creation of an

instance, followed by the preparation of the runtime environment, the execution of the function, and finally, the release of resources. This characteristic brings cost-efficiency by allocating resources on an on-demand basis, eliminating the waste associated with idle resources. However, this efficiency comes with its set of challenges. The unpredictable execution of lambda functions means that the corresponding lambda instances must be allocated dynamically, which can lead to inefficiencies in scenarios where the functions are called repeatedly. The overhead of provisioning and releasing resources as well as the runtime environment for each function call can introduce resource inefficiency and latency in function execution. Since each invocation of a lambda function potentially requires a fresh lambda instance and a runtime – the provisioning process is commonly referred to as "cold start" – it can significantly impact the overall execution performance. To address these challenges, various studies are actively conducted on caching or pre-provisioning of lambda instances. The primary aim of these studies is to mitigate the frequency of cold starts by employing warm-up policies that either cache lambda instances or preload instances set to execute shortly.

In a study [10] that aims to provide lower latency and better resource efficiency to serverless computing, the authors proposed application-level sandboxing and a hierarchical message bus for lambda functions. The application-level sandboxing uses containers to isolate different applications but lighter mechanisms like processes to isolate functions within the same application. Hierarchical message queues analyze lambda functions' locality to ensure lambda instances are deployed closer to the client application.

In [11], the authors proposed a lightweight container scheme, extending the Zygote process mode, originally utilized on Android, to enable the rapid forking of new containers from pre-initialized template containers (Zygotes) that have preloaded common language runtimes and libraries. They identified that the existing layered file systems and per-container network namespaces were creating scalability bottlenecks. To address these challenges, the authors utilized bind mounts for shared filesystems, UNIX sockets for shared networks, and a reusable cgroup pool to enhance resource management.

Nuka [12] addressed the cold start problem in existing serverless platforms that use Docker containers through an isolation pool that reuses Linux's isolation configurations, local package caching to avoid pulling remote images, and a self-adaptive container reuse strategy to reduce the frequency of cold starts. The concept of lightweight pause container was proposed to reuse Linux namespaces and cgroups to scalability bottleneck associated with creating and destroying Docker containers in high concurrency scenarios.

Recent trends aim to enhance the accuracy of workload predictions to mitigate cold starts. The SSC framework [13] leverages gradients derived from workload analysis to pre-allocate instances or reclaim resources by disposing of them accordingly. Another study [14] explored the feasibility of machine learning-based serverless workload prediction, demonstrating the effectiveness of LSTM models in forecasting time-series workloads. The MASTER framework [15] explored the use of machine learning and deep learning models to predict cold start latency. Upon evaluating the performance of various predictive models, it was reported that the XGBoost model exhibited the best performance in both single-stage and multi-stage latency predictions.

III. Conclusion

Serverless computing has improved the inefficiencies from traditional cloud computing paradigm. However, there is still room for further improvement, and various research efforts are ongoing in this area. One of the major issues that need to be resolved is the cold start problem. Researches have been conducted exploring lightweight containers and caching techniques. Another recent trend is the use of machine learning to predict cold start latency. It has shown the potential to mitigate cold start latency. Accurate workload prediction includes precisely monitoring the current workload. However, it is challenging to estimate the underlying resource usage caused by dependencies at the instance or function level. In the future, we plan to conduct research on precise resource monitoring and understand its impact on workload prediction

References

[1] K. Kumar and Y. -H. Lu, “Cloud Computing for Mobile Users: Can Offloading Computation Save Energy?,” in Computer, vol. 43, no. 4, pp. 51-56, April 2010.

[2] E. van Eyk, L. Toader, S. Talluri, L. Versluis, A. Uță and A. Iosup, "Serverless is More: From PaaS to Present Cloud Computing," in IEEE Internet Computing, vol. 22, no. 5, pp. 8-17, Sep./Oct. 2018.

[3] AWS Lambda, [Online]. Available: https://aws.amazon.com/pm/lambda/

[4] Google Cloud Functions,
[Online]. Available: https://cloud.google.com/functions

[5] Microsoft Azure, [Online]. Available: https://azure.microsoft.com

[6] Kubeless, [Online]. Available: https://github.com/vmware-archive/kubeless

[7] OpenFaaS, [Online]. Available: https://github.com/openfaas/faas

[8] Apache OpenWhisk,
[Online]. Available: https://github.com/apache/openwhisk

[9] Open Lambda, [Online].
Available: https://github.com/open-lambda/open-lambda

[10] I. E. Akkus et al., “SAND: Towards High-Performance Serverless Computing,” in 2018 USENIX Annual Technical Conference (USENIX ATC 18), Jul. 2018, pp. 923–935.

[11] E. Oakes et al., “SOCK: Rapid Task Provisioning with Serverless-Optimized Containers,” in 2018 USENIX Annual Technical Conference (USENIX ATC 18), Jul. 2018, pp. 57–70.

[12] S. Qin et al., "Nuka: A Generic Engine with Millisecond Initialization for Serverless Computing," 2020 IEEE International Conference on Joint Cloud Computing, Oxford, UK, 2020, pp. 78-85.

[13] S. Pan, H. Zhao, Z. Cai, D. Li, R. Ma, and H. Guan, “Sustainable Serverless Computing with Cold-start Optimization and Automatic Workflow Resource Scheduling,” IEEE Transactions on Sustainable Computing, pp. 1–12, 2023.

[14] N. T. Noble, Y. P. Dev and C. T. Joseph, "Machine Learning Based Techniques for Workload Prediction in Serverless Environments," 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM), Roorkee, India, 2023, pp. 1-6.

[15] M. Golec et al., "MASTER: Machine Learning-based Cold Start Latency Prediction Framework in Serverless Edge Computing Environments for Industry 4.0," in IEEE Journal of Selected Areas in Sensors, early access, 2024

Short-Packet Based Power Allocation Using Quantum Neural Network in Downlink NOMA Systems for Military Communication

Won Jae Ryu
ICT Convergence Research Center
Kumoh national Institute of Technology
Gumi, Republic of Korea
wj0828@kumoh.ac.kr

Dong-Seong Kim*
Dept. of IT Convergence Engineering
Kumoh national Institute of Technology
Gumi, Republic of Korea
dskim@kumoh.ac.kr

***Abstract*—The integration of the Internet of Things (IoT) into military operations represents a transformative advancement in leveraging technology for strategic military purposes. This study introduces a cutting-edge approach that enhances IoT functionalities in military applications by utilizing Non-Orthogonal Multiple Access (NOMA), short-packet communications, and Quantum Neural Networks (QNNs) for optimization tasks. With the imminent deployment of 6G networks, which promise enhanced connectivity and support for real-time, complex data analytics and machine learning, our research focuses on utilizing these advancements to improve communication systems crucial for military operations. We present a novel QNN-based power allocation strategy within NOMA systems, optimized specifically for the conditions of military communications within the Finite Block Length (FBL) regime. This method addresses the critical need for real-time radio resource optimization, while also ensuring massive connectivity in dynamic military environments. By effectively managing power distribution among NOMA user pairs, our approach aims to maximize spectral efficiency and enhance user connectivity, ensuring robust, reliable, and secure communications in line with the performance metrics required by military operations. Our findings indicate that the QNN-enhanced NOMA framework offers improvements in energy efficiency, delay reduction, and reliability—key factors in the operational success of military IoT systems in the 6G era.**

***Index Terms*—Downlink, non-orthogonal multiple access, quantum neural network, finite blocklength regime**

I. Introduction

In an era dominated by rapid technological advancements, the integration of the Internet of Things (IoT) into military applications has emerged as a pivotal development, driven by escalating demands for data-driven decision-making and enhanced computational capabilities [1], [2]. The need for appropriate devices, software configurations, and robust infrastructure has become increasingly urgent, especially as massive amounts of IoT applications are deployed on cloud platforms, requiring efficient solutions to manage these complexities [3], [4].

As military operations increasingly depend on high connectivity and rapid communication, traditional IoT systems must evolve to meet these critical needs. This study introduces a sophisticated approach that enhances IoT performance in military contexts by incorporating Non-Orthogonal Multiple Access (NOMA), short-packet communications, and optimizations through Quantum Neural Networks (QNNs).

The advent of 6G networks marks a significant shift towards connected intelligence, enabling not just enhanced machine-to-machine and person-to-person connections but also supporting complex data analytics and machine learning processes in real-time. These capabilities are crucial in supporting advanced services like virtual reality (VR), augmented reality (AR), mixed reality (MR), and the tactile Internet, which are becoming increasingly relevant in military operations for training and real-time tactical decision-making [5], [6]. Furthermore, the burgeoning field of quantum technologies provides unique advantages in optimizing radio resources in real-time, which is essential for the high data requirements and dynamic environments typical of 6G networks [6].

Moreover, Quantum Neural Networks (QNNs) represent an innovative amalgamation of artificial neural networks and quantum information processing principles, offering breakthrough potentials in computational speed and problem-solving capabilities that are particularly beneficial in optimizing complex systems [7]–[10].

Military communication systems require robust and secure transmission capabilities to support essential mission domains. The efficiency of IoT applications, particularly in terms of energy consumption and reliable data transmission under stringent conditions, becomes critical [11], [12]. Utilizing IoT in military scenarios not only demands adherence to strict performance metrics such as delay, reliability, and security classification but also mandates a sustainable approach to energy usage, given the often resource-constrained environments of military operations [13]. Furthermore, we explore the role of quantum computing in solving complex optimization problems inherent in military applications, aligning our use-cases with the strategic agenda of advanced defense technologies [16].

To address these challenges, our research leverages the potential of Quantum Neural Networks (QNNs) to optimize power allocation in NOMA systems, a method that markedly enhances spectral efficiency and supports dense user connec-

tivity crucial for military IoT applications [15]. Specifically, this paper focuses on the application of QNNs to optimize power distribution among NOMA user pairs within the Finite Block Length (FBL) regime [17]–[21]. This approach realistically considers the constraints associated with transmission times and decoding error probabilities, crucial for ensuring reliable communication in military scenarios [22].

II. Proposed System Model

As outlined in [11], military applications demand high reliability, energy efficiency, and minimal latency. To address these requirements, it is proposed to employ FBL-based power allocation through QNN in downlink NOMA systems. This section delineates the system model and introduces the proposed power allocation scheme leveraging QNNs within the FBL regime. The system model for the downlink NOMA setup comprises a Near User (NU), a Far User (FU), and a Base Station (BS), all operating in a Rayleigh fading channel. The BS is designed to simultaneously transmit superposed signals to both NU and FU. Due to the stronger signal strength intended for FU compared to NU, NU employs Successive Interference Cancellation (SIC) to eliminate the signals directed at FU before decoding its own. Conversely, FU directly decodes its signals without necessitating SIC, owing to its stronger signal strength. Nonetheless, NU's signals serve as interference to FU, affecting its Signal-to-Interference-plus-Noise Ratio (SINR). The SINRs x_n and x_f for NU and FU are mathematically represented as follows:

$$x_n = \frac{p_n |h_n|^2}{N_{0,n}}, \qquad x_f = \frac{p_f |h_f|^2}{(p_n |h_f|^2 + N_{0,n})}. \tag{1}$$

Here, the power ratios p_n and p_f satisfy the condition:

$$1 = p_n + p_f. \tag{2}$$

Addressing the FBL regime, it is employed to ascertain the block length, indicative of the radio resource blocks allocated for transmission. The block length calculation is provided by:

$$n(x, \epsilon, T) = \left(\frac{\sqrt{4TC(x)Q(\epsilon)^2 + V(x)} + \sqrt{V(x)}}{2C(x)Q(\epsilon)} \right)^2, \tag{3}$$

where n denotes the block length, Q represents the Q-function, ϵ is the target error rate, T is the packet size, $C(x) = \log_2(1 + x)$ signifies the conventional capacity, and $V(x) = 1 - \frac{1}{(1+x)^2}$ denotes the channel dispersion.

Based on the above, the optimal power ratio p is determined by minimizing the total block length for both NU and FU, as given by:

$$p^* = \underset{p}{\operatorname{argmin}} \left[n(x_n, \epsilon, T_n) + n(x_f, \epsilon, T_f) \right], \tag{4}$$

where the objective function is to minimize the aggregate block lengths for both users. Utilizing this formulation, datasets to train and evaluate the QNN model shown in Figure 1 are generated, with the optimal power ratio p^* identified through iterative search. Upon dataset creation, the QNN model undergoes training. The model utilizes three qubits as inputs: q_0 and q_1 corresponding to the channel gains h_n and h_f, respectively, and q_2 representing the normalized noise. In the QNN framework, a qubit is defined by the superposition state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, with the state collapsing upon measurement based on α and β. The model employs a Variational Quantum Regressor (VQR) approach [23], optimizing parameterized quantum gates against given quantum states to derive the power allocation factor α.

VQR is a type of quantum machine learning algorithm designed to tackle regression problems using a quantum computer. Regression tasks involve predicting continuous outcomes based on input variables. VQR utilizes a variational approach, employing parameterized quantum circuits, also known as Variational Quantum Circuits (VQC), to model and predict continuous data [24], [25]. It leverages the inherent capabilities of quantum computers to efficiently handle high-dimensional data. This attribute is particularly beneficial for managing large-scale or complex datasets. The optimization strategies used in VQR can be combined with classical algorithms, facilitating an efficient quantum-classical hybrid approach. This hybrid methodology optimally utilizes the strengths of quantum algorithms while maintaining the stability and scalability of classical computing. The VQR model consists of 1 ZZ feature map and 1 Real amplitudes. The ZZ feature map consists of three ZZ gates, and the Real amplitudes circuit is composed of 15 RY gates arranged in five layers, with three gates on each layer. ZZ feature map is to encode input datas to quantum states that utilizes ZZ gates to create entanglement between qubits. Real amplitudes circuit is used to parameterize rotation gates on each qubit, making it easier to optimize the parameters of the circuit.

To optimize the parameters in the VQR model, constrained optimization by linear approximation (COBYLA) optimizer is adopted [26], [27]. COBYLA is a numerical optimization algorithm designed to solve constrained optimization problems without requiring derivative information of the objective function. This derivative-free characteristic makes it particularly valuable for functions where derivatives are difficult or impossible to compute. COBYLA is best suited for problems with a small number of variables and effectively handles both continuous, smooth, or even non-smooth objective functions and constraints. The algorithm operates by using linear approximations to model the objective function and constraints at each iteration, generating new trial points based on these approximations. In the case of quantum computing, particularly in Variational Quantum Algorithms (VQA), COBYLA is used to optimize parameters effectively.The non-requirement of derivatives makes COBYLA especially beneficial for applications like quantum circuits, where outputs can be complex and difficult to differentiate. This attribute helps in effectively navigating the parameter space to minimize the objective function under given constraints. This sophisticated method showcases the potential of QNNs and the FBL regime in enhancing downlink NOMA system performance through



Fig. 1. Quantum Circuit

TABLE I
SIMULATION PARAMETERS

Qiskit Version	1.0.1
Quantum Neural Network Model	Variational Quantum Regressor
Quantum Encoder	ZZFeatureMap (3 Inputs)
Parameterized Quantum Circuit	RealAmplitudes (3 Gates $\times$ 5 Layers)
Optimizer	COBYLA
Iterations	250

dynamic and efficient power allocation.

III. SIMULATION RESULTS

This section details the training outcomes of the Quantum Neural Network (QNN) model employed for power allocation in the downlink Non-Orthogonal Multiple Access (NOMA) system. The model used is the Variational Quantum Regressor (VQR), which is tasked with finding a sub-optimal power allocation factor between the Near User (NU) and the Far User (FU). To evaluate the performance of the VQR model, a dataset specifically generated for power allocation tasks was utilized. Figure 2 illustrates the progression of training, capturing the model's performance across successive iterations.

As depicted in Figure 2, there is a noticeable decrease in the value of the objective function with each iteration, demonstrating the VQR model's capability to optimize the power allocation effectively in a downlink NOMA system.

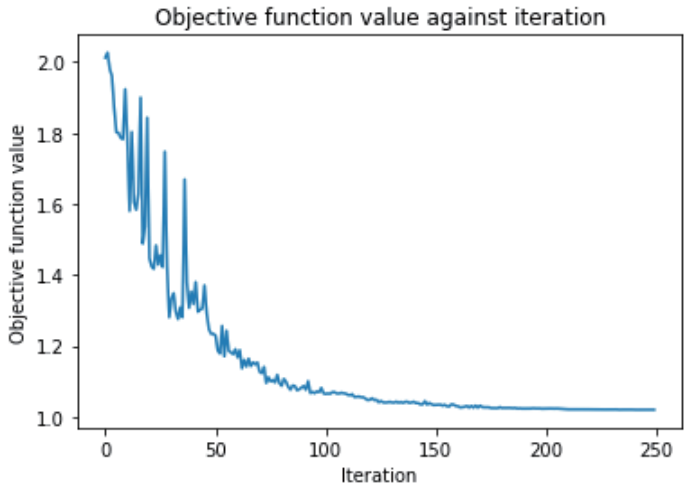


Fig. 2. Training results

This decreasing trend is indicative of the model's ability to progressively refine its predictions and enhance its accuracy in power distribution between users, which is critical for the efficiency and reliability of communication systems.

The simulation results underscore the QNN model's effectiveness in solving complex optimization problems inherent to power allocation in NOMA systems. The observed trend in the objective function value suggests not only the model's capacity for improvement but also its potential to converge towards an optimal solution. Furthermore, these results highlight the practical applicability of quantum machine learning models like VQR in telecommunications, where dynamic and efficient resource allocation is vital.

Additionally, the robustness of the VQR model in handling real-world scenarios is reflected in its ability to adapt and learn from a specially curated dataset that mimics actual operating conditions. This adaptation is critical as it suggests that quantum-enhanced machine learning can move beyond theoretical applications to practical, real-world implementations in complex network environments. The continual decrease in the objective function with each training iteration reassures the potential of quantum computing techniques to provide substantial improvements over traditional algorithms, especially in systems where the allocation of resources needs to be dynamically adjusted based on fluctuating network demands and user requirements.

Overall, the extended analysis of the VQR model within this simulation framework not only demonstrates its effectiveness but also reinforces the viability of integrating quantum computing approaches into the realm of network optimization and telecommunications.

IV. CONCLUSION

This study has successfully demonstrated the integration of the Variational Quantum Regressor (VQR) within a downlink Non-Orthogonal Multiple Access (NOMA) system for optimizing power allocation, addressing key military application needs such as high reliability, energy efficiency, and minimal latency. The employment of Quantum Neural Networks (QNNs) in the sophisticated military communication landscape signifies a substantial enhancement in handling dynamic resource allocation effectively and efficiently. The use of the VQR in the NOMA system marks an advancement in the field of quantum machine learning, leveraging quantum computational capabilities to tackle complex optimization problems in real-time communication scenarios. This integration aligns with the strategic demands of advanced military technologies, which require robust, secure, and efficient communication systems to operate in increasingly complex environments. Our simulation results have convincingly shown that the QNN model, specifically the VQR, not only meets but exceeds expectations in refining power distribution strategies between users, marked by a consistent decrease in the objective function value with each iteration. This trend underscores the model's capacity to converge towards optimal solutions, highlighting its potential to significantly outperform traditional

optimization methods in critical communication systems. Furthermore, the adaptability of the VQR model to real-world conditions, as demonstrated by its ability to learn and optimize from a specially curated dataset, suggests a promising avenue for future deployments in other complex network environments. The ability of the VQR to capture and utilize non-linear patterns and complex correlations within data further enhances its utility in military scenarios where quick and reliable decision-making is paramount. In conclusion, this research not only substantiates the effectiveness of integrating advanced quantum computing techniques into network optimization but also sets the stage for further exploration into their practical applications in military and civilian telecommunications. The positive outcomes observed in the NOMA system model advocate for broader implementation and testing of quantum-enhanced machine learning models across various domains, potentially revolutionizing how communication networks are optimized for performance and reliability. By pushing the boundaries of quantum computing and telecommunications, this study contributes to the foundational knowledge required to drive forward the next generation of military communication systems, ensuring that they are capable of meeting the stringent demands of modern warfare and security.

ACKNOWLEDGMENT

This work was supported in part by the Ministry of Science and ICT (MSIT), Korea, under the Innovative Human Resource Development for Local Intellectualization program, supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP) (IITP-2024-2020-0-01612, 33%) in part by the Priority Research Centers Program (2018R1A6A1A03024003, 33%), and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (2022R1I1A1A01069334, 34%).

REFERENCES

[1] E. P. DeBenedictis, “A future with quantum machine learning,” Computer, vol. 51, no. 2, pp. 68–71, Feb. 2018.

[2] A. Shahini and N. Ansari, ”NOMA Aided Narrowband IoT for Machine Type Communications With User Clustering,” in IEEE Internet of Things Journal, vol. 6, no. 4, pp. 7183-7191, Aug. 2019.

[3] N. Y. Philip, J. J. P. C. Rodrigues, H. Wang, S. J. Fong and J. Chen, ”Internet of Things for In-Home Health Monitoring Systems: Current Advances, Challenges and Future Directions,” in IEEE Journal on Selected Areas in Communications, vol. 39, no. 2, pp. 300-310, Feb. 2021.

[4] S. Kumar, P. Tiwari and M. Zymbler, ”Internet of Things is a revolutionary approach for future technology enhancement: A review”, J. Big Data, vol. 6, no. 1, Dec. 2019.

[5] E. A. Tuli, J. -M. Lee and D. -S. Kim, ”Integration of Quantum Technologies into Metaverse: Applications, Potentials, and Challenges,” in IEEE Access, vol. 12, pp. 29995-30019, 2024.

[6] T. Q. Duong, L. D. Nguyen, B. Narottama, J. A. Ansere, D. V. Huynh and H. Shin, ”Quantum-Inspired Real-Time Optimization for 6G Networks: Opportunities, Challenges, and the Road Ahead,” in IEEE Open Journal of the Communications Society, vol. 3, pp. 1347-1359, 2022.

[7] Altaisky, M. V. ”Quantum neural network.” arXiv preprint quant-ph/0107012 (2001).

[8] S. Choudhury, A. Dutta and D. Ray, ”Chaos and complexity from quantum neural network. a study with diffusion metric in machine learning”, J. High Energy Phys., vol. 2021, no. 4, pp. 1-33, 2021.

[9] N. H. Nguyen, E. C. Behrman, M. A. Moustafa and J. E. Steck, ”Benchmarking neural networks for quantum computations”, IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 7, pp. 2522-2531, Jul. 2020.

[10] N. H. Nguyen, E. C. Behrman, M. A. Moustafa and J. E. Steck, ”Benchmarking neural networks for quantum computations”, IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 7, pp. 2522-2531, Jul. 2020.

[11] R. Bajracharya, R. Shrestha, S. A. Hassan, H. Jung and H. Shin, ”5G and Beyond Private Military Communication: Trend, Requirements, Challenges and Enablers,” in IEEE Access, vol. 11, pp. 83996-84012, 2023

[12] V. Gotarane and S. Raskar, ”IoT Practices in Military Applications,” 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 891-894.

[13] M. Tortonesi et al., ”Leveraging Internet of Things within the military network environment — Challenges and solutions,” 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), Reston, VA, USA, 2016, pp. 111-116.

[14] W. J. Ryu, J.-W Kim, and D.-S. Kim, ”Resource allocation in downlink VLC-NOMA systems for factory automation scenario.” Sensors 22.23 (2022): 9407.

[15] S. Bhardwaj, D.-H. Kim and D.-S. Kim, ”Backscatter-enabled CR-NOMA based cooperative V2X communication with imperfect CSI”, Veh. Commun., vol. 42, Aug. 2023.

[16] N. M. P. Neumann, M. P. P. van Heesch, F. Phillipson and A. A. P. Smallegange, ”Quantum Computing for Military Applications,” 2021 International Conference on Military Communication and Information Systems (ICMCIS), The Hague, Netherlands, 2021, pp. 1-8.

[17] Y. Polyanskiy, H. V. Poor and S. Verdu, ”Channel Coding Rate in the Finite Blocklength Regime,” in IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307-2359, May 2010

[18] N. Hesham, A. Chaaban, H. ElSawy and M. J. Hossain, ”Finite Blocklength Regime Performance of Downlink Large Scale Networks,” in IEEE Transactions on Wireless Communications, vol. 23, no. 1, pp. 479-494, Jan. 2024.

[19] Y. Hu, M. Ozmen, M. C. Gursoy and A. Schmeink, ”Optimal Power Allocation for QoS-Constrained Downlink Multi-User Networks in the Finite Blocklength Regime,” in IEEE Transactions on Wireless Communications, vol. 17, no. 9, pp. 5827-5840, Sept. 2018.

[20] R. Ma, W. Yang, L. Tao, X. Lu, Z. Xiang and J. Liu, ”Covert Communications With Randomly Distributed Wardens in the Finite Blocklength Regime,” in IEEE Transactions on Vehicular Technology, vol. 71, no. 1, pp. 533-544, Jan. 2022.

[21] W. J. Ryu, D. Seong Kim and S. Y. Shin, ”Bio-inspired Relay Selection Algorithm Using Finite Blocklength Regime for URLLC,” 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2020, pp. 955-957

[22] Ryu, Won Jae, and Soo Young Shin. ”Performance evaluation of a power allocation algorithm based on dynamic blocklength estimation for URLLC in the multicarrier downlink NOMA systems.” Turkish Journal of Electrical Engineering and Computer Sciences 29.1 (2021): 310-320.

[23] Qi, Jun, et al. ”Theoretical error performance analysis for variational quantum circuit based functional regression.” npj Quantum Information 9.1 (2023): 4.

[24] Jun Qi, Chao-Han Huck Yang, Pin-Yu Chen, and Min-Hsiu Hsieh, “Theoretical error performance analysis for variational quantum circuit based functional regression,” npj Quantum Information, vol. 9, no. 1, pp. 4, 2023.

[25] R. Huang, X. Tan and Q. Xu, ”Learning to Learn Variational Quantum Algorithm,” in IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 11, pp. 8430-8440, Nov. 2023.

[26] Mateusz Ostaszewski et al., ”Reinforcement learning for optimization of variational quantum circuit architectures”, Advances in Neural Information Processing Systems, vol. 34, 2021.

[27] T. Miki, D. Tsukayama, R. Okita, M. Shimada and J. -i. Shirakashi, ”Variational Parameter Optimization of Quantum-classical Hybrid Heuristics on Near-term Quantum Computer,” 2022 IEEE International Conference on Manipulation, Manufacturing and Measurement on the Nanoscale (3M-NANO), Tianjin, China, 2022, pp. 415-418.

Federated Learning-Based Joint Radar-Communication mmWave Beamtracking for V2X-Communications

1st Sanjay Bhardwaj, 2nd Dong-Seong Kim
ICT Convergence Research Center, Department of IT Convergence Engineering
Kumoh National Institute of Technology, Gumi, South Korea
sanjay.b1976@gmail.com, dskim@kumoh.ac.kr

***Abstract*—Robust beamforming in millimeter-wave (mmWave) communication, vital for Vehicle-to-everything communication, faces challenges like short-range links due to path loss and obstacles. Privacy concerns and bandwidth constraints make transmitting the entire dataset impractical. To tackle this, a federated learning (FL)-based joint radar and communication mmWave beamtracking approach, FL-JRC, is proposed. FL-JRC employs a federated convolutional neural network (CNN) for vehicle users (local clients) and road side user (RSU) (main server) to estimate arrival/departure angles, optimizing beam tracking. Simulation results demonstrate superior performance compared to contemporary methods and two Kalman filter versions (extended and unscented).**

***Index Terms*—Federated Learning, V2X communications, Beamtracking, mmWave, Joint Radar-Communication.**

I. Introduction

Federated learning (FL) assisted mmWave communications and radar are crucial for next-generation applications, including vehicle-to-everything communications (V2X-C), as they provide high-resolution sensing with a wide field of view, along with high-data-rate solutions for connected vehicles [1]. Combination these two technologies into a joint radar and communication (JRC), hardware can be reused and a similar signaling waveform is used, leading to advantages in power usage, spectrum efficiency [2], [3].

Dual-functional radar-communication (DFRC), is a cost-efficient solution for combining communication and radar functionalities. Authors of [4] proposed an approach called as JRC-PB, where beamtracking benefited from the DFRC transmission system's signal-to-noise ratio (SNR) gain, but without considering Doppler shift (DS). Previously, mmWave beamtracking techniques for V2X-C relied on communication-only feedback-based protocols, which had high overhead due to pilot symbols and feedback procedures [5]. To address this, a DFRC-assisted beamtracking design was proposed for V2X-C by [6], [7], which used echoes from the vehicles to estimate their angle of arrival/departure (AoA/AoDs) and distance.The state space model based on extended Kalman filter (EKF) by [8] uses path gains and AOA/AoD as parameters. But EKF converts a nonlinear model into a linear approximation using current state estimates, covariance, and transformation, and thus its performance varies due to non-consistency in the Jacobian matrices, and it has a significant computing burden. The unscented Kalman filter (UKF) overcomes these limitations by accurately propagating means and covariance through nonlinearity [9], but has similar computational complexity to the EKF and does not guarantee convergence beyond local stability as well as in constrained conditions.

To improve precision in V2X-C, a FL-based mmWave beamtracking technique based on cooperative radar communication, called FL-JRC is proposed. The approach evaluates DS and AoA/AoDs parameters simultaneously, without requiring pilots or feedback loops, using federated convolutional neural network (CNN) modules for vehicle users and road side users (RSUs). A detailed comparison with auxiliary beam pair (ABP) approach [10], JRC-PB [4], EKF, and UKF beamtracking methods are discussed.

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (IITP-2024-2020-0-01612, 50%) and Priority Research Centers Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2018R1A6A1A03024003, 50%)

II. System and Radar-Communication Model

A federated CNN as the DL model is used for parameter estimation, Table I. In first Conv1D, $c_l = 1$, the input vectors are $x^{c_l} = y \in \mathbb{R}^{\left[\rho_{c_l-1} \times \varrho_{c_l-1}\right]}$, where ρ and ϱ are the dimensions and vector number, respectively. The layer c_l with one dimension has c_{l-1} input and c_l output vectors, thereby providing $c_l \times c_{l-1}$ convolution kernels and c_l biases for a kernel size of k_{c_l}. The current layer's output vector for the results of the previous layer $x_i^{c_l-1}, i \in \left\{1, ..., \rho_{c_{l-1}}\right\}$ is written as $x_i^{c_l} = f\left(\sum_{j=1}^{\rho_{c_l-1}} x_j^{c_l-1} * k_{i,j}^{c_l} + b_i^{c_l}\right)$, where ReLU is the nonlinear activation function $f(x) = max(0, x)$ and $*$ is convolution operation. The average pooling layer reduces dimensionality by aggregating the average value of the elements present in the vector; thus, for the i^{th} output vector, $x_i^{c_l+1} = down(x_i^{c_l})$. The output vectors are then restricted to a vector by the FC layer. Subsequently, Dense layer d_l encodes the input vector x^{d_l-1} into an output vector (extracted high-level information) given by $x^{d_l} = f\left(W^{d_l} x^{d_l-1} + b^{d_l}\right)$, where $W^{d_l} \in \mathbb{R}^{\rho_{d_l} \times \rho_{d_l-1}}$ and $b^{d_l} \in \mathbb{R}^{\rho_{d_l}}$. Dense layer with the

Softmax activation function yields the normalized likelihood vector $\hat{\varpi}^{\rho} = [\hat{\varpi}_1^{\rho}, ...\hat{\varpi}_l^{\rho}]$, where $\hat{\varpi}_i^{\rho} = \frac{e^{x_i}}{\sum_{j=1}^{l} e^{x_j}}$. Thus, final output is $\hat{\varpi}^{\rho} = \mathbb{F}(y, \beta)$, where β is the network parameter, and $\mathbb{F}(\cdot)$ is the overall CNN-based output.

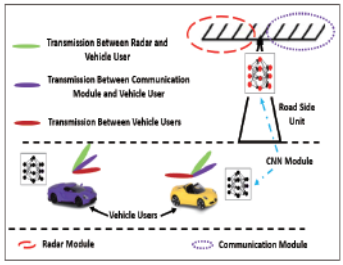


Fig. 1: Elucidation of System Model

Structure	Number of Layers
Input (samples, categories)	–
Conv1D(128,16)+ ReLU + BN + Dropout (0.1)	2
AveragePool	1
Conv1D(64,8)+ ReLU + BN + Dropout (0.1)	1
Fully Connected	1
Dense (1024) + Softmax + Dropout (0.5)	1

TABLE I: CNN Structure Description

Fig. 1 depicts a mmWave MIMO communication system (communication module) assisted by a MIMO radar system (radar module) in a V2X scenario. FL architecture using CNN modules is also shown for each vehicle user (local clients) and at the RSU (main server). For a MIMO radar system with N_t and N_r transmit and receiving antennas, respectively, and ϕ and θ, $\in [0\ \pi]$ are the AoA and AoD, respectively, then for a transmitted signal x_m by the m^{th} transmitted antenna, the target localization can be given as $\sum_{m=1}^{M_t} e^{-j2\pi f_c \tau_m(\theta)} x_m(n) \triangleq \mathbf{a}^*(\theta)\mathbf{x}(n)$, $n = 1, ..., N$, where f_c is the carrier frequency of the radar, $\tau_m(\theta)$ is the time needed by the signal to arrive at the target, $(\cdot)^*$ denotes the conjugate transpose, N are the number of samples of each transmitted pulse, $\mathbf{x}(n) = [x_1(n), x_2(n), ...x_{N_t}(n)]^T$ $\mathbf{a}(\theta) = \left[e^{j2\pi f_c \tau_1(\theta)}, e^{j2\pi f_c \tau_2(\theta)}, ...e^{j2\pi f_c \tau_{N_t}(\theta)}\right]^T$, where $(.)^T$ is the transpose. Assuming that the transmit array of the radar is calibrated, $\mathbf{a}(\theta)$ is a known function of θ. The spatial correlation matrix of the transmitted data sample is $\mathbf{R}_{xx} = E\{x(n)x^H(n)\} \approx \frac{1}{N}\sum_N x(n)x^H(n)$. The received echoed signal received by the m^{th} receive antenna for $\mathbf{y}_k(n) = [y_1(n), y_2(n), ...y_{N_r}(n)]^T$, $n = 1, 2, ..., N$ and $\mathbf{b}(\theta) = \left[e^{j2\pi f_c \tilde{\tau}_1(\theta)}, e^{j2\pi f_c \tilde{\tau}_2(\theta)}, ...e^{j2\pi f_c \tilde{\tau}_{N_r}(\theta)}\right]^T$, where $\tilde{\tau}_m(\theta)$ is the time required for the signal reflected by the target located at θ to arrive at the m^{th} receiving antenna. Thus, for the transmitted DFRC signal $x(n)$ on the n^{th} transmission block, the corresponding echo received data vector is given as

$$\mathbf{y}_k(n) = \sum_{k=1}^{K} \sqrt{p_k}\beta_k \mathbf{b}^c(\theta_k)\mathbf{a}^*(\theta_k)\mathbf{x}_k(n) + \boldsymbol{\gamma}(n), n = 1, 2, ...N, \quad (1)$$

where p is the power associated with n^{th} transmission block, K is the number of vehicle users (targets) that reflect the signal back to the radar receiver, $\{\beta_k\}$ are the complex amplitudes proportional to the radar cross-sections (RCSs) of the target, $\{\theta_k\}$ is the target location parameter, $(\cdot)^c$ is the complex conjugate, and $\boldsymbol{\gamma}(n) \sim \mathscr{CN}\left(0, \sigma_{\gamma}^2 I_M\right)$ denotes the noise term for I_M as a unit matrix. The unknown parameters to be estimated from $\{\mathbf{y}(n)\}_{n=1}^{N}$ were $\{\beta_k\}_{k=1}^{K}$ and $\{\theta_k\}_{k=1}^{K}$.

III. Problem Formulation and Proposed Approach

The signal matrix at the receiving array output for the k^{th} vehicle user, with θ_k as the local parameter (target location), can be expressed as.

$$\mathbf{Y}_k(\theta_k) = \sqrt{p_k}\mathbf{b}^c(\theta_k)\beta(\theta_k)\mathbf{a}^*(\theta_k)\mathbf{X}_k + \mathbf{Z}_k(\theta_k), \quad (2)$$

where $\mathbf{X}_k = [\mathbf{x}_k(1), \mathbf{x}_k(2), ...\mathbf{x}_k(N)]$, the columns of $\mathbf{Y}_k \in \mathbb{C}^{N_r \times N}$ are the reception data samples $\{\mathbf{y}(n)\}_{n=1}^{N}$, and $\beta(\theta_k) \in \mathbb{C}$ denotes the complex amplitude of the reflected signal from θ_k, which is proportional to the RCS of the focal point θ_k. Matrix $\mathbf{Z}_k(\theta_k) \in \mathbb{C}^{N_r \times N}$ denotes the residual term inclusive of noise and interference from targets at locations other than θ_k. For target location θ_k, where $k = 1, 2, ...K$, the spatial spectrum as receive steering vector, transmit steering vector and complex amplitude is given as $\mathbf{A}_r(\theta_k) = [\mathbf{b}(\theta_1), \mathbf{b}(\theta_2), ...\mathbf{b}(\theta_K)]$, $\mathbf{A}_t(\theta_k) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), ...\mathbf{a}(\theta_K)]$ and $\beta(\theta_k) = [\beta(\theta_1), \beta(\theta_2), ...\beta(\theta_K)]$.

Then Eq. (2) can be written as [11]

$$\mathbf{Y}_k(\theta_k) = \sqrt{p_k} diag(\beta(\theta_k))\mathbf{A}_r^c(\theta_k)(\mathbf{A}_t^*(\theta_k)\mathbf{X}_k) + \tilde{\mathbf{Z}}_k(\theta_k), \quad (3)$$

where $diag(\beta(\theta_k))$ is the diagonal matrix and the estimation of $\beta(\theta_k)$ [12] is given as Eq. (4) {at the top of next page}, for ease of representation θ_k is not written, where $vecd(\cdot)$ is the column vector formed by the diagonal elements of the matrix, $\odot$ is the Hadamard product [13], $\hat{\mathbf{R}}_{xx} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^*(n)$ and for the received data samples is $\hat{\mathbf{R}}_{yy} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^*$.

In the presence of K target vehicle users and I_s interference sources, Eq. (3) can be extended to

$$\mathbf{Y}_k(\theta_k) = \kappa + \sum_{k=1}^{K}\eta_k e^{-jf_{d_k}}\mathbf{A}_r^c(\theta_k) + \sum_{i=1}^{I_s}(\mathbf{A}_t^*(\theta_i)\mathbf{X}_i) + \mathbf{w}, \quad (5)$$

where $\kappa = \sqrt{p_k} diag(\beta(\theta_k))\mathbf{A}_r^c(\theta_k)(\mathbf{A}_t^*(\theta_k)\mathbf{X}_k)$, η is the complex attenuation, f_d is the DS and $\mathbf{w}$ is the additive noise vector. Let $\boldsymbol{\eta} \triangleq [\eta_1, ..., \eta_K]^T$, and $\boldsymbol{\iota} \triangleq \left[\boldsymbol{\iota}_i^T, ...\boldsymbol{\iota}_K^T\right]^T$, where $\boldsymbol{\iota} \triangleq [\theta, f_d]^T$, and the interference direction $\boldsymbol{\Theta}_i = \left[\theta_{i_1}, ..., \theta_{i_{I_s}}\right]^T$. Therefore, maximum-Likelihood estimator for the target parameters is given by [14] as $(\hat{\boldsymbol{\iota}}, \boldsymbol{\Theta}_i) = arg\ \max_{\boldsymbol{\iota}, \boldsymbol{\Theta}_i} \mathscr{ML}(\boldsymbol{\iota}, \boldsymbol{\Theta}_i) = \left(\hat{\theta}, \hat{f}_d\right) = arg\ \max_{\theta, f_d}$ $\left(\hat{\theta}, \hat{f}_d\right) = arg\ \max_{\theta, f_d}\left(\left\{\frac{\left|\mathbf{A}_r^H(\theta_k)\mathbf{A}_t^*(\theta_k)\mathbf{A}_t^H(\theta_k)\right|^2}{N_r\mathbf{A}_t^H(\theta_k)\mathbf{A}_t(\theta_k)}\right\}\{\mathbb{E}(f_d)\}\right)$, where $\mathbb{E}(f_d) = \hat{f}_d$ {explained in next paragraph}. Thus, the received communication signal at the l^{th} vehicle user is

$$\mathbf{Y}_l^V(\phi_l) = \sqrt{p_l}\mathbf{b}^c(\phi_l)\beta(\phi_l)\mathbf{v}(\phi_l)\mathbf{a}^*(\theta_l)\mathbf{X}_l + \mathbf{Z}_l(\phi_l), \quad (6)$$

where $\mathbf{v}(\phi_l)$ is the receive steering vector of the vehicle's ULA antenna, ϕ is the AoA of the DFRC signal, $\mathbf{Z}_l$ is the additive white Gaussian noise (AWGN) with variance σ_l. The received SNR [2] for the vehicle users is given as $\varsigma^V = \frac{\zeta|\chi|^2 p}{\sigma_l^2}$. where $\chi = \mathbf{b}^c(\phi_l)\beta(\phi_l)\mathbf{v}(\phi_l)\mathbf{a}^*(\theta_l)$ and ζ is the array gain. Thus, the rate of vehicle users can be written as $Th^V = log(1+\varsigma^V)$.

Authors in [15] established the original DS estimation as $\hat{f}_d = f_d\sqrt{\frac{1-\sqrt{\frac{\varsigma^V}{1+\varsigma^V}}J_0(2\pi f_d S_t)}{1-\sqrt{J_0(2\pi f_d S_t)}}}$, where two sampling intervals, $S_{t_1} = AS_t$ and $S_{t_2} = BS_t$, A, B are integers, and $A < B$.

$$\hat{\beta} = \frac{1}{N}\left[\left(\mathbf{A}_r^T\left[N\hat{\mathbf{R}}_{yy} - \frac{1}{N}\mathbf{Y}\mathbf{X}^*\mathbf{A}_t\left(\mathbf{A}_t^*\hat{\mathbf{R}}_{xx}\mathbf{A}_t\right)^{-1}\mathbf{A}_t^*\mathbf{X}\mathbf{Y}^*\right]^{-1}\mathbf{A}_r^c\right)\odot\left(\mathbf{A}_t^T\hat{\mathbf{R}}_{xx}^c\mathbf{A}_t^c\right)\right]^{-1} vecd\left(\mathbf{A}_r^T\left[N\hat{\mathbf{R}}_{yy} - \frac{1}{N}\mathbf{Y}\mathbf{X}^*\mathbf{A}_t\left(\mathbf{A}_t^*\hat{\mathbf{R}}_{xx}\mathbf{A}_t\right)^{-1}\mathbf{A}_t^*\mathbf{X}\mathbf{Y}^*\right]^{-1}\mathbf{Y}\mathbf{X}^*\mathbf{A}_t\right), \quad (4)$$

Two equations are based on two original DS estimations in which, a fourth order approximation of the Bessel function J_0 is adopted to enable the DS estimator to be practicable in high-speed scenarios, as $\hat{f}_{d_1} = \sqrt{\frac{2-\sqrt{\frac{\varsigma^V}{1+\varsigma^V}\left(2-(A\pi f_d S_t)^2\right)}}{(A\pi S_t)}}$, and the estimation is given as $\hat{f}_d = \frac{1}{\pi}\sqrt{\frac{2\left(\frac{2-\left(A\pi f_{\hat{d}_1} S_t\right)^2}{2-\left(B\pi f_{\hat{d}_2} S_t\right)^2}-1\right)}{\frac{2-\left(A\pi f_{\hat{d}_1} S_t\right)^2}{2-\left(B\pi f_{\hat{d}_2} S_t\right)^2}B^2S_t^2 - A^2S_t^2}}$, with the condition that the sampling rate must be twice the bandwidth and the inequality $f_{d_m} < \frac{0.5}{max(A,B)}\left(F_{d_m} = f_d S_t\right)$ is satisfied, where f_{d_m} is maximum value of normalized DS. **Federated Learning Explication:** A wireless multi-user system includes one server (R) and C local clients (vehicle users), where the total data size (T_d) for each client, c, with data size d_c, is given by $T_d = \sum_{c=1}^{C} d_c$. The set of input-output pairs for data samples d_c is defined as $\{a_j, b_j\}_{j=1}^{d_c}$, where $a_j \in \mathbb{R}^d$ and $b_j \in \mathbb{R}$ represent input and output vector, respectively, for sample a_j. The output characterization of a global model parameter G_m for input a_j with loss function $\mathscr{L}(f_j(G_m))$ can be represented as follows: **A)** For linear regression: $\mathscr{L}(f_j(G_m)) = \frac{1}{2}\left(a_j^T(G_m) - b_j\right)^2, b_j \in \mathbb{R}$. **B)** For support vector machine: $\mathscr{L}(f_j(G_m)) = \left\{0, 1 - b_j a_j^T(G_m)\right\}, b_j \in \{-1, 1\}$, where $a_j^T(G_m)$ is linear mapping function. Therefore, the global optimization problem for G_m is $\underset{G_m \in \mathbb{R}^d}{min}\, \mathscr{P}(G_m) = \sum_{c=1}^{C} \frac{d_c}{T_d}\mathscr{L}(\mathscr{P}_c(G_m))$, where $\mathscr{L}(\mathscr{P}_c(G_m)) = \frac{1}{d_c}\sum_{j=d_c}\mathscr{L}(f_j(G_m))$ is the local loss function for user c and $\mathscr{L}(f_j(G_m))$ captures the model parameter error on the input-output pair $\{a_j, b_j\}$. The solution is obtained by an iterative FL-based approach indexed by I with three steps: **1). Scheduling of participants:** The main server, R selects the initial model parameters G_{m_p} and schedules R_d for its training and then broadcasts these details to each client participant $C_{p_i}, i \in 1, 2..N_c$, where N_c are a number of clients for an iterative round I. C_{p_i} receives the current global model parameter $G_{m_{p_{I-1}}}^{S_{d_{I-1}}}$ from S on the previous scheduling decisions given as $R_{d_{I-1}} = [R_{d_1}, R_{d_2}, ...R_{d_{I-1}}]$. **2). Schedule sharing:** Each C_{p_i} has its own local data set, thus using the gradient descent algorithm updates its local model parameter, $\tilde{G}_{m_p} \in G_{m_p}$ for a learning rate of μ, as $\tilde{G}_{m_p}(I+1) = \tilde{G}_{m_p}(I) - \mu\tilde{\nabla} f_j\left(G_{m_p}(I)\right)$, where for a mini-batch $d_{C_p} \subseteq D_{C_p}$ of $C_p \in c$ data samples, the estimate of the gradient $\nabla f_j\left(G_{m_p}(I)\right)$ is given as $\tilde{\nabla} f_j\left(G_{m_p}(I)\right) = \frac{1}{|d_{C_p}|}\sum_{j \in d_{C_p}} \nabla\mathscr{L}(f_j\left(G_{m_p}(I)\right))$. The exchange of the local model parameter is scheduled and repeated α times, represented as $f_j(G_{m_p}(\alpha))$ is uploaded and shared to R. **3). Update global model:** The weighted average combination of the received local parameters is carried out by R, which aggregates them to obtain a new global model parameter, which is represented as $G_{m_p}(I)^{R_{d_I}} = \frac{\sum_{c \in R_{d_I}} d_c G_{m_p}(I)^{R_{d_I}}}{\sum_{c \in R_{d_I}} d_c}$.

IV. Performance Evaluation

The proposed approach is compared with ABP and JRC-PB in terms of average throughput, Fig. 2 at SNR = 0 dB. The FL-JRC significantly outperforms contemporary approaches as it is able to provide consistent better throughput, which clearly shows its competency and capability.

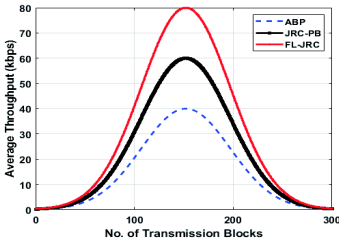


Fig. 2: Average Throughput Comparison

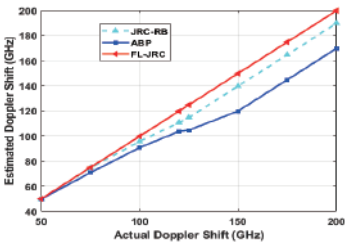


Fig. 3: DS Estimation Comparison

A comparison is drawn between the FL-JRC, ABP, and JRC-RB in terms of the actual and estimated DS estimation Fig. 3. The estimated DS meticulously pursues the actual DS for the proposed approach as compared to contemporary approaches, thereby establishing the superior accomplishments and advantages of the proposed approach.

The mean AoD estimation error was compared between the proposed approach, ABP, and JRC-RB, Fig. 4. For FL-JRC, the angle estimation error not only starts from the lower value but also shows an exponential decrease, as compared with the other two approaches, which present an almost horizontal decrease as the number of transmit antennas increases. This validates and establishes the efficacy of the proposed method in overcoming impairments caused by amalgamation of DS.

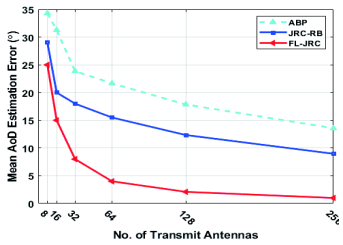


Fig. 4: Mean AoD Estimation Error Comparison

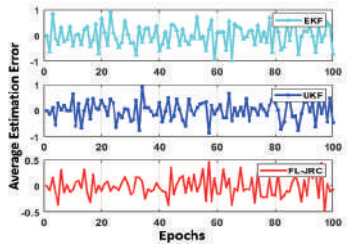
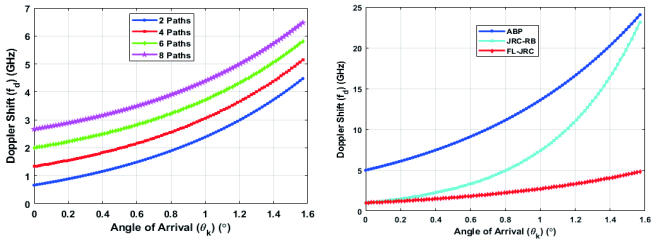


Fig. 5: Average Estimation Error Comparison

Fig. 5 shows a comparison of the average estimation error for the AoD between the UKF and EKF approaches that were suggested. In this case, the average is calculated over five runs, each of 100 epochs. The proposed approach's mean estimation

error is 0.00216, while the UKF's and EKF's are 0.01073, 0.01101, and 0.00216, respectively. The maximum value for the FL-JRC is nearly 50% lower than that of the other two approaches. As a result, the error performances of the UKF and EKF are identical. This unequivocally demonstrates that proposed approach reduces the estimation error.



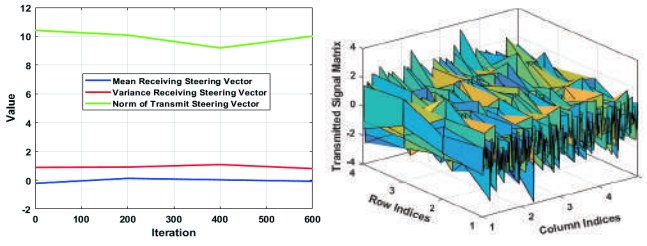
(a) Effect of Different Paths on FL-JRC

(b) Comparison of FL-JRC, ABP, JRC-RB

Fig. 6: Impact of AoA on the f_d under Multipath

Fig. 6a depicts the variation in f_d for FL-JRC under the number of paths: 2, 4, 6, and 8. In all cases, there is a marginal increase in f_d as the AoA increases. Additionally, as the number of paths increases from 2 to 8, the f_d also increases. Fig. 6b compares AoA and f_d for the three approaches, with the number of paths fixed at 2. The proposed approach exhibits a negligible increase in f_d, whereas the other two approaches show an almost exponential increase. This substantiates that the proposed approach is more resilient against multipath.

Fig. 7 provides quantitative measures to assess the efficacy, stability, and characteristics of the involved vectors and matrices for the proposed approach. The term "value" in this figure represent the received steering vector's mean, variance, and norm, while the indices indicate the rows and columns in the transmitted signal matrix. In addition to a stable and



(a) Steering Vector

(b) Signal Matrix

Fig. 7: Analysis of the Proposed Approach

consistent behavior, Fig. 7a illustrates that the value exhibits minimal variations, indicating a high degree of stability. This allows the proposed approach to perform consistently and can effectively handle variations in the channel conditions.

To identify any inconsistencies or anomalies in the communicated data, Fig. 7b shows the values of the transmitted signal matrix for each element (cell) in the matrix. Since there are no noticeable changes or irregularities in the transmitted signal matrix, it can be further inferred that the proposed approach is effective in mitigating the channel conditions. The consistency in color of each cell indicates the magnitude of the corresponding element in the matrix.

V. Conclusion

A FL-based joint radar-communication mmWave beamtracking scheme for V2X-C is proposed that is efficient, accurate, and provides a robust link. The CNN module is federated between the vehicle users (local clients) and the RSU (main server), which estimates the AoA/ AoDs, thus providing for efficient beamtracking in V2X-C. Simulation results validate that the proposed approach significantly outperforms the contemporary approaches along with EKF and UKF in terms of better throughput, lower estimation error, and DS estimation, thereby augmenting the strength of the FL-JRC.

References

[1] S. Bhardwaj, D.-H. Kim, and D.-S. Kim, "Backscatter-Enabled CR-NOMA Based Cooperative V2X Communication with Imperfect CSI," *Vehicular Communications*, p. 100611, 2023.

[2] P. Kumari, N. J. Myers, and R. W. Heath, "Adaptive and fast combined waveform-beamforming design for mmWave automotive joint communication-radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 996–1012, 2021.

[3] S. Bhardwaj and D.-S. Kim, "Federated Learning-Based Joint Radar-Communication mmWave Beamtracking with Imperfect CSI for V2X Communications," in *2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2023, pp. 201–206.

[4] F. Liu and C. Masouros, "A tutorial on joint radar and communication transmission for vehicular networks—Part III: Predictive beamforming without state models," *IEEE Communications Letters*, vol. 25, no. 2, pp. 332–336, 2020.

[5] A. M. Nor, S. Halunga, and O. Fratu, "Survey on positioning information assisted mmwave beamforming training," *Ad Hoc Networks*, vol. 135, p. 102947, 2022.

[6] M. A. P. Putra, D.-S. Kim, and J.-M. Lee, "Adaptive LRFU replacement policy for named data network in industrial IoT," *ICT Express*, vol. 8, no. 2, pp. 258–263, 2022.

[7] S. Bhardwaj and D.-S. Kim, "Deep Q-learning based sparse code multiple access for ultra reliable low latency communication in industrial wireless networks," *Telecommunication Systems*, vol. 83, no. 4, pp. 409–421, 2023.

[8] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, "Fast channel tracking for terahertz beamspace massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 5689–5696, 2016.

[9] S. G. Larew and D. J. Love, "Adaptive beam tracking with the unscented kalman filter for millimeter wave communication," *IEEE Signal Processing Letters*, vol. 26, no. 11, pp. 1658–1662, 2019.

[10] D. Zhu, J. Choi, and R. W. Heath, "Auxiliary beam pair enabled AoD and AoA estimation in closed-loop large-scale millimeter-wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4770–4785, 2017.

[11] H.-Y. Hsu, J. J. Lin, S. T. Skidmore, and M. Kim, "Evaluating fit indices in a multilevel latent growth curve model: A Monte Carlo study," *Behavior Research Methods*, vol. 51, no. 1, pp. 172–194, 2019.

[12] L. Xu, P. Stoica, and J. Li, "A diagonal growth curve model and some signal-processing applications," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3363–3371, 2006.

[13] J.-H. Kim, K.-W. On, W. Lim, J. Kim, J.-W. Ha, and B.-T. Zhang, "Hadamard product for low-rank bilinear pooling," *arXiv preprint arXiv:1610.04325*, 2016.

[14] J. Tabrikian, J. Li, and P. Stoica, "Performance bounds and techniques for target localization using MIMO radars," *MIMO Radar Signal Processing, J. Li and P. Stoica, Eds*, pp. 153–191, 2008.

[15] J. Wen, L. Yan, X. Feng, Z. Ni, J. Hua, and Z. Xu, "Optimization of ACF-DSR-based joint Doppler shift and SNR estimator for Internet of Vehicle system," *EURASIP Journal on Advances in Signal Processing*, vol. 2022, no. 1, pp. 1–11, 2022.

Node Election Scheme for Energy-Efficient Autonomous Underwater Vehicles

Young-June Kwon, Jae-Min Lee, Tae-Soo Jun, Dong-Seong Kim*
Dept. of IT Convergence Eng.
Kumoh National Institute of Technology
Gumi, Republic of Korea
{enuj_young2, ljmpaul, taesoo.jun, dskim*}@kumoh.ac.kr

***Abstract*— Recent research is being conducted on the operational concepts of the Korean-style unmanned maritime system. In particular, the operation of the Autonomous Underwater Vehicle (AUV), which is a component of the unmanned maritime system, is emphasized. AUVs require high energy efficiency because they perform underwater operations and must be able to transmit data to warships without tampering or alteration for military detection purposes. Recently, an energy-efficient blockchain-based multi-AUVs cooperative operation system has been proposed, featuring a leader node selection method based on residual energy. However, this method does not consider the distance to the sink node, which can affect energy efficiency, nor does it account for interoperability among heterogeneous devices. Therefore, this paper proposes a leader node selection method based on the residual energy and distance to the sink node, as well as communication based on Data Distribution Service (DDS). The proposed method shows higher energy efficiency compared to existing systems.**

Keywords—Blockchain, Energy-Efficiency, AUV

I. INTRODUCTION

In Korea, middleware research for defense distributed systems is currently underway[1]. Additionally, research in the defense sector, such as blockchain-based data sharing techniques to enhance the reliability and security of naval combat systems, is ongoing[2][3]. Globally, particularly in the United States, the development of unmanned maritime systems is progressing, and the concept of their operation at an actual operational level is being systematized, with research being conducted for future warfare applications of these systems[4]. North Korea has been continuously conducting tests of Submarine-Launched Ballistic Missiles (SLBM) since 2015, necessitating a deterrence strategy to counter the threat posed by North Korea's SLBMs. One deterrence strategy involves forward surveillance using Autonomous Underwater Vehicles (AUV). AUVs are used for military detection purposes due to their high degree of mobility and precision in attitude control. The information detected by these AUVs includes enemy submarine detection data, which must be transmitted to warships without tampering or alteration. Furthermore, AUVs operate in underwater environments and rely on battery power, requiring high energy efficiency[5]. Currently, research is being conducted on an energy-efficient blockchain-based multi-AUVs cooperative operation system that uses the Raft consensus algorithm to select a leader node based on energy[6]. However, this system does not consider the distance to the mothership, which is a sink node and a crucial parameter for the energy-efficient selection of leader nodes within a cluster. Additionally, it does not address the interoperability of underwater unmanned systems that operate heterogeneous devices such as autonomous underwater mine seekers and reconnaissance unmanned submarines. Therefore, this paper proposes an energy-efficient leader node selection method for unmanned submarines using blockchain technology. The proposed system enhances interoperability among heterogeneous AUVs through Data Distribution Service (DDS) by facilitating data communication, such as consensus algorithms, and increases system energy efficiency by selecting leader nodes based on residual energy and the distance to the mothership, thereby securing process execution time.

II. ANALYSIS OF REQUIREMENTS FOR UNMANNED MARITIME SYSTEMS AND UNDERWATER CLUSTER HEADS

A. Analysis of Unmanned Maritime Systems and Requirements

The unmanned maritime system is divided into unmanned surface systems and unmanned aerial vehicles. Unmanned surface systems are further categorized into unmanned surface vehicles for reconnaissance, mine warfare, and combat. Unmanned underwater systems are divided into autonomous underwater mine search vehicles, reconnaissance unmanned submarines, and combat unmanned submarines. Since the unmanned maritime system consists of heterogeneous devices, interoperability must be considered[7]. Additionally, the missions of AUVs are divided into surveillance/reconnaissance, mine countermeasures, and anti-submarine warfare. For effective surveillance/reconnaissance missions, AUVs must have long-distance communication capabilities with the mothership in hazardous coastal areas and the ability to stay underwater for extended periods. Therefore, AUVs, which operate on batteries in underwater environments, require high energy efficiency. Moreover, they must be able to transmit detected information about enemy submarines to the mothership without tampering or alteration.

B. Analysis of Requirements for Cluster Head Election for Energy Efficiency

In underwater environments that demand high energy efficiency, clusters are formed, and a cluster head is designated to transmit the detected data. The cluster head sends the data received from the nodes in the cluster group to the sink node. Therefore, when electing a cluster head, it is important to consider the remaining energy and the distance to the sink node,

which represents the data transmission distance. The farther the data needs to be transmitted in an underwater environment, the more energy is required for transmission. Consequently, in underwater AUVs networks, when forming clusters and determining the cluster head, it is essential to consider both the remaining energy and the distance to the sink node, which is the mothership[8].

III. Related Research and Problem Analysis

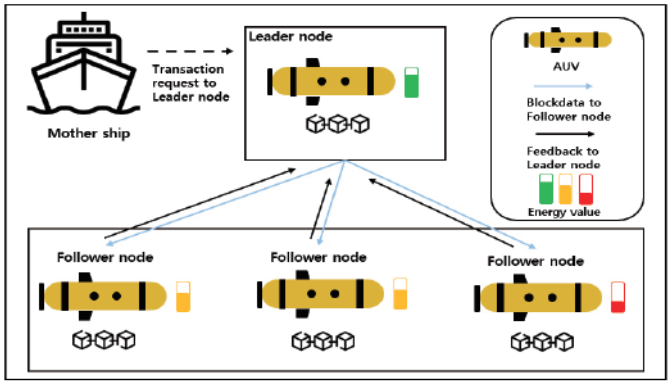


Fig.1. Multiple AUVs Operating System Model based on E-Raft Agreements Algorithm

Due to the limited resources of AUVs in open and unreliable underwater environments, malicious users can tamper with network communication information. Therefore, it is essential to combine multi-AUVs networks with blockchain technology to prevent data tampering. However, the use of blockchain and consensus algorithms can lead to excessive energy consumption by the leader node, causing the AUVs distributed network's processes to terminate prematurely and reducing availability. Thus, energy balance among AUVs is required for blockchain-based multi- AUVs cooperation. In this regard, [6] proposes an E-Raft consensus algorithm-based multi- AUVs cooperative operation system. Figure 1 illustrates the E-Raft consensus algorithm-based multi- AUVs operation system model. The system model consists of a distributed network composed of a mothership and multiple AUVs, forming a private network that uses the E-Raft consensus algorithm. The mothership transmits task data to the leader AUV node, which creates a block with the received task data and propagates it to follower nodes. In the leader node re-election process of the E-Raft consensus algorithm, nodes are categorized as leader, follower, candidate, or also-ran nodes. The leader node processes transactions and sends heartbeats to indicate normal operation, while follower nodes receive heartbeats from the leader and attempt to re-elect a leader node if a problem arises. Candidate nodes are elected as leader nodes through votes from follower nodes, and if they do not receive votes, they revert to follower nodes. Also-ran nodes act as follower nodes when their energy is insufficient, updating the energy threshold, and after a leader node is elected, each node calculates its remaining energy. If a follower node does not receive the leader node's heartbeat within a certain period, all follower nodes check if their residual energy exceeds the energy threshold. If their residual energy is above the threshold, they become candidate nodes; if it is below, they switch to also-ran nodes and update the threshold. If there are multiple candidate nodes, the leader node is chosen randomly. In [6], it is defined that if even one node among the multiple AUVs nodes exhausts its energy, the operating process is terminated. According to [6], the re-election of the leader node occurs only when the leader node fails to send heartbeats to follower nodes, i.e., when an error occurs with the leader node. However, this can lead to premature process termination for the following reasons. The leader node, receiving transaction requests from the mothership, creates blocks and transmits them to follower nodes. It also continuously sends heartbeats and receives feedback from follower nodes, resulting in the highest energy consumption per hour among the nodes in the AUVs network. Replacing such a leader node only when an error occurs can cause premature termination of the AUVs network processes. Moreover, in [6], the leader node of the AUVs network is the cluster head of the AUVs cluster, which exchanges data with the mothership, the sink node. However, [6] does not consider the distance between candidate nodes and the mothership during the re-election of the leader node. In underwater cluster environments, not considering the distance to the sink node, which determines the transmission energy consumption of the cluster head, increases the energy consumption of the cluster head, thus reducing the energy efficiency of the AUVs network. Lastly, the underwater unmanned system operational fleet includes autonomous underwater mine seekers, reconnaissance unmanned submarines, and combat unmanned submarines. However, the interoperability of these heterogeneous devices was not considered. [2] applied DDS to the Practical Byzantine Fault Tolerance (PBFT) consensus algorithm in distributed systems to enhance its real-time capabilities. However, due to the inherent nature of PBFT, it requires three phases (Pre-prepare, Prepare, Commit) for consensus, and the communication complexity of the PBFT consensus algorithm is $O(N^2)$. Therefore, it can lead to excessive energy consumption in underwater environments.

IV. Proposed System

A. *DDS-based Consensus Algorithm Architecture for Interoperability among Heterogeneous Devices*

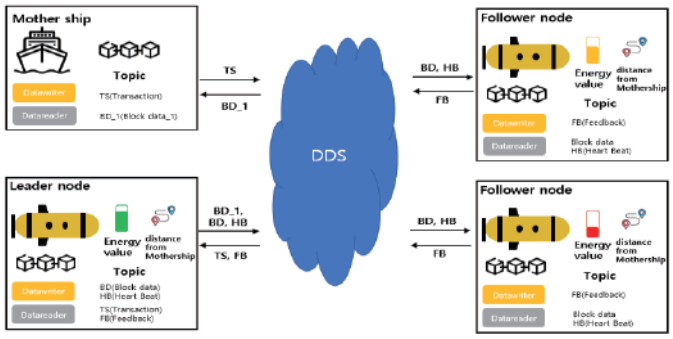


Fig.2. Multiple AUVs Operating System Model based on E-Raft Agreements Algorithm

The architecture of the DDS-based consensus algorithm for interoperability among heterogeneous devices proposed in this paper is illustrated in Figure 2. The proposed system implements a DDS-based Raft consensus algorithm. The unmanned maritime system consists of a ship and numerous heterogeneous AUVs, where these numerous AUVs are organized with one leader node and multiple follower nodes. When the leader node

receives a transaction request from the ship, it verifies whether the ID included in the transaction request matches the ship's ID. If they match, the leader node generates a block and transmits it to the follower nodes. Additionally, the leader node periodically sends heartbeats within the follower node's election timeout period, which ranges from 150ms to 300ms. All communication between the ship and the AUVs is facilitated through DDS, with each entity acting as both Publisher and Subscriber. An AUV, upon being elected as the leader node, activates the DataReader for the topic that receives transaction requests from the ship and the topic that receives feedback from the follower nodes. In the follower node state, the AUVs activates the DataReader for the topic that receives heartbeats from the leader node. This DDS-based communication, utilizing topic-based middleware, ensures interoperability among heterogeneous devices when the leader node converts the transaction requests received from the ship into blocks and transmits them to the follower nodes. Furthermore, all AUVs add the newly generated block to their respective blockchains, ensuring that the data is distributed and stored across the network. This process enables interoperability among heterogeneous AUVs via DDS-based communication, while the Raft consensus algorithm maintains the data reliability of the unmanned maritime system.

B. Leader Node Election Technique Based on Remaining Energy and Distance to the Mothership

The node states in this paper are divided into four categories: leader, follower, candidate, and low-energy node. The leader node receives transaction requests from the ship, verifies them using the ship's private key, creates blocks, and sends these blocks to the follower nodes. Additionally, the leader node periodically sends heartbeats within the election timeout period of the follower nodes. Follower nodes receive heartbeats from the leader node to verify its proper functioning, check the blocks sent by the leader node, and send feedback to the leader node. Candidate nodes are follower nodes that have not received heartbeats from the leader node within the election timeout period and have remaining energy above the energy threshold; these nodes send vote requests to the follower nodes. Low-energy nodes are follower nodes that have not received heartbeats from the leader node within the election timeout period and have remaining energy below the energy threshold; they update their threshold and respond to the vote requests from the candidate nodes just like follower nodes. It is assumed that all AUVs nodes start with the same initial energy value. Because all AUVs have the same initial energy, the leader node is initially selected at random. After the leader node is selected, subsequent leader node elections occur under two conditions: when the leader node fails to send heartbeats to the follower nodes, and when the leader node successfully sends heartbeats to the follower nodes but its remaining energy falls below the threshold. Figure 3 illustrates the leader node election technique based on remaining energy and the distance to the mothership during the process. If the remaining energy of the leader node falls below the threshold, the leader node updates the energy threshold. It also notifies all follower nodes that a re-election of the leader node is required. Upon receiving this notification, follower nodes transition to candidate nodes if their remaining energy is higher than the threshold, or to low-energy nodes if their remaining energy is lower than the threshold. A candidate node transitions to the leader node if it receives (total number of nodes / 2) + 1 votes; otherwise, it transitions back to a follower node. If there are multiple candidate nodes during the leader node election process, the candidate node closer to the mothership is selected as the leader node. After the leader node is elected, all nodes recalculate their remaining energy, and the process concludes. This detailed election process ensures the energy-efficient operation of the AUVs network by considering both energy levels and the distance to the mothership. This approach not only optimizes the network's energy consumption but also enhances the sustainability of the unmanned maritime system network by ensuring that the most suitable node assumes the leader node role.

C. Energy Calculation Equation of the Proposed System

In [6], it is assumed that all AUVs move along the same path, so the power consumption for movement is the same for all AUVs. Therefore, this paper only considers the energy consumption of nodes during data transmission, reception, and block creation. It is also assumed that each AUVs consumes the same amount of energy for transmitting and receiving data, as well as for creating blocks, and that the distance between any two nodes in the network is the same.

$$E_{tr} = lP_{tr},\ E_{re} = lP_{re},\ E_{ge} = S_bP_{ge} \tag{1}$$

$$E_{lea} = E_{tr}(S_h+S_b) + E_{ge} + E_{re}(S_{fee}) \tag{2}$$

$$E_{can} = E_{tr}(S_{req}) + E_{re}(S_b+S_{thr}+S_{vot}) \tag{3}$$

$$E_{fol} = E_{tr}(S_{fee}+S_{vot}) + E_{re}(S_b+S_{thr}+S_{req}) \tag{4}$$

d represents the distance between the leader node and the follower node, while d_{mother} represents the distance between the leader node and the mother ship. Let S_b, S_{thr}, S_{fee}, S_{vot} and S_h represent the size of the block, the size of the threshold message, the size of the feedback sent by the follower, the size of the vote request message sent by the candidate, and the size of the heartbeat message, respectively. E_{tr} is the transmission power of the AUVs, P_{re} is the reception power of the AUVs, and E_{ge} is the power required to generate a unit-size block. Additionally, E_{tr}, E_{re}, and E_{ge} represent the energy consumption when transmitting a packet, receiving a packet, and generating a block, respectively. According to [9][10], the energy consumption for data transmission in underwater acoustic communication is related to the transmission distance and packet size. The energy consumption of a node transmitting an l-bit packet is defined by Equation 1, and the energy consumption of a node receiving an l-bit packet is defined by Equation 1.

Therefore, Equation 1 represents the energy consumption of a node transmitting an l-bit packet, the energy consumption of a node receiving an l-bit packet, and the energy consumption when the leader node generates a block. Equations 2, 3, and 4 define the energy consumption formulas for the leader, candidate, and follower nodes, respectively.

V. Simulation Results

According to [11], when the transmission energy of the AUVs acoustic model for distances less than 1 km is set to 1, the energy for transmissions over distances of 1 km to 2 km is approximately 2.8 times that amount. Therefore, the

transmission energy for distances less than 1 km is defined as E_{tr}, and the transmission energy for distances between 1 km and 2 km is defined as $2.8 \cdot E_{tr}$.

TABLE I. SIMULATION VALUE

Symbol	Definitions
m	5, 7, 9
E_{init}	$500J$
d	1000 m
d_{mother}	[500, 1500] m
P_{tr}	$1 \cdot 10^{-3}$ $J/bits$
P_{re}	$0.2 \cdot 10^{-3}$ $J/bits$
P_{ge}	$1 \cdot 10^{-3}$ $J/bits$
S_b	10 $bits$
S_h	1 $bits$
S_{vot} , S_{thr} , S_{fee}	1.5 $bits$

m represents the number of nodes comprising the AUVs network, and the definitions of the remaining values are as mentioned in Chapter 3.

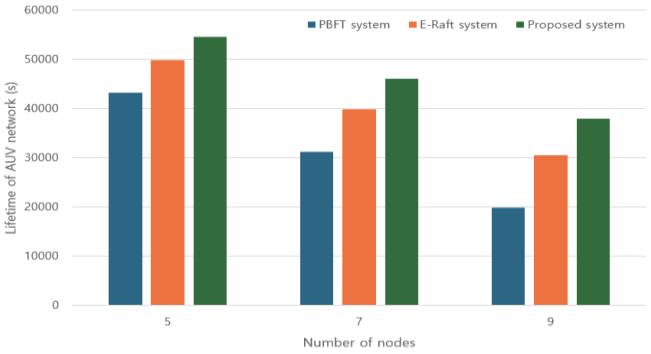


Fig.3. Process operation time by number of nodes

Figure 3 is a graph showing the operating time in each system. The E-Raft system re-elects the leader node based on residual energy. Therefore, since the distance between the leader node and the mother ship is not considered, when a block is generated according to the transaction request of the mother ship and then transmitted to the mother ship for distributed storage, the increase in transmission energy due to the distance results in a shorter overall operating period compared to the proposed system. The PBFT system does not receive regular feedback from follower nodes but requires more communication during the consensus process compared to the RAFT system. Additionally, it does not consider the distance to the mother ship, resulting in the shortest operating time among the systems. The proposed system has a relatively long operating period by re-electing a leader considering the distance to the mother ship and the remaining energy. However, as the number of nodes in the proposed system increases, the leader node receives heartbeats, sends feedback, and transmits blocks to more nodes. The frequent changes of the leader due to the increase in the number of nodes lead to higher energy consumption, resulting in a decrease in lifetime as the number of nodes in the proposed system increases. However, it can be seen that the lifetime is longer than that of the existing two systems.

VI. CONCLUSION

In this paper, we design and implement a leader node selection scheme for AUVs that require high energy efficiency and interoperability through DDS-based communication and leader node selection based on remaining energy and distance to the mother ship. The proposed system achieves higher energy efficiency compared to the existing two systems. This means that AUV networks monitoring underwater can operate for longer periods. Future work will focus on applying machine learning to predict the distance to the mother ship based on the movement patterns of autonomous unmanned submarines and integrating this prediction into the leader node selection process.

ACKNOWLEDGMENT

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 33.3%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33.3%) and by the MSIT, Korea, under the ICAN support program(IITP-2024-00156394, 33.3%) supervised by the IITP

REFERENCES

[1] J.-H. Cha, H.-J. Kim, J.-M. Lee and D.-S. Kim, “Trends in Message-Oriented Middleware Research for Smart Defense Distributed Systems”, KIISE, vol. 41, pp. 32-39, Mar. 2023

[2] J.-W. Lee, H.-J. Kim, J.-M. Lee, T.-S. Jun and D.-S. Kim, “Blockchain-Based Data Sharing Scheme to Enhance Reliabilityand Security for Naval Combat”, J-KICS, vol. 47, pp. 809-817, June 2022

[3] J.-W. Kim, W.-J. Ryu, J.-M. Lee and D.-S. Kim, "An Analysis of the Network Simulation Tools for Military Network Design", KIISE, vol. 41, pp. 40-50, Mar. 2023

[4] H.-Y. Bae, “A Study on the Operational Concept of K-Ghost Fleet and Force Development”, KADIS, vol. 29, pp. 77-88, Aug. 2022

[5] X. Zhuo, M. Liu, Y. Wei, G. Yu, F. Qu, and R. Sun, "AUV-Aided Energy-Efficient Data Collection in Underwater Acoustic Sensor Networks", IEEE Internet of Things Journal, vol. 7, pp. 1-13 , Oct. 2020

[6] X. Li, Y. Guo, L. Yan and X. Xu, "Energy-aware blockchain for multiple autonomous underwater vehicles cooperative operation", 2021 40th Chinese Control Conference, pp. 26-28, July 2021

[7] H.-S. Shin, " Operational Concepts and Development Directions of Unmanned maritime systems ", Korea Maritime Security Forum, vol. 20, Feb. 2017

[8] J. Amutha, S. Sharma and J. Nagar, “WSN Strategies Based on Sensors, Deployment, Sensing Models, Coverage and Energy Efciency: Review, Approaches and Open Issues”, Wireless Personal Communications, vol. 111, pp. 1089-1115, Oct. 2019

[9] N. Javaid, S. Cheema, M. Akbar, N. Alrajeh, M. Souheil Alabed and N. Guizani, "Balanced Energy Consumption Based Adaptive Routing for IoT Enabling Underwater WSNs", IEEE Access, vol.5, pp. 10040-10051, May 2017

[10] K. Wang, H. Gao, X. Xu, J. Jiang and Dong Yue, "An Energy-Efficient Reliable Data Transmission Scheme for Complex Environmental Monitoring in Underwater Acoustic Sensor Networks", IEEE Sensors Journal, vol. 16, pp. 4051-4062, May 2015

[11] P. Danielis, H. Parzyjegla, M. A. M. Ali and F. S. Torres, "Simulation model for energy consumption and acoustic underwater communication of autonomous underwater vehicles", WMU Journal of Maritime Affairs, vol. 21, pp. 89-107, Nov. 2021

Performance Analysis for Next Generation Integrated Naval Ship System Network

Min-Hui Jang, Neul-Som Lee, Jae-Min Lee, Tae-Soo Jun and Dong-Seong Kim*
Dept. of IT Convergence Eng.
Kumoh National Institute of Technology
Gumi, Republic of Korea
{jmhee, alwaysom2, ljmpaul, taesoo.jun, dskim*}@kumoh.ac.kr

Hyeong-Jin Kim
Technology Research Institute
NSLab Co., Ltd.
Gumi, Republic of Korea
haengg@nslab.tech

Young-Keun Go and Nak-Jung Choi
Maritime Technology Research Institute
Agency For Defense Development
Jinhae, Republic of Korea
{ykgo, njchoi}@add.re.kr

Jae-Ho Lee and Jae-Hak Yu
Naval System
Naval R&D Center, *Hanwha Systems*
Gumi, Republic of Korea
{jh8028.lee, jahak90}@hanwha.com

***Abstract*—The naval ship system includes functions to perform various missions and achieve military objectives through weapon systems, radar and sensors, and communication equipment. These systems can vary depending on the mission and operational environment of the ship, and they continue to develop with the advancement of technology. However, complexity is increasing due to the increase in various sensors and software and systems as the requirements increase due to technological advances. This may improve the ability to perform various tasks but may lead to difficulties in management and operation. Therefore, the need for changing the structure of trap networks is emerging to understand and manage this complexity. In this paper, an analysis of the new network structure was conducted to improve the combat and operational efficiency of the naval ship by systematically integrating the naval ship's system. This was simulated by modeling in consideration of the node and network characteristics of the naval ship's combat system, and reliable performance analysis was conducted by comparing values measured in a real system environment with the same structure and reference values.**

Keywords— Integrated Naval Ship System Network, Network Simulation, Performance Analysis

I. INTRODUCTION

The naval ship system combines various sensors, weapons and defense systems to efficiently perform armed control and engagement. This is accomplished by integrating and operating various subsystems such as C4I(Command, Control, Communications and Intelligence), communication, sensor and weapon systems [1]. However, with the advancement of technology, these combat systems are facing new requirements to ensure interoperability and improve performance and efficiency. Due to the complexity of the network and the increase in the amount of data, the performance of the system is concerned, and thus the introduction of a new network structure is required. In addition, the process of verifying the network in advance is essential in a naval ship system involving a complex structure and multiple nodes [2]. To this end, it is necessary to verify and evaluate network performance by simulating an environment similar to a real system.

Therefore, in this paper, a simulation was conducted to evaluate the research and performance of the next-generation integrated naval ship system network. A network simulation environment was constructed by selecting an appropriate tool suitable for the requirements and characteristics of the naval ship system among various previously developed network simulator tools. This should take into account the node, traffic, and network characteristics that make up the integrated naval ship system. In addition, these characteristics and structures were configured identically in the actual system and the same scenario was implemented. As a result, the tendency was confirmed through the comparison of the simulation results and the values measured in the real system environment, and the reliability of the built simulation results was verified by reviewing the error rate. In addition, the performance analysis of the next-generation integrated naval ship system network was conducted through the simulation results obtained through this.

II. ANALYSIS OF NETWORK REQUIREMENTS FOR INTEGRATED NAVAL SHIP SYSTEMS

TABLE I. CLASSIFICATION OF NAVAL COMBAT SHIP SYSTEM

Sortation	Subsystems
Combat System	Combat management, sensors, armaments, navigation sensors, training, etc.
C4I System	Naval Tactical Information, Naval Command and Control System, etc.
Communication System	Integrated communication system, data link, etc.
Integrated Engine Control System	Propulsion control and monitoring, view control and monitoring, power control and monitoring, damage control, integrated bridge, integrated status evaluation, training system, etc
Administrative System	National Defense Network, Internet, etc.
Others	Mine resource, special operation support, unmanned system control, education and training system, etc.

The computing system of the control system mounted on the naval ship is implemented for each purpose, operated on an independent server, and follows different standards. Accordingly, the control system classified into combat, C4I, integrated institution control, communication, and administrative systems is organized in detail as shown in Table 1 [3]. However, such an environment is completely physically and functionally separated, and it is controlled individually, making it less interoperable and compatible between systems. It is also disadvantageous in terms of viability, such as failing to utilize the computing resources of other systems if a particular system fails or damages as each system is independently configured. Therefore, the concept of ship control is being changed from an independent operation structure for each control system to an integrated mission performance structure in order to cope with the reduction of response time, reduction of crew and operating costs, and technological development [4]. These integrated naval ship system networks require redundant deployment of network equipment to ensure high availability and high reliability. In addition, efficient processing techniques through middleware-based data communication are required to secure stability and reliability. At this time, DDS, a MOM(Message Oriented Middleware) that supports Topic-based publication and subscription standardized by OMG, is used [5]. This allows for the management of numerous data and messages in a large system inside the naval ship [6]. In addition, it should be possible to identify and operate the characteristics of data used in combat systems and the amount of data for each data.

III. Modeling Naval Ship System Nodes and Network Structures

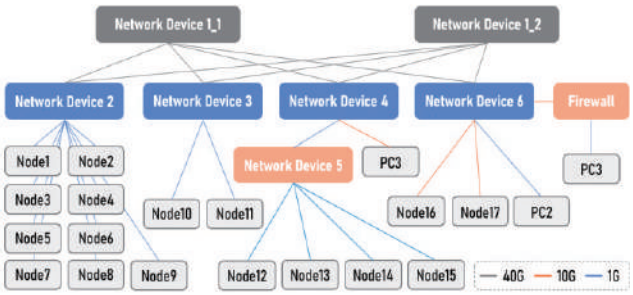


Fig. 1. Integrated Ship System Node and Network Structure

This paper mainly organized the integration of computing resources including networks, nodes, and processing devices. The naval ship system is large-scale system and consists of various weapon systems and equipment. It includes various detection and tracking sensors and processes and fuses tactical information collected through a tactical data link in real time to generate a system target and display a synthesized tactical situation. This process generates a large amount of message traffic. As shown in Figure 1, the integrated naval ship system consists of nodes such as PCs, radars, and various sensors, network devices that manage nodes, and integrated network devices that integrate and manage these devices. At this time, the network device consists of Ethernet, LAN, L3 switches, etc., and the node was modeled based on the benchmark for each switch. In addition, nodes such as PCs and various sensors were modeled based on benchmarks. Therefore, network links suitable for each network device, node, and data capacity were arranged. The link size of the nodes that make up the ship's combat system network is a unit of Gbyte, and the amount of data handles large amounts of data of about 35Mbyte, so it is deployed at a capacity of 1, 10, and 40 Gbit/s. In addition, the network structure is centered on the backbone of the integrated network device consisting of redundancy, and subnets for each river area such as weapon control, radar, surveillance, and object tracking are connected. In this paper, a part of the integrated naval ship system was modeled and constructed as a simulated environment. Therefore, it consists of a structure that communicates with a total of five network devices and nodes connected to them, centering on the redundant integrated network device.

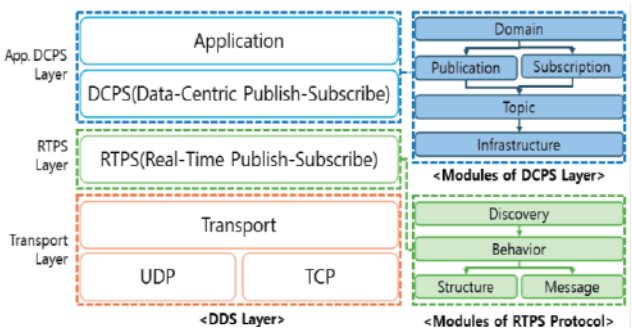


Fig. 2. Publishing/Subscription Technology of DDS Structure

In addition, the naval ship system controls and manages messages and data based on DDS and shown in Figure 2 [7]. DDS is a communication middleware standardized by OMG and is used in distributed system environments in various fields such as military, medical care, and robots. It consists of a structure that issues or subscribes to messages based on Topic, providing an environment in which participants in the same domain issue and subscribe to messages without direct connection with each other. At this time, DDS is divided into transport, Real Time Publicity Subscribe (RTPS), Data Centric Publicity Subscribe (DCPS), and Application layers, and supports the Unbrokered-P2P communication method through RTPS [8,9]. In this paper, ns3, an open-source network simulation tool, was used to analyze and evaluate the next-generation integrated trap network structure. However, ns3 does not support DDS. However, it provides a variety of module libraries, network models, routing algorithms, etc., and is easy to expand the custom model and is highly flexible, making it easy to build a model that simulates the core functions of DDS. Therefore, in the process of publishing/subscription-based data transmission of DDS, the main steps of the core protocol, RTPS, were separated and the DDS was modeled in consideration of the ship system network environment requirements. RTPS is roughly classified as Partisan Discovery, Endpoint Discovery, Data Exchange, and Heartbeat Exchange, although it is not accurately classified [10]. At this time, all discovery and data transmission processes operate in a UDP multicast manner. Therefore, a simulation was performed by simulating DDS by appropriately combining UDP applications operating as multicast in NS3.

IV. Create Scenario for Integrated Naval Ship System Network Simulation

The purpose of this integrated ship system network simulation is to check the behavior and functionality before building a real network. At this time, simulation was attempted using a virtualization tool to verify in an environment similar to the actual system. In addition, it is necessary to verify, evaluate, and optimize the normal operation and performance of the built network environment [11]. Therefore, the amount of traffic considering the characteristics of combat traffic was calculated, and the scenario for transmitting such traffic was selected and the simulation was conducted.

TABLE II. Combat Traffic Transfer Scenario

No.	COMM.	Send	Receive	Data Size	Data Type
1	DDS	Node11	Node1	327Mbps	Video Information
			Node2		
			Node3		
			Node4		
			Node12		
			Node13		
			Node14		
2	DDS	Node13	Node1	72Mbps	Tactical Information
			Node2		
			Node3		
			Node4		
			Node11		
			Node12		
			Node14		

Table 2 shows a scenario with traffic volume and communication method of combat system applied to the simulation. The traffic was constructed based on data that could occur in the subsystem of the ship system network. At this time, the communication method was constructed based on the modeled DDS as mentioned in Section 3. At this time, multiple nodes with the same Topic were created and configured as a communication scenario in which data was transmitted from a specific node. For example, Scenario 1 constructed a process of exchanging video information about a combat situation in a ship's combat system to observe and display its surroundings [12]. And considering the characteristics of combat traffic, the amount of data traffic for this was calculated as a data capacity of 327 Mbps. In addition, the node names were arbitrarily numbered and listed for security, but data was transmitted from node 11, which can be relayed through systems and applications, to a node that requires video information. That is, it has a process of receiving video information data of the data capacity calculated by nodes 1, 2, 3, 4, 12, 13, and 14 having video information topics. The results that can be verified and executed in this scenario are throughput, Goodput, data transmission amount, and number of transmitted packets for each node for transmission. In addition, based on the scenario, normal communication was confirmed through the values of the data throughput and total amount transmitted by each node that can be inferred theoretically

V. Integrated Naval Ship System Network Simulation Results And Preformance Analysis

Through the simulation results, the simulation was conducted based on the scenario generated in consideration of the characteristics of combat traffic, and the generated results and the reference value to be generated theoretically were compared. In addition, we compared the actual data results of the real system environment consisting of the same structure and the same traffic scenario as the simulation. Through this result comparison, it is possible to evaluate the network performance established and determine whether it communicates normally.

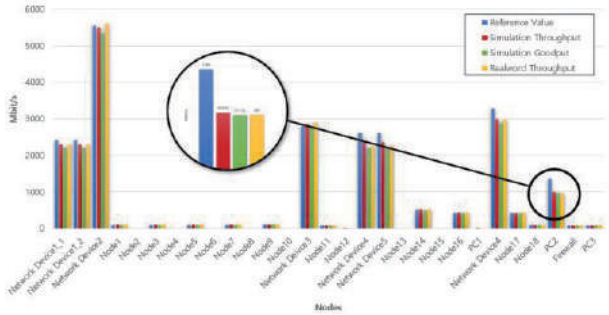


Fig. 3. Compare the Results of Bottleneck Communication Scenario

Figure 3 shows these results listed by nodes constituting the naval ship system network. Therefore, it was confirmed that the results derived from the simulation through the graph showed similarity with the reference value to be theoretically generated and the result of the actual system. Here, Goodput should theoretically be the same as the data size that should be sent from the node if the communication has been completed normally by handling only the data that has been substantially transmitted. However, there is a difference between the reference values of some nodes and the Goodput of the simulation. The average error rate for this reaches 18.62%. The reason for this is that the network capacity of PC2 is 1 Gbit/s, but due to the scenario, 1365Mbit/s had to be transmitted, causing bottlenecks. Therefore, it was confirmed through simulation that normal communication was not performed, and in addition, it was confirmed that normal communication was not performed due to overloading in the real system as well. From this, it can be concluded that PC2 needs higher network capacity.

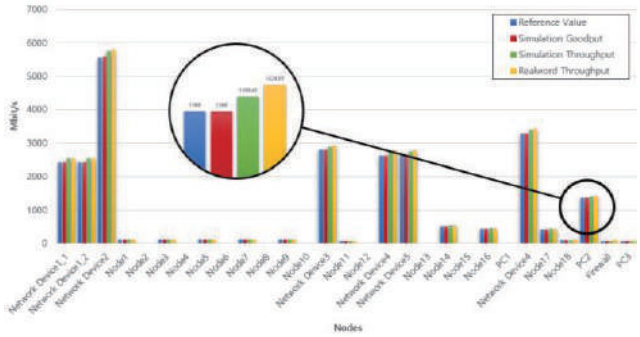


Fig. 4. Compare the Results of Normal Communication Scenario

Therefore, the simulation was performed by increasing the network capacity of PC2 to 10 Gbit/s in the next step. As a result, the network capacity for PC2, which needs to transmit 1365Mbit/s, will not be exceeded due to the scenario. The figure 4 shows these results listed by nodes constituting the ship system network. Therefore, it was confirmed that the simulation Goodput result was similar to the reference value. The error rate for this was 0.004%. Therefore, it was confirmed through simulation that normal communication was performed, and it was confirmed that the same normal communication was performed in the real system. This allows to test the essential traffic scenarios in advance to determine changes to the network structure or suitability for the capacity and specification of the configuration node and suggest directions for them.

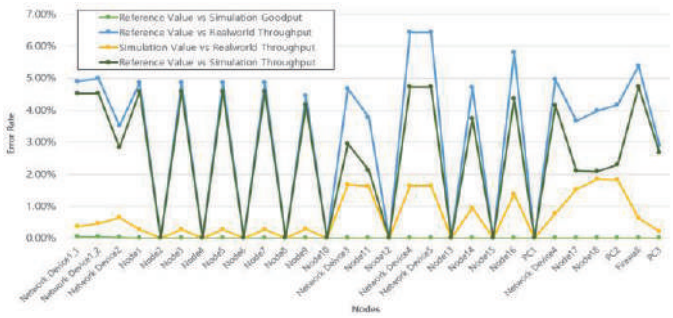


Fig. 5. Comparison of Error Rates for Three values

Figure 5 shows the error rate by comparing the actual system results and comparing them with theoretical reference values to verify the normal communication and reliability of these simulation results. At this time, in order to verify the normal operation of the simulation, the error rate for the Goodput result value and the reference value of the simulation was checked. In addition, in order to verify the reliability of the simulation, the measured throughput result value including the information necessary to transmit the data and the error rate for the throughput of the real system environment were checked. Also, the tendency was confirmed by comparing the error rate for the actual system throughput and the reference value and the error rate for the simulation throughput and the reference value to confirm the reliability and significance of the simulation. Therefore, it was confirmed that the error rate of the goodput and the reference value of the simulation averaged 0.004%, and the error rate of the simulation throughput and the actual system throughput averaged 0.64%. Since this means that the similarity between the simulation values of the actual system and the results is observed, the simulations performed in this paper are considered meaningful and reliable. It is also believed that this provided a way to design new structures and evaluate and optimize their performance before building a real naval ship system network.

VI. Results

In this paper, we identified the need for an integrated ship system network and modeled and implemented simulations for it. Integrated naval ship system network simulation was implemented and analyzed through scenarios considering combat traffic characteristics based on the nodes and network structure constituting the ship. At this time, the normal communication and reliability of the network structure were analyzed by comparing the simulation results with the reference value to be theoretically generated. And the reliability of the simulation was analyzed by comparing the results of the simulation with the results in the actual system. As a result of the analysis, it was confirmed that a simulation with a behavior like that of a real system was secured in traffic processing. This can provide an environment to check behavior and functionality through simulation before actual building an integrated ship system network. In addition, it is possible to verify, evaluate, and optimize the normal behavior and performance of the built network environment.

In future research, we intend to implement a simulator that can configure and implement a node placement and network environment in consideration of user convenience. In addition, the performance is currently evaluated by configuring only one system, but when designing a simulator, two or more systems are configured to evaluate the performance and reflect the error rate to secure high reliability.

Acknowledgment

This work was supported by the Agency for Defense Development Grant funded by the Korean Government.

References

[1] G.-S. Kim, M.-H. Jang, J.-Y. Ryu, B.-I. Jung, J.-M. Lee and D.-S. Kim, “Edge Computing Based Frameworkfor Next Generation Naval Ship Combat System,” The Journal of Korean Institute of Communications and Information Sciences, vol. 47, pp. 2078-2085, Dec. 2022.

[2] J.-G. Lee, “A Study on the Standard Architecture of Weapon Control Software on Naval Combat System,” Journal of The Korea Society of Computer and Information, vol. 26, pp. 101-110, Nov. 2021.

[3] S.-W. Oh, “Integrated Control and Monitoring Systems on Naval Ships,” Journal of Korean Institute of Information Technology, vol. 17, pp. 47-54, June 2019.

[4] J.-Y. Im and D.-S. Kim, “Performance Evaluation of Virtualization Solution for Next Generation Naval Combat Systems,” Journal of The Institute of Electronics and Information Engineers, vol. 56, pp. 41-49, Feb. 2019

[5] J.-H. Cha, J.-W. Lee, J.-M. Lee and D.-S. Kim, “A Study on the Application and Trends of Real-Time Middleware for Civil and Military ICT Convergence Technology,” Information and Communications Magazine, vol. 37, pp. 41-49, Sep. 2020.

[6] Y.-G. Jung, “A Study on the Verification of Integrity of Message Structure in Naval Combat Management System”, Journal of The Korea Society of Computer and Information, vol. 27, pp. 209-217, Dec. 2022.

[7] C.-G. Yi and Y.-G. Kim, “Security Testing for Naval Ship Combat System Softwarel”, IEEE Access, vol. 9, pp. 66839-66851, April 2021.

[8] J.-H. Cha, H.-J. Kim, J.-M. Lee and D.-S. Kim, “Trends in Message-Oriented Middleware Research for Smart Defense Distributed Systems,” Korean Institute of Information Scientists and Engineers, pp. 32-39, Mar. 2023.

[9] J.-W. Lee, J.-H. Cha, J.-W. Lee, G.-S. Kim and D.-S. Kim, “PFDP-Based DDS Security Lightweight Discovery Scheme,” The Journal of Korean Institute of Communications and Information Sciences , pp. 2123-2131, Dec. 2020.

[10] B. Almadani, N. Alshammari and Anas A.-R., “Adaptive Cruise Control Based on Real-Time DDS Middleware,” IEEE Access, vol.11, pp. 75407-75423, July 2023.

[11] C.-W. Seo, “A Study on the Improved Post-Analysis Development System in the Naval Combat System,” Journal of The Korea Society of Computer and Information, vol.27, pp. 197-207, Dec. 2022.

[12] J.-W. Son, “A Study on the Design Plan of Naval Combat System Software to Reduce Cost of Hardware Discontinuation Replacement,” Journal of The Korea Society of Computer and Information, vol.28, pp. 71-78, Jan. 2023.

Intent Inference Method for an Intent-based System in Military Communications

So-Yeon Lee
Dept. of Software Convergence
Soonchunhyang University
Asan-si, South Korea
lsy8647@sch.ac.kr

Jungwook Choi
LIG Nex1 Co. Ltd.
Seongnam, 13488, South Korea
jungwook.choi2@lignex1.com

Soobum Park
LIG Nex1 Co. Ltd.
Seongnam, 13488, South Korea
sbpark93@lignex1.com

Dae-Young Kim°
Dept. of Computer Software Engineering
Soonchunhyang University
Asan-si, South Korea
dyoung.kim@sch.ac.kr

***Abstract*—The rapid expansion of networked devices and the complexity of managing modern networks have led to the need for more sophisticated and automated solutions. This paper introduces Intent-based System (IBS), which extends the capabilities of traditional Intent-based Networking to cover a wide range of systems and applications beyond simple network management, and discusses in detail the intent inference methodology targeting military communications. The methodology automates the data-driven decision-making process based on user needs, which is especially beneficial in areas that require high reliability and adaptability, such as military communications. We expect that the IBS proposed in this paper will not only transform system management into a more proactive and preventive approach but also result in higher user satisfaction.**

***Index Terms*—Artificial Intelligence, Intelligent System Management, Intent-based Systems, Intent Inference, Military Communications**

I. Introduction

Recent advances in wireless communications and artificial intelligence have increased the number of devices connected to the network, increasing the load and size of the network. Traditional network management approaches rely on a manual approach by knowledgeable administrators, making network management a difficult task for network users. To address this, there is a need for an abstraction layer on top of the network infrastructure that hides technical details from network users and allows them to express their goals in a declarative manner [1].

Intent-based Networking (IBN) is a paradigm for creating fully autonomous networks that are capable of self-adaptation, facilitating network usage and reducing the knowledge gap between end-users and network operators. Users express their desires to the network through high-level requests, or "Intents", which are then translated into technical policies by controllers and orchestrators and deployed to the network, thereby reducing the knowledge gap and simplifying network management [2]–[4].

Traditional IBNs are mainly focused on the networking domain and have been studied to automate and optimize network management and operations. However, in this paper, we propose an Intent-based System (IBS) structure that extends the existing IBN for broader application. It aims at intent-based automation and optimization not only in networking but also across various systems and applications.

The main characteristics of IBS are:

- **Intent-centered system operation:** When users communicate high-level intentions to the system, it is designed to automatically understand them and take necessary actions to ensure that the entire system operates in accordance with the intentions.
- **Automated intent inference and policy enforcement:** Intent Analyzer and Policy DB are used to infer user intent, convert it into system operation policies, and automatically enforce them.
- **Intelligent closed-loop control:** Continuous optimization of the system through network monitoring and self-learning.
- **Broad applicability:** It can be applied in various fields such as military communications, autonomous vehicle networking, and the Industrial Internet of Things (IIoT).

IBSs with these characteristics can provide real benefits, especially in areas of increasing complexity, such as military communications, autonomous vehicle networking, and the Industrial Internet of Things (IIoT). For example, in military communications, an IBS can change network configurations in real time in response to operational orders, enabling faster and more flexible responses. In autonomous vehicle networking, instant data exchange based on traffic conditions can increase safety and efficiency. In the Industrial Internet, productivity can be maximized through efficient data flow between different

This work was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MIST) (No.2021R1C1C1013133) and the Korea Research Institute for Defense Technology Planning and Advancement-Grant funded by the Defense Acquisition Program Administration (DAPA) (KRIT-CT-21-030).

devices and process automation.

Considering these various application cases, this paper proposes an IBS architecture to meet the complex system requirements and explores how IBS can be achieved, focusing on intent inference methodologies, including intent analysis and policy DB design methods, to improve the limitations of existing research, especially the abstract approach to intent inference.

The remainder of this paper is organized as follows. Section 2 covers the main concepts and technical structure of Intent-based Systems. Section 3 focuses on the intent inference methodology of IBS, describing how to analyze intents and design the Policy DB. Section 4 concludes the paper with a summary of findings and a discussion on future research directions.

II. Intent-based System (IBS)

By introducing intelligence into a closed-loop platform, IBN provides an enhanced control solution for the operation of today's networks. It abstracts the complexity of the network, autonomously configures and continuously optimizes the network through a closed-loop mechanism that enables network monitoring and intelligence to be dynamically updated. It can also control the network by adapting to the high-level intent of network users and operators. This approach reduces the errors that can occur with traditional manual control methods and enables more proactive and dynamic network operations [5].

On the other hand, the Intent-based System (IBS) approach proposed in this paper, in addition to the basic IBN process, learns data related to user intent and system state in advance and automatically recognizes and processes incoming intent from the actual system. IBS implements highly automated system management with the explicit intent of the user as the key driver of system operation. The overall structure is shown in Fig. 1 and is accomplished through the following key components.

A. Intent Inference

This module is embedded within the IBS Service and analyzes pre-trained results and incoming user requirements in real-time to derive system rules based on intent. This module comprises two sub-modules, and the final intent-based operation is determined by integrating the results of these two modules. This module is essential to the IBS Service, as it manages the appropriate processing and optimization of all intents before they are input into the system.

1) Intent Analyzer: It is a key step to derive specific intent based on user requirements and low-level data in the system. The data undergoes the Translation, Resolution, Activation, and Assurance processes during this stage. This module performs the important function of translating user input into system-executable intent and verifies the accuracy and feasibility of the intent through four processes.

2) Policy DB: It is involved in the interpretation and alignment of the intentions produced in the *Intent Analyzer* with the operational policies of the system. By referring to the *Policy DB*, which stores profiles, policies, and parameters required for system operation, the Policy Parser searches for appropriate system policies based on the intent and derives specific system rules.

B. Intent Controller

It is responsible for configuring and managing the system according to rules derived from inferred intent and queried policies. This can include changing network settings, allocating resources, virtualization, and more.

C. IBS Service

An environment containing real hardware or software resources, including the *Intent Inference*, *Intent Controller*, and *System Monitor* modules. This environment provides the basis for each module to operate effectively, executing and monitoring the intent in the actual system environment, dynamically configuring and optimizing it according to requirements. In particular, IBS Service can be applied in various fields such as military communications, autonomous vehicles, and IIoT environments to maximize efficiency and responsiveness.

D. User Requirement

The user's requirements that are actually entered into the IBS Service. This is the starting point for the system and is passed to the *Intent Inference* module.

E. System Monitor

It is responsible for continuously monitoring and collecting the status of the dynamic system as well as user requirements in the IBS Service. It performs the critical function of detecting and reporting issues as they arise, and suggests necessary adjustments. Monitoring results are fed back to *Intent Inference*, enabling the system to self-learn, optimize, and continuously improve.

In this way, IBS turns explicit user requests into automated system responses, efficiently improving upon traditional manual and static system operations. This not only transforms system management into a more proactive and preventive approach, but also ensures a high level of user satisfaction. Fig. 2 below shows the IBS components and their interactions.

III. Intent Inference

The Intent Inference module, the core component of an Intent-based System, consists of the *Intent Analyzer* and *Policy DB*, which are essential for accurately identifying intent and translating it into appropriate system policies. In this section, we will explore the various environments in which IBS can be applied, specifically the scenarios in which IBS is applied in military communications, and what system state data is relevant to the intent that can be inferred. IBS facilitates efficient decision-making in a military communication environment where timely and accurate system operation is crucial

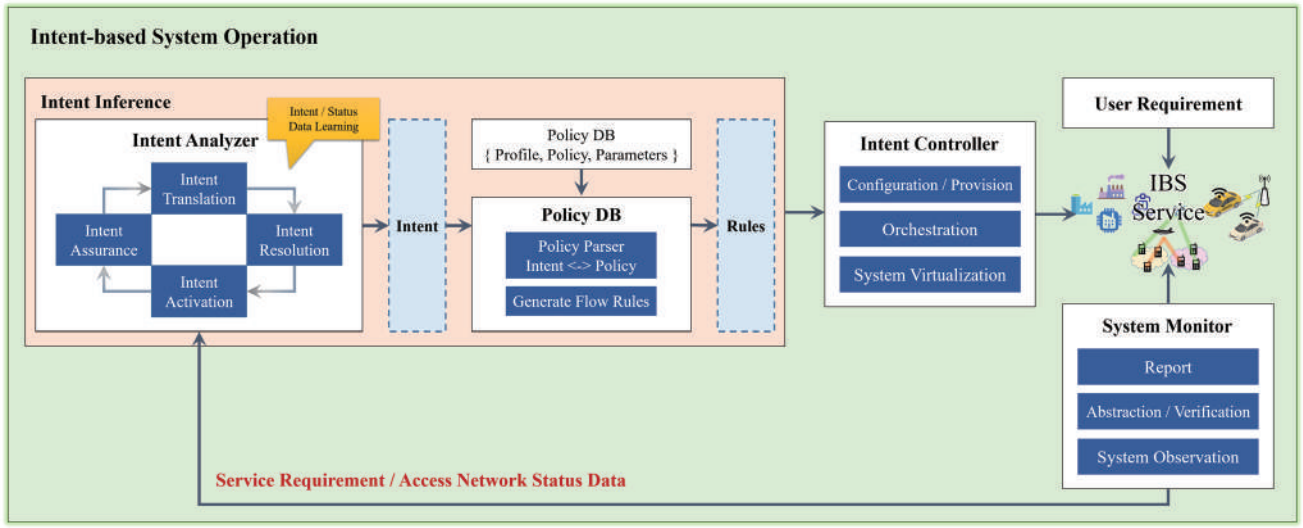


Fig. 1. Intent-based System Architecture

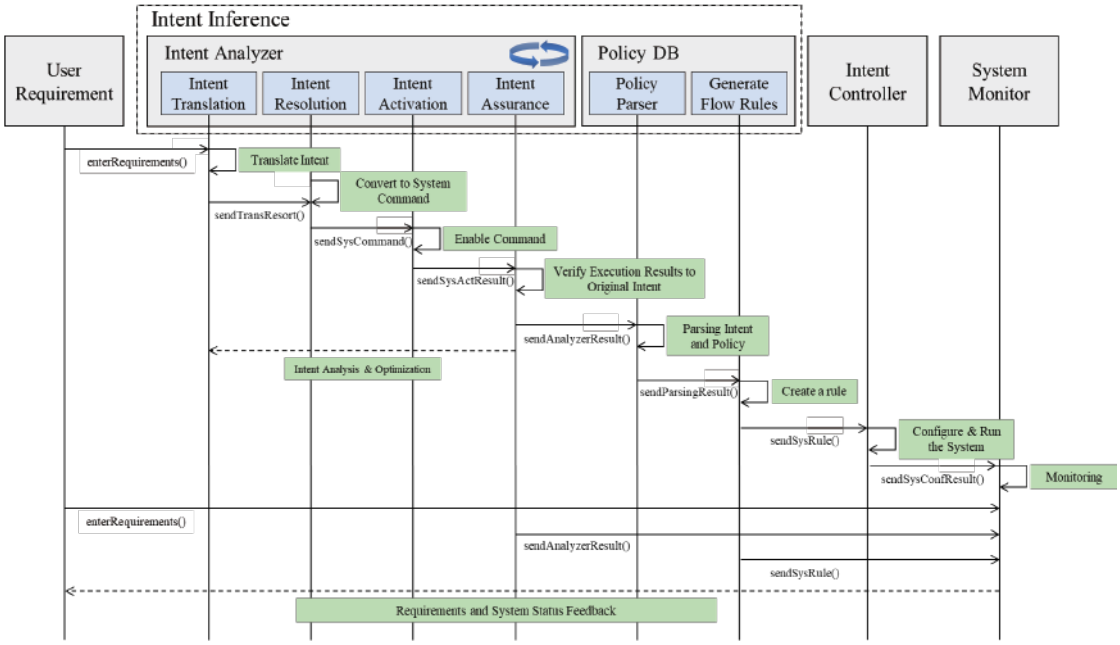


Fig. 2. The operational flow of IBS

by accurately interpreting user intentions and aligning them with real-time system state data.

[Scenario] Military communications networks are composed of a complex of heterogeneous networks, each of which is categorized into three types: ground, satellite, and UAV networks. Link quality differs depending on the three types of networks, and accommodates six distinct military traffic characteristics. Optimizing system operation by considering the specific attributes of military communication networks.

A. Intent Analyzer

For each heterogeneous network type, there are quality conditions and operational requirements, which are the requirements of the military communications [6] user and lead directly to the Intent. The Intent Analyzer has the steps necessary to translate and activate to infer intent from user requirements. This requires learning data between system state and intent by mapping the user's intent based on various system state data. The intentions that can be generated for data learning are shown in Table 1 below.

B. Policy DB

The *Policy DB* performs the role of analyzing the inferred intent through the *Intent Analyzer* and generating actual system rules. Policies are developed based on prioritization considering the operational requirements and quality of each

TABLE I
MILITARY COMMUNICATIONS SYSTEM STATE DATA AND SYSTEM PARAMETERS BY INTENT

System Status Data				Network Type	System Parameters by Intent				
Delay (ms)	Jitter (ms)	Throughput (kbps)	Success Rate (%)		Bandwidth (Mbps)	Service Traffic	E2E Latency (MS)	E2E Packet Loss (%)	E2E Jitter (MS)
0.00084	0.00017	10	1	Ground	10	Short Message	1000	0.5	20
0.10036	0.001	1.83	0.916	Satellite	2	Assured Voice	220	1	20
0.00301	0.00216	2	1	UAV	2	Broadcast Video	1000	0.1	20

heterogeneous network type, ensuring they match the traffic characteristics of military communication. For example, real-time operational data takes higher priority than non-urgent service data. This maximizes the overall efficiency of the system and ensures that important information is transmitted quickly. The two algorithms below are examples of the process of developing policies based on the priorities set. According to these policies, the *Policy DB* generates flow rules suitable for each network type and situation.

Algorithm 1 Priority Setting in Heterogeneous Networks for Military Communications

Data: $networkType \in$ {"ground", "satellite", "UAV"}
Result: Assigns priorities to different network types and creates network flow rules.
Initialize $priorityMap$ as an empty dictionary
Define $priorityValues$: { "ground" $\rightarrow$ 1, "UAV" $\rightarrow$ 2, "satellite" $\rightarrow$ 3 }
function ASSIGNPRIORITY($networkType$)
    $priority \leftarrow priorityValues[networkType]$
    $priorityMap[networkType] \leftarrow priority$
    return $priority$
end function
function CREATEFLOWRULES($networkType$)
    $priority \leftarrow$ ASSIGNPRIORITY($networkType$)
    if $priority = 1$ **then**
        // Rules for high priority ground network
        DEPLOYRULES("Optimize bandwidth and minimize latency")
    else if $priority = 2$ **then**
        // Rules for medium priority UAV network
        DEPLOYRULES("Ensure robust signal strength and mobility support")
    else if $priority = 3$ **then**
        // Rules for lower priority satellite network
        DEPLOYRULES("Manage high latency and synchronization")
    end if
end function
CREATEFLOWRULES($networkType$)

IV. CONCLUSION

In this paper, we provide a detailed explanation of the concept and structure of the Intent-based System (IBS) and investigate the methodology of intent inference in military communication. In addition to the basic IBN process, IBS provides the ability to automatically recognize and process intentions in real-world systems by learning user intentions and system state data in advance. IBS shifts the current passive and static system management methodology by prioritizing the user's explicit intent as the primary factor in system operation, thus facilitating a more automated system management process. In future research, we plan to further develop the technical implementation of IBS and validate the system through performance testing in real-world environments. This will extend the scope of application of IBS technology and explore the application of the theoretical model to real-world operational systems. The methodology presented in this paper

Algorithm 2 Military Traffic Priority Assignment Based on Traffic Type

Data: Traffic type $\in$ {"voice", "text message", "video"}
Result: Assign network settings based on priority and type-specific requirements.

Define $trafPriority$: { "Assured voice" $\rightarrow$ 1, "Short Message" $\rightarrow$ 2, "Broadcast Video" $\rightarrow$ 3 }

Define $networkRequirements$:
{
    "Assured voice" $\rightarrow$ { "latency" $\rightarrow$ <150 ms, "jitter" $\rightarrow$ <1 ms, "success rate" $\rightarrow$ >99.7% },
    "Short Message" $\rightarrow$ { "latency" $\rightarrow$ <400 ms, "jitter" $\rightarrow$ <2 ms, "success rate" $\rightarrow$ >97% },
    "Broadcast Video" $\rightarrow$ { "latency" $\rightarrow$ <600 ms, "jitter" $\rightarrow$ <4 ms, "success rate" $\rightarrow$ >99% }
}

procedure APPLYNETWORKSETTINGS($trafType$)
    $priority \leftarrow trafPriority[trafType]$
    $requirements \leftarrow networkRequirements[trafType]$
    ADJUSTNETWORKSETTINGS($requirements$)
    return "Settings applied for " + $trafType$ + " with priority " + $priority$
end procedure
APPLYNETWORKSETTINGS($trafType$)

is expected to serve as a springboard for future applications, and the development and application of IBS technology will maximize the efficiency of system operation and enable more intelligent and responsive system management.

REFERENCES

[1] M. Gharbaoui, B. Martini and P. Castoldi, "Intent-Based Networking: Current Advances, Open Challenges, and Future Directions," 2023 23rd International Conference on Transparent Optical Networks (ICTON), Bucharest, Romania, 2023, pp. 1-5.

[2] A. Leivadeas and M. Falkner, "Autonomous Network Assurance in Intent Based Networking: Vision and Challenges," 2023 32nd International Conference on Computer Communications and Networks (ICCCN), Honolulu, HI, USA, 2023, pp. 1-10.

[3] H. Yu, H. Rahimi, C. Janz, D. Wang, C. Yang and Y. Zhao, "A Comprehensive Framework for Intent-Based Networking, Standards-Based and Open-Source," NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium, Miami, FL, USA, 2023, pp. 1-6.

[4] D. Wang, R. Su, S. Zhang and Y. Xing, "Trends and Challenges of Policy Verification for Intent-based Networking towards 6G," 2022 IEEE/CIC International Conference on Communications in China (ICCC Workshops), Sanshui, Foshan, China, 2022, pp. 134-139.

[5] T. A. Khan, A. Muhammad, W. Akbar, A. Mehmood, A. Rafiq and W. -C. Song, "Intent-based Networking Approach for Service Route and QoS control on KOREN SDI," 2021 IEEE 7th International Conference on Network Softwarization (NetSoft), Tokyo, Japan, 2021, pp. 24-30.

[6] Teresa, M. T. "Unified Capabilities Requirements 2013 (UCR 2013)." Department of Defense: Washington, DC, USA, 2013.

Low Complexity MU-MIMO Scheduling for B5G Cellular Network

Satoshi Denno, Akihito Yamada, and Yafei Hou
Faculty of Environmental, Life, Natural Science and Technology
Okayama University
Okayama, Japan
denno@okayama-u.ac.jp

***Abstract*—This paper proposes a low complexity scheduling for uplink multi-user multiple-input-multiple-output (MU-MIMO) that takes an approach of iterative heuristic search. The proposed scheduling randomly selects some of all the users in a cell, and searches the best combination of users in the possible combinations of the selected users for the MU-MIMO communication. The proposed scheduling iterates the above user search, while the best combination of the user is carried over to the user combination search at the next iteration. The performance of the proposed scheduling is evaluated in a three-dimensional wireless network by computer simulation. Even though the proposed scheduling can be implemented with much less complexity than the optimum scheduling, the performance of the proposed scheduling is shown to be near as superior as that of the optimum scheduling.**

***Index Terms*—Multi-user multiple-input-multiple-output, scheduling, throughput maximization, minimum mean square error (MMSE) filter, low complexity**

I. Introduction

Network throughput has been demanded to raise in wireless access networks, for instance, the cellular networks and the IEEE802.11 wireless local area networks (WLAN). Many cutting edge technologies have been applied for increasing network throughput. Among those technologies, multi-input-multiple-output (MIMO) has been playing the most important role in network throughput enhancement [1]–[3]. Since MIMO spatial multiplexing can raise the transmission speed of a wireless link in wireless networks by increasing the number of antennas on the transmitter and the receiver increases, many antennas has been installed in wireless access networks. Especially, the 5th generation cellular system (5G) applies lots of antennas on the base station, which is called "Massive MIMO" [4]–[6]. For further network throughput enhancement, multiuser MIMO (MU-MIMO) has been considered, because MU-MIMO enables to establish concurrent communications with several users at the same time in the same frequency band [7], which multiplies the network throughput. User scheduling is known to enable systems with MU-MIMO to exploit the user diversity gain, which improves the transmission performance and the network throughput [8], [9]. Because precoding is indispensable in the downlink MU-MIMO, precoding has been investigated in conjunction with scheduling [10]–[12]. However, base stations have to collect channel matrices with all the users. Because that scheduling imposes not only the overhead to collect the channel matrices but also complexity to select the best combination in all the candidate combinations on the base stations, the complexity reduction techniques for scheduling have been considered [13], [14]. Since MU-MIMO can be also implemented in uplinks of wireless access systems, scheduling for MU-MIMO in uplinks has been considered [15]–[18]. Moreover, uplink scheduling techniques are extended to multi-cell wireless networks [19]–[23]. On the other hand, higher frequency bands tend to be applied to emerging wireless systems. To compensate severe attenuation of wireless signals in such higher frequency band, massive MIMO is going to be introduced in such systems. The massive MIMO causes user scheduling more complex [24]–[27]. To combat with that difficult problem, new approaches, for instance, application of machine learning, have been proposed [28]–[30]. For the beyond 5th generation cellular system, a new concept called "Cell-free radio access network (RAN)" has been proposed [31], where users select base stations to connect despite of a cell architecture of cellar systems. Because MIMO is indispensable even in systems with the cell-free RAN [32], user scheduling has also been investigated for cell-free massive MIMO systems [33]. As is described above, the optimum user scheduling is too complex, because it applies the brute force search that needs prohibitive high computational complexity. Two stage scheduling is applied in most MIMO scheduling techniques. For instance, users are divided into several user groups, and the best group is selected at the first stage. In the second stage, the optimum beamforming weight is searched among all the possible beamforming weights in the group. Heuristic algorithms are applied to the first stage for complexity reduction in some of the conventional techniques. However, those approaches need channel state information between all the users and the base station, which increases the complexity of the scheduling. Conventional techniques using heuristic algorithms are still complex to implement in wireless communication systems.

This paper proposes a low complexity scheduling for uplink MU-MIMO that takes an approach of iterative heuristic search. We apply a utility function for the maximum sum throughput, which helps us reduce the computational complexity to select the users. The proposed scheduling requires only

This work has been supported by JSPS KAKENHI JP21K04061, the support center for advanced telecommunications technology research (SCAT), and Softbank Co. ltd.

several users to send short packets for the sounding, and selects some promising users out of the users having sent the sounding packets. The proposed scheduling iterates the above simple signal processing. The performance is evaluated in 3-dimensional (3D) wireless networks which are going to be provided in the next generation wireless communication systems.

Throughout the paper, $\arg\min_{n\in B}[f(n)]$ denotes a value of n included in a set B that minimizes a function $f(n)$. Superscript T and H indicate transpose and Hermitian transpose of a matrix or a vector. $\mathbf{A}^{-1}$, and $\mathrm{tr}[\mathbf{A}]$, denote inverse and a trace of a matrix $\mathbf{A}$, respectively.

II. System Model

We assume an uplink of a wireless access system where users with N_{T} antennas are scattered around a base station with N_{R} antennas. N_{U} users are randomly moving in a cell with the diameter $d_{\mathrm{H}} \in \mathbb{R}$ where the base station is located at the center. Though the user locations are drawn in the 2-dimensional picture, we assume that the user locations are also distributed in the vertical direction. In fact, the users are located below the height of $d_{\mathrm{V}} \in \mathbb{R}$ in the cell. The distribution of the users' location is assumed to be uniform in the 3D cell. When the user locations are distributed in the 3D-cell, the injection angles of the signals received at the base station are distributed vertically and horizontally. To cope with the received signals, we apply a uniform planar array antenna with N_{V} by N_{H} elements where N_{V} and N_{H} represent the number of antenna elements in the vertical axis and in the horizontal axis, respectively.

The information bit stream is encoded with a convolutional code and fed to a modulator via a bit interleaver. The modulator output signals are divided into N_{T} streams and every stream is provided to the inverse Fourier transform (IFFT) for the orthogonal frequency division multiplexing (OFDM). The IFFT output signal stream is emitted from each antenna after the cyclic prefix (CP) is added as a header of the IFFT output signals. While N_{U} users exist in the 3D cell, the N_{P} users are allowed to simultaneously send their packets for the base station. Let $\mathbf{x}_p(n) \in \mathbb{C}^{N_{\mathrm{T}}}$ denote an nth transmission signal vector in the packet sent from the pth user, which is defined as $\mathbf{x}_p(n) = (x_{p,0}(n) \cdots x_{p,N_{\mathrm{T}}-1}(n))^{\mathrm{T}}$ where $x_{p,m}(n) \in \mathbb{C}$ denotes a transmission signal sent from the mth antenna of the pth user at time index n, a signal vector received at the base station is represented as,

$$\mathbf{y}(n) = \sum_{p=0}^{N_{\mathrm{P}}-1}\sum_{i=0}^{N_\tau-1} \mathbf{T}_p(i)\,\mathbf{x}_p(n-i) + \mathbf{n}(n) \quad (1)$$

In (1), $N_\tau \in \mathbb{Z}$, $\mathbf{y}(n) \in \mathbb{C}^{N_{\mathrm{R}}}$, $\mathbf{n}(n) \in \mathbb{C}^{N_{\mathrm{R}}}$ and $\mathbf{T}_p(i) \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$ denote a channel impulse response length, a received signal vector at the base station, an additive white Gaussian noise (AWGN) vector, and a channel matrix of the ith path between the base station and the pth user. The received signal vector is fed to the fast Fourier transform (FFT) after the CP is removed. Let $N_{\mathrm{F}} \in \mathbb{Z}$ indicate an FFT point, the FFT output signal vector $\mathbf{z}(m) \in \mathbb{C}^{N_{\mathrm{R}}}$ is described as,

$$\begin{aligned}\mathbf{z}(m) &= \frac{1}{\sqrt{N_{\mathrm{F}}}}\sum_{n=0}^{N_{\mathrm{F}}-1}\mathbf{y}(n)\mathrm{e}^{-\mathrm{j}2\pi\frac{nm}{N_{\mathrm{F}}}}\\ &= \sum_{p=0}^{N_{\mathrm{P}}-1}\dot{\mathbf{T}}_p(m)\,\dot{\mathbf{x}}_p(m) + \dot{\mathbf{n}}(m)\\ &= \dot{\mathbf{T}}_{\Omega_0}(m)\,\dot{\mathbf{X}}_{\Omega_0}(m) + \dot{\mathbf{n}}(m). \end{aligned} \quad (2)$$

$\dot{\mathbf{T}}_p(m) \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$, $\dot{\mathbf{x}}_p(m) \in \mathbb{C}^{N_{\mathrm{T}}}$, and $\dot{\mathbf{n}}(m) \in \mathbb{C}^{N_{\mathrm{R}}}$ in (**??**) represent an mth frequency response matrix in the channel between the base station and the pth user, an mth frequency spectrum of the transmission signals sent from the pth user, and the mth instantaneous spectrum of the AWGN vector in the mth subcarrier, which are defined as $\dot{\mathbf{T}}_p(m) = \sum_{i=0}^{N_\tau-1}\mathbf{T}_p(i)\,\mathrm{e}^{-\mathrm{j}2\pi\frac{im}{N_{\mathrm{F}}}}$, $\dot{\mathbf{x}}_p(m) = \frac{1}{\sqrt{N_{\mathrm{F}}}}\sum_{n=0}^{N_{\mathrm{F}}-1}\mathbf{x}_p(n)\,\mathrm{e}^{-\mathrm{j}2\pi\frac{nm}{N_{\mathrm{F}}}}$, and $\dot{\mathbf{n}}(m) = \frac{1}{\sqrt{N_{\mathrm{F}}}}\sum_{n=0}^{N_{\mathrm{F}}-1}\mathbf{n}(n)\,\mathrm{e}^{-\mathrm{j}2\pi\frac{nm}{N_{\mathrm{F}}}}$, respectively. In addition, $\dot{\mathbf{T}}_{\Omega_0}(m) \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}N_{\mathrm{P}}}$ and $\dot{\mathbf{X}}_{\Omega_0}(m) \in \mathbb{C}^{N_{\mathrm{T}}N_{\mathrm{P}}}$ represent a composite frequency response matrix and a composite transmission signal vector in the mth subcarrier, which are defined as $\dot{\mathbf{T}}_{\Omega_0}(m) = \left(\dot{\mathbf{T}}_0(m)\cdots\dot{\mathbf{T}}_{N_{\mathrm{P}}-1}(m)\right)$ and $\dot{\mathbf{X}}_{\Omega_0}(m) = \left(\dot{\mathbf{x}}_0(m)^{\mathrm{T}}\cdots\dot{\mathbf{x}}_{N_{\mathrm{P}}-1}(m)^{\mathrm{T}}\right)^{\mathrm{T}}$ where Ω_0 indicates a user combination index. The received signal $\mathbf{z}(m)$ is provided for the frequency response matrix estimation, called "sounding". The sounding period precedes the signal detection period in the proposed scheduling.

A minimum mean square error (MMSE) filter is applied for the signal detection for the proposed scheduling being practical. As is well-known, the MMSE filter can detect N_{R} signal streams with high detection performance. When every user sends N_{T} streams without any precoding, the number of the users to send the streams should be limited to $N_{\mathrm{S}} = \frac{N_{\mathrm{R}}}{N_{\mathrm{T}}}$. This implies that the number of the users to send the signal simultaneously during the sounding period is possibly different from that during the detection period.

Let B_{Ω_k} denote a set of the user indexes for the MMSE filter, e.g, $B_{\Omega_k} = \{p_{\Omega_k}(0), \cdots, p_{\Omega_k}(N_{\mathrm{S}}-1)\}$, where Ω_k and $p_{\Omega_k}(i)$ indicate an index of the set and an ith entry in the set, the MMSE filter $\mathbf{W}_{\Omega_k}(m) \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{R}}}$ for the mth subcarrier is defined as,

$$\mathbf{W}_{\Omega_k}(m) = \left(\bar{\dot{\mathbf{T}}}_{\Omega_k}(m)\,\bar{\dot{\mathbf{T}}}_{\Omega_k}(m)^{\mathrm{H}} + \frac{2\sigma^2}{\sigma_{\mathrm{D}}^2}\mathbf{I}\right)^{-1}\bar{\dot{\mathbf{T}}}_{\Omega_k}(m). \quad (3)$$

In (3), σ^2, σ_{D}^2 $\mathbf{I} \in \mathbb{R}^{N_{\mathrm{R}}\times N_{\mathrm{R}}}$, and $\bar{\dot{\mathbf{T}}}_{\Omega_k}(m) \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}N_{\mathrm{S}}}$ denote a single side band noise density, power of the modulation symbols, the N_{R}-dimensional identity matrix, and an estimated composite response matrix in the mth subcarrier between the base station and the users included in the set B_{Ω_k}, which is defined in the following section. The signal detection is carried out as,

$$\bar{\dot{\mathbf{X}}}_{\Omega_k} = \mathbf{W}_{\Omega_k}(m)^{\mathrm{H}}\,\mathbf{z}(m), \quad (4)$$

where $\bar{\dot{\mathbf{X}}}_{\Omega_k} \in \mathbb{C}^{N_{\mathrm{T}} N_{\mathrm{P}}}$ denotes an estimate of the composite transmission signal vector in the mth subcarrier, $\dot{\mathbf{X}}_{\Omega_0}(m)$.

We apply a utility function for maximizing the sum throughput in this paper. In other words, the scheduler searches the best combination of the users that maximizes the sum throughput. When the MMSE filter is used for the signal detection, the sum throughput maximization can be approximately achieved by selecting the user combination that maximizes the sum of the signal to noise power ratio (SNR) of the MMSE output signals, because the MMSE filter is expected to orthogonalize the output signals. This approximation can make the scheduling working without the logarithm function, which reduces the complexity of the scheduling, while the logarithm function is usually necessary in the scheduling for the sum throughput maximization. Even though the optimum scheduling achieves the best transmission performance, on the other hand, the high computational complexity is necessary to implement the optimum scheduling, because the optimum scheduling executes the brute force search to search the best combination of the users among all the possible combinations.

This paper propose a scheduling technique to find sub-optimum combination of the users with small computational complexity, which is presented in the following.

III. Low Complexity Scheduling

While all the users send their packets for the base station to estimate the frequency response matrices with all of them, only N_{P} users are randomly selected to send the packets for the sounding in the proposed scheduling, because the base station has no knowledge about the frequency response with the users. We assume that the packet conveys some control bits, only one eighth of all the subcarriers are occupied by pilot signals for channel estimation. The base station carries out the channel estimation for the subcarriers conveying the pilot signals. In the proposed scheduling, the least square channel estimation is applied. Let an estimated composite frequency response matrix in the mth subcarrier be denoted by $\bar{\mathbf{T}}_{\Omega}(m) \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}} N_{\mathrm{S}}}$, the proposed scheduling searches the best combination of the users as follows.

$$\bar{\Omega}(n) = \underset{\Omega_k \in B_{\mathrm{C}(n)}}{\arg\min} \left[\mathrm{tr} \left[\sum_{m=0}^{N_{\mathrm{F}}-1} \left(\bar{\dot{\mathbf{T}}}_{\Omega_k}(m) \bar{\dot{\mathbf{T}}}_{\Omega_k}(m)^{\mathrm{H}} + \frac{2\sigma^2}{\sigma_{\mathrm{D}}^2} \mathbf{I} \right)^{-1} \right] \right] \quad (5)$$

In (5), $\bar{\Omega}(n)$ and $B_{\mathrm{C}(n)}$ indicate the best combination of the users at the nth round [1] and a set accommodating the indexes of the possible user combinations, Ω_k $k = 0 \cdots {}_{N_{\mathrm{P}}+N_{\mathrm{S}}}\mathrm{C}_{N_{\mathrm{S}}} - 1$ at the nth round [2].

Since the N_{P} users are not probably the optimum from the view point of the transmission performance, the proposed scheduling repeats the channel estimation and the best user combination selection many rounds, where one sequence of the channel estimation and the user combination selection is counted as a round. Actually, the proposed scheduling iterates the sequence with a procedure described below, many rounds, for improving the transmission performance.

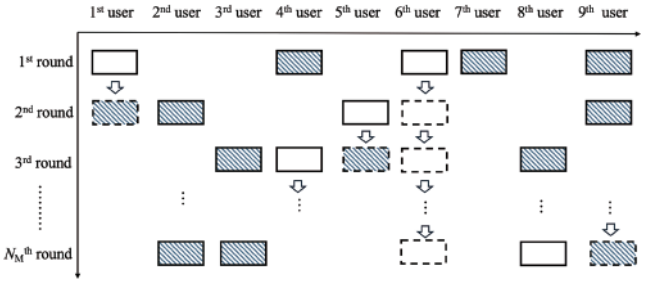


Fig. 1. Multi Round Packet Transmission

The user indexes searched in (5) are stored in the user set $B_{\mathrm{u}(n+1)}$. In addition to those, the indexes of the N_{P} users sending their packets at the $n+1$ round are also stored in the user set $B_{\mathrm{u}(n+1)}$, which increases the entries of the set $B_{\mathrm{u}(n+1)}$ to $N_{\mathrm{P}} + N_{\mathrm{S}}$. For the search in (5) at the $n+1$ round, the N_{S} user indexes out of the $N_{\mathrm{P}} + N_{\mathrm{S}}$ indexes are grouped in a candidate combination index. Since the $N_{\mathrm{P}} + N_{\mathrm{S}}$ users are stored in the user set $B_{\mathrm{u}(n+1)}$, ${}_{N_{\mathrm{P}}+N_{\mathrm{S}}}\mathrm{C}_{N_{\mathrm{S}}}$ candidate indexes can be generated, and all of them are stored in $B_{\mathrm{C}(n+1)}$, which is applied to the search in (5) at the $n+1$ round. As is described above, the packet transmission and the selection are regarded as one round, and they are repeated N_{M} rounds. Finally, the proposed scheduler selects the user combination searched at the final round as its output. However, to achieve better performance with the same complexity, the $N_{\mathrm{P}} + N_{\mathrm{S}}$ users are allowed to simultaneously send their packets at the first round, because the set $\mathbf{B}_{\mathrm{u}(1)}$ is null before the first round. The repetitive sounding packets transmission in the proposed scheduling is named "Multi round packet transmission" in this paper, which is illustrated in figure 1. The horizontal axis and the vertical axis are the user indexes and the round numbers, respectively. This is an example of $N_{\mathrm{P}} = 3$, $N_{\mathrm{S}} = 2$, and $N_{\mathrm{U}} = 9$. The 1st, 4th, 6th, 7th, and 9th users are permitted to send the packets at the first round. The 1st and 6th users are selected by the procedure in (5). The shaded blocks mean the packets that are not selected. The selected user indexes, 1st and 6th user indexes, are carried over to the next round by storing them in the set $B_{\mathrm{u}(2)}$. The box with broken line edges means a packet carried over from the previous round. While the 2nd, 5th, and 9th users send their packets, only the combination of the 5th and 6th users is selected among the ${}_5\mathrm{C}_2$ combinations, the other user indexes are discarded at the 2nd round. Those signal transmission and the selection are repeated up to the final round of the N_{M}th round. As a result, the combination of the 6th and 8th users is selected, which is scheduled to communicate with the base station.

IV. Simulation

The performance of the proposed scheduling is confirmed by computer simulations in a 3D cellular system. While the base station is assumed to be located at the height of 6 meters,

[1] The round is defined below.

[2] we describe the reason why the number of the entries in the set $B_{\mathrm{C}(n)}$ is ${}_{N_{\mathrm{P}}+N_{\mathrm{S}}}\mathrm{C}_{N_{\mathrm{S}}}$ when the proposed scheduling is summarized afterwards.

TABLE I
SIMULATION PARAMETERS

Multicarrier transmission	OFDM
The carrier frequency f	4 GHz
The number of subcarriers,N_{F}	64
The number of the round N_{M}	10
The number of antennas on a terminal and a base station, $(N_{\mathrm{T}}, N_{\mathrm{R}})$	$(2,\ 6)$
The BS planar array antenna configuration and spacing,$(N_{\mathrm{H}}, N_{\mathrm{V}}, d_{\mathrm{h}}, d_{\mathrm{v}})$	$(3, 2, 0.5, 0.5)$
The height of the antennas on the base station, d_{BS}	6 m
Antenna spacing (d_{v}, d_{h})	$(\frac{\lambda}{2},\ \frac{\lambda}{2})$
Channel model	NLOS and LOS
NLOS channel model	4-path Rayleigh fading
LOS channel model	Rice fading with K-factor of 5 dB
Cell size, d_{H}	100 m
The maximum height of the users, d_{V}	10 m
The distribution of the users in the Z axis and X-Y plane	uniform
The number of users sending the packet par round, N_{P}	6
The number of the users to select, N_{S}	3
The number of all users, N_{U}	66
The normalized maximum Doppler frequency, $f_{\mathrm{D}}T$	2.3×10^{-5}

the user location is distributed uniformly in the 3D cell. The channel model depends on the height of the users. When the users are located higher than the base station, because the line of sight (LOS) signals can arrive at the base station, the instantaneous signal fading can be modeled as the rice fading. That channel is called "LOS channel". Otherwise, the instantaneous signal fading is reduced to the Rayleigh fading based on Jakes' model [36], because LOS signals from the user can be difficult to arrive at the base station. That channel is called "NLOS channel". The pass loss models are also dependent on the height of the users. The following path loss models are applied to the LOS channel and the NLOS channel in the performance evaluation [34], [35].

$$L_{\mathrm{NLOS}} = 39.08 \log_{10}\left(\frac{d}{d_0}\right) + 20 \log_{10}(f) - 0.6\,(h_{\mathrm{UE}} - 1.5)\,[\mathrm{dB}] \quad (6)$$

$$L_{\mathrm{LOS}} = 25 \log_{10}\left(\frac{d}{d_0}\right) + 20 \log_{10}(f)\,[\mathrm{dB}] \quad (7)$$

In (6) and (7), $f \in \mathbb{R}$, $d \in \mathbb{R}$, L_{NLOS}, and L_{LOS}, represent carrier frequency, distance between the base station and one of the users, path loss in dB for the NLOS channel and that for the LOS channel, respectively. In addition, $d_0 \in \mathbb{R}$ indicates distance from the base station, at which the energy per bit to noise density ratio (E_{b}/N_0) is defined, instead of defining the transmission power and the noise figure. We assume two types of user distributions. The one is that all the users are located below the base station, which means that all the channels are NLOS. The other is that some of the users are located above the base station and the other users are below the base station, which channel is described as (LOS α%, NLOS $100-\alpha$ %) in this paper. On the other hand, we evaluate the performance of the optimum scheduling as a reference. Simulation parameters are listed in Table 1.

A. SNR Distribution

Fig. 2 shows the performance of the proposed scheduling in (LOS 20 % and NLOS 80 %) channels. The performances of the random scheduling and the optimum scheduling are added for comparison in the figure. Obviously, the proposed scheduling achieves better SNR performance than the random scheduling. Similar as the performance in Fig. **??**, the performance of the proposed scheduling comes close to that of the optimum scheduling as the number of the rounds increases. The proposed scheduling with the 10 rounds is only 2dB inferior to the optimum scheduling at the CDF of 10^{-2}.

The performance in Fig. 2 is better than that in Fig. **??**, which means that the performance can be improved as the percentage of the LOS channels is increased. Fig. 3 confirms the performance where the SNR is evaluated with respect to the percentage of the LOS channels. The CDF of the SNR is shown in the figure when the percentage of the LOS is changed from 0 % (NLOS channel) to 100 % (complete LOS channel). As is inferred in the previous observation, the performance is improved as the percentage of the LOS channels increases.

B. Complexity

Fig. 4 shows the complexity of the proposed scheduling with respect to the number of the rounds. In the figure, the complexity is evaluated in terms of the number of the complex multiplications needed for the channel estimation and the selection. The complexity of the optimum scheduling is also added in the figure. Although the complexity of the proposed scheduling is increasing in proportion to the number of the rounds, the complexity is much less than that of the optimum scheduling as long as the number of the rounds is less than 10. In fact, the complexity of the proposed scheduling with the 10 rounds is approximately one hundredth as much as that of the optimum scheduling. Since the optimum scheduling collects all the frequency response matrices and search the best combination of the users among all the combinations, i.e., ${}_{N_{\mathrm{U}}}\mathrm{C}_{N_{\mathrm{S}}}$ combinations, with the brute force search technique,

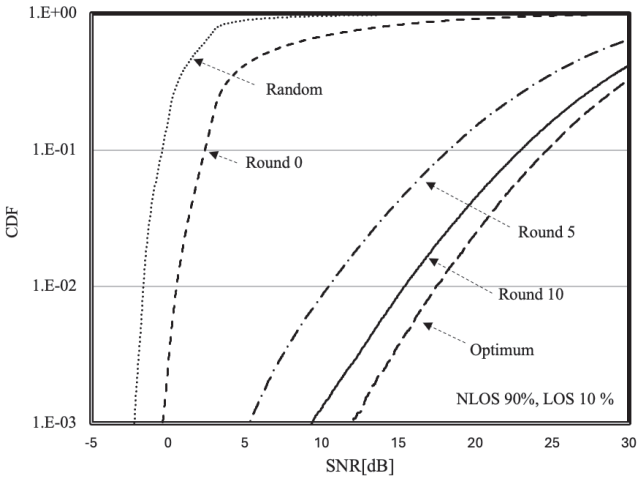


Fig. 2. CDF of SNR in (LOS 20 % and NLOS 80 %) channels

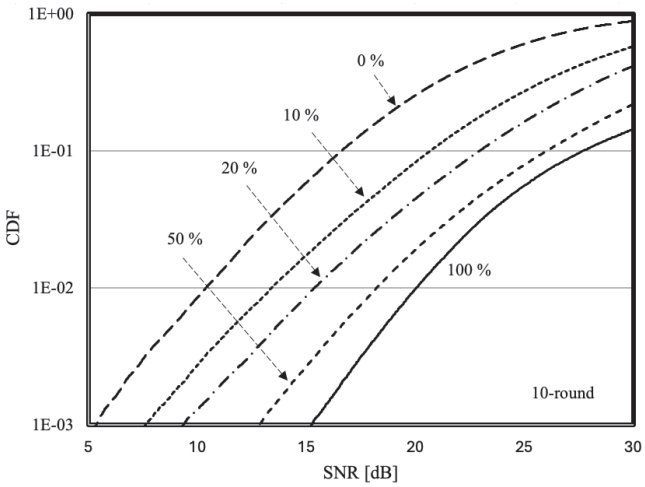


Fig. 3. SNR Distribution With Respect to User Percentage in LOS Channels

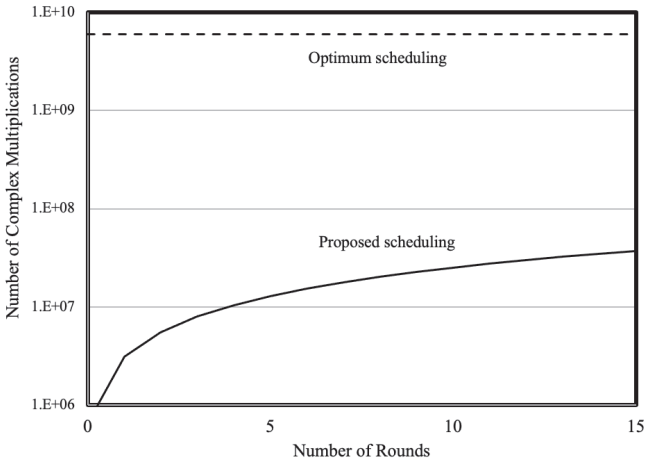


Fig. 4. Complexity

the complexity is quite high. On the other hand, although the proposed scheduling applies the brute force technique to search the best combination, the number of the possible combinations a round is reduced to ${}_{N_{\mathrm{S}}+N_{\mathrm{P}}}\mathrm{C}_{N_{\mathrm{S}}}$. This is the reason why the complexity of the proposed scheduling is much less than that of the optimum scheduling, even if the proposed scheduling iterates the searches many rounds.

V. Conclusion

This paper proposes a low complexity scheduling for uplink MU-MIMO that takes an approach of iterative heuristic user selection. If a sequence of a random user selection, channel estimation for the users, and the best user combination search is defined as a round, the proposed scheduling iterates the round many times. However, the user indexes in the best combination at a round are carried over to the next round. The base station searches the best combination of the users from the carried over users and the uses selected for this round.

The performance of the proposed scheduling is evaluated by computer simulations in a 3D cellular system where the users are located in a cell where the users are distributed vertically and horizontally.. When the number of the users with 2 antennas is 66 and only the base station with 6 antennas exists in the center of the cell at the height of 6 m, and only 6 users are permitted to simultaneously send the packet a round in the proposed scheduling, the proposed scheduling achieves the performance only $3dB$ inferior to the optimum scheduling for an MU-MIMO communication system with the MMSE filter. However, the complexity of the proposed scheduling is approximately one hundredth as much as that of the optimum scheduling.

Acknowledgment

The work has been supported by JSPS KAKENHI JP21K04061 and JP24K07475, the support center for advanced telecommunications technology research (SCAT), and Softbank Co. ltd.

References

[1] G. J. Foschini and M. J. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” Wireless Pers. Commun., vol.6, no.3, pp.311—335, 1998.

[2] I. E. Telater, “Capacity of multi-antenna Gaussian channels,” European Trans. Telecommun., vol. 10, no. 6, pp. 585–595. 1999.

[3] S. Yang and L. Hanzo, “Fifty years of MIMO detection: The road to large-scale MIMOs,” IEEE Commun. Surveys & Tutorials, vol.17, no.4, pp.1941–1988, Fourthquarter 2015.

[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Communications Magazine, vol. 52, no. 2, pp.186–195, Feb. 2014.

[5] M. Sakai, K. Kamohara, H. Iura, H. Nishimoto, K. Ishioka, Y. Murata, M. Yamamoto, A. Okazaki, N. Nonaka, S. Suyama, J. Mashino, A. Okamura, and Y. Okumura, “Experimental field trials on MU-MIMO transmissions for high SHF wide-band massive MIMO in 5G,” IEEE Trans. Wireless Commun., vol. 19, no. 4, pp.2196–2207, April 2020.

[6] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” IEEE J. of Select. Topics in Signal Process., vol.8, no.5, pp.742–758, Oct. 2014.

[7] Q. H. Spencer, A. L. Swindlehurst, M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” IEEE Trans. Signal Process., vol.52, no.2, pp.461–471, 2004.

[8] G. Dimić and N. D. Sidiropoulos, "On Downlink Beamforming With Greedy User Selection: Performance Analysis and a Simple New Algorithm," IEEE Trans. Signal Process., vol.53, no.10, pp.3857–3868, 2005.

[9] M. Torabzadeh and W. Ajib, "Packet Scheduling and Fairness for Multiuser MIMO Systems," IEEE Trans. Veh. Technol., vol.59, no.3, pp.1330–1340, 2010.

[10] R. Kudo, Y. Takatori, K. Nishimori, A. Ohta, S. Kubota and M. Mizoguchi, "A New User Selection Measure in Block Diagonalization Algorithm for Multiuser MIMO Systems," IEICE Trans. Commun., vol.E92-B, no.10, pp.3206–3218, 2009.

[11] M.-O. Pun, V. Koivunen, and H. V. Poor, "Performance Analysis of Joint Opportunistic Scheduling and Receiver Design for MIMO-SDMA Downlink Systems," IEEE Trans. Commun., vol.59, no.1, pp.268–280, 2011.

[12] C.-M. Yen, C.-J. Chang, and L.-C. Wang, "A Utility-Based TMCR Scheduling Scheme for Downlink Multiuser MIMO-OFDMA Systems," IEEE Trans. Veh. Technol., vol.59, no.8, pp.4105–4115, 2010.

[13] T. Nabetani, N. Madhavan, H. Mori, and T. Aoki, "A Novel Low-Overhead Channel Sounding Protocol for Downlink Multi-User MIMO in IEEE 802.11ax WLAN," IEICE Trans. Commun., vol.E101-B, no.3, pp.924–932, 2018.

[14] H. Suganuma, Y. Shimbo, T. Maruko, H. Tomeba, T. Onodera, and F. Maehara, "An Efficient Method for Combining Multi-User MIMO Tomlinson-Harashima Precoding With User Selection Based on Spatial Orthogonality," IEEE Access, vol.9, pp.148449–148458, 2021.

[15] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, Jr., and B. L. Evans, "Low Complexity User Selection Algorithms for Multiuser MIMO Systems With Block Diagonalization," IEEE Trans. Signal Process., vol.54, no.9, pp.3658–3663, 2006.

[16] L. Bai, C. Chen, J. Choi, and C. Ling, "Greedy User Selection Using a Lattice Reduction Updating Method for Multiuser MIMO Systems," IEEE Trans. Veh. Technol., vol.60, no.1, pp.136–147, 2011.

[17] C.-K. Wen, S. Jin, and K.-K. Wong, "On the Sum-Rate of Multiuser MIMO Uplink Channels with Jointly-Correlated Rician Fading," IEEE Trans. Commun., vol.59, no.10, pp.2883–2895, 2011.

[18] T. Datta, N. A. Kumar, A. Chockalingam, and B. S. Rajan, "A Novel Monte-Carlo-Sampling-Based Receiver for Large-Scale Uplink Multiuser MIMO Systems," IEEE Trans. Veh. Technol., vol.62, no.7, pp.3019–3038, 2013.

[19] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The Multicell Multiuser MIMO Uplink with Very Large Antenna Arrays and a Finite-Dimensional Channel," IEEE Trans. Commun., vol.61, no.6, pp.2350–2361, 2013.

[20] N. Krishnan, R. D. Yates, and N. B. Mandayam, "Uplink Linear Receivers for Multi-Cell Multiuser MIMO With Pilot Contamination: Large System Analysis," IEEE Trans. Wireless Commun., vol.13, no.4, pp.4360–4373, 2013.

[21] G. O. Okeke, W. A. Krzymień, Y. Jing, and J. Melzer, "A Novel Low-Complexity Joint User-Relay Selection and Association for Multi-User Multi-Relay MIMO Uplink," IEEE Wireless Commun. Lett., vol.4, No.3, pp.309–312. 2015.

[22] W.-L. Shen, K. C.-J. Lin, W.-J. Cheng, L. Qiu, and M.-S. Chen, "Concurrent Packet Recovery for Distributed Uplink Multiuser MIMO Networks," IEEE Trans. Mobile Comut., vol.15, no.12, pp.3014–3027, 2016.

[23] M. Ge and D. M. Blough, "High Throughput and Fair Scheduling for Multi-AP Multiuser MIMO in Dense Wireless Networks," IEEE/ACM Trans. Network, vol.26, no.5, pp.2414–2427, 2018.

[24] J. Choi, J. Mo, and R. W. Heath Jr., "Near Maximum-Likelihood Detector and Channel Estimator for Uplink Multiuser Massive MIMO Systems With One-Bit ADCs," IEEE Trans. Commun., vol.64, no.5, pp.2005–2018, 2016.

[25] G. Lee and Y. Sung, "A New Approach to User Scheduling in Massive Multi-User MIMO Broadcast Channels," IEEE Trans. Commun., vol.66, no.4, pp.1481–1495, 2018.

[26] S. Sharifi and S. Shahbazpanahi, "A POMDP-Based Approach to Joint Antenna Selection and User Scheduling for Multi-User Massive MIMO Communication," IEEE Trans. Commun., vol.71, no.3, 1691–1706, 2023.

[27] J. Li, L. Xiao, X. Xu, and S. Zhou, "Robust and Low Complexity Hybrid Beamforming for Uplink Multiuser MmWave MIMO Systems," IEEE Commun. Lett., vol.20, no.6, pp.1140–1143, 2016.

[28] J. Seo and Y. Sung, "Beam Design and User Scheduling for Nonorthogonal Multiple Access With Multiple Antennas Based on Pareto Optimality," IEEE Trans. Signal Process., vol.66, no.11, pp.2876–2891, 2018.

[29] M. Xie, T.-M. Lok, and Q. Yang, "User Association and Scheduling Based on Auction in Multi-Cell MU-MIMO Systems," IEEE Trans. Wireless Commun., vol.17, no.6, pp.4150–4162, 2018.

[30] C. Feres and Z. Ding, "An Unsupervised Learning Paradigm for User Scheduling in Large Scale Multi-Antenna Systems," IEEE Trans. Wireless Commun., vol.22, no.5, pp.2932–2945, 2023.

[31] K. Yamazaki, T. Ohseki, Y. Amano, H. Shinbo, T. Murakami, and Y. Kishi, "Proposal for a User-Centric Ran Architecture Towards Beyond 5G," Proc. ITU Kaleidoscope: Connecting Physical and Virtual Worlds (ITU K), Geneva Switzerland, Dec.6-10, 2021.

[32] J. Zheng, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-Free Massive MIMO: A New Next-Generation Paradigm," IEEE Access, vol.7, pp.99878–99888, 2019.

[33] J. Denis and M. Assaad, "Improving Cell-Free Massive MIMO Networks Performance: A User Scheduling Approach," IEEE Trans. Wireless Commun., vol.20, no.11, pp.7360–7374, 2021.

[34] 3GPP TR 38.901, "Study on channel model for frequencies from 0.5 to 100GHz," V14.3.0, Dec., 2017.

[35] H. Sawada, K. Ishizu, and F. Kojima, "Study on propagation model for advanced utilization of millimeter- and terahertz waves," Journal of the National Institute of Information and Communication Technology, vol.64, no.2, pp.19–24, 2017.

[36] W. C. Jakes, "Microwave mobile communications," IEEE Press, 1994.

Channel Modeling Based on Ray Tracing for Wireless Transmission on Wood Materials

Nurul Fahmi Arief Hakim
Dept. of Electrical Eng. Education
Universitas Pendidikan Indonesia
Bandung, Indonesia
nurulfahmi@upi.edu

Alyani Ismail
Wireless and Photonic Networks R C
Universiti Putra Malaysia
Serdang, Malaysia
alyani@upm.edu.my

Nur Luqman Saleh
Wireless and Photonic Networks R C
Universiti Putra Malaysia
Serdang, Malaysia
nurluqman@upm.edu.my

Aduwati Sali
The Institute for Mathematical Research
Universiti Putra Malaysia
Serdang, Malaysia
aduwati@upm.edu.my

Nabil Taufiq Nur Izzat
Dept. of Electrical Eng. Education
Universitas Pendidikan Indonesia
Bandung, Indonesia
nabiltni@upi.edu

***Abstract*—This article discusses the application of ray tracing techniques to channel modeling for MIMO-OFDM wireless communication systems in complex indoor environments. Signal propagation characteristics are analyzed using ray tracing methods, specifically the Shooting Bouncing Ray (SBR) method, which considers the variations in position between the transmitter (Tx) and receiver (Rx) as well as the impact of reflector material on the channel. The article aims to determine the interactions between propagation distance and path loss, power delay profile, and BER of 64-QAM modulation using ray tracing modeling. The method involves simulation with LDPC coding at a frequency of 3.5 GHz and AWGN channels. The modeling results show a path loss value of 92.88 dB for R_x1 and 97.20 dB for R_x2. BER increases with transmission distance due to increased degradation and interference encountered by the transmitted signal. R_x1 has a lower BER value compared to R_x2. Long-range signals are more susceptible to attenuation and possible interference, including multipath fading, noise, and interference, all of which can contribute to increased BER.**

***Index Terms*—BER, LDPC, Ray tracing, SBR**

I. Introduction

In the field of wireless communications, the ability to understand and predict the properties of electromagnetic waves as they travel through different materials is important [1]. Considerable attention has been devoted to the potential of Multiple-Input Multiple-Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) to improve the capacity and spectral efficacy of wireless communications networks [2]. For the practical implementation of wireless communications, a thorough understanding of the characteristics of wireless networks is essential [3]. These characteristics are susceptible to various influences, such as the spatial arrangement of the transmitting and receiving apparatus and the characteristics of nearby construction materials [4]. This information is especially critical when designing efficient wireless systems for complex environments, including industrial settings, residential structures, and urban landscapes. Channel modeling serves as an essential tool for clarifying the characteristics of the propagation of wireless signals [5]. Among the various methodologies, ray tracing has surfaced as a highly effective technique for simulating and analyzing wave propagation [6].

Ray tracing is a computational methodology that incorporates the interactions of electromagnetic waves with adjacent surfaces and materials to simulate their trajectory [7]. Through the clarification of phenomena such as reflection, refraction, diffraction, and scattering, this approach provides a comprehensive and accurate illustration of the characteristics of waves as they propagate through various media. The incorporation of material physical properties into the ray tracing algorithm enhances its effectiveness in channel modeling, enabling precise simulation of signals in a wide range of scenarios. The conveyance of wireless signals through various materials, such as concrete and wood, poses unique challenges. The material employed can impact various signal characteristics, such as polarization, phase shift, and attenuation [8], [9]. For example, the signal attenuation caused by concrete walls within structures is significant, while multipath propagation can occur due to reflections induced by glass.

Precise channel models are necessary for the effective operation of emerging technologies such as 5G and the Internet of Things (IoT) in a variety of environments, including those with wooden structures [10]. An understanding of the impact of wood on signal propagation drives ray-tracing-based channel modeling for wireless transmission in wood materials. The behavior of wireless signals is influenced by wood, a common material in buildings and furniture, through attenuation, reflection, and scattering. Unique physical properties of various wood species can substantially impact the quality and intensity of signals [11]. Another primary motivation for this form of channel modeling is to enhance indoor wireless communications. Reliable and efficient communication within buildings is of utmost importance as wireless technology becomes more ingrained in daily life. The interaction between wireless signals and wooden structures, including walls, floors, and furniture,

can be closely simulated through ray tracing.

The purpose of this article is to investigate the use of ray tracing in wireless transmission channel modeling across wood materials. This course will provide an overview of the fundamental principles that support ray tracing, analyze the impact of material properties on signal transmission, and illustrate the practical applications of this modeling methodology through the use of case studies. By conducting an exhaustive examination, the objective is to underscore the potential of ray tracing in furthering the development of wireless communications technologies and resolving the difficulties presented by various propagation environments.

II. Channel Modeling in Wireless Communications

Channel modeling in wireless communications refers to the process of characterizing the behavior of wireless channels to accurately predict how electromagnetic waves propagate through different environments. This involves creating mathematical functions that describe signal attenuation, delays, phase shifts, and other factors as the signal moves from the transmitter to the receiver. The main goal of channel modeling is to simulate real-world conditions to improve the design, optimization, and performance evaluation of wireless communication systems.

Typical components of channel models consist of the following: path loss, which quantifies the degradation in signal strength as distance increases; shadowing, which induces substantial fading attributable to obstacles; and multipath fading, which arises when a signal traverses multiple routes to the receiver as a consequence of reflection, diffraction, and scattering. The aforementioned factors are incorporated into models that depict environments with substantial multipath effects, such as the Rayleigh and Rician fading models, and the Free Space Path Loss (FSPL) model for optimum open environments. In order to represent channels through ray tracing, the *propagationModel* function is implemented. The function parameters consist of the ray tracing channel modeling method implemented, the specified form of channel modeling, the coordinate system utilized, the angular separation, the building surface material, and the maximum number of reflected signals. The channel modeling process is then initiated by invoking the raytrace function with the *propagationModel* and (T_x) parameters.

The parameters of the experiment are detailed in Table 1. The encoder and decoder channels employ Low-density parity-check (LDPC) in conjunction with 64-QAM modulation at a frequency of 3.5 GHz. An LDPC (Low Density Parity Check) coding and decoding method was employed for error coding or channel encoder in this study. The low-density parity matrix structure of LDPC is among its most distinguishing characteristics. In other words, an exceedingly minuscule fraction of the matrix elements are nonzero. This results in a matrix with numerous empty elements and, in most cases, extremely brief rows. By virtue of this architecture, LDPC is capable of delivering exceptional efficacy in error detection and correction. Implementing ray tracing image Bouncing Ray (SBR) is utilized in path loss and multipath modeling with AWGN propagation channels. A valid path will be corrected by the SBR method to ensure that it has the proper geometric accuracy. Signal path determination is performed by the SBR method via double precision floating point computing. The Gaussian distribution, which is prevalent in nature and numerous physical systems, is utilized to depict noise in AWGN channel modeling due to its distinctive characteristics. In AWGN channel modeling, interference or noise is typically represented by a Gaussian distribution due to the fact that, in many real-world scenarios, it tends to occur arbitrarily and independently in a complex and sinusoidal form. Thermal interference, electromagnetic interference, and flaws in electronic components of communication devices are all potential sources of this cacophony. The central limit theorem posits that when numerous minor factors contribute to the noise, the sum of numerous independent random variables will approach the value of a normal distribution. Using the outcomes of the ray-tracing modeling, a deterministic multipath channel model is subsequently constructed [12]. In addition, the receiver speed is specified in this section so that the minimum device movement within a confined space can be accurately reflected. The velocity of the receiver is denoted in Cartesian coordinates, with the X and Y axes both set to 0.1.

TABLE I
RAY TRACING MODELLING PARAMETER

Parameters	Specifications
Channel encoder	LDPC encoding
Channel decoder	LDPC decoding
Modulation	64-QAM
Frequency	3.5 GHz
Path loss and multipath	Ray tracing SBR
Channel modelling	AWGN
Maximum reflexion	1
Maximum difraction	1
Tx and Rx antenna	ULA 4 x 1 spacing 2 λ

The result or dependent variable derived from this study is represented by parameters that serve as indicators of the efficacy of the channel model developed. The research utilizes the Power Delay Profile (PDP) and graphs depicting the correlation between propagation distance, number of interactions, and path loss as the criteria for comparison. In order to derive conclusions about the impact of alterations in T_x and R_x positions, as well as material effects, on route loss and the quality of a telecommunications channel model, these parameters are then examined and contrasted.

Free Space Path Loss (FSPL) can be calculated using equation (1), which is the drop in signal strength that occurs when a radio signal or electromagnetic wave travels through an environment that is devoid of any obstructions, such as buildings or other types of obstructions. The signal frequency is denoted by f, the distance between the transmitter and receiver is denoted by d, the constant for the speed of light is c (3×10^8), the gain of the transmitting antenna is denoted

by G_T, the gain of the receiving antenna is denoted by G_R, and the reflection coefficient is denoted by R. In order to determine the parameter R, one can utilize the relative permittivity value (ε_r). Equation (2) represents the widely employed equation representing R. This equation serves as the foundational paradigm for planning wireless networks and can be identified using software such as Matlab [13].

$$FSPL(dB) = 20\log_{10}\left(\frac{4\pi fd}{c}\right) - G_t - G_R + R \quad (1)$$

$$R = \left|\frac{\sqrt{\varepsilon_r} - 1}{\sqrt{\varepsilon_r} + 1}\right|^2 \quad (2)$$

III. Results and Analysis

The experimental configuration is illustrated in Fig. 1. The receiver representation is in blue, while the transmitter is depicted in red. As seen in Fig. 1, R_x1 was positioned 36 meters from T_x in the initial experiment, while R_x2 was positioned 48 meters from T_x. T_x is 2.6 meters above floor level. The results of this propagation path modeling consist of a PDP graph, which serves as a site viewer of the modeled propagation path, and a graph of the relationship between path loss, number of interactions, and propagation distance.

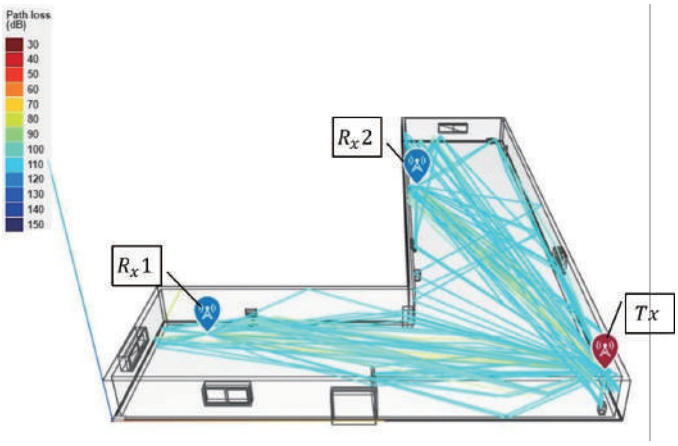


Fig. 1. Modeling scheme using 2 receivers

The ray tracing simulation yields additional information regarding the path loss for each signal, the number of signal interactions with the environment, and the distance the signal travels. Figs. 2 and 3 depict the information in the form of two-dimensional scatter diagrams for R_x1 and R_x2, respectively. The number of interactions that the signal has undergone with its environment is denoted by the color of each point. The path loss value for each propagation path is represented on the X-axis, whereas the signal's propagation distance to reach Rx is indicated on the Y-axis. The relationship between the path loss value and the signal propagation distance is linear, as illustrated by the graph. Likewise, the number of interactions appears to be proportional to the path loss value.

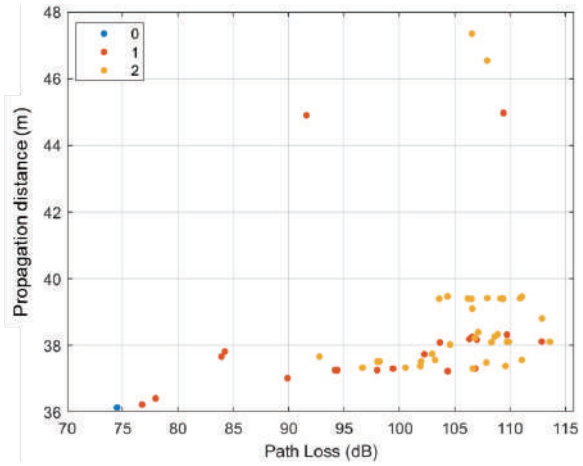


Fig. 2. The interaction between propagation distance, path loss, and number of interactions for R_x1

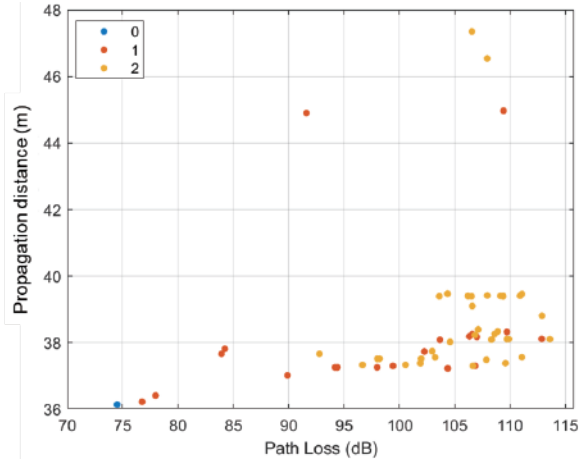


Fig. 3. The interaction between propagation distance, path loss, and number of interactions for R_x2

The PDP graph illustrates the varying arrival times of each multipath path at the receiver. PDP measures the time-dependent variation in the strength of a signal that has been received by a device. PDP is a mathematical function that characterizes the relationship between signal intensity and time. PDP determines which peak component or peak power traverses the channel's strongest path. Delay spread quantifies the degree of dispersion exhibited by the power profile within a given channel. As illustrated in Figs. 4 and 5, the PDP graphs derived from R_x1 and R_x2 are presented. The provision of this data is critical for the development of wireless systems, as the delay spread signifies the degree to which the passage of time impacts the signal. Increasing the delay spread poses a greater challenge in terms of accurately recovering the signal at the receiver. The Y axis represents the signal power received by Rx from each existing multipath path, while the X axis represents the delay dispersion of each multipath path. Experimental findings indicate that wood material has the potential to impact the channel quality. The calculated average

path loss value for R_x1 92.88 dB and R_x2 is 97.20 dB.

The variation in the reflection coefficient value is the cause of this change in the path loss value. Additionally, the reflection coefficient of the reflecting surface that engages in signal-to-reflective interaction impacts the path loss value. The average path loss value is determined through the application of the average formula, which involves adding the path loss values for each path and subsequently dividing the sum by the number of extant paths. Materials also have an impact on the graph of the power delay profile. The signal strength received by Rx1 is 8.274×10^{-6} dB and for R_x2 is 1.296×10^{-5} dB. Experimental results suggest that wood material may have an effect on the condition of the channel. The average path loss values for R_x1 and R_x2 are 92.88 dB and 97.20 dB. This variation in the path loss value can be attributed to the variation in the reflection coefficient value. Furthermore, the path loss value is influenced by the reflection coefficient of the reflecting surface that is engaged in signal-to-reflective interaction. The average path loss value is calculated using the average formula, which entails the sum of the path loss values for all existing paths and a significant division of the result by the total number of paths. Additionally, materials influence the power delay profile graph. The amplitude of the signal received by R_x1 is 8.274×10^{-6} dB. On the contrary, the signal intensity received for R_x2 is 1.296×10^{-5} dB.

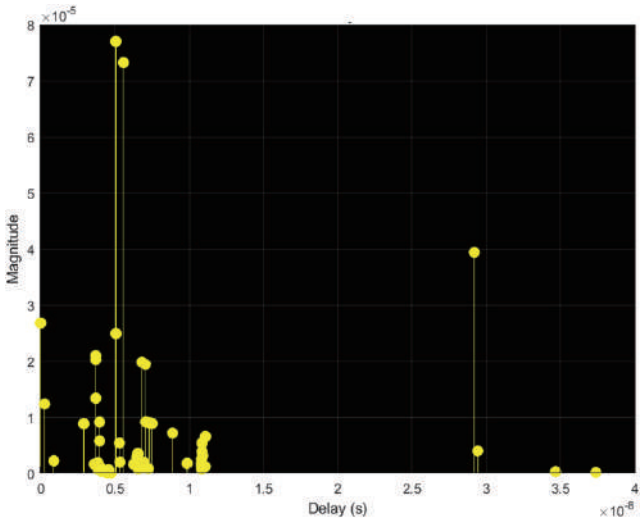


Fig. 4. Initial power delay profile for R_x1

The data delivery for R_x1 is illustrated in Figs. 6 and 7 through a BER graph. R_x1 has a reduced BER value in comparison to R_x2. BER increases as the transmission distance increases due to the heightened degradation and interference encountered by the transmitted signal. Under the assumption that all other factors remain constant (including transmission environment, transmit power, modulation technology, and interference), the BER of T_x and R_x2 is expected to be greater when they are separated by 40 meters as opposed to T_x and R_x1 which are separated by 30 meters. Longer-distance signals are susceptible to greater attenuation (signal

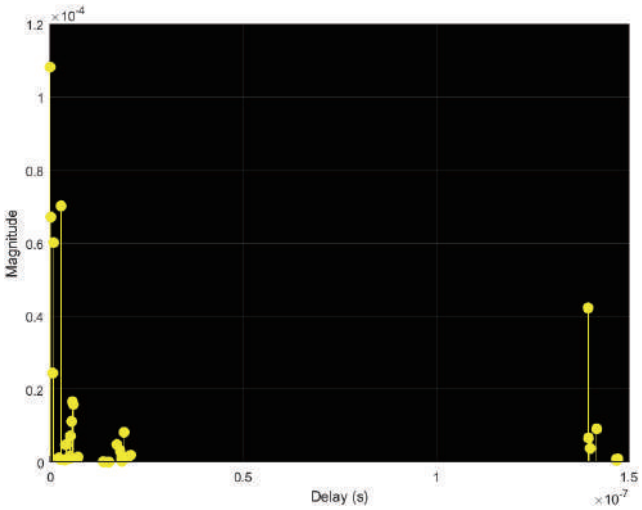


Fig. 5. Initial power delay profile for R_x2

deterioration) and possible interference, including multipath fading, noise, and interference, all of which can contribute to an increase in BER. A signal that is weakened at the receiver is additionally prone to bit errors that may occur during the processes of demodulation and decoding. Wood is also capable of attenuating wireless signals considerably, particularly at higher frequencies utilized in contemporary 5G networks. The extent of attenuation is contingent upon variables including the moisture content, density, and thickness of the wood. Wet wood, for instance, has a greater signal loss due to its higher dielectric constant in comparison to dried wood. It is imperative to comprehend these interactions in order to develop efficient indoor wireless networks in environments with a high prevalence of wooden structures.

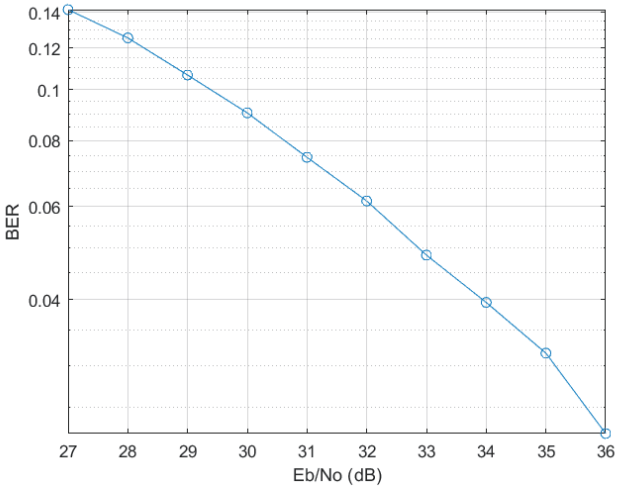


Fig. 6. Value of BER compared to Eb/No for R_x1

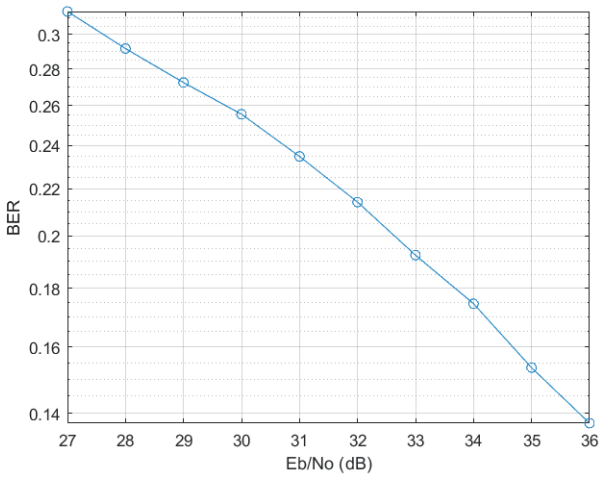


Fig. 7. Value of BER compared to Eb/No for R_x2

IV. Conclusion

Wireless communications rely heavily on ray tracing techniques, particularly in environments containing wooden structures. By providing precise forecasts of signal behavior, it aids in the development and enhancement of wireless networks in order to surmount obstacles presented by signal attenuation, absorption, and dispersal. The continued development of wireless technology underscores the criticality of ray tracing in maintaining dependable and effective communications within intricate settings. An essential discovery derived from investigations utilizing ray tracing in wood environments pertains to the significance of attenuation and absorption. The simulation yielded path loss values of 92.88 dB for R_x1 and 97.20 dB for R_x2. In contrast, the value of BER on R_x1 is less than that of R_x2. Long-distance signals are more vulnerable to attenuation and potential interference, such as multipath fading, noise, and interference, all of which can contribute to an increase in BER, as demonstrated by the results obtained. The optimization of transmitter and receiver placement in wood-based environments is facilitated by ray tracing. Through precise modeling of reflection and scattering phenomena, it is possible to ascertain the optimal location that minimizes signal attenuation and guarantees robust connectivity. Maintaining a dependable wireless network is particularly critical in environments such as smart residences or offices, as it dictates the seamless operation of numerous devices and applications. Additional inquiry into these results may be conducted on a variety of different substances, including Styrofoam, metal, or glass.

References

[1] L. Wu et al., "Artificial Neural Network Based Path Loss Prediction for Wireless Communication Network," in *IEEE Access*, vol. 8, pp. 199523-199538, 2020, doi: 10.1109/ACCESS.2020.3035209.

[2] T. Kebede, Y. Wondie, J. Steinbrunn, H. B. Kassa and K. T. Kornegay, "Multi-Carrier Waveforms and Multiple Access Strategies in Wireless Networks: Performance, Applications, and Challenges," in *IEEE Access*, vol. 10, pp. 21120-21140, 2022, doi: 10.1109/ACCESS.2022.3151360.

[3] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Cambridge, United Kingdom: Cambridge University Press, 2024.

[4] S. Nandi, N. N. Pathak, and A. Nandi, "Efficacy of channel estimation and efficient use of spectrum using optimised cyclic prefix CP in MIMO OFDM," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 2, pp. 3032–3038, Dec. 2019. doi:10.35940/ijeat.b4093.129219

[5] C. Han et al., "Terahertz Wireless Channels: A Holistic Survey on Measurement, Modeling, and Analysis," in *IEEE Communications Surveys Tutorials*, vol. 24, no. 3, pp. 1670-1707, thirdquarter 2022, doi: 10.1109/COMST.2022.3182539

[6] S. Kasdorf, B. Troksa, C. Key, J. Harmon and B. M. Notaroš, "Advancing Accuracy of Shooting and Bouncing Rays Method for Ray-Tracing Propagation Modeling Based on Novel Approaches to Ray Cone Angle Calculation," in *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 8, pp. 4808-4815, Aug. 2021, doi: 10.1109/TAP.2021.3060051.

[7] M. M. Taygur and T. F. Eibert, "A Ray-Tracing Algorithm Based on the Computation of (Exact) Ray Paths With Bidirectional Ray-Tracing," in *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 8, pp. 6277-6286, Aug. 2020, doi: 10.1109/TAP.2020.2983775

[8] M. Qin, L. Zhang, and H. Wu, "Dielectric loss mechanism in electromagnetic wave absorbing materials," *Advanced Science*, vol. 9, no. 10, Feb. 2022. doi:10.1002/advs.202105553

[9] M. B. Mollah, S. Zeadally, and Md. A. Azad, "Emerging wireless technologies for internet of things applications: Opportunities and challenges," *Encyclopedia of Wireless Networks*, pp. 390–400, 2020. doi:10.1007/978-3-319-78262-1_328

[10] A. N. Uwaechia and N. M. Mahyuddin, "A Comprehensive Survey on Millimeter Wave Communications for Fifth-Generation Wireless Networks: Feasibility and Challenges," in *IEEE Access*, vol. 8, pp. 62367-62414, 2020, doi: 10.1109/ACCESS.2020.2984204.

[11] S. Bakirtzis, T. Hashimoto and C. D. Sarris, "FDTD-Based Diffuse Scattering and Transmission Models for Ray Tracing of Millimeter-Wave Communication Systems," in IEEE Transactions on Antennas and Propagation, vol. 69, no. 6, pp. 3389-3398, June 2021, doi: 10.1109/TAP.2020.3044373

[12] A. W. Mbugua, Y. Chen, L. Raschkowski, L. Thiele, S. Jaeckel and W. Fan, "Review on Ray Tracing Channel Simulation Accuracy in Sub-6 GHz Outdoor Deployment Scenarios," in *IEEE Open Journal of Antennas and Propagation*, vol. 2, pp. 22-37, 2021, doi:10.1109/OJAP.2020.3041953.

[13] S. K. Hinga and A. A. Atayero, "Deterministic 5G mmWave Large-Scale 3D Path Loss Model for Lagos Island, Nigeria," in *IEEE Access*, vol. 9, pp. 134270-134288, 2021, doi: 10.1109/ACCESS.2021.3114771.

A DDPG-based Framework to Optimize Hierarchical Rate Splitting Multiple Access

Anh-Tien Tran*, Thanh-Phung Truong*, Manh Cuong Ho *, Tung-Son Do*, Quang Tuan Do*, Dongwook Won*, Nhu-Ngoc Dao†, and Sungrae Cho*
*School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea.
†Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea.
Email: {attran, tptruong, hmcuong, tsdo, dqtuan, dwwon}@uclab.re.kr, nndao@sejong.ac.kr, srcho@cau.ac.kr.

***Abstract*—Rate-splitting multiple access (RSMA) is recognized as a promising solution for strengthening the performance of densely populated downlink networks that seek to meet the demanding needs of networks beyond 5G.These networks may need the capacity to accommodate a substantial number of users concurrently using a single bandwidth spectrum.RSMA employs rate splitting during transmission and successive interference cancellation (SIC) during reception as its core principles. As the user count rises, the number of broadcast streams and possible decoding instructions grows exponentially. This study introduces a rate splitting method rooted in learning principles. It classifies users into hierarchical clusters and attaches a shared stream to each cluster. This hierarchical grouping ensures user clusters are of equal size, thus negating the necessity for user prioritization. The objective of this work is to enhance the efficiency of beamforming vectors, power allocations for transmission streams, and rate allocation of common messages by using the deep deterministic policy gradient (DDPG) approach.**

***Index Terms*—weighted sum-rate, rate-splitting multiple access, downlink, 2-layer RSMA, hierarchical RSMA.**

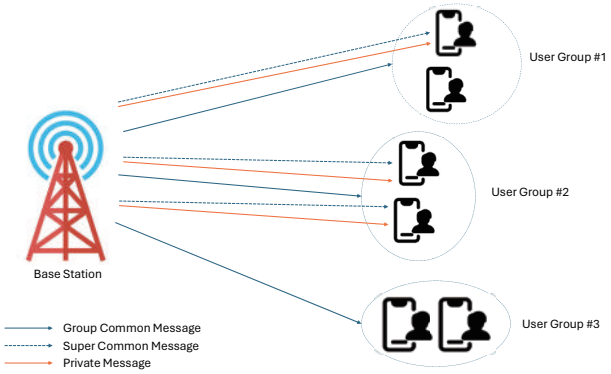


Fig. 1. System Model

I. Introduction

RSMA [1] serves as a means to close the technological disparity between NOMA [2], [3] (which fully decodes interference) and SDMA (where interference is treated as noise). This ability arises from the inherent characteristic of rate splitting. At the transmitter side, every communication is divided into two parts: a shared portion and an individual portion. Using a public codebook, known as generalized RS, all users can encode their shared components into a single stream that is decodable by everyone. Alternatively, with RS-CMD [4], users can encode their shared components into separate streams, accessible only to specific subsets of users. Users' personal data is encrypted and stored in private streams, accessible solely to intended recipients. Power allocation techniques for transmitting both shared and private streams facilitate rate optimization. Moreover, solving the optimization function is made more challenging by the non-linear nature of the interconnected rate expressions for individual and shared streams across each user.

RSMA improves the attainable rate compared to NOMA, while also reducing the computational complexity for both transmitters and receivers, particularly in terms of the number of SIC layers needed. RSMA allows for the use of portable devices with minimal computing capability. The versatility of RSMA allows it to operate well in the presence of various amounts of interference. RSMA quickly transitions to either NOMA or SDMA by fine-tuning the power levels of both the shared and individual data streams when there is either a little or significant level of interference.

The hierarchical RSMA strategy was first introduced in [5] in the massive MIMO context. In our paper, the idea is to dynamically group users based on their channel statistics to exploit the benefits of spatially correlated channels. The adopted hierarchical RSMA is also known as 2-layer RSMA scheme [6]. Besides the common stream sending to all users and private streams intended to each user in traditional 1-layer RSMA [7], the adopted hierarchical RSMA also transmit group common streams to different group of users. It increases the flexibility of proposed scheme by adding more optimization variables. However, the complexity of the problem is also increased. We attempted to solve the problem using a deep reinforcement learning framework named DDPG [8].

II. System Model

As illustrated in Fig.1, the considered scenario has one multiple-antenna base station with M antenna serving K single-antenna users assuming that perfect CSI is available at the BS. The channel matrix is given by $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in C^{M\times K}$, where $\mathbf{h}_k \in C^{M\times 1}$ is the channel vector between the BS and k-th user. These K users are divided into G groups

$\mathcal{G} = \{1, \ldots, G\}$ with same number of users in each group following an user grouping order $\mathbf{a} = \{a_1^1, \ldots, a_k^g, \ldots, a_K^G\}$ where $a_k^g \in \{0, 1\}$ is the binary group assignment variable if k-th user belong to g-th group. Each user is required to decode 3 streams $s_{\mathcal{K}}, s_{\mathcal{K}_g}, s_k$ sequentially where $s_{\mathcal{K}}$ is the super common stream to be interpreted by all users, $s_{\mathcal{K}_g}$ is the group common stream to be solved by all users within the g-th group, and s_k is the private stream to be decoded independently by k-th user only.

The overall data streams to be transmitted from the BS based on hierarchical RSMA is $\mathbf{s} = \left[s_{\mathcal{K}}, s_{\mathcal{K}_1}, \ldots, s_{\mathcal{K}_G}, s_1, \ldots, s_K\right]$. The data streams are linearly precoded via precoder $\mathbf{P} = \left[\mathbf{p}_{\mathcal{K}}, \mathbf{p}_{\mathcal{K}_1}, \ldots, \mathbf{p}_{\mathcal{K}_G}, \mathbf{p}_1, \ldots, \mathbf{p}_K\right]$. By utilizing multiple antennas at the BS, these streams are transmitted together to all users. In this work, we divide the power action into 2 independent variables, i.e. $\mathbf{p}_k = \sqrt{\rho_k}\mathbf{w}_k$,

$$||\mathbf{w}_k||^2 = 1, k \in \{0, 1, 2, \ldots, K\} \quad \text{AND} \quad \sum_{k \in \mathcal{K}} \rho_k \leq 1 \tag{1}$$

Let P_0 stand for the transmission power at the BS and $\rho_{\mathcal{K}}, \rho_{\mathcal{K}_g}, \rho_k$ stand for the power ratios allocated to streams $\mathbf{s}_{\mathcal{K}}, \mathbf{s}_{\mathcal{K}_g}, \mathbf{s}_1$, respectively. The transmitting signal at the BS has the following components:

$$\mathbf{x} = \sqrt{P_0}\left(\sqrt{\rho_{\mathcal{K}}}\mathbf{s}_{\mathcal{K}} + \sum_{g \in \mathcal{G}} \sqrt{\rho_{\mathcal{K}_g}}\mathbf{s}_{\mathcal{K}_g} + \sum_{k \in \mathcal{K}} \sqrt{\rho_k}\mathbf{s}_k\right) \tag{2}$$

The collected signal at k-th user is:

$$y_k = \mathbf{h}_k^H \mathbf{x} + \omega_k \tag{3}$$

where ω_k is the Additive White Gaussian Noise (AWGN) noise following a multi-variate circularly-symmetric complex Gaussian distribution that $\omega_k \sim \mathcal{CN}\left(0, \sigma_k^2\right)$.

At every user, decoding occurs sequentially for the super common stream, group common stream, and private streams. The SINR for decoding the super common stream at the k-th user is:

$$\gamma_k^{\mathcal{K}} = \frac{|\mathbf{h}_k|^2 P_0 \rho_{\mathcal{K}}}{|\mathbf{h}_k|^2 \sum_{g \in \mathcal{G}} P_0 \rho_{\mathcal{K}_g} + |\mathbf{h}_k|^2 \sum_{k \in \mathcal{K}} P_0 \rho_k + \omega_k} \tag{4}$$

because user k treats interference from other streams as noise. Following the decoding of the super common stream, it is eliminated from the received signal through SIC. User k deciphers its group common stream $s_{\mathcal{K}_g}$, where $a_g^k = 1$, considering interference from other group common streams and other private streams as noise. The SINR for deciphering the group common stream is expressed as:

$$\gamma_k^{\mathcal{K}_g} = \frac{|\mathbf{h}_k|^2 P_0 \rho_{\mathcal{K}_g}}{|\mathbf{h}_k|^2 \sum_{g' \in \mathcal{G}, g' \neq g} P_0 \rho_{\mathcal{K}_g'} + |\mathbf{h}_k|^2 \sum_{k \in \mathcal{K}} P_0 \rho_k + \omega_k} \tag{5}$$

The next step is to remove the group common streams out of received signal using SIC technique and then decode its own private stream. The SINR of deciphering the private stream is given by:

$$\gamma_k^k = \frac{|\mathbf{h}_k|^2 P_0 \rho_k}{\sum_{k' \in \mathcal{K}, k' \neq k} |\mathbf{h}_k|^2 P_0 \rho_{k'} + \omega_k} \tag{6}$$

The obtainable rate of k-th user is calculated by:

$$R_k = R_k^k + C_k^g + C_k^{\mathcal{K}} \tag{7}$$

where $R_k^k = B \log_2\left(1 + \gamma_k^k\right)$ is the achievable rate of decoding the private stream where B is the downlink bandwidth and C_k^g and $C_k^{\mathcal{K}}$ are the quantity of rate R^g and $R^{\mathcal{K}}$ distributed to k-th user.

$$R^g = B \min_{k \in \mathcal{K}, a_k^g = 1} \left\{\log_2\left(1 + \gamma_k^{\mathcal{K}_g}\right)\right\}, \tag{8}$$

$$R^{\mathcal{K}} = B \min_{k \in \mathcal{K}} \left\{\log_2\left(1 + \gamma_k^{\mathcal{K}}\right)\right\}, \tag{9}$$

$$\sum_{k \in \mathcal{K}, a_k^g = 1} C_k^g = R^g, \tag{10}$$

$$\sum_{k \in \|} C_k^{\mathcal{K}} = R^{\mathcal{K}} \tag{11}$$

III. Problem Formulation

In this study, our aim is to collectively enhance the assignment of user groups, beamforming vector, and rate allocation to maximize the total weighted achievable data rate for all users. This issue is formally expressed through the following mathematical formulation:

$$\max_{\mathbf{P}, \mathbf{a}, \mathbf{w}, \mathbf{C}} \sum_{k \in \mathcal{K}} \mu_k R_k, \tag{12}$$

$$\text{s.t. } (1), (10), (11), \tag{13}$$

$$\sum_{k \in \mathcal{K}} a_k^g = 1 \tag{14}$$

where $\mathbf{C} = \left\{\mathbf{C}^{\mathcal{K}}, \{\mathbf{C}^g\}_{g \in \mathcal{G}}\right\}$ is the common rate matrix formed by super common rate vector $\mathbf{C}^{\mathcal{K}} = \left\{C_k^{\mathcal{K}}\right\}_{k \in \mathcal{K}}$ and group common rate vectors $\mathbf{C}^g = \left\{C_k^g\right\}_{k \in \mathcal{K}, a_k^g = 1}$. The constraint (14) is added to ensure that each user is categorized into at most 1 group. Note that the values of μ_k also represents the priority of users. The given problem may be identified as a mixed-integer nonlinear programming problem, which presents difficulty in solving due to the presence of both binary and numerical variables. Unlike traditional methods of multiple access, the division of messages into separate portions increases the difficulty of the task. This is because power allocation must be done in a manner that allows all messages to be decoded. In order to address this issue, we suggest using a sophisticated deep reinforcement learning system called deep deterministic policy gradient (DDPG).

IV. DDPG-based Framework

To model the dynamic of the channel state, we first formulate the problem by using the MDP framework [9]. An MDP is denoted by $\mathcal{M} = \{\mathcal{X}, \mathcal{A}, P, R, \tau\}$, with $\mathcal{X}$ representing the space of states, $\mathcal{A}$ representing the feasible actions, P representing the transition probability matrix, R representing the reward function evaluating actions based on the current state, and $\tau \in (0, 1)$ serving as the discount factor .

A. Components

- **State:** The state space is defined as $\mathcal{X} = \{\mathbf{h}_k\}_{k\mathcal{K}}$ as the collected current channel state between the BS and associated users.
- **Action:** The action space is described as $\mathcal{A} = \{\mathbf{P}, \bar{\mathbf{a}}, \mathbf{w}, \mathbf{C}\}$. Note that due to the binary nature of assignment values, we relax the assignment variables into its continuous version to match the output of neural networks, i.e. $0 \leq \bar{a}_k^g \leq 1$. For each user, to select which group it should associate to, we use an heuristic rule,

$$a_k^g = \begin{cases} 1 \text{ if } g = \max_{g\in\mathcal{G}} \{\bar{a}_k^g\} \\ 0 \text{ otherwise} \end{cases} \tag{15}$$

  to convert the relaxed value into the binary value.
- **Reward:** The reward function is designed to maximize the weighted sumrate of all users, i.e. $R = \sum_{k\in\mathcal{K}} \mu_k R_k$.

B. DDPG Algorithm

The DDPG algorithm is derived from the actor-critic approach, which consists of an actor network (AN) and a critic network (CN) [10]. The AN generates the action according to the observed state. By letting $\upsilon^{\theta_\upsilon}(s)$ denote the AN with the parameter θ_υ, the action at each time slot is expressed as

$$a[t] = \upsilon^{\theta_\upsilon}(s[t]) + \mathcal{N}[t], \tag{16}$$

where $\mathcal{N}[t]$ is the additional noise for exploration, following the Ornstein–Uhlenbeck process. The CN is applied to evaluate the action-state value for training the AN. The AN's parameter is revised by calculating the policy gradient, expressed as

$$\nabla_{\theta_\upsilon} J = \frac{1}{B} \sum_b \nabla_a Q^{\theta_Q}(s_b, a)|_{a=\upsilon^{\theta_\upsilon}(s_b)} \nabla_{\theta_\upsilon} \upsilon^{\theta_\upsilon}(s_b), \tag{17}$$

where B is the training batch size, s_b is the state of b-th sample, and $Q^{\theta_Q}(s, a)$ denotes the CN with the parameter θ_Q. The parameter of the CN is updated by calculating the gradient descent of the loss function, expressed as

$$L = \frac{1}{B} \sum_b (y_b - Q^{\theta_Q}(s_b, a_b)), \tag{18}$$

where $y_b \triangleq r_a + \gamma Q'^{\theta_{Q'}}(s'_b, \upsilon'^{\theta_{\upsilon'}}(s'_b))^2$ denotes the target value, a_b, s'_b and r_b are the action, next state and reward at b-th sample, γ is the discount factor, $Q'^{\theta_{Q'}}(s, a)$ and $\upsilon'^{\theta_{\upsilon'}}$ are the target networks for improving the training performance. The status of the target networks are revised using ς-soft update, expressed as

$$\begin{aligned} \theta_{Q'} &\leftarrow \varsigma\theta_Q + (1-\varsigma)\theta_{Q'}, \\ \theta_{\upsilon'} &\leftarrow \varsigma\theta_\upsilon + (1-\varsigma)\theta_{\upsilon'}. \end{aligned} \tag{19}$$

V. Conclusion

In this study, we introduce a resource allocation method designed for hierarchical RSMA MISO downlink networks. Our objective is to enhance the weighted sum-rate of all users by concurrently optimizing user group assignments, beamforming vectors, rate allocation matrices, and power allocations for each stream. The formulated problem is hard to be solved straightforwardly and then converted into a MDP problem. The issue is solved using the DDPG method. The next study will provide simulation results to showcase the dependability and efficacy of the proposed framework in comparison to previous benchmark methods.

Acknowledgment

This work was supported in part by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2022-00156353) supervised by the IITP (Institute for Information Communications) and in part by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (RS-2023-00209125).

References

[1] Y. Mao, B. Clerckx, and V. O. Li, "Rate-splitting for multi-user multi-antenna wireless information and power transfer," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2019, pp. 1–5.

[2] T.-H. Nguyen, T. P. Truong, A.-T. Tran, N.-N. Dao, L. Park, and S. Cho, "Intelligent heterogeneous aerial edge computing for advanced 5g access," *IEEE Transactions on Network Science and Engineering*, 2024.

[3] T. P. Truong, T.-V. Nguyen, W. Noh, S. Cho *et al.*, "Partial computation offloading in noma-assisted mobile-edge computing systems using deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 196–13 208, 2021.

[4] A. A. Ahmad, H. Dahrouj, A. Chaaban, A. Sezgin, and M.-S. Alouini, "Interference mitigation via rate-splitting and common message decoding in cloud radio access networks," *IEEE Access*, vol. 7, pp. 80 350–80 365, 2019.

[5] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive mimo with imperfect csit," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4611–4624, 2016.

[6] Y. Mao, B. Clerckx, and V. O. Li, "Rate-splitting multiple access for downlink communication systems: bridging, generalizing, and outperforming sdma and noma," *EURASIP journal on wireless communications and networking*, vol. 2018, no. 1, pp. 1–54, 2018.

[7] J. Zhang, B. Clerckx, J. Ge, and Y. Mao, "Cooperative rate splitting for miso broadcast channel with user relaying, and performance benefits over cooperative noma," *IEEE Signal Processing Letters*, vol. 26, no. 11, pp. 1678–1682, 2019.

[8] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[9] M. C. Ho, A. T. Tran, D. Lee, J. Paek, W. Noh, and S. Cho, "A ddpg-based energy efficient federated learning algorithm with swipt and mc-noma," *ICT Express*, 2023.

[10] T. P. Truong, V. D. Tuong, N.-N. Dao, and S. Cho, "Flyreflect: Joint flying irs trajectory and phase shift design using deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 4605–4620, 2023.

Design of an Antenna for Electronic Article Surceillance System Capable of Receving all Polarizations

Jinkyu Bang
Maritime ICT & Mobility Research Department
Korea Institute of Ocean Science & Technology
Busan, South Korea
jinkyu.bang@kiost.ac.kr

Jae Hee Kim
Electrical Electronics and Commnunication Engineering
KOREATECH
Cheonan, South Korea
jaehee@koreatech.ac.kr

***Abstract*—This paper introduces the design of a new transmission (Tx) antenna for an Electronic Article Surveillance (EAS) system aimed at maximizing tag detection from various directions. Traditionally, Tx antennas used in EAS systems are designed to generate magnetic fields in specific directions. However, the orientation of tags attached to items is difficult to predict accurately, leading to areas where tag detection does not occur. To address this issue, this study proposes an antenna that uses three Tx coils to apply signals of different phases, thereby creating a magnetic field distribution that varies over time.**

Keywords—Electronic Article Surveillance (EAS), antennas, magnetic fields

I. Introduction

Electronic Article Surveillance (EAS) is a technology widely used in commercial facilities to prevent the theft of goods. EAS systems can be broadly classified into three types based on their frequency and operational mechanisms: Acousto-Magnetic (AM), Frequency Modulation (FM), and Radio Frequency (RF) systems[1-4]. Among these, the RF system primarily uses the frequency band around 8.2 MHz and is designed to trigger an alarm when it receives a signal corresponding to the resonance frequency via a resonant circuit installed inside the tag. The tags attached to the merchandise can be oriented in various directions, making it difficult to predict the orientation of tags as they pass through the monitoring system in advance.

Considering the varied orientations, to maximize the detection efficiency of the EAS system, the transmission antenna must be capable of generating effective magnetic fields corresponding to all directions of polarization. Most transmission antennas currently used in EAS systems are mainly composed of loop coils, designed to maintain a specific form of the magnetic field. However, this design has the limitation of creating 'undetected areas' where tags cannot be detected. To solve this issue, various structural modifications have been proposed, including designing longer transmission coils considering the direction of human movement and adjusting the spacing between coils to enable tag detection from various directions. However, these approaches have not completely resolved the issue of undetected areas, and increasing the length of the transmission coils can lead to constraints on installation space.

This paper proposes a new approach to address these problems. By developing a new coil structure for the EAS transmission antenna that can effectively form magnetic fields in all directions, even within a short installation area.

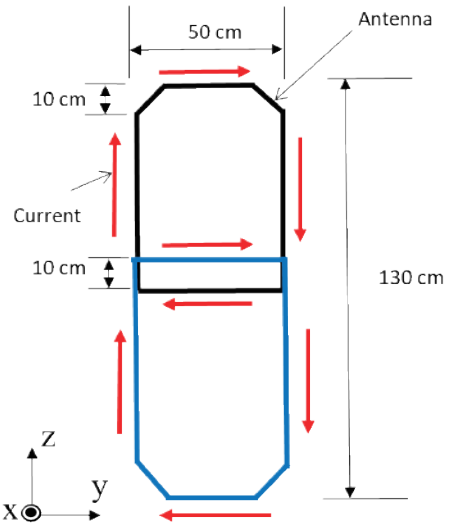


Fig. 1. Structure of a conventional EAS Tx Antenna

II. Conventional Antnena Structure and Problems

Figure 1 presents the structure of a transmission antenna previously proposed for EAS systems [5, 6]. A simulation was conducted by applying a signal of 8.2 MHz to this structure, using the electromagnetic software CST Studio Suite. The simulation results are shown in Figure 2. Figure 2(a) displays the distribution of the current, Figure 2(b) shows the magnetic field distribution in the x-z plane, and Figure 2(c) illustrates the magnetic field distribution in the x-y plane. As a person passes by the transmission antenna, examining the magnetic field distribution in the x-z plane reveals that only the x component

exists at the midpoint, with no z component present. Observing Figure 2(c), as a person moves, the components of the magnetic field exist as varying y and x components. Thus, considering all x, y, z components, the region where components fail to emerge as a person moves is the middle area of the transmission antenna, where the z component is absent. This area can lead to detection errors.

To eliminate such shadow areas, a structure like that shown in Figure 3 has been used [7]. Employing a structure similar to Figure 3 results in the movement of the middle region, where the z-direction component of the magnetic field is not generated, from below to above as the person moves. Thus, using this structure allows people to pass through regions that include some z component at the same height, thereby increasing the tag recognition rate. However, this structure also does not fundamentally eliminate shadow areas; it merely shifts their location depending on the position. Additionally, there is a problem that the transmission antenna needs to be larger in the y-axis direction, which can increase the required installation space.

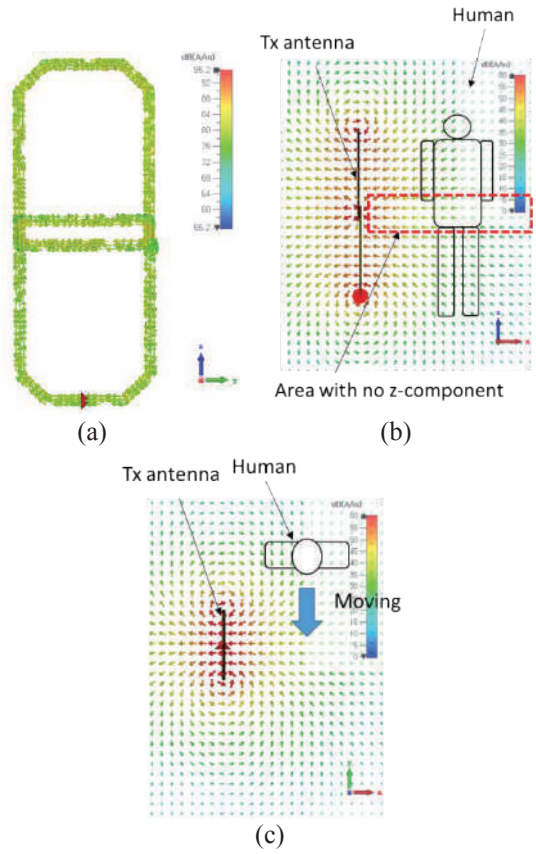


Fig. 2. Simulation results of the conventional EAS antenna: (a) current distribution, (b) magnetic field in the x-z plane, (c) magnetic field in the x-y plane

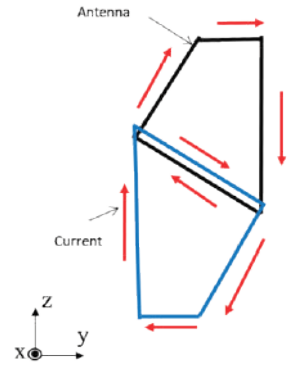


Fig. 3. Previously proposed structure of EAS Antenna to enhance z-component of magnetic field

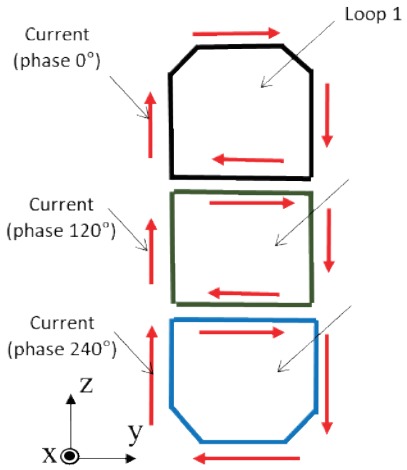


Fig. 4. Proposed antenna structure of the EAS

III. Proposed Antenna Structure and Simulation Results

The structure of the proposed antenna designed to enhance the z component of the magnetic field is illustrated in Figure 4. The proposed antenna utilizes three loops. Signals are applied to each loop antenna, with each signal phased 120 degrees apart from the others. By applying signals in this manner, the magnetic field over the three loops forms not as standing waves but as traveling waves. This means the direction of the magnetic field at a fixed position changes over time. The simulation results are depicted in Figure 5, which shows the magnetic field distribution of the transmission antenna over time periods. Since the area of interest is primarily the z component, the magnetic field distribution is illustrated in the x-z plane. Figure 5(a) corresponds to the 0 cycle, (b) to the 1/3 cycle, and (c) to the 2/3 cycle. From Figure 5(a), it can be seen that, similar to the existing transmission antenna, there are areas in the middle of the antenna where the z component of the magnetic field does not occur. However, as time progresses to 1/3 of the cycle, it can be observed that a z component of the magnetic field arises in the same region. By constructing the transmission antenna using the proposed method, magnetic field components can be generated in all directions over time, thereby reducing the rate

of missed detections of tags. Additionally, since there is no need to extend the length in the y direction, this approach offers the advantage of reducing the size of the transmission antenna. Reducing the size also means that it can be installed in smaller spaces, thus solving space issues during installation and offering a significant advantage.

In this paper, it is necessary to further validate the proposed method through experiments verifying the magnetic field distribution. To achieve this, it is essential to fabricate a high-frequency three-phase inverter for experimentation. The addition of new components not previously used may lead to a slight increase in costs. Additionally, for practical application, modulation of the high-frequency waveform is required, which will need to be addressed in future research.

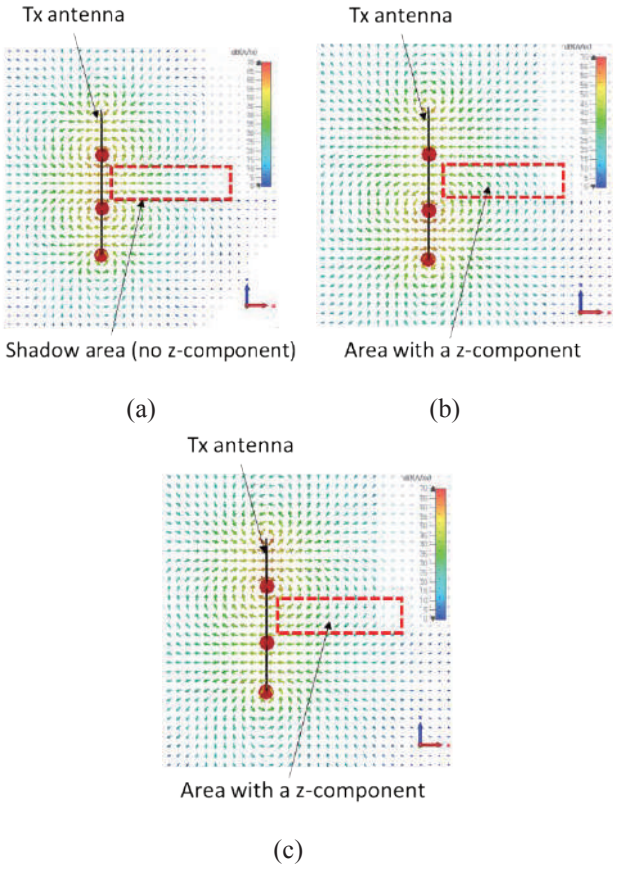


Fig. 5. Magnetic field distribution in the x-z plane over time periods: (a) 0 Cycle, (b) 1/3 Cycle, (c) 2/3 Cycle

IV. Conclusion

This study proposed a new antenna structure to address the issue of shadow areas for tag detection within Electronic Article Surveillance (EAS) systems. By vertically arranging three transmission coils and applying signals with a 120-degree phase difference to each coil, this structure induces temporal changes in the magnetic field, enabling the detection of tags from various directions. Notably, this research presents a method for effectively forming the z-axis component of the magnetic field, a critical aspect of the magnetic field, and validates the effectiveness of this approach through simulation results. The proposed antenna structure overcomes the limitations of existing EAS systems and significantly reduces shadow areas that can occur due to magnetic field polarization, thereby substantially improving the system's detection range and reliability. The findings of this study provide important guidelines for the design and optimization of EAS systems and are expected to contribute significantly to improving theft prevention efficiency in actual commercial environments. Future research should evaluate the applicability and efficiency of this approach in various real-world settings through experimental testing.

References

[1] B.P.Bhatt, J..M. Rathod, Y.D. Parmar, “A Comparati ve Stu dy of Electron ic Artic leSurveillance (EAS) System,” 2nd International Conference on Signals, Systems & Automation (ICSSA-11), Gujarat, India, 2011, pp. 1-4.

[2] A. Gulczyński and J. Biernacki, "Algorithm of Detection and Location of the Tag for Acousto- Magnetic Electronic Article Surveillance System," 2018 Progress in Applied Electrical Engineering (PAEE), Koscielisko, Poland, 2018, pp. 1-5

[3] M. Takeda, Y. Bannno, H. Hirata, Y. Okita, M. Kimura and T. Sugiura, "Electromagnetic field simulation around implantable cardiac pacemakers caused by EAS system [electronic article surveillance]," 2004 Asia-Pacific Radio Science Conference, 2004. Proceedings., Qingdao, China, 2004, pp. 444-446

[4] O. P. Gandhi, G. Lazzi, Jun Sun and Gang Kang, "Use of the impedance method for ICNIRP-related compliance testing of electronic article surveillance devices," IEEE Antennas and Propagation Society International Symposium. Transmitting Waves of Progress to the Next Millennium. 2000 Digest. Held in conjunction with: USNC/URSI National Radio Science Meeting, C, Salt Lake City, UT, USA, 2000, pp. 1092-1095

[5] C. Li, T. Wu, “Numerical Evaluation of Exposure to the Electromagnetic Fields of an Electronic Article Surveillance System with Postured Infant Model,” nternational Symposium on Electromagnetic Compatibility 2014, Tokyo , Japan, 2014, pp. 14P1-H4,

[6] Gang Kang and O. P. Gandhi, "Comparison of various safety guidelines for electronic article surveillance devices with pulsed magnetic fields," in IEEE Transactions on Biomedical Engineering, vol. 50, no. 1, pp. 107-113, Jan. 2003

[7] Seungin No, “A Study on Z-component H-Field Enhancement of Coil for Electronic Article Surveillance System,” MS dissertation, Dept. Elect. Com. Eng., Seoul National University, Seoul, Korea, 2016.

Blockchain-Inspired GPS Spoofing Detection for Zero-Trust Mission-Critical Military IoT Networks

Ahmad Zainudin[*†], Revin Naufal Alief[§], Made Adi Paramartha Putra[‡], Dong-Seong Kim[§], and Jae-Min Lee[§]
[*]Department of Electronic Engineering, Kumoh National Institute of Technology, Gumi 39177, South Korea
[§]Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi 39177, South Korea
[†]Department of Electrical Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia
[‡]Faculty of Information Technology and Design, Primakara University, Denpasar, Indonesia
(zai*, dskim, ljmpaul)@kumoh.ac.kr and adi@primakara.ac.id

***Abstract*—Advanced battlefields utilize the Internet of Military Things (IoMT) devices with accurate global positioning system (GPS) information for communication and navigation. However, these modern battlefields are vulnerable to GPS signal spoofing attacks, which create an incorrect GPS signal as the legitimate navigation point. Furthermore, in the IoMT environment, one does not trust others, making it more challenging to develop secured military services. To address those issues, this study introduces a blockchain-inspired collaborative GPS spoofing detection (CGPS-SD) for zero-trust mission-critical military IoT networks. A blockchain-based decentralized authentication is employed to provide a trusted CGPS-SD framework. Based on the measurement results, the proposed CGPS-SD system outperforms existing approaches with enhanced security capabilities.**

***Index Terms*—Collaborative GPS spoofing detection (CGPS-SD), Zero-trust architecture (ZTA), blockchain-based Internet of military things (IoMT)**

I. Introduction

The network of Internet of Miltary Things (IoMT) devices utilizes contemporary warfare by intelligently gathering critical data, including navigation, positioning, and surrounding environmental information [1]. Most IoMT devices (military vehicles, soldiers, tanks, naval, aircraft, and UAVs) facilitate autonomous operations in military scenarios by utilizing a global positioning system (GPS). Nevertheless, GPS data lacks encryption, rendering it susceptible to advanced spoofing attacks. A GPS spoofing attack misleads the user target by broadcasting falsified signals that mimic the attributes of legitimate GPS satellite signals. This disruption poses a critical threat to the safety of automated IoMT devices. Here, accomplishing mission and security is a priority in military services. The classical defense methods rely on real-time analysis of GPS signals, including clock data and arrival angles [2]. This approach may lack adaptability, be tailored to specific threats, and require significant investment in installing specialized equipment. To address this challenge, current approaches leverage machine learning (ML) and deep learning (DL) advancements to distinguish and classify falsified locations [3]. The model classifies the type of falsification based on the GPS signals feature. However, a central-centric learning technique in developing a GPS spoofing detection framework poses data privacy and communication overhead issues.

To address the limitation of the centralized learning technique, a decentralized learning scenario using the federated learning (FL) paradigm was adopted in GPS spoofing detection method development [4]. The edge devices collaboratively conduct local training using their own data and utilize a model aggregation mechanism. This decentralized technique provides privacy-preserving, high scalability, and better model performance [5]. However, the conventional FL technique utilizes a centralized aggregation technique vulnerable to single-point-of-failure (SPoF) issues. Furthermore, the untrusted IoMT devices can participate in the FL environment and affect the model's reliability. Integrating blockchain with the FL technique provides fully decentralized and trustworthy learning scenarios. This advanced technique allows a decentralized aggregation technique by employing blockchain to store the aggregated model through a consensus mechanism. However, the IoMT environment does not trust other devices, making it challenging for collaborative military services [6]. A zero-trust mechanism is required to provide a trust mechanism by joining blockchain, collaborative learning, and digital signature techniques. Considering the necessity of privacy, security, robustness, and zero-trust in detecting GPS spoofing within IoMT device networks, this paper provides potential contributions:

1) We implement an integrated blockchain-FL-based collaborative GPS spoofing detection system to identify falsified GPS locations and provide a zero-trust architecture (ZTA) within IoMT networks.
2) We introduce a blockchain-based decentralized authentication and an Elliptic Curve Digital Signature Algorithm (ECDSA)-based privacy-preserving technique to enable a trusted and anti-tampering CGPS-SD framework.
3) We develop an efficient collaborative GPS spoofing detection framework to ensure fast detection with a low-complexity model structure for constrained IoMT devices.

The remainder of this study is structured as follows: Section II literature review of existing GPS spoofing detection approach. The outline of the proposed system for blockchain-based GPS spoofing detection for zero-trust IoMT network is

TABLE I
COMPARISON BETWEEN PROPOSED SYSTEM AND EXISTING GPS SPOOFING DETECTION TECHNIQUES

Ref	Decentralized Learning	Lightweight Model	Privacy Preserving	Distributed Aggregation	Verifiable & Trustworthiness	Decentralized Authentication	Zero-Trust Architecture
[4]	✓	✗	✓	✗	✗	✗	✗
[7]	✗	✓	✗	✗	✗	✗	✗
[8]	✗	✓	✗	✗	✗	✗	✗
[9]	✗	✓	✗	✗	✗	✗	✗
Ours	✓	✓	✓	✓	✓	✓	✓

presented in Section III. Section IV contains a detailed presentation of the experimental results and discussions. Lastly, Section V summarizes the conclusions and suggests directions for future works.

II. RELATED WORKS

The GPS spoofing detection system based on the ML model was investigated, and an artificially generated GPS dataset was employed to assess the model's detection speed and memory usage in [7]. To analyze the system's robustness, this work has compared various ML models, such as K-Near Neighbor (KNN), Support Vector Machine (SVM), Nu-SVM, and C-SVM. A strategy employing an ensemble model was implemented to safeguard unmanned aerial vehicles (UAVs) against GPS signal spoofing attacks [8]. This method employs a stacked SVM and convolutional neural network (CNN) SVM-CNN approach to enhance performance reliability. Furthermore, a GPS spoofing detection framework supported by transportation infrastructure as the secured data source was implemented [9]. This approach utilized roadside units (RSU) and GPS data as the input of the ML model to decide the error location during GPS spoofing attacks. However, these approaches [7]–[9] operated using a centralized learning setting, raising concerns about data privacy within IoMT services.

The federated learning (FL) concept allows decentralized learning setups using edge devices for secure local training without sharing the raw data with third parties [5]. The authors [10] developed a decentralized GPS spoofing detection for vehicular ad hoc networks. A statistical technique based on spatial diversity and interconnectivity of vehicles is used to detect spoofed GPS signals. However, this technique has scalability issues and high false detection capability performance. Furthermore, in a cooperative intelligent transportation system (C-ITS) context, FL-based spoofing detection was utilized [4]. This method ensures both the privacy of data users and the effective detection of misconduct while facilitating swift training. Nevertheless, the basic FL method is susceptible to manipulation attacks during the aggregation phase [11]. Table I compares the proposed system and the existing GPS spoofing detection techniques. The proposed system provides various features, such as decentralized learning, lightweight model, privacy-preserving, distributed aggregation, verifiable and trustworthiness, decentralized authentication, and zero-trust architecture to address the limitations of the state-of-the-art.

III. PROPOSED BLOCKCHAIN-BASED ZERO-TRUST GPS SPOOFING DETECTION FRAMEWORK

Figure 1 presents the proposed collaborative GPS spoofing detection framework for blockchain-enabled IoMT networks. This setup involves terrestrial, aerial, and space network layers. The IoMT devices e_i (such as military vehicles, tanks, aircraft, soldiers, and UAVs), equipped with computational capabilities, perform securely distributed learning tasks. A permissioned blockchain-based authentication mechanism has been implemented to prevent unauthorized access. To provide a zero-trust mechanism, this system employs blockchain-based decentralized authentication (BDA), trusted model aggregation (TMA), and ECDSA-based privacy-preserving (ECDSA-PP) techniques.

A. Blockchain-based Decentralized Authentication

This system implements blockchain-based decentralized authentication to provide a zero-trust mechanism in the IoMT environment. Blockchain offers decentralized authentication, mitigating the risks associated with centralized authentication methods like public key infrastructure (PKI) employing a certificate authority (CA) and safeguarding against SPoF attacks [11]. Firstly, BDA defines the hash function $\mathbb{H} : \{0,1\}^* \rightarrow \mathbb{Z}_q^*$ and distributes the global parameter with the blockchain. The key generator node KGN randomly chooses the master private key msk, and the corresponding master public key is derived as $mpk = msk.P$. The KGN shares both the private and public keys to the IoMT devices during the registration phase through a secure communication channel. To access the FL environment network, the IoMT devices send a request message to aggregator As. The A_s verify, create the authentication parameters, and sign using their private keys $M = Sign\left(SK_{E_s}, \left(\left\{PID, PK_q\right\}, ID_{d_i}\right)\right)$. Subsequently, calculate $\alpha_p = \mathbb{H}\left(PID_r \,||\, PK_q \,||\, ID_{d_i}\right)$ and store α_p to the blockchain. Access the α_p; if available, the authentication is successful.

B. ECDSA-based Privacy Preserving Technique

The authorized IoMT devices conduct local training using the gathered GPS data alongside the initial model. The resulting updated local model is archived in a distributed storage system, the InterPlanetary File System (IPFS), generating a unique content identifier (CID). Before integration into the blockchain, the CID of the local model (CID_{local}) necessitates signing using an ECDSA-based signature mechanism, ensuring the integrity of model data and safeguarding against

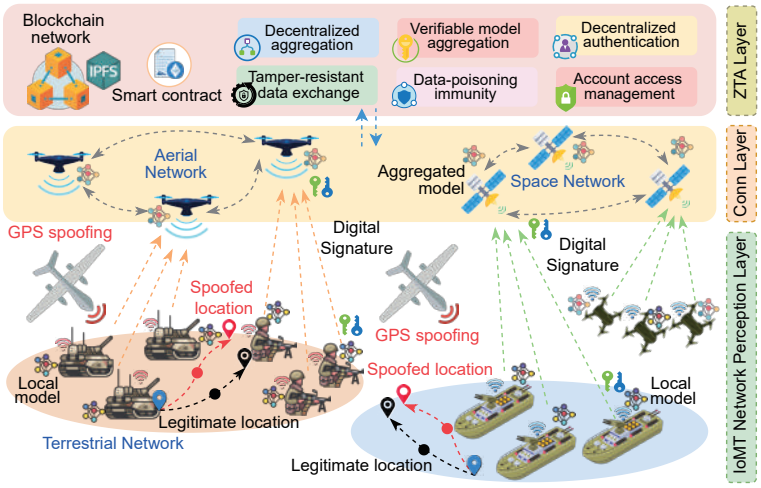


Fig. 1. Proposed blockchain-based collaborative GPS spoofing detection framework for zero-trust IoMT networks

TABLE II
COMPARISON OF EXISTING GPS SIGNAL SPOOFING DETECTION WITH PROPOSED FRAMEWORKS

Model	Accuracy	Precision	Recall	F1-Score	Loss	Average AUC Score	Trainable Parameter	Model Size	MFLOPs
DNN	91.65%	89.27%	88.09%	89.73%	0.1561	0.9239	6,137	22.82 KB	0.0135
CNN [8]	93.69%	91.28%	91.74%	91.58%	0.1176	0.9458	93,488	365.19 KB	0.1863
LSTM [8]	93.42%	91.06%	91.18%	91.12%	0.1516	0.9327	17,882	69.85 KB	0.0641
BiLSTM	92.96%	91.35%	90.96%	90.23%	0.1558	0.9214	21,713	83.17 KB	0.0832
GRU	93.14%	91.25%	91.36%	91.11%	0.1274	0.9389	13,622	53.21 KB	0.0481
CNN-MLP	92.36%	90.25%	89.64%	89.73%	0.1304	0.9317	8,474	33.10 KB	0.0165
Proposed	**95.12%**	**92.24%**	**92.46%**	**92.28%**	**0.0798**	**0.9547**	**4,682**	**18.29 KB**	**0.0091**

tampering during aggregation. The IoMT devices FL clients produce a 256 key-pair and sign the $\text{CID}_{n,r}^{local}$ utilize ECDSA private key $\{p, d\}$, where p and d are typically big integers. In sequence, calculate their local model hash h_{local} using SHA-256 and encrypt h_{local} to execute the signature $sign_{local}^{\text{CID}}$, as follows:

$$h_{local} = \text{hash}(\text{CID}_{n,r}^{local}), \tag{1}$$

$$sign_{local,n}^{\text{CID}} = h_{local}^{d} \,(\text{mod } p)\,, \tag{2}$$

where mod represents a modulus operation. The UAV aggregator A_n scrutinizes all local models consolidates the authenticated models, and deposits the CID of the global model (CID_{global}) into the permissioned blockchain via the Hyperledger Besu framework. The system allows only authorized IoMT devices to participate in the collaborative learning environment. The ECDSA signature technique facilitates a verification mechanism for model-transferring integrity and anti-tampering model aggregation.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed model utilizes a multi-flow multilayer perceptron (multi-MLP). Furthermore, real-time GPS signals are used to evaluate the model's capability. This dataset [7] employs an N200 universal software radio peripheral (USRP) device as a GPS receiver to gather the GPS signals in stationary and moving scenarios for an autonomous UAV. The gathered data comprises 13 extracted features from eight channels with $158,170$ data samples. The GPS spoofing attacks are simulated in the indoor testbed environment based on the complexity and sophistication, such as simplistic, intermediate, and sophisticated attacks.

Table II presents a comparison of the performance's proposed GPS spoofing detection model with existing approaches, such as DNN, CNN [8], LSTM [8], BiLSTM, GRU, and CNN-MLP models. The CNN [8] model achieves an accuracy of 93.69%, a precision of 91.28%, a recall of 91.74%, and an F1-score of 91.58%. This model has $93,488$ trainable parameters, 365.19 KB model size, and 0.1863 MFLOPs calculation. Furthermore, the existing LSTM [8] model performs an accuracy, precision, recall, and F1-score of 93.42%, 91.06%, 91.18%, and 91.12%, respectively. For model efficiency measurement, this model has trainable parameters of $17,882$, a model size of 69.85 KB, and an MFLOPs calculation of $0/0641$. Based on this comparison, the proposed model outperforms existing approaches, achieving an accuracy of 95.12%, a precision of 92.44%, a recall of 92.46%, and an F1-score of 92.28% while maintaining a low-complexity model structure. The proposed model comprises $4,682$ trainable parameters, occupies a model

TABLE III
TRANSACTION TIME OF SETTING AND ACCESSING CID DATA TO THE BLOCKCHAIN WITH DIFFERENT CONSENSUS

Consensus	Transaction Time (s)			
	Setting GM	Setting LM	Accessing GM	Accessing LM
QBFT	**1.891**	**1.815**	**0.007**	**0.008**
IBFT 2.0	1.992	1.957	0.008	0.008
Clique	2.021	1.995	0.008	0.008
Ethash	3.584	3.157	0.009	0.009

*__GM__: Global Model, *__LM__: Local Model

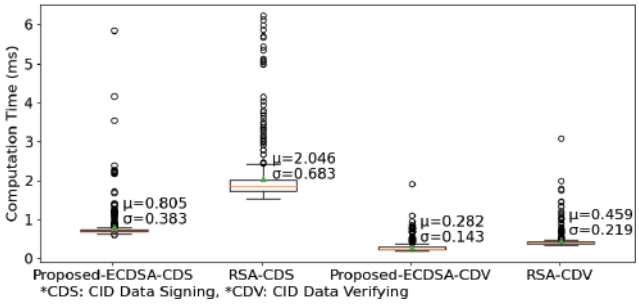


Fig. 2. Computation time of proposed ECDSA-based signature for privacy-preserving technique in distributed blockchain-based IoMT

size of 18.29 KB, and has an MFLOPs calculation of 0.0091.

The transaction time of setting and accessing CID data to the blockchain with different consensus algorithms, such as QBFT, IBFT 2.0, Clique, and Ethash consensuses, is presented in Table III. This comparison evaluates *SettingGM()*, *SettingLM()*, *AccessingGM()* and *AccessingLM()* functions in decentralized aggregation. Based on the comparison results, the blockchain-based decentralized aggregation performs exceptionally well with QBFT consensus, exhibiting transaction times of 1.891 seconds for the average setting global model, 1.815 seconds for the average setting local model, 0.078 seconds for the average access global model, and 0.008 seconds for the average access local model. Table IV presents the performance of the proposed blockchain-based decentralized authentication. Based on the evaluation results, superior performance is achieved when using QBFT consensus. The transaction time of the *Registration()* function is 1.744 seconds, and for *Authentication()* function is 0.085 seconds. The comparative performance of integrating digital signatures with a decentralized aggregation mechanism is depicted in Figure 2. The proposed system uses an ECDSA-based signature, which entails a computation time of 0.805 milliseconds for CID data signing and 0.282 milliseconds for CIDs data verification.

V. CONCLUSION

This study implements a reliable and trusted collaborative GPS spoofing detection framework, integrating FL, blockchain, and signature-based privacy-preserving techniques for zero-trust mission-critical military IoT networks. The system provides a zero-trust mechanism by leveraging blockchain-based decentralized authentication and ECDSA-based signature for model exchange immunity. The proposed model outperforms by achieving high accuracy and compact model size while providing a secure learning mechanism.

TABLE IV
TRANSACTION TIME OF DECENTRALIZED AUTHENTICATION PERFORMANCE

Consensus	Transaction Time (s)	
	Registration()	Authentication()
QBFT	**1.744**	**0.085**
IBFT 2.0	1.853	0.093
Clique	1.892	0.097
Ethash	3.457	0.128

ACKNOWLEDGMENT

This work was partly supported by the Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 33.3%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33.3%) and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(2018R1A6A1A03024003,2022R1I1A1A01071058, 33.3%).

REFERENCES

[1] R. Akter, M. Golam, V.-S. Doan, J.-M. Lee, and D.-S. Kim, "IoMT-Net: Blockchain-Integrated Unauthorized UAV Localization using Lightweight Convolution Neural Network for Internet of Military Things," *IEEE Internet of Things Journal*, vol. 10, no. 8, pp. 6634–6651, 2022.

[2] M. Ceccato, F. Formaggio, N. Laurenti, and S. Tomasin, "Generalized Likelihood Ratio Test for GNSS Spoofing Detection in Devices with IMU," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3496–3509, 2021.

[3] M. Kamal, A. Barua, C. Vitale, C. Laoudias, and G. Ellinas, "GPS Location Spoofing Attack Detection for Enhancing the Security of Autonomous Vehicles," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*. IEEE, 2021, pp. 1–7.

[4] K. A. Bonfim, F. D. S. Dutra, C. E. T. Siqueira, R. I. Meneguette, A. L. Dos Santos, and L. A. P. Júnior, "Federated Learning-based Architecture for Detecting Position Spoofing in Basic Safety Messages," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. IEEE, 2023, pp. 1–5.

[5] A. Zainudin, R. Akter, D.-S. Kim, and J.-M. Lee, "Federated Learning Inspired Low-Complexity Intrusion Detection and Classification Technique for SDN-Based Industrial CPS," *IEEE Transactions on Network and Service Management*, 2023.

[6] E. C. Nkoro, J. N. Njoku, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Zero-Trust Marine Cyberdefense for IoT-Based Communications: An Explainable Approach," *Electronics*, vol. 13, no. 2, p. 276, 2024.

[7] G. Aissou, S. Benouadah, H. El Alami, and N. Kaabouch, "Instance-based Supervised Machine Learning Models for Detecting GPS Spoofing Attacks on UAS," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022, pp. 0208–0214.

[8] M. Umer, I. Ashraf, Y. Park *et al.*, "Enhanced Machine Learning Ensemble Approach for Securing Small Unmanned Aerial Vehicles from GPS Spoofing Attacks," *IEEE Access*, 2024.

[9] F. Wang, Y. Hong, and X. Ban, "Infrastructure-Enabled GPS Spoofing Detection and Correction," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[10] F. A. Milaat and H. Liu, "Decentralized detection of GPS spoofing in vehicular ad hoc networks," *IEEE Communications Letters*, vol. 22, no. 6, pp. 1256–1259, 2018.

[11] A. Zainudin, M. A. P. Putra, R. N. Alief, R. Akter, D.-S. Kim, and J.-M. Lee, "Blockchain-Inspired Collaborative Cyber-Attacks Detection for Securing Metaverse," *IEEE Internet of Things Journal*, 2024.

Blockchain-enabled Token for Military Equipment Logistics

Ikechi Saviour Igboanusi[1], Cosmas Ifeanyi Nwakanma[1], Mirian Odigomma Igboanusi[1], and Dong-Seong Kim[2]

[1]ICT Convergence Research Center, Kumoh National Institute of Technology, South Korea
[2]Department of IT Convergence Engineering, Kumoh National Institute of Technology, South Korea
ikechisaviour@kumoh.ac.kr, cosmas.ifeanyi@kumoh.ac.kr, mirian@kumoh.ac.kr, dskim@kumoh.ac.kr

***Abstract*—This research explores the integration of blockchain technology into military equipment logistics, aiming to enhance security, transparency, and efficiency. By leveraging blockchain, a decentralized and tamper-resistant ledger, we establish a robust system for tracking and managing military assets throughout their lifecycle. Digital identity tokens, IoT-Driven updates, smart contracts, and security are the components of this blockchain-enabled system model. Each military equipment item is assigned a unique digital token on the blockchain. This token serves as its verifiable identity, containing essential information such as origin, maintenance history, and current location. Blockchain-enabled Internet of Things (IoT) devices are strategically deployed at various checkpoints (e.g., supply depots, and transport hubs). These devices continuously update the equipment's token as it moves from one location to another. Immutable records ensure data integrity. We implement smart contracts on the Ethereum blockchain. These self-executing contracts automate processes such as equipment transfer, maintenance scheduling, and compliance checks. Smart contracts enhance efficiency and reduce manual intervention. The blockchain's cryptographic security mechanisms protect against unauthorized access, fraud, and data manipulation. Additionally, redundancy across nodes ensures resilience even in challenging operational environments. Our empirical evaluation demonstrates that this blockchain-based approach achieves 98% security, making it a promising solution for modernizing military logistics.**

***Index Terms*—Blockchain, Military, Token, Ownership, smart contract.**

I. Introduction

With the introduction of decentralized and immutable ledger systems, blockchain technology integration has become a disruptive force in some industries in recent years, completely transforming traditional processes [1]. The use of blockchain to secure military computing systems has been explored [2], [3], [4]. Of the many uses, the use of blockchain technology in military equipment logistics has attracted a lot of interest since it can resolve issues with supply chain management that have long plagued defense facilities across the globe [5]. An exciting opportunity to improve the robustness and efficacy of military logistical operations is to investigate blockchain technologies as contemporary warfare develops and expectations for efficiency, transparency, and security increase [1], [5].

Managing military equipment presents unique logistical challenges due to its large inventory, complex supply networks, and strict security measures. Inefficiencies in traditional logistical systems, such as data inconsistencies and supply chain weaknesses, might jeopardize national security and operational preparedness. Researchers and defense organizations have looked to blockchain technology in recognition of these difficulties as a way to improve resource allocation, strengthen data integrity, and expedite procedures in military logistics [6], [7], [8].

Blockchain technology provides a decentralized, immutable ledger system that makes transactions between various stakeholders transparent and auditable. Blockchain improves the visibility and traceability of military equipment throughout its lifecycle, from acquisition and maintenance to deployment and decommissioning, by securely and irreversibly documenting each transaction. Transparency promotes accountability and reduces the possibility of fraudulent activity, illegal alterations, and the infiltration of counterfeit parts into the supply chain.

Furthermore, military logistics networks are made more resilient and durable by the decentralized structure of blockchain, which lessens reliance on centralized authorities and single points of failure. Blockchain makes it possible to automate contracts and compliance procedures using smart contracts [1], which speeds up transactional procedures while guaranteeing conformity to operational and regulatory requirements. Furthermore, real-time monitoring and data-capturing capabilities are further enhanced by the integration of Internet of Things (IoT) devices with blockchain, allowing for predictive maintenance and optimal resource usage [5].

The implementation of blockchain technology in military equipment logistics presents several obstacles and considerations, notwithstanding its potential advantages. With an emphasis on case studies, theoretical frameworks, real-world implementations, and future possibilities, this publication seeks to offer a thorough examination of blockchain technology's potential applications in military equipment logistics. The aim of this work is to propose:

- the use of blockchain smart contracts to manage military logistics data, and
- to practically implement the proposed system on a blockchain network.

The rest of this work is arranged as follows: Section II presents the Related Works. The proposed System Model is presented in Section III. In Section IV, the results Section is shown and finally, the Conclusion is presented in Section V.

II. Related paper

The authors in paper [9] introduced the conceptual model of military blockchain for repair parts supply chain management. The model aims to address the limitations of traditional supply chain management, particularly in the military context, by leveraging blockchain technology to ensure the authenticity and traceability of spare parts. It emphasizes the importance of data integrity and transparency in military supply chain management and outlines the key components of the proposed military blockchain model, including the involvement of three main organizations: the supplier, service provider, and military depot. The model focuses on ensuring the authenticity of spare parts, detecting counterfeit parts, and maintaining the transparency and immutability of data. Additionally, the section highlights the potential application of blockchain technology in military defense applications and emphasizes the significance of traceability in the military blockchain model.

In the paper [10], the authors discussed the application of blockchain technology in military supply chain management, focusing on addressing issues such as trust management, privacy protection, and information transparency. It highlights the advantages of blockchain in enhancing collaboration within the military supply chain, emphasizing the need for a more efficient and secure system. The study analyzes the current status of blockchain technology, supply chain collaboration, and military supply chain management, providing insights into the potential integration of blockchain to improve the shortcomings in military supply chain operations.

In the context of military maintenance, a specialized framework is imperative to tackle issues such as limited access to maintenance data, equipment traceability, and monitoring maintenance activities by higher authorities. By authors in [1] incorporating the PoET consensus algorithm, the framework ensures transparent validation of transactions without the need for extensive computational power or tokens, promoting efficiency and accuracy. Smart contracts streamline the validation process, reducing human intervention and enhancing the overall maintenance workflow. This blockchain-based solution not only safeguards critical military data but also empowers data-driven decision-making, enabling the integration of cutting-edge technologies like artificial intelligence and machine learning to revolutionize military capabilities. The framework's emphasis on security and data integrity addresses the challenges posed by third-party vendors and service providers, ensuring controlled and monitored access to military equipment usage data. Ultimately, this comprehensive maintenance management framework propels military organizations towards modernization and operational excellence through advanced technological tools and secure data management practices.

The authors in [11] proposes the integration of blockchain technology, smart contracts, and artificial intelligence to enhance the management of dual-use goods and technologies in international trade. By utilizing blockchain, the process gains autonomy, trust, and transparency, reducing the risk of fraud and errors. Smart contracts streamline authorization procedures and ensure compliance with regulations, while AI enhances decision-making capabilities. The combination of these technologies aims to improve the accuracy and efficiency of export control processes, benefiting both international security and commercial competitiveness. The proposed system offers a secure and decentralized approach to managing cross-border trade, providing a reliable and transparent framework for stakeholders involved in the export process.

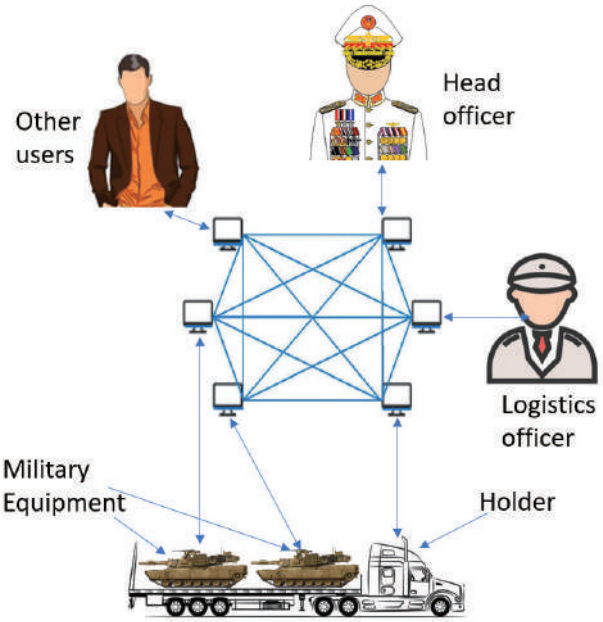


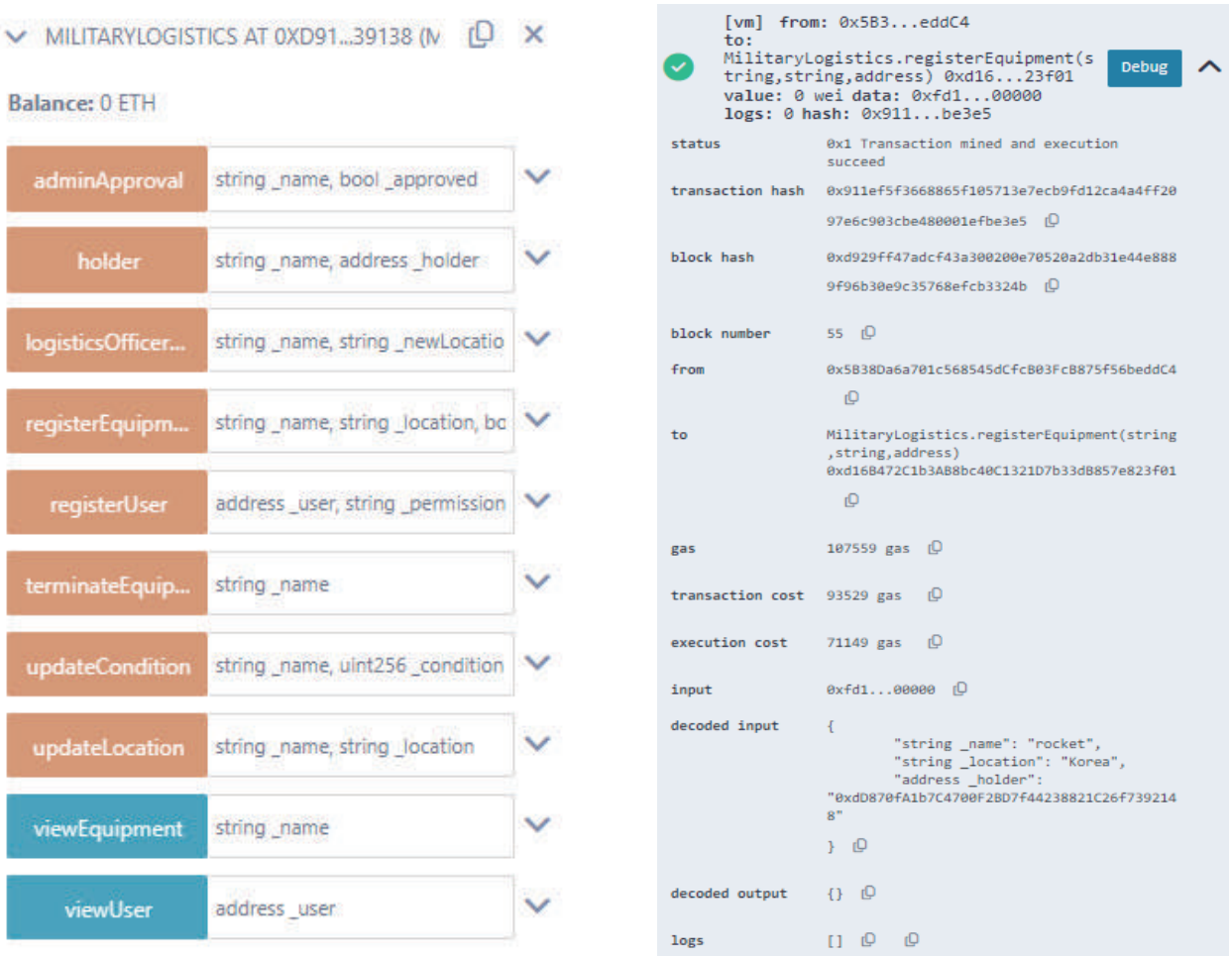
Fig. 1: Diagram of the proposed blockchain-enabled military logistics system model

This paper unlike the above paper attempts to focus on the practical implementation of a blockchain for military logistics.

III. System Model

This system interaction revolves around three primary categories of user accounts: the holder, the logistics officer, and the head officer as shown in Fig. 1. The holder is an essential user of this system. They are associated with specific military equipment and play a crucial role in maintaining its information. The holder's responsibilities include updating the equipment's location, status, and condition. The logistics officer is responsible for registering military equipment into the blockchain-enabled logistics network. When a piece of military equipment is registered, a unique token is generated. This token serves as a key for updating the equipment's information. The head officer oversees the entire logistics process. Before any movement of military equipment begins, the logistics officer sets the destination. The head officer then approves the movement.

Users (holders, logistics officers, and head officers) are registered into the blockchain network by an authorized entity. This ensures that only authorized personnel can participate in the logistics system. The blockchain network provides security, transparency, and immutability for the recorded data. The logistics officer registers military equipment within the blockchain network. This registration process generates a unique token associated with the equipment.



(a) (b)

Fig. 2: The interaction interface to the smart contract is shown in (a) and the blockchain activity is shown in (b)

The information linked to a piece of military equipment includes:

- Location: The current geographical position of the equipment. Usage Status: Whether the equipment is still in use or decommissioned.
- Registration Status: Whether the equipment is officially registered in the system.
- Holder's Address: The address of the authorized holder. Last Location Update Time: Timestamp indicating when the location information was last updated.
- Equipment Condition: Details about the equipment's state (e.g., operational, under maintenance, damaged). Updating Holder Information:

The logistics officer is responsible for maintaining accurate information about the equipment's holder. If there are any changes (e.g., a new holder), the logistics officer updates this information in the system. When military equipment is preparing to move, the logistics officer sets the destination. This step ensures that the equipment reaches the intended location. The head officer's approval is required before the movement begins. This ensures proper coordination and authorization. An Internet of Things (IoT) device associated with the holder continuously updates the equipment's location. Whenever the equipment moves or its location changes, the IoT device updates the system with the latest information.

IV. Results

The implementation of the blockchain system was carried out through a Solidity smart contract deployed on a private network based on an Ethereum Virtual Machine (EVM) blockchain network. In Fig. 2(a), we can observe the interaction interface of the smart contract within the Remix Ethereum Integrated Development Environment (IDE). Fig. 2(b) depicts the blockchain network activities during the use of the blockchain smart contract for equipment registration.

To update equipment information, the corresponding function is utilized. During our testing, the smart contract effectively updated the records associated with the equipment. Furthermore, the information pertaining to both the equipment and all users was successfully retrieved from the blockchain smart contract.

The security assessment of our smart contract implementation was conducted using the Solidcheck platform. This evaluation aimed to identify any vulnerabilities within our system. As depicted in Fig. 3, the results indicate an impressive

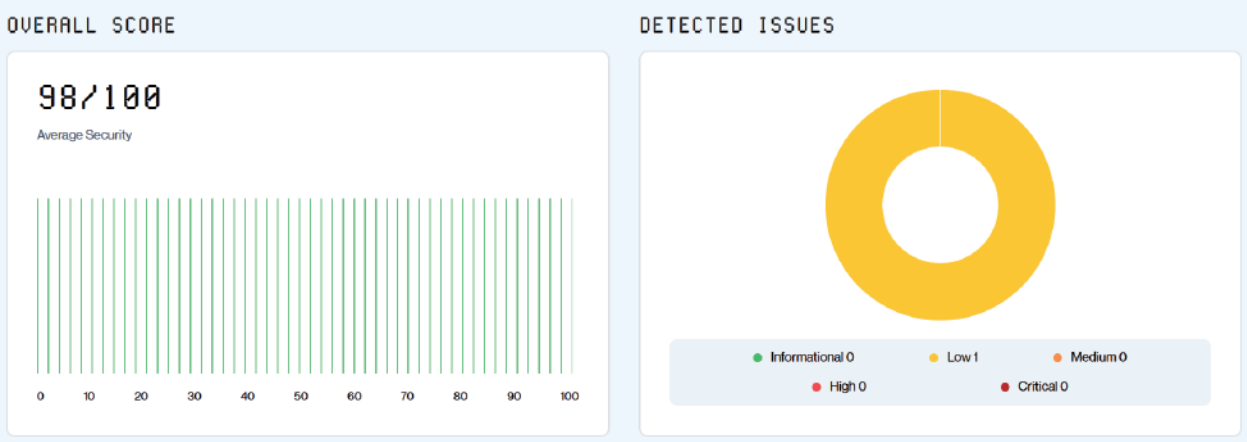


Fig. 3: The security analysis result of the implemented system

average security score of 98%. However, it's essential to note that a low-risk warning was flagged during the assessment.

V. Conclusion

The project has achieved successful implementation of a blockchain-enabled solution for military equipment logistics, utilizing an Ethereum Virtual Machine (EVM) smart contract. Through this system, equipment-related transactions and data are securely managed on the blockchain.

In terms of security, a comprehensive analysis was conducted, resulting in an impressive security score of 98%. This high rating reflects the robustness of the implemented solution and its ability to safeguard sensitive information and processes.

In the future, we intend to extend this work to include equipment tracking and ownership and achieve 100% security of the implementation.

Acknowledgement

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korean government(MSIT) (IITP-2024-2020-0-01612, 50%) and Priority Research Centers Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology"(2018R1A6A1A03024003, 50%)

References

[1] R. S. Mohril, B. S. Solanki, B. K. Lad, and M. S. Kulkarni, "Blockchain enabled maintenance management framework for military equipment," *IEEE Transactions on Engineering Management*, vol. 69, no. 6, pp. 3938–3951, 2022.

[2] R. Kumar and R. Ahmad Khan, "Securing military computing with the blockchain," *Computer Fraud & Security*, vol. 2024, no. 2, p. null, 2024. [Online]. Available: https://doi.org/10.12968/S1361-3723(24)70007-4

[3] D. Sonchita, "Application of blockchain technology in military affairs," in *Emerging Digital Technologies and India's Security Sector*, 2024, p. 16.

[4] M. Golam, R. Akter, R. Naufal, V.-S. Doan, J.-M. Lee, and D.-S. Kim, "Blockchain inspired intruder uav localization using lightweight cnn for internet of battlefield things," in *MILCOM 2022 - 2022 IEEE Military Communications Conference (MILCOM)*, 2022, pp. 342–349.

[5] M. Khan, S. Imtiaz, G. S. Parvaiz, A. Hussain, and J. Bae, "Integration of internet-of-things with blockchain technology to enhance humanitarian logistics performance," *IEEE Access*, vol. 9, pp. 25 422–25 436, 2021.

[6] S. B. Rahayu and A. Azahari, "Military supply chain management and blockchain development," *Journal of Education and Social Sciences*, 2019.

[7] H. Tran-Dang and D.-S. Kim, "Physical internet for military logistics: Perspectives," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, 2019, pp. 755–757.

[8] S. S. Et.al, "Consortium blockchain for military supply chain," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:234854412

[9] S. B. Rahayu, N. Jusoh, M. H. M. Halip, S. M. Taib, and M. G. Lee, "A conceptual model of military blockchain for repair parts supply chain management," in *2021 International Conference on Computer & Information Sciences (ICCOINS)*, 2021, pp. 146–150.

[10] Y. Zhang and Y. Tang, "Analysis of the application of blockchain in the militarysupplychain," *Advances in Economics and Management Research*, pp. 402–407, 2022.

[11] C. Coscarella and A. Minotti, "An institutional blockchain as a tool to control the export of dual-use and military goods," *STRATEGIC TRADE REVIEW*, vol. 6, no. 9, pp. 93–112, 2020.

Detecting DDoS in Internet of Military Things (IoMT) Data Caching: Machine Learning Approach

Love Allen Chijioke Ahakonye †, Cosmas Ifeanyi Nwakanma †, Dong-Seong Kim
IT Convergence Engineering, † ICT Convergence Research Center,
Kumoh National Institute of Technology Gumi, South Korea
(loveahakonye, cosmas.Ifeanyi, dskim)@kumoh.ac.kr

***Abstract*—With its networked military equipment, the Internet of Military Things (IoMT) is bothered by severe security issues, hampering data accessibility and service delivery. Distributed Denial of Service (DDoS) incidents are risks that disrupt mission-critical activities. To detect DDoS assaults within IoMT data caching systems, this article investigated machine learning algorithms. Leveraging machine learning techniques, it adapts dynamically to evolving attack patterns, enhancing the resilience and availability of the cached data in the intricate IoMT networks. Extensive experiments demonstrate the significance of ensemble learning in addressing the intricacies of multiclass DDoS instances in IoMT networks.**

***Index Terms*—AI, Data Caching, DDoS, Deep Learning, IoMT, ML**

I. Introduction

The Internet of Military Things (IoMT) integrates military devices, sensors, and communication channels for real-time data exchange and decision-making in military operations [1]. This technology offers advantages such as asset tracking, resource optimization, and enhanced situational awareness [2]. IoT sensors improve logistics and maintenance, while drones and sensors aid surveillance missions [3], [4]. Wearable IoT devices monitor soldier health and communication, enhancing security and mission readiness [1].

IoMT enables dynamic responses through real-time learning, utilizing edge networks, cached data, and wireless nodes [5]. The volumes of data exchanged between these networked devices and the strategic storage and retrieval of critical information within the military network infrastructure are paramount [2]. The IoMT ensures swift access to crucial data at strategic locations, such as command instructions, surveillance data, and intelligence reports by caching information from edge devices, base stations, and other network nodes. Moreover, IoT networks constantly face security breaches, which disrupt service availability [6].

Studies have shown that edge caching enhances system efficiency and reduces network latency in various contexts, including the Internet of Vehicles and Information-Centric Networks [7]. Edge caching in IoT framework optimizes system performance, reduces latency, and improves scalability [2]. However, the prevalence of DDoS attacks necessitates efficient detection mechanisms to safeguard sensitive data [8]. Proposed solutions include access control and encryption, but careful implementation is crucial to avoid latency overheads [5].

Data caching is essential in resource-constrained devices and remote locations, where DDoS attacks could lead to catastrophic security situations [2]. Thus, a robust and efficient intrusion detection mechanism is critical for maintaining data availability and continuous service delivery in IoMT environments [9]. IoMT poses distinct challenges compared to civilian IoT due to its impact on human lives and national security. In hostile environments like battlefields and border surveillance, adversaries use hoaxes, physical attacks, and cyber threats to target systems [5]. Secure interconnectivity for real-time data collection and transmission among soldiers, vehicles, satellites, and UAVs is critical but challenging. Data caching ensures data availability and accessibility despite its security limitations, requiring a holistic, intelligent, secure, and adaptable approach [1].

Machine learning has proven effective for intrusion detection in various IoT networks [6]. This study explores the efficiency of machine learning for detecting DDoS attacks in IoMT data caching systems, addressing unique security challenges and marking a pioneering effort in military IoT. This research underscores the need for specialized detection approaches due to IoMT's high mobility, diverse device types, and stringent security requirements. Real-time detection is crucial to prevent severe consequences from DDoS attacks. Machine learning-based IDS can significantly enhance IoMT security with intelligent, adaptive, and automated detection capabilities integrated into military communication and information systems for practical deployment.

Specific contributions of the study are as follows:

1) Exploration of representative machine learning models for accurate DDoS attack detection in IoMT data caching.
2) To demonstrate the feasibility of the machine learning techniques based on accuracy, precision, recall, error loss, and computation time for DDoS detection in IoMT data caching.
3) Highlight the possible machine learning candidates for DDoS detection in IoMT networks.

Section I is Section II, reviewing existing works data caching and DDoS attack detection in IoT networks. Section III discusses the system methodology for IoMT DDoS attack detection. Section IV is the experimentation and performance evaluation of explored machine learning algorithms.

Section V concludes the study.

II. Related Works

Various studies have explored the advantages of data caching, particularly in improving content delivery speed in network systems [2], [10]. However, these investigations often overlook the security aspects associated with caching, especially concerning IoT applications vulnerable to disruptive security threats [2]. While George et al. [11] examined vulnerabilities in Industrial IoT (IIoT) networks and proposed risk mitigation strategies, security-aware caching and content transfer received little attention in models such as that of Chaudhary et al. [12], focused on safe content distribution. Zahed et al. [2] introduced trusted caching nodes in IIoT networks, optimizing content breach costs and energy consumption. Despite Fu et al. [13] and Xu et al. [14] investigating IIoT data processing and content delivery efficiency, security-aware caching algorithms still need to be incorporated. Establishing trustworthy systems within IIoT environments, as emphasized by Zahed et al. [2], remains pivotal for ensuring security.

Exploring machine learning for DDoS attacks, primarily focusing on traditional networks, Aljuhani et al. [15] investigated DDoS attacks within IoT and SDN environments. Najafimehr et al. [6] extensively reviewed DDoS defence techniques within the IoT scenario, outlining attack methods and defence systems and suggesting implementation criteria for IoT security against DDoS. Nevertheless, this review overlooked the role of machine learning in mitigating DDoS threats. Developing effective DDoS defence necessitates consideration of various attack types as attackers continually evolve strategies. Today, leveraging modern technologies like machine learning offers advantages such as flexibility, scalability, and cost reduction [16], crucial in combatting diverse DDoS threats across dynamic environments.

However, contemporary networking domains such as IoMT remain susceptible to DDoS attacks. Sharafaldin et al. [17] proposed a new DDoS taxonomy for the application layer following their study of several DDoS attack categorizations, analyzed the most widely used DDoS datasets, and identified common flaws and deficiencies. A new DDoS dataset, CICDDoS2019, evaluated IDS/IPS algorithms and systems in response to these limitations. They also availed the most essential attributes for detecting various DDoS attacks. This study assists the researchers with the practical development of new approaches for tackling DDoS intrusions. Following the developmental study of DDoS attack detection in IoT networks, this study, as a preliminary research, reviews, analyzes and classifies DDoS attacks in IoMT data caching.

III. System Methodology

This section briefly overviews machine learning schemes for DDoS detection in IoMT networks with recourse to detection precision, error loss, and computation time. Fig. 1 is the process flow of the system methodology in this study.

A. Characterization of DDoS Attacks in IoMT

DDoS attacks on cached data in IoMT networks pose significant challenges for military operations, necessitating robust cybersecurity measures, proactive threat detection, and rapid response capabilities to ensure network resilience. Understanding various DDoS attack types, methods, targets, and impacts is crucial for developing effective detection and mitigation strategies. Attack types such as UDP flood and TCP SYN flood exploit vulnerabilities to inundate systems with traffic, categorized by target communication protocols [15]. Given the evolving nature of DDoS attack trends, continuous monitoring is necessary to adapt defence mechanisms, develop proactive detection strategies, and effectively assuage the effect of DDoS attacks on networks and systems.

B. Machine Learning for DDoS Detection in IoMT Data Caching

The machine learning architecture for IoMT leverages traffic data from actual IoT devices, as detailed in [17]. These models analyze network flow traffic to detect attack patterns and trigger countermeasures effectively. Utilizing data mining techniques to identify DDoS traffic in IoMT environments, popular machine learning methods such as Extreme Gradient Boosting (XGBoost), Decision Trees, and Random Forests were employed. This study's system architecture involves extracting IoMT traffic data and measuring the variable correlation utilizing the Pearson Correlation Coefficient (PCC). It quantified the variable's direct linkage to one another with a value range of -1 to 1 for 1, denoting an ideal positive linear relation, no association at all by a value of 0 and a perfect negative relationship by -1. Equation 1 represents the PCC formula, with r as the PCC, ζi and εi are independent data locus for X and Y factors, and $\bar{\zeta}$ and $\bar{\varepsilon}$ indicating the means of factors ζ and ε, respectively.

$$r = \frac{\sum(\zeta i - \bar{\zeta})(\varepsilon i - \bar{\varepsilon})}{\sqrt{\sum(\zeta i - \bar{\zeta})^2 \sum(\varepsilon i - \bar{\varepsilon})^2}}. \tag{1}$$

Numerical features are scaled to a fixed range between 0 and 1 using Min-Max scaling, as shown in Equation 2 to prevent features with large dimensions from dominating the model training process.

$$\varphi_{scaled} = \frac{\varphi - \varphi_{min}}{\varphi_{max} - \varphi_{min}}. \tag{2}$$

where φ is the initial feature value, φ_{min} is the minimal feature value, φ_{max} is the maximum value of the feature, and φ_{scaled} is the scaled feature value between 0 and 1.

IV. Performance Evaluation

A. Simulation Scenario and Dataset description

The CICDDoS2019 dataset [17] [1], a well-known dataset in cybersecurity research centred on DDoS attacks, is used in this work to address the shortcomings of previously available DDoS datasets. A new detection and classification method

[1] http://www.unb.ca/cic/datasets/CICDDoS2019

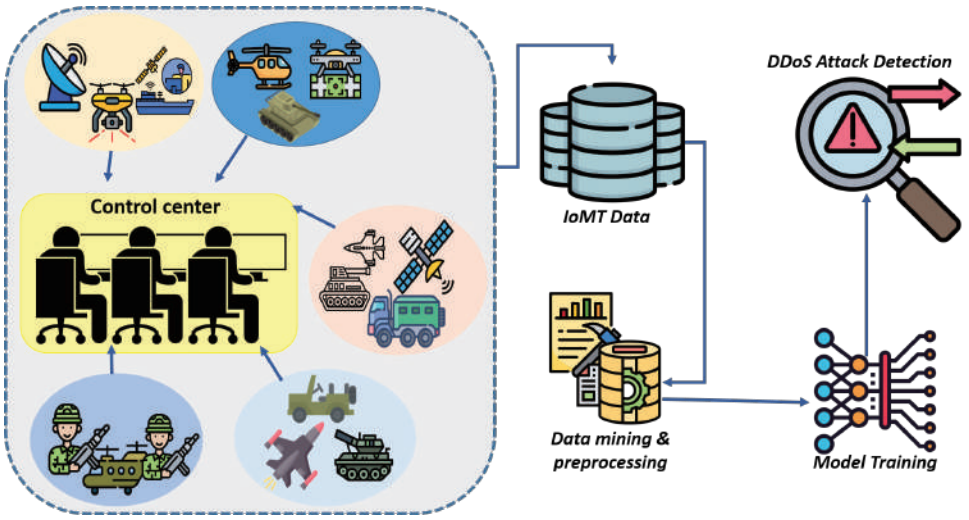
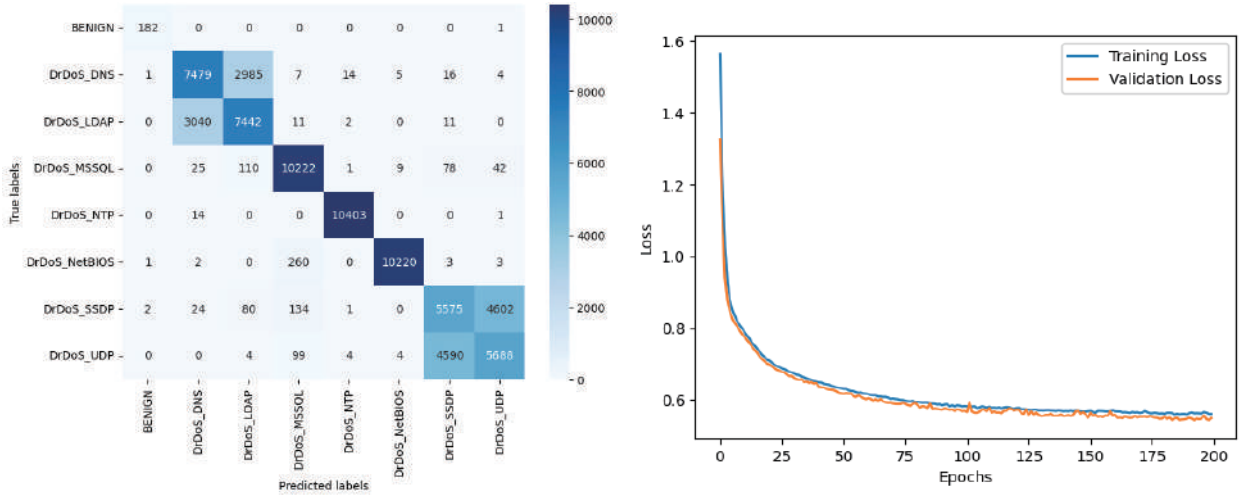


Fig. 1: System Process flow of the DDoS attack detection in IoMT networks data caching.



(a) Confusion matrix of the multiclass DDoS attack detection.

(b) Learning process path of the DDoS attack detection.

Fig. 2: Demonstration of the achievement of the evaluated machine learning models.

based on network flow is made possible by the created dataset, which comprises DDoS attacks of NTP, DNS, LDAP, MSSQL, NetBIOS, SNMP, SSDP, and UDP. With their associated weights, the dataset offers the most crucial feature sets for identifying various DDoS attacks. The dataset was divided into three sets: validation (15%), testing (25%), and training (60%). The choice of the dataset is due to its significance in analyzing new attacks executed using UDP/TCP-based rules at the application layer. The simulation environment was Python on Google Colaboratory, running on a system with an 11th Gen Intel(R) Core(TM) i7-11700K @ 3.60GHz and 64GB RAM, operating on Windows 10 Pro.

B. Machine learning Model Performance and Evaluation Metrics

This study performed methodological experiments to assess classifier effectiveness in detecting DDoS attacks. Although

TABLE I: Comparison performance of evaluated machine learning models for DDoS attack detection in IoMT data caching

Models	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
Random Forest	77.94	77.94	77.94	77.94
Logistic Regression	65.81	65.81	65.81	65.81
Decision Tree	75.51	75.51	75.51	75.51
LSTM	71.03	71.68	71.03	70.41
XGBoost	99.52	99.52	1.00	99.52

big data analytics and machine learning have bolstered IoT network security, IoMT data caching introduces diverse cyber risks due to distinct functionalities from traditional networks. Recent research underscores machine learning algorithms' efficacy in IoT attack detection [18], driving their investigation here. The assessed algorithms successfully classified DDoS attacks (see Table I). However, improving multiclass classifi-

cation accuracy is crucial due to varied DDoS attack types. Table I highlights ensemble classifiers' suitability for multi-class DDoS scenarios. It significantly addressed the intricacies of multiclass DDoS detection due to collective intelligence, diversity, error correction capabilities, and stability across varying imbalanced classes or overlapping attack patterns. Fig. 2a is the confusion matrix showing its detection ability; similarly, Fig. 2b is the learning path performance, illustrating the error reduction rate in DDoS attack detection. This study suggests an ensemble mechanism for effective DDoS detection in critical heterogeneous IoT networks with the possibility of varying degrees of imbalance in attack scenarios.

V. Conclusions

This study focuses on the challenges of securing the Internet of Military Things, which requires specialized detection methods. Real-time detection is crucial for military applications to prevent severe consequences from DDoS attacks. Machine learning-based IDS can enhance security with intelligent, adaptive, and automated detection capabilities. The study evaluated machine learning approaches for detecting recent DDoS attacks, emphasising accuracy, precision, recall, and f-score metrics. The ensemble model demonstrated superior performance across these metrics, with the ability to handle the complexity of IoMT network data caching. Ensemble methods improve generalization and robustness against unseen attack variations by aggregating predictions across models. They also enhance overall accuracy, which is crucial for managing complex multiclass DDoS scenarios. The results show the effectiveness of ensemble techniques in detecting minor instances of DDoS attacks, which is vital for maintaining data availability and accessibility in mission-critical environments. Future research will explore federated learning for distributed detection and continuous model updates to counter evolving attack strategies.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (IITP-2024-2020-0-01612, 50%) and Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2018R1A6A1A03024003, 50%)

References

[1] L. Mishra, Vikash, and S. Varma, "Internet of Things for Military Applications," in *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2020, pp. 118–123. [Online]. Available: https://doi.org/10.23919/INDIACom49435.2020.9083730

[2] M. I. A. Zahed, I. Ahmad, D. Habibi, and Q. V. Phung, "Content Caching in Industrial IoT: Security and Energy Considerations," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 491–504, 2019.

[3] M. Al-Helalat, "INTERNET OF THINGS: Revolutionizing Military Operations and Techniques," *International Journal of Progressive Research in Engineering Management and Science (IJPREMS)*, vol. 03, pp. 214–231, 2023. [Online]. Available: https://doi.org/10.58257/IJPREMS31814

[4] F. T. Johnsen, Z. Zieliński, K. Wrona, N. Suri, C. Fuchs, M. Pradhan, J. Furtak, B. Vasilache, V. Pellegrini, M. Dyk, M. Marks, and M. Krzysztoń, "Application of IoT in Military Operations in a Smart City," in *2018 International Conference on Military Communications and Information Systems (ICMCIS)*, 2018, pp. 1–8. [Online]. Available: https://doi.org/10.1109/ICMCIS.2018.8398690

[5] S. Li, W. Sun, H. Zhang, and Y. Zhang, "Physical Layer Security for Edge Caching in 6G Networks," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6. [Online]. Available: https://doi.org/10.1109/GLOBECOM42002.2020.9322524

[6] M. Najafimehr, S. Zarifzadeh, and S. Mostafavi, "DDoS Attacks and Machine-Learning-Based Detection Methods: A Survey and Taxonomy," *Engineering Reports*, vol. 5, no. 12, p. e12697, 2023. [Online]. Available: https://doi.org/10.1002/eng2.12697

[7] S. S. Musa, M. Zennaro, M. Libsie, and E. Pietrosemoli, "Mobility-Aware Proactive Edge Caching Optimization Scheme in Information-Centric IoV Networks," *Sensors*, vol. 22, no. 4, p. 1387, 2022. [Online]. Available: https://doi.org/10.3390/s22041387

[8] T. Zhang, Y. Li, and C. P. Chen, "Edge Computing and its Role in Industrial Internet: Methodologies, Applications, and Future Directions," *Information Sciences*, vol. 557, pp. 34–65, 2021. [Online]. Available: https://doi.org/10.1016/j.procs.2022.09.017

[9] L. A. C. Ahakonye, G. C. Amaizu, C. I. Nwakanma, J. M. Lee, and D.-S. Kim, "Classification and Characterization of Encoded Traffic in SCADA Network using Hybrid Deep Learning Scheme," *Journal of Communications and Networks*, vol. 26, no. 1, pp. 65–79, 2024.

[10] Z. Zhao, Y. Shi, B. Diao, and B. Wu, "Optimal Data Caching and Forwarding in Industrial IoT with Diverse Connectivity," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2288–2296, 2018.

[11] G. George and S. M. Thampi, "A Graph-Based Security Framework for Securing Industrial IoT Networks From Vulnerability Exploitations," *IEEE Access*, vol. 6, pp. 43 586–43 601, 2018.

[12] R. Chaudhary, G. S. Aujla, S. Garg, N. Kumar, and J. J. P. C. Rodrigues, "SDN-Enabled Multi-Attribute-Based Secure Communication for Smart Grid in IIoT Environment," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2629–2640, 2018.

[13] J.-S. Fu, Y. Liu, H.-C. Chao, B. K. Bhargava, and Z.-J. Zhang, "Secure Data Storage and Searching for Industrial IoT by Integrating Fog Computing and Cloud Computing," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4519–4528, 2018.

[14] Q. Xu, Z. Su, Q. Zheng, M. Luo, and B. Dong, "Secure Content Delivery With Edge Nodes to Save Caching Resources for Mobile Users in Green Cities," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2550–2559, 2018.

[15] A. Aljuhani, "Machine Learning Approaches for Combating Distributed Denial of Service Attacks in Modern Networking Environments," *IEEE Access*, vol. 9, pp. 42 236–42 264, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3062909

[16] A. Seifousadati, S. Ghasemshirazi, and M. Fathian, "A Machine Learning Approach for DDoS Detection on IoT devices," *arXiv preprint arXiv:2110.14911*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2110.14911

[17] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy," in *2019 International Carnahan Conference on Security Technology (ICCST)*, 2019, pp. 1–8.

[18] L. A. C. Ahakonye, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Trees Bootstrap Aggregation for Detection and Characterization of IoT-SCADA Network Traffic," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 5217–5228, 2024.

Revolutionizing Naval Logistics through NFT and Hyperledger Besu Integration

Gifar Arif Haryadi*, Revin Naufal Alief†, Jae-Min Lee*, Dong-Seong Kim*
*Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
†NS Lab Co., Ltd, Gumi, South Korea
*{gieworld, ljmpaul, dskim}@kumoh.ac.kr
†revinnaufalalief@nslab.tech

***Abstract*—The integration of blockchain technology holds significant promise in addressing the complex challenges of naval logistics. Centralized systems often fall short in ensuring data integrity, privacy, and traceability. This paper introduces a blockchain-based framework leveraging Hyperledger Besu and Non-Fungible Tokens (NFTs) to enhance naval logistics operations through improved asset management and operational efficiency. The system design incorporates privacy groups and permissioning to secure sensitive military data. Performance testing indicates that the system can effectively handle high transaction rates, with throughput capabilities of up to 191 transactions per second, while maintaining a latency of 2.38 seconds under high-load scenarios. This framework offers a scalable and secure approach to managing naval logistics, ensuring compliance with the stringent demands of military environments and operations.**

***Index Terms*—blockchain, digital twin, naval logistics, non-fungible tokens, privacy groups**

I. Introduction

Blockchain technology has seen widespread implementation across various sectors, significantly enhancing operational efficiencies and security. For instance, in the industrial sector, research has focused on improving fault detection in industrial IoT systems by integrating federated learning with blockchain, thereby securing data privacy and trust in IIoT environments [1]. In healthcare, a novel NFT-Integrated E-Prescription Management model leverages blockchain to enhance prescription traceability and security, demonstrating the efficacy of blockchain in addressing healthcare challenges [2]. In the entertainment sector, blockchain has been used to protect copyright in decentralized music sharing, ensuring secure and transparent audio file management [3]. Similarly, in the logistics sector, efficient logistics support systems are crucial for naval combat operations, emphasizing the need for advanced technologies to optimize logistics processes and enhance combat readiness [4]. Building on these advancements, this paper proposes a blockchain-based framework tailored for naval logistics, integrating privacy groups, permissioning, and NFTs to enhance asset management and operational efficiency.

Naval logistics management is associated with challenges that reduce the efficiency and security of military operations. Key issues include privacy concerns, where sensitive information must be securely managed and made accessible only to authorized personnel to maintain confidentiality. The integrity of data within the supply chain is also crucial to prevent data tampering and ensure record authenticity. Additionally, the naval logistics sector is vulnerable to counterfeit products, which pose a risk of introducing unauthorized materials into the supply chain. Other persistent problems include insufficient logistics information and a lack of visibility over personnel and materials, which can lead to miscommunications and inadequate planning for port visits. These challenges underscore the pressing need for an enhanced security framework and more sophisticated management systems within naval logistics.

TABLE I
List of Acronyms Used in This Study

Acronym	Description
DLA	Defense Logistics Agency
FISC	Fleet Industrial Supply Centers
NAVICP	Navy Inventory Control Point
NAVSUP	Naval Supply Systems Command
OC	Operational Commands
OB	Operational Bases

The proposed solution leverages a private blockchain network to address these issues effectively. This paper details several key contributions of our framework:

- **Private Blockchain Network**: Hyperledger Besu creates a secure, decentralized platform that greatly improves privacy and data integrity.
- **Privacy Groups and Permissioning**: These features ensure that only authorized personnel can access sensitive information, thereby enhancing data security and confidentiality.
- **Integration of NFT Technology**: Non-Fungible Tokens enhance the traceability and verification of assets and are linked with digital twin technology for real-time monitoring.
- **Evaluation Using Benchmark Tools**: Performance is assessed using Hyperledger Caliper, which checks if the blockchain meets the operational standards for naval logistics.

This approach is aimed at transforming naval logistics by providing a more secure, transparent, and efficient man-

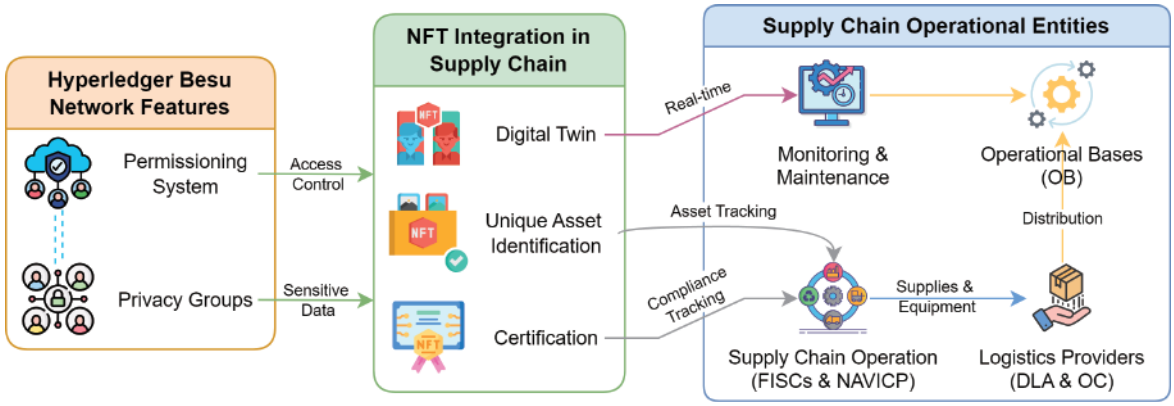


Fig. 1. System Architecture.

agement system. Table I provides a list of acronyms used throughout this paper, detailing various entities and terms relevant to naval logistics and blockchain implementation.

II. Related Works

The integration of blockchain into naval logistics is informed by several pivotal studies. A noteworthy framework from the MBA Professional Report at the Naval Postgraduate School suggests a Web-Based Logistics Information System designed to centralize naval supply chain management and enhance overall logistics effectiveness by streamlining processes and improving knowledge management within naval operations [5]. While this model offers substantial improvements in logistics planning, it remains inherently centralized, which could pose limitations in adaptability and resilience.

Building on military logistics, another significant study introduces a military blockchain model aimed at enhancing the traceability and authenticity of spare parts within the military supply chain [6]. This model improves transparency and performance while reducing risks associated with counterfeit products. However, it lacks specific mechanisms to address privacy concerns and does not incorporate features to manage ownership, which are crucial for maintaining operational secrecy and security.

The role of Non-Fungible Tokens (NFTs) in supply chain management has also been explored, particularly in the healthcare sector during the COVID-19 pandemic [7]. The proposed NFT-based system enhances the traceability and ownership of healthcare products, ensuring data integrity and facilitating the secure transfer of product information. Despite these advancements, the implementation faces challenges such as privacy concerns and integration with existing systems.

Furthermore, the application of NFTs to manage digital twins in a decentralized environment is discussed [8]. This approach uses blockchain to secure digital twins and their physical counterparts, ensuring accurate verification and lifecycle management. Although promising for asset management, this study does not address privacy concerns in depth, nor does it cater to the unique operational scenarios found in naval logistics.

These foundational works contribute significantly to the proposed framework by highlighting the potential of blockchain to transform naval logistics through enhanced security, efficiency, and traceability. The current proposal seeks to address the limitations identified in these studies by employing a private blockchain network enhanced with privacy groups, permissioning, and the strategic use of NFTs tailored to the naval context.

III. System Design

A. System Architecture

The proposed architecture for enhancing naval logistics integrates Hyperledger Besu and NFTs across multiple operational dimensions. The architecture features several key components. First, the NFT integration in the supply chain includes unique asset identification NFTs, certification NFTs, and digital twin NFTs for tracking, compliance, and management of physical assets. These NFTs enhance operational efficiency by providing precise asset tracking, compliance monitoring, and real-time asset management. Core blockchain functionalities such as a permissioning system and privacy groups are incorporated to secure interactions with NFTs and manage sensitive data. The permissioning system controls access, while privacy groups ensure that sensitive data remains protected.

Secondly, the supply chain operational entities encompass supply chain operations, logistics providers, and operational bases. NFTs are utilized for asset tracking and real-time monitoring, facilitating the efficient distribution of supplies and equipment. The permissioning system and privacy groups ensure that all interactions with NFTs are secure and compliant, supporting critical operations from asset tracking to maintenance at operational bases. This structured approach not only boosts security but also enhances the overall efficiency and traceability of naval logistics management.

B. Blockchain Private Network Design

Hyperledger Besu is used for the proposed naval logistics system due to its robustness in handling private network and support various enterprise-grade consensus mechanism making it suitable for the private network environment. For

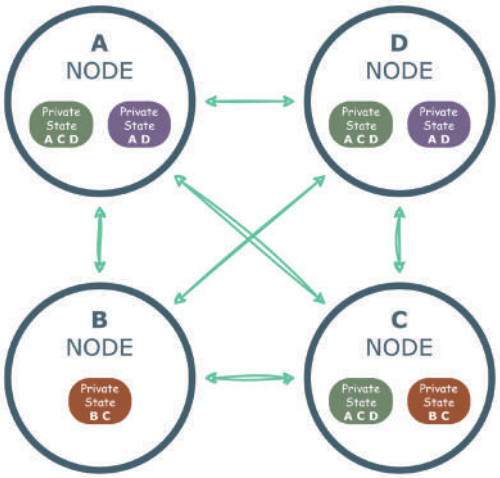


Fig. 2. Private transaction storage within Enterprise Ethereum Alliance (EEA) using privacy groups.

```
{
  "config": {
    "chainId": 1338,
    "contractSizeLimit": 2147483647,
    "qbft": {
      "blockperiodseconds": 2,
      "epochlength": 30000,
      "requesttimeoutseconds": 4
    }
  },
  "gasLimit": "0x1fffffffffffff",
  "alloc": {
    "fe3b557e8fb62b89f4916b721be55ceb828dbd73": {
      "balance": "0xad78ebc5ac6200000"
    }
  }
}
```

Fig. 3. Genesis Block Configurationn.

this system, QBFT (Quorum Byzantine Fault Tolerance) is chosen as the consensus mechanism based on its performance advantages over IBFT 2.0 and Clique, as mentioned in [9]. QBFT is particularly suited for environments where forkless consistency is critical, ensuring reliable transaction execution without the risks associated with potential forks. This choice underscores the system's focus on maintaining a secure, consistent, and private transactional environment in naval logistics.

C. Privacy Group and Permissioning

In Hyperledger Besu, privacy is ensured through private transaction management and encryption using Tessera. Each node involved in a private transaction communicates with its associated Tessera node, which manages encryption and distribution [9]. This ensures that only the designated participants can see the transactions, keeping them confidential from other network members.

Privacy groups in Besu are defined through unique identifiers that organize nodes into distinct groups managed by Tessera. This configuration supports the separation of transaction data into a public world state and private states specific to each group, as outlined in the Figure 2. Interactions within these privacy groups are tightly controlled, contracts can interact with other contracts within the same group or read from the public state but cannot access contracts outside their group. In addition, permissioning mechanisms is used to implement level of control in the network so that only explicitly authorized nodes and accounts can access certain network functions and data. This mecahnism is crucial for maintaining the security and integrity of operations within the naval logistics, allowing for secure and controlled access to the blockchain resources [10].

D. Non-Fungible Token and Digital Twin

Integrating ERC-721 NFTs with the Hyperledger Besu platform significantly improves asset management for naval logistics by providing unique identification and secure ownership of assets such as ships and equipment. This integration ensures precise lifecycle tracking and secure data handling within a private blockchain environment [8].

Furthermore, NFTs facilitate verifiable proof of delivery (PoD), ensuring that each transaction related to the delivery of naval assets is immutably recorded on the blockchain. This reduces disputes and enhances accountability. Smart contracts automate essential processes, maintaining high security and integrity in naval operations such as asset transfers, maintenance schedules, and governance logic.

Additionally, Digital Twins (DTs) are leveraged alongside NFTs to create virtual replicas of physical assets. Each digital twin is linked to a unique NFT that encapsulates detailed information about the physical asset, providing a secure and traceable connection between the digital and physical realms. This synergy between DTs and NFTs enables effective monitoring, control, and analysis of assets, further enhancing the security, privacy, and efficiency of the naval supply chain.

IV. System Evaluation

A. Simulation Environment

The simulation environment for evaluating the proposed blockchain system comprises an Intel Core i9-10980XE CPU 3.00GHz, 256 GB RAM, Ubuntu 20.04.6 LTS operating system, and a Samsung SSD 870. The system operates on Hyperledger Besu v24.3.0, with JDK 21 and Hyperledger Caliper v0.6.0 used for benchmarking. A total of 12 nodes were set up, with 4 designated as validators running the QBFT consensus algorithm.

The genesis block configuration for the system as shown in the Figure 3 includes key parameters such as the $chainId$, $contractSizeLimit$, and specific settings for the QBFT protocol like $blockperiodseconds$, $epochlength$, and $requesttimeoutseconds$. The gasLimit is set exceptionally high to facilitate a large number of transactions without limitations, and the $--min-gas-price = 0$ command line is used to create a free gas network, eliminating transaction costs within this private blockchain setup.

The system was tested with 1,000 transactions at different send rates to assess average latency and throughput, as shown in the Table II, using Hyperledger Caliper [11]. The results

TABLE II
PERFORMANCE OF THE PROPOSED SYSTEM UNDER TEST (SUT) WITH 1000 TRANSACTIONS AT VARIED RATE CONTROL

Send Rate (tx/s)	Average Latency					Throughput (tx/s)				
	Open (s)	Query (ms)	Transfer (s)	Mint NFT (s)	Transfer NFT (s)	Open	Query	Transfer	Mint NFT	Transfer NFT
50	1.66	1.60	1.81	1.83	1.86	43	50	46	48	48
100	1.79	1.44	1.91	1.78	1.84	88	100	73	88	91
150	2.30	1.41	1.92	2.31	1.69	132	150	108	133	94
200	2.74	1.31	2.26	2.14	3.74	145	200	175	128	140
250	2.84	1.30	3.73	3.14	2.38	151	250	145	175	191

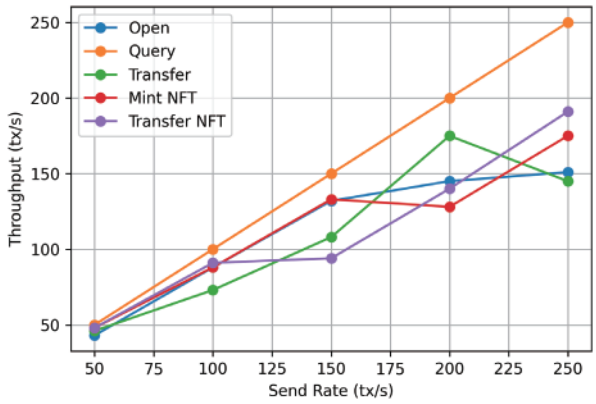


Fig. 4. Throughput Performance of The Proposed Network.

indicate a correlation between increased transaction send rates and changes in latency and throughput. For example, as the send rate increases from 50 tx/s to 250 tx/s, latency generally increases while throughput shows a varied response, suggesting the system's ability to handle higher loads with some efficiency fluctuations. Figure 4 further illustrate throughput performance, showcasing the system's effectiveness in managing secure and high-load transactions within a controlled military setting.

V. CONCLUSION

This paper has presented a robust blockchain-based framework for naval logistics, utilizing Hyperledger Besu and Non-Fungible Tokens (NFTs) to enhance asset management and operational efficiency. The system's architecture, incorporating privacy groups and permissioning features, ensures secure and private transactions within the naval environment. Performance evaluations using Hyperledger Caliper have demonstrated the system's capability to handle varying transaction rates effectively while maintaining data integrity and security. The integration of digital twins with NFTs further augments the traceability and real-time monitoring of naval assets. This framework not only meets the stringent requirements of military logistics but also offers scalable solutions adaptable to future technological advancements and operational needs. Future research will focus on testing the system's scalability in greater detail by utilizing a higher number of transactions.

ACKNOWLEDGEMENT

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government (MSIT) (IITP-2024-2020-0-01612, 50%) and by Priority Research Centers Program through the NRF funded by the MEST (2018R1A6A1A03024003, 50%).

REFERENCES

[1] M. A. P. Putra, R. N. Alief, S. M. Rachmawati, G. A. Sampedro, D.-S. Kim, and J.-M. Lee, "Proof-of-authority-based secure and efficient aggregation with differential privacy for federated learning in industrial IoT," *Internet of Things*, vol. 25, p. 101107, 2024.

[2] G. A. Haryadi, Allwinnaldo, J. M. Lee, and D.-S. Kim, "Enhanced Healthcare System with NFT-Prescription," in *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*, 2023, pp. 942–946.

[3] M. R. R. Ansori, Allwinnaldo, R. N. Alief, I. S. Igboanusi, J. M. Lee, and D.-S. Kim, "HADES: Hash-Based Audio Copy Detection System for Copyright Protection in Decentralized Music Sharing," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 2845–2853, 2023.

[4] D. Schrady and D. Wadsworth, "Naval Combat Logistics Support System," *Journal of the Operational Research Society*, vol. 42, no. 11, pp. 941–948, November 1991.

[5] Y. Li, "Web-Based Naval Fleet Logistics Information System," MBA Professional Report, Naval Postgraduate School, Monterey, California, Dec. 2007, available: Naval Postgraduate School Library.

[6] S. B. Rahayu, N. Jusoh, M. H. M. Halip, S. M. Taib, and M. G. Lee, "A Conceptual Model of Military Blockchain for Repair Parts Supply Chain Management," in *2021 International Conference on Computer Information Sciences (ICCOINS)*, 2021, pp. 146–150.

[7] A. Musamih, I. Yaqoob, K. Salah, R. Jayaraman, M. Omar, and S. Ellahham, "Using NFTs for Product Management, Digital Certification, Trading, and Delivery in the Healthcare Supply Chain," *IEEE Transactions on Engineering Management*, vol. 71, pp. 4480–4501, 2024.

[8] H. R. Hasan, M. Madine, I. Yaqoob, K. Salah, R. Jayaraman, and D. Boscovic, "Using NFTs for ownership management of digital twins and for proof of delivery of their physical assets," *Future Generation Computer Systems*, vol. 146, pp. 1–17, 2023.

[9] R.-V. Tkachuk, D. Ilie, R. Robert, V. Kebande, and K. Tutschku, "On the Performance of Consensus Mechanisms in Privacy-Enabled Decentralized Peer-to-Peer Renewable Energy Marketplace," in *2023 26th Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, 2023, pp. 179–186.

[10] M. Uddin, K. Salah, R. Jayaraman, S. Pesic, and S. Ellahham, "Blockchain for drug traceability: Architectures and open challenges," *Health Informatics Journal*, vol. 27, no. 2, p. 14604582211011228, 2021.

[11] W. Choi and J. W.-K. Hong, "Performance Evaluation of Ethereum Private and Testnet Networks Using Hyperledger Caliper," in *2021 22nd Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 2021, pp. 325–329.

Deterministic Bandwidth Allocation in Passive Optical Network with a Preemptive Approach

A.M.Zin[1,2], S.M Idrus[1], A.Ramli[1,2], F.M.Atan[1,2], N.A. Ismail[1,2]
[1]*Lightwave Communication Research Group (LCRG), Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*
[2]*Electrical Engineering Studies, Universiti Teknologi MARA Cawangan Johor, 81750 Masai, Pasir Gudang, Johor, Malaysia.*
sevia@utm.my, affida0253@uitm.edu.my

***Abstract*— Automation has become a prevalent change for many industrial systems nowadays. With increased real-time Human-to-Machine (H2M) and Machine-to-Machine (M2M) communications, stringent networking performance in terms of low latency and jitter are required. Hence, advanced optical access technologies must be capable of supporting the bursty development and connection of IoT sensors, microchips, data centers, and cloud technologies. The rethinking of newest networking policy has proposed the deterministic Passive Optical Network (PON) with deterministic performance. For this goal, we propose a deterministic bandwidth allocation in PON with a preemptive approach. We model, validate and analyze the latency performance of each transmission container (T-CONT) class with the use of M/G/1 preemptive priority queue in an Optical Network Unit (ONU). The results indicate that preemption guarantees the transmission of high-priority traffic while achieving network latency requirements.**

Keywords—Preemptive bandwidth allocation, deterministic PON, low-latency PON, M/G/1 Priority queue

I. INTRODUCTION

The Passive Optical Network (PON) has matured in providing the gigabit connections to customers for years at low cost of deployment. The legacy Fiber-to-the-x technology, where x stands for home, curb, building, and node has been enabled by the development of low-attenuation and high-capacity fiber optic links. With emerging data applications, fixed optical access networks are also becoming the predominant fronthaul technology for 5G mobile access networks, carrying traffic between the central office and remote antenna sites. Hence, optical transport is highly responsible for the capacity and quality-of-service (QoS) requirements of the 5G services. The advancement of todays' networking services requires stringent QoS standards. Time-critical provisions require low latency, and some, such as Internet-of-Things (IoT) and Tactile Internet require low peak-to-peak (p2p) latency variation [1].

Current dynamic bandwidth allocation (DBA) schemes in PON allocates the bandwidth according to the traffic classifications such as Transmission Container (T-CONT) in ITU-based PON and Service Differentiation in IEEE-PON. These allocations are governed by the Optical Line Terminator (OLT) at the PON Central Office (CO) which based on requests by each Optical Network Unit (ONU). In practice, the upstream bandwidth efficiency and QoS specifications, are all decided by the network provider as per users subscriptions. However, industrial networking demand are more critical with many real-time data services. In industrial automation for example; printing, tooling, and packaging machines have control cycles on the order of only milliseconds [2]. With large processing overheads, and algorithm complexity, the current Dynamic Bandwidth Allocation (DBA) scheme in PON need upgrades to transport those low-latency services within PON network.

The Time-Sensitive Networking (TSN) working group has proposed the frame preemption method which allows high-priority traffic preempt the lower priority traffic to utilize the service [3]. Preemptive scheduling has been long practiced in computing devices (CPU, server, router, switch etc.) to prioritize more importance tasks [4]. There is now increasing effort to adopt this mechanism into networking to allow or provide service for higher priority traffic at relatively low implementation and management complexity. A considerable number of efforts are presented in Ethernet-based networks. An increasing number of works are also noted in optical networks. However, more analytical and validation works are needed at the MAC layer (layer 2 in PON) especially related to bandwidth allocation.

In this paper, we present a deterministic bandwidth allocation scheme in PON using a preemptive approach. Section II reviews the literature related to the concepts and architectures of a deterministic PON. Section III presents the proposed analytical works that models T-CONT queues with the preemptive M/G/1 priority queueing system. We limit our work to an ITU-based PON. Simulation work to validate the proposed model is presented in Section IV. Concluding results and remarks are emphasized in Section V.

II. RELATED WORKS

A. Industrial Requirements

Latency is a metric used to determine the time duration of a data transmission. In this IoT era, the Ultra-Reliable Low-Latency Communications (URLLC) class also raises stringent requirements on the completion time of mission-critical applications [5]. For example, real-time control features in factory automation (e.g., assembling, packaging, and printing machines) may need a latency requirement of 250µs. Smart grid IoT applications also need latency requirements in the range of 3−20ms in order to achieve the real-time condition monitoring function. A latency range of 10−100ms is required for autonomous driving in the Intelligent Transport System (ITS), which provides traffic information and real-time coordination of actions. Moreover, a suitable latency for professional audio applications (e.g., live concerts) is 2ms since the latency between sound generation and the tonal perception exceeding 4ms will disturb the audio effect.

Jitter is the time deviation of transmission latency. This jitter occurs due to various reasons in the transmission network, which varies the frames' latency. In 5G architecture, the traffic between the radio unit and distributed unit needs to

be transported over a fronthaul interface with a low jitter requirements. In "isochronous real-time" systems, both latency and jitter must be tightly controlled to maintain the required performance and reliability. It is reported to require millisecond range of latency and a maximum of 1μs of jitter [2].

B. Time Sensitive Networking

TSN has been developed by IEEE 802.1 standards to provide deterministic and reliable data transmission over Ethernet networks. TSN has several mechanisms in ensuring or improving the real-time transmission of Ethernet traffic and time-triggered message delivery in switched Ethernet networks [3]. The key feature of the TSN standard is to allow flows with various requirements, such as ultra-low latency, low-jitter, and zero-loss reliability, to be controlled on an integrated bridged network. TSN and deterministic PON are interrelated in their goal to provide deterministic, low-latency, and reliable network performance. By combining TSN's capabilities in Ethernet networks with the high-bandwidth, low-latency benefits of deterministic PON, they enable end-to-end deterministic communication for a wide range of demanding applications. Hence, serving multiple time-critical applications in PON requires synchronization and co-scheduling, which typically requires their convergence over, TSN-IEEE 802.1Qbv-based network [3].

C. Deterministic PON Concept

A time-division-multiplexed PON (TDM-PON) that operates like multiple point-to-point (P2P) links can accommodate applications with strict timing requirements, while preserving the cost advantages of the point-to-multipoint (P2MP) topology. This necessitates the TDM-PON to deliver a deterministic performance: lossless transmission, low and constant (with low variations/jitter) latency to selected flows, while also serving other types of data such as best-effort traffic. In TDM-PON, all ONUs are virtually located at the same distance as a result of the ONU's individually assigned equalization delay. However, this puts challenges in properly designing the deterministic PON architecture because sometimes, asymmetric latency is added by the PON.

In a CPRI-over-XGS-PON proof of concept experiment reported in [6], the maximum fiber distance used is only 6 km due to additional PON protocol related delays. The LTE equipment in that experiment tolerated round trip delays of 200μs also been added. The work also identified possible use cases for future fronthaul or intra-DC networks. First, with modified passive optical splitters, and second by attaching the servers over a PON architecture to the intra-pod aggregation switches. This work also reported that, with 4 bursts per frame the minimum end-to-end (E2E) queuing delay is reduced to 31.25 μs. Hence, smart assignment of packet and frame sizes and synchronization at the CPRI, Ethernet and XGEM level also been proposed to avoid fragmentation, and reduces delays to a minimum [6].

The statistical multiplexing features in TDM-PON allows connections with multiple antenna sites, additional higher layer split, or with additional fixed line services on the same PON. Furthermore, the transport capacity at the lower layer split fronthaul interfaces can be dynamically varied with the amount of user data, contrary to the static traffic volumes in mobile access. However, the conventional upstream DBA mechanisms in TDM-PON are not capable of meeting ultra-low latency requirements of mobile access. With synchronized bandwidth assignment in both the wireless network and the PON, it can reduce additional delays to a minimum. Hence, deterministic bandwidth allocation is promising for aligning with available transport capacities and ensuring timely bandwidth assignment.

III. M/G/1 PREEMPTIVE PRIORITY QUEUE IN PON

The standard or traditional DBA process in any PON network has been modelled as a cyclic polling system for N users, where time slots for each ONU are particularly following the round robin arrangement. In each cycle, time slots are allocated to T-CONT1, T-CONT2, T-CONT3 and T-CONT4. Thus, an ITU-based PON system can be modelled as a polling system but with gated service, exponentially distributed frame interarrival time, and generic service time with the queue at both OLT and ONU are infinite. This is due to, in PON, time slots per T-CONT are varied according to the location (distance) of an ONU from OLT, and size of the granted bandwidth for each T-CONT in the ONU. Each time slot consists of two intervals, which are a reservation interval followed by the data interval. The reservation interval is used for transmitting control information of the system. However, in PON, the reservation interval only consists of guard time between frames because all the control messages between OLT and ONUs are integrated in the frame bursts.

With a non-preemptive priority queue in PON, for an incoming or a new arrival of a particular T-CONT traffic frame which want to be served in the next scheduling cycle, it needs to wait the delay from other low-priority T-CONTs, W_k in their corresponding queue to finish transmission during previous scheduling cycle. This is where k represents the number of T-CONT class in PON. The waiting time for these low-priority T-CONTs to finish transmission increases the W_k for that incoming T-CONT frame.

An example of preemption scenario is shown in Fig. 1 where a new arrival of T-CONT1 frame preempts the T-CONT4 frame which is currently being served. After all T-CONT1 transmitted according to its allocated bandwidth, the balance of T-CONT4 frame is resumed for transmission. Then, the arrival of other higher priority T-CONTs practiced the same manner until achieving the Allocated Bytes, AB_{min} of the DBA for all respective T-CONTs. By referring to [7], with preemption and considering 4 T-CONT types in PON, their corresponding queueing delay are presented in Equ. 1 until Equ. 4.

$$\overline{W_1} = \frac{(1-\rho_1)\bar{X} + \overline{W_{F1}}}{(1-\rho_1)} \tag{1}$$

$$\overline{W_2} = \frac{(1-\rho_1-\rho_2)\bar{X} + \overline{W_{F2}}}{(1-\rho_1)(1-\rho_1-\rho_2)} \tag{2}$$

$$\overline{W_3} = \frac{(1-\rho_1-\rho_2-\rho_3)\bar{X} + \overline{W_{F3}}}{(1-\rho_1-\rho_2)(1-\rho_1-\rho_2-\rho_3)} \tag{3}$$

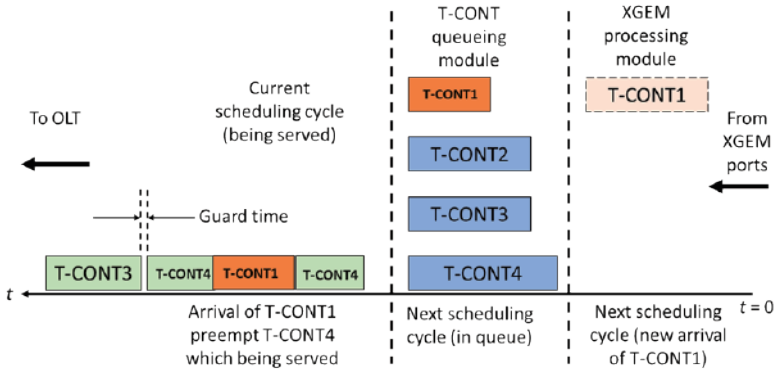


Fig. 1 Preemptive M/G/1 queue where new arrival of T-CONT1 frames preempt T-CONT4 frames which currently being served

$$\overline{W_4} = \frac{(1-\rho_1-\rho_2-\rho_3-\rho_4)\bar{X} + \overline{W_{F4}}}{(1-\rho_1-\rho_2-\rho_3)(1-\rho_1-\rho_2-\rho_3-\rho_4)} \quad (4)$$

Accordingly, the mean residual service time, $\overline{W_{Fk}}$ appearing in each $\overline{W_k}$ is derived as in Equ. 5 where the T_{AB_k} is the delay from the execution of the AB_{min} of the DBA for all T-CONTs.

$$\overline{W_{Fk}} = \frac{1}{2}\sum_{k=1}^{4} \lambda_k \overline{X_k^2} + T_{AB_k} \quad (5)$$

From the above $\overline{W}$ definitions, and by referring to work in [8] which has considered the whole requirements of a physical PON network, the mean end-to-end latency or delay, $\overline{D_{US}}$ for a traffic frame in PON network can be stated as follows:

$$\overline{D_{US_k}} = \overline{W_k} + \frac{SI_k}{2} + \frac{RTT}{2} \quad (6)$$

In the above equation, the RTT is the Round-Trip-Time for a T-CONT packet. $\frac{SI_k}{2}$ is the average service interval for a particular T-CONT type where it is the maximum delay in sending, the bandwidth grant to an ONU by a DBA scheme at the OLT. Therefore, a newly arrived T-CONT frame may get bandwidth grant immediately or in worse case may get it after a period of SI.

IV. VALIDATION WORKS

In order to test and perform the verification procedure for the preemptive bandwidth allocation, an OMNeT++ simulation model comprising an ONU and 4 users as sources of various T-CONT traffic generators is constructed. To achieve a comparable performance to the deterministic allocation model, this verification phase only involved the US direction. The results from this verification process are compared to its theoretical performance. Table I displays the configuration parameters for this procedure.

Average delay per frame, D_{avg} for each T-CONT class is shown in Fig. 2. The number of frames per T-CONT class which arrived at the OLT node is recorded. In overall, the

TABLE I. SIMULATION PARAMETERS

Parameters	Symbol	Values
Number of ONU	N	1
Number of users	n	4
ONU buffer size	β	2 Mbyte
Simulation time	T_{sim}	1 s
Network load	ρ_i	0.1 – 0.9
Service time	$\bar{X}$	125 µs
ONU to OLT line rate		200 Mbps
Fiber delay	D_L	20 µs
Frame sizes	FS	64, 1500, 500 Bytes
BW assignments for each T-CONTs		Follow values from [8]

average delay per frame, (D_{avg} for each T-CONT frame) is decreasing towards higher network load. This is due to an increasing number of frames generated at the User node. Thus, an increased number of served packets in each transmission window are recorded and a high number of T-CONT frame received at OLT. From Fig. 2, it is obviously seen that D_{avg} for T-CONT1 frame has the least and almost constant which is less than 1 ms. The first reason for this outcome is the proposed preemption allocation's ability to calculate the queue length and its ability to serve T-CONT1 immediately after the ONU received the BW grant. D_{avg} for T-CONT2 has the second highest values but with some fluctuations due to preemption that need to serve T- CONT1 first. D_{avg} for T-CONT3 and T-CONT4 are quite similar. D_{avg} for T-CONT3 decrease sharply at $\rho = 0.2$ and

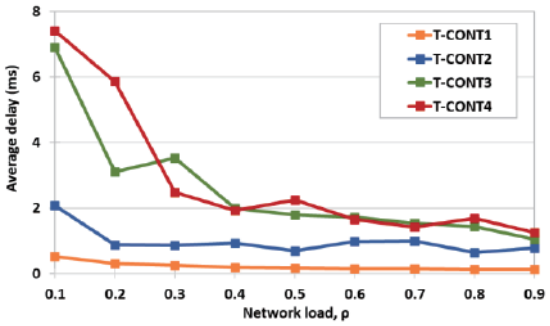


Fig.2 Comparison of average T-CONT delay, D_{avg} when testing the preemptive allocation in $T_{sim} = 1\ sec$

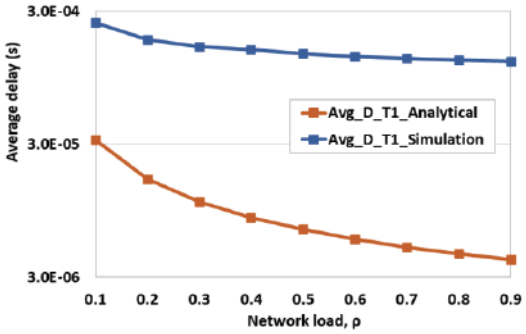


Fig.3 Average frame delay for T-CONT1

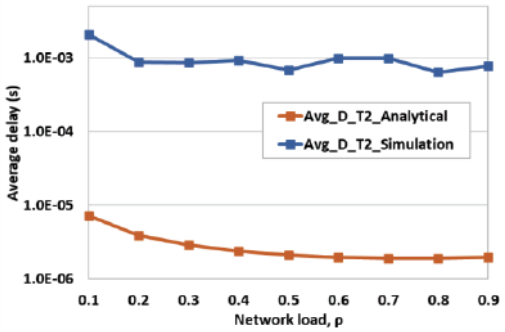


Fig.4 Average frame delay for T-CONT2

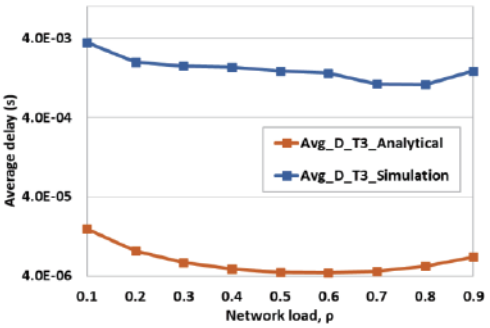


Fig.5 Average frame delay for T-CONT3

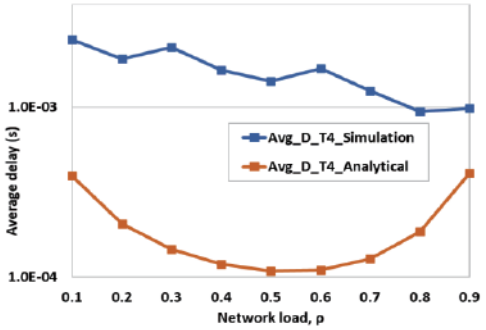


Fig.6 Average frame delay for T-CONT4

similar situation occurred at $\rho = 0.3$ for T-CONT4. This indicates that the number of frames for each T-CONT has significantly increased at the particular network load.

Fig. 3 until Fig. 6 shows the comparative analysis of D_{avg} for each T-CONT class between the analytical model and the simulated performance. Analytical D_{avg} is computed from division with the number of frames available for each T-CONT class for a particular network load. D_{avg} computed for T-CONT1 and T-CONT2 are decreasing towards higher network load. However, analytical D_{avg} for T-CONT3 and T-CONT4 are decreasing first but then increasing towards $\rho = 0.9$. At higher load, higher priority T-CONTs dominate the transmissions with guaranteed bandwidth. Hence, more T-CONT3 and T-CONT4 left in their respective queue at ONU, increased their queueing time and less of them arrived at the OLT node.

For all T-CONTs, the analytical average delay per frame is much lower than the simulation values. Higher D_{avg} are recorded from the simulation because of the object-oriented simulation involves many node processings such as creation of the traffic frame at the User, checking for each object ID, frame fragmentation, updating event status and etc. T-CONT2, T-CONT3 and T-CONT4 have a fluctuation trend of D_{avg} due to the random process of the simulation in generating the total amount of traffic for each class.

V. CONCLUSION

We successfully proposed and demonstrated a deterministic bandwidth allocation structure with an M/G/1 preemptive priority queueing system. We proposed its analytical latency model with the inclusion of traffic probabilities according to allocation bytes per T-CONT class. We also included other physical PON delay in the latency. We demonstrated results for T-CONT1 latency less than 1ms which is very promising for time-sensitive networking demands nowadays. The latencies for other T-CONTs are generally less than 5ms. Hence, these findings are beneficial towards the implementation work of deterministic PON for industrial automation with many real-time data services.

ACKNOWLEDGMENT

The authors would like to thank Universiti Teknologi Malaysia for providing the facilities of this project. Authors also thank Universiti Teknologi MARA for the financial support for this project.

REFERENCES

[1] L. Ruan, "Low-Latency Communication over Heterogeneous Fiber-Wireless Networks for Human-to-Machine Applications," no. September, pp. 0–2, 2019.

[2] K. Christodoulopoulos, S. Bidkar, W. Lautenschlaeger, T. Pfeiffer, and R. Bonk, "Deterministic passive optical networking," *J. Opt. Commun. Netw.*, vol. 16, no. 2, pp. 180–192, 2024.

[3] Y. Seol, D. Hyeon, J. Min, M. Kim, and J. Paek, "Timely Survey of Time-Sensitive Networking: Past and Future Directions," *IEEE Access*, vol. 9, pp. 142506–142527, 2021.

[4] B. High-priority, "Preemptive Scheduling Behavior," *IBM Spectr. LSF*, vol. 10.1.0, 2024.

[5] Z. Zhou, "Latency-Critical Networking," 2020.

[6] T. Pfeiffer, "Considerations on transport latency in passive optical networks," *IET Conf. Publ.*, vol. 2019, no. CP765, 2019.

[7] S. K. Bose, "Priority Operation of The M / G / 1 Queue," *Queue*, pp. 1–11, 2002.

[8] R. Aslam Butt, S. Mahdaliza Idrus, K. Naseer Qureshi, P. M. A. P. M. A. Shah, and N. Zulkifli, "An energy efficient cyclic sleep control framework for ITU PONs," *Opt. Switch. Netw.*, vol. 27, no. January 2017, pp. 7–17, 2018.

Performance Evaluation Transporting Multiservice RoF WDM System

R Mohamad
Telekom Research & Development (TMR&D), TM Innovation Centre, 63000 Cyberjaya, Malaysia

romli@tmrnd.com.my

SM Idrus*
Department of Communication Engineering, Faculty of Electrical Engineering Universiti Teknologi Malaysia (UTM), Johor, Malaysia
sevia@utm.my (corresponding)

NH Awang
Department of Communication Engineering, Faculty of Electrical Engineering Universiti Teknologi Malaysia (UTM), Johor, Malaysia

T Kanesan
Telekom Research & Development TM Innovation Centre 63000, Cyberjaya, Malaysia

S Yaakob
WIPNET, Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), Serdang, Selangor, Malaysia

Adam Wong Yoon Khang
Center for Telecommunication Research & Innovation, Fakulti Teknologi Dan Kejuruteraan Elektronik Dan Komputer, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia.

***Abstract*— Radio over Fiber (RoF) technology capable to carry multiples services as mobile backhaul technology. Experimental of multiservice RoF using Wavelength Division Multiplexing (WDM) system are achieved and performance evaluation was done. Two services were considered are Wireless Fidelity (WiFi) 802.11ac at 5.8GHz center frequency with 40MHz bandwidth, using 64QAM (2/3) data modulation and Long Term Evolution (LTE) Frequency Division Duplexing (FDD) at 875MHz carrier frequency with 20MHz bandwidth and 64QAM data modulation. Quality signal Error Vector Magnitude (EVM) over 10km SMF distances is evaluated for both services respectively. An optimum range of received optical power at Photodetector (PD) to ensure EVM value meet 3GPP and IEEE standard was recorded. The findings shows that transporting multiservice RoF using WDM system with 10km distance can be operated optimally at received optical power -11dBm to -17dBm to ensure EVM for both signals are meet 3GPP and IEEE standard requirement (below 8%). This technique is experimentally proven to solve offloading of WiFi and LTE signals.**

Keywords*— Multiservice Radio over Fiber Wavelength Division Multiplexing (RoF WDM) system, Long Term Evolution Frequency Division Duplex (LTE FDD), WiFi 802.11ac, Offloading WiFi and LTE, Error Vector Magnitude (EVM)

I. INTRODUCTION

Internet connectivity is now an essential part of modern life, indispensable no matter where people live. Everyone and everything is connected to the internet, simplifying both life and work. As we enter the era of the fourth industrial revolution (IR4.0), the world is rapidly embracing digitalization. To support this transformation, a fully connected network is crucial. In other words, internet connectivity must be available anytime and anywhere to sustain this new era of innovation.

The concept of a distributed antenna system (DAS) is particularly useful for meeting network requirements, especially in areas with signal blockages or dead spots and in lower ground buildings with poor signal penetration. Radio over fiber (RoF) technology is an effective solution for DAS, allowing signals to be distributed over long distances to remote antenna units.[1], [2], [3].

The RoF technology can offer large transmission capacity, transparent to bandwidth or modulation format (for multi-service application), allow small remote stations and its potential for centralized operation [4][5]. Another benefit of the RoF system is the simplification of remote antenna unit (RAU) resulting in small, light, easily-installed units. Furthermore, the switching center (SC) for remote access unit (RAU) can be centralized to avoid the complexity of network design providing more effective network management and cheaper operational cost.

The beauty of RoF technology is that the transparency of technology upgrade including high-frequency transformation. Technology upgrading such as LTE to 5G, Wireless Fidelity (WiFi) 802.11ac to WiFi 802.11ax, or upgrading frequency to higher can be done swiftly. Moreover, RoF technology is capable to carry multiple wireless services[6], [7], [8], [9], [10] as a mobile backhaul, also mixed between wireless and wireline services as mobile fronhaul[11], [12], [13]. Various methods have been studied and experimented to carry out multiservice RoF technology comprising deployment physical layer Wavelength Division Multiplexing (WDM) technique[14], [15], Subcarrier multiplexing (SCM) or Dual Sideband Optical Carrier Suppression technique[6], [9], [13],[16] and Polarization Multiplexing (PolMux) Technique [8].

Previous technical forums such as Broadband Forum and TM Forum reported that 5G transport is an enabler for IR4.0 [17][18]. This technology enables the development of internet-based applications and tools toward digital lifestyle. As predicted in Cisco Annual Internet Report, IP traffic will keeps increasing in double per year by 2023[19] [20][21], [22]. Again RoF is a promising solution with strategic planed bringing multi-service wireless platform Long Term Evolution (LTE) and Wireless Fidelity (WiFi) to serve the current need. In some cases, mobile coverage (LTE) tend to loss of signal from Evolved NodeB (eNB) due to blockage issue such underground building[9]. In that scenario, seamless offloading between LTE and WiFi technology services is suggested as possible solution.

In this paper, we demonstrated experimentally and performed performance evaluation of multiservice RoF using Wavelength Division Multiplexing (WDM) technique represented mentioned scenario. This experiment is part of full duplex transmission multiservice RoF system which deployed Polarization Multiplexing (PolMux) technique for downlink and WDM method for uplink. Part of that experiment, performance architecture based on WDM technique for uplink had been characterised. In this manuscript, we presented performance evaluation based to carry both LTE and WiFi signals using WDM technique.

II. Experimental Setup

An experimental setup of uplink architecture transporting multiservice RoF signal using WDM system is illustrated as in Figure 1. The multiservice signals comprise of WiFi 802.11ac and LTE Frequency Division Duplex (FDD) format. The WiFi 802.11ac signal source is generated from NI Universal Software Radio Peripheral (USRP) device at 5.8GHz central frequency with 40MHz bandwidth, using 64QAM (2/3) data modulation. Meanwhile, the Vector Signal Generator (VSG) generates LTE FDD signal at 875MHz carrier frequency with 20MHz bandwidth and 64QAM data modulation.

At the transmitter part, two laser sources with 1548.51nm and 1547.72nm central wavelength are connected to Mach Zenhder modulators 1(MZM 1) and MZM 2 separately. The optical signal from both lasers are modulated by WiFi 802.11ac and LTE FDD signals in dedicated in-line MZM 1 and MZM 2 respectively. The outputs of both MZMs are linked to the DWDM multiplexer via specific wavelength channel. In the DWDM Mux, modulated signals are combined into a single fiber and then propagated, simultaneously thru 10km Single Mode Fiber (SMF) cable to the DWDM demultiplexer at the receiver.

The DWDM demultiplexer used to split the signals with respective wavelength. The output optical signal at each DWDM Demux is amplified by post-amplifiers Erbium Doped Fiber Amplifier 1 (EDFA 1) and EDFA 2 to compensate the attenuation that occurs in the fiber before it will be detected at the receiver Photodetector 1 (PD1) and PD2 accordingly. PDs translates the optical signal into RF signal of WiFi 802.11ac at 5.8GHz and LTE FDD at 875MHz. These signals are then amplify using RF amplifier before fed into Agilent Performance Extreme Analyzer (PXA) Signal Analyzer for signal performance observation and measurement.

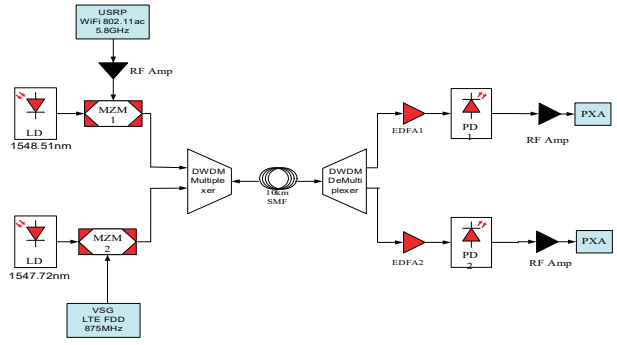
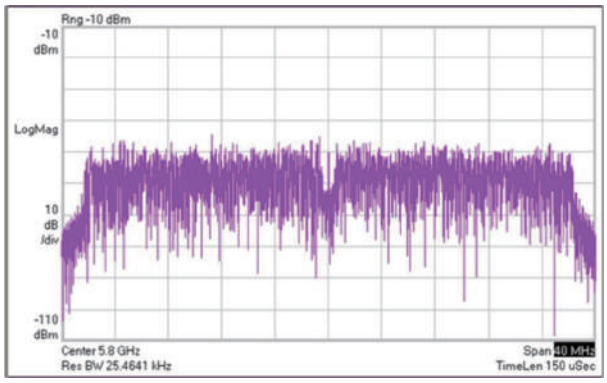


Figure 1. Multiservice RoF using WDM system experimental setup

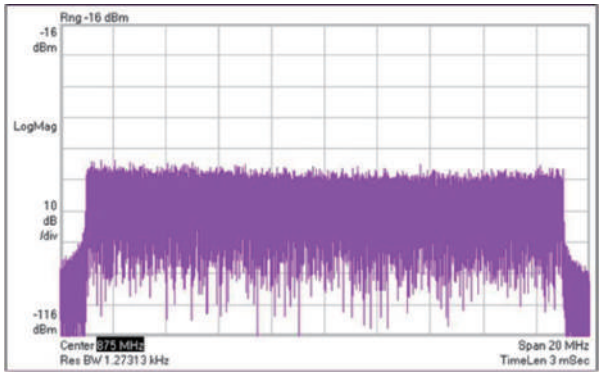
III. Result And Analysis

The proposed setup is evaluated by two different scenarios as a comparison, which are back-to-back (BTB) without 10km SMF and performance of the signals propagates after 10km SMF. In this work, the performance measurement of Error Vector Magnitude (EVM) is carried out.

Figure 2(a) and 2(b) depicted a spectrum characteristic of Radio Frequency (RF) spectrums from WiFi 802.11ac signal and LTE FDD signal sources as input into MZMs as illustrated in Figure 1. The WiFi 802.11ac spectrum at 5.8GHz center frequency with 40MHz bandwidth is shown in Figure 2(a) and spectrum for LTE FDD at 875MHz with 20MHz bandwidth indicated in Figure 2(b).



a) WiFi 802.11ac Bandwidth:40MHz 64 QAM (2/3) modulation @ 5.8GHz



b) LTE FDD Bandwidth:20MHz 64 QAM modulation @ 875MHz

Figure 2. RF spectra for WiFi (a) and LTE (b)

And Figure 3 presents a characterization measurement of EVM at MZM 1 and MZM 2 for WiFi and LTE carried out as a reference. RF input power that is controlled in the range of -8dBm to +7dBm. The best and optimum EVM achieve for WiFi is 1.4% at 0dBm RF input power**.** For LTE, -1dBm RF input power contributes to the finest EVM at 0.98%. The constellation points for both signals are also included in Figure 3. The constellation points for 64QAM data modulation can be seen clearly. For 64QAM signal measurement, this is well below 8% for both 3GPP [23]and IEEE standard requirements.

Measurement of EVM as a function of optical received power at the photodetector for BTB is recorded, with corresponding constellation points for WiFi and LTE signals

respectively. This is then followed with 10km SMF connections inserted in between MUX and DeMUX to determine signal quality over the optical fiber distance.

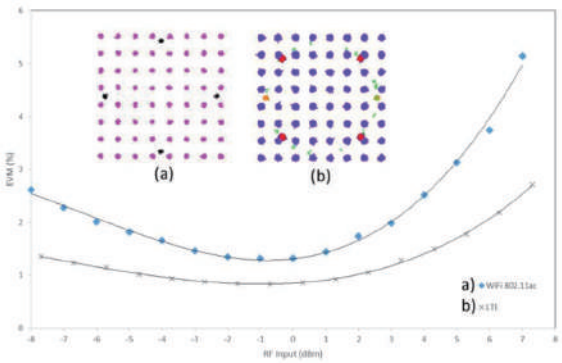


Figure 3. EVM vs. RF input power to MZM

Post EDFAs amplifier is used to compensate the received optical signal at PDs. Results in Figure 4 shows a comparison of EVM between BTB and after 10km propagation for both services. Optimum EVM value for BTB WiFi is 2.21% was recorded at -11dBm optical received power, meanwhile 1.36% was recorded at -11dBm received optical power for LTE.

The EVM quality of WiFi signal obviously degrades as received optical signal degraded. Maximum EVM for BTB WiFi was recorded is 6.8% at -18dBm received optical power, however, the trend is not similar for LTE signal i.e 2.56% at -18dBm received an optical signal.

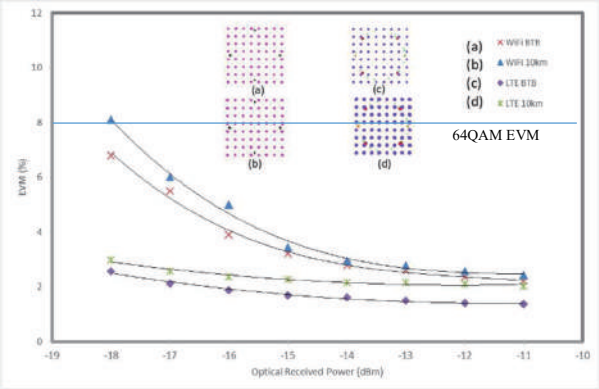


Figure 4. EVM vs. Optical received power

Consequently, the quality of signal over the distance at 10km is measured for the signals. Again, as shown in Figure 4, the optimum EVM for WiFi and LTE was achieved at -11dBm received optical power with 2.4% and 2.0% respectively. Due to low power received an optical signal, its effecting signal to noise ratio that contributes to the quality of a signal. Maximum EVM for WiFi is 8.1% recorded instead of 2.9% for LTE at -18dBm received optical power. This is important to ensure the optimum received optical power for the system performing below 8% of EVM to meet the standard 3GPP LTE[23] and IEEE WiFi for 64QAM transmission.

EVM performance between 10km and BTB are around 0.75% and 0.53% on average for WiFi and LTE, respectively. Table 1 shows the EVM variance between BTB and 10km distances for both services WiFi and LTE respectively. This indicates error free transmission of multiservice RoF using WDM technique.

Service	BTB EVM (%)	10km Distance EVM (%)	Variance (%)
WiFi 802.11ac (Min)	2.21	2.41	0.2
WiFi 802.11ac (Max)	6.8	8.1	1.3
LTE FDD (Min)	1.36	2.01	0.65
LTE FDD (Max)	2.56	2.98	0.42

Table 1: EVM variance for WiFi and LTE services over the distance

IV. Conclusion

In this paper, we demonstrate a performance evaluation of transporting multiservice (WiFi 802.11ac and LTE FDD) radio over fiber technique using WDM system to determine the optimum value of EVM and received optical power. The findings show that transporting multiservice RoF using WDM system with 10km distance can be operated optimally at received optical power -11dBm and up to maximum -17dBm to ensure EVM for both signals meet IEEE and 3GPP standard requirement of below 8%. The finding is importance for ensure of successful offloading of WiFi and LTE signals.

Acknowledgment

The author would like to acknowledge TM Research & Development and Telekom Malaysia for supporting and funding this work.

References

[1] C. Lim and A. Nirmalathas, "Radio-Over-Fiber Technology: Present and Future," *Journal of Lightwave Technology*, vol. 39, no. 4, pp. 881–888, Feb. 2021, doi: 10.1109/JLT.2020.3024916.

[2] B. Kaur and N. Sharma, "Radio over Fiber (RoF) for Future Generation Networks," in *Broadband Connectivity in 5G and Beyond*, Cham: Springer International Publishing, 2022, pp. 161–184. doi: 10.1007/978-3-031-06866-9_9.

[3] M. Kamoun, Sheng Yang, and M. De Courville, "Multi-RAU pilots for ROF enabled distributed antenna systems," in *2009 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology*, IEEE, May 2009, pp. 177–181. doi: 10.1109/WIRELESSVITAE.2009.5172446.

[4] H. Al-Raweshidy and S. Komaki, *Radio over fiber technologies for mobile communications networks*. London: Artech House, 2002.

[5] G. Chang, "Converging Broadband Access Networks : Enabling Technologies," in *ICCSC*, 2008.

[6] M. Zhu, L. Zhang, J. Wang, L. Cheng, C. Liu, and G. K. Chang, "Radio-over-fiber access architecture for integrated broadband wireless services," *Journal of Lightwave Technology*, vol. 31, no. 23, pp. 3614–3620, 2013, doi: 10.1109/JLT.2013.2286564.

[7] T. Su *et al.*, "Multiservice wireless transport over RoF link with colorless BS using PolM-to-IM convertor," *IEEE Photonics Technology Letters*, vol. 27, no. 4, pp. 403–406, 2015, doi: 10.1109/LPT.2014.2372032.

[8] M. H. M. Nor, T. Kanesan, F. Maskuriy, A. Yusof, F. A. Fatah, and S. B. A. Anas, "Experimental Realization of Multi-Service RoF system using OCS-PolMux Techniques," pp. 5–7.

[9] S. Yaakob, M. Z. A. Kadir, S. M. Idrus, N. M. Samsuri, R. Mohamad, and N. Emran Farid, "On the carrier generation and HD signal transmission using themillimeter-wave radio over fiber system," *Optik (Stuttg)*, vol. 124, no. 23, 2013, doi: 10.1016/j.ijleo.2013.04.096.

[10] T. H. Dahawi, Z. Yusoff, M. S. Salleh, and J. M. Senior, "Low-cost MIMO-RoF-PON architecture for next-generation integrated wired and wireless access networks," *Journal of Optical Communications and Networking*, vol. 13, no. 3, pp. 41–52, Mar. 2021, doi: 10.1364/JOCN.413596.

[11] W. Ji and J. Chang, "Design of WDM-RoF-PON for wireless and wire-line access with source-free ONUs," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 5, no. 2. pp. 127–133, 2013. doi: 10.1364/JOCN.5.000127.

[12] C. Liu, L. Zhang, M. Zhu, J. Wang, L. Cheng, and G. K. Chang, "A novel multi-service small-cell cloud radio access network for mobile backhaul and computing based on radio-over-fiber technologies," *Journal of Lightwave Technology*, vol. 31, no. 17, pp. 2869–2875, 2013, doi: 10.1109/JLT.2013.2274193.

[13] P. P. Monteiro *et al.*, "Mobile fronthaul RoF transceivers for C-RAN applications," *2015 17th International Conference on Transparent Optical Networks (ICTON)*. pp. 1–4, 2015. doi: 10.1109/ICTON.2015.7193452.

[14] W. Ji and Z. Kang, "Design of WDM RoF PON based on OFDM and optical heterodyne," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 5, no. 6. pp. 652–657, 2013. doi: 10.1364/JOCN.5.000652.

[15] W. Ji, X. Li, Z. Kang, and X. Xue, "Design of WDM-RoF-PON Based on Improved OFDM Mechanism and Optical Coherent Technology," vol. 7, no. 2, pp. 74–82, 2015.

[16] R. Mohamad, S. M. Idrus, A. S. Supaat, S. Yaakob, A. K. Zamzuri, and S. N. A. Sukito, "Millimeter wave carrier generation based on Brillouin fiber laser with improved tuning capability," *Optik (Stuttg)*, vol. 125, no. 1, pp. 205–207, 2014, doi: https://doi.org/10.1016/j.ijleo.2013.06.019.

[17] V. La Virtualization, "Beyond the hype – Broadband Forum Executive Workshop looks to the future of broadband," 2019.

[18] M. Sahni, "The Disruptive Impact of 5G-Enabled Business Services The road to 5G," pp. 1–17, 2018.

[19] CISCO, "White paper Cisco public: Annual Report, Internet," 2018.

[20] CISCO, "The Zettabyte Era: Trends and Analysis," *Cisco*, no. May 2015, pp. 1–29, 2015.

[21] T. J. Barnett, A. Sumits, S. Jain, and U. Andra, "Cisco Visual Networking Index (VNI) Update Global Mobile Data Traffic Forecast," *Vni*, pp. 2015–2020, 2015, doi: 1465272001663118.

[22] Fredrik Jejdling, "Ericsson Mobility Report, June 2018," no. June, 2018.

[23] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception (Release 13)," 2021, Accessed: Jul. 03, 2024. [Online]. Available: http://www.3gpp.org

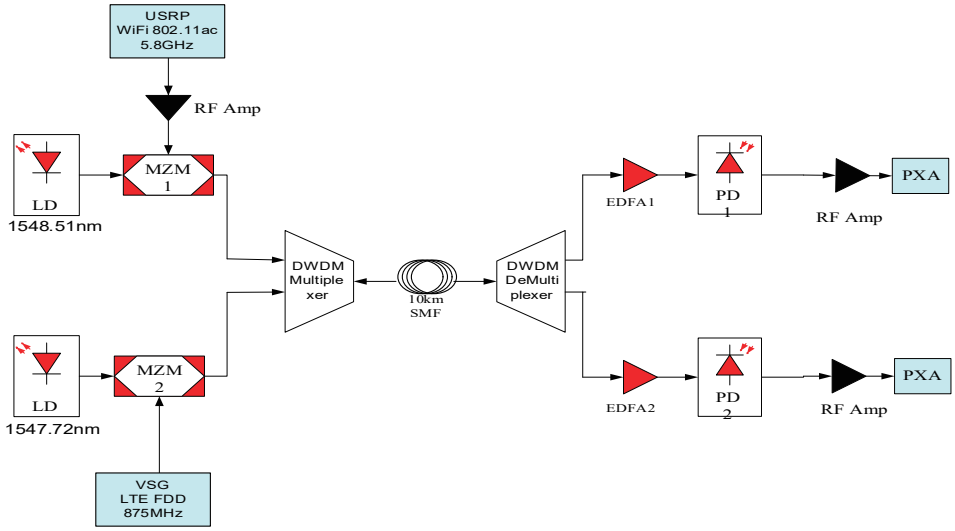
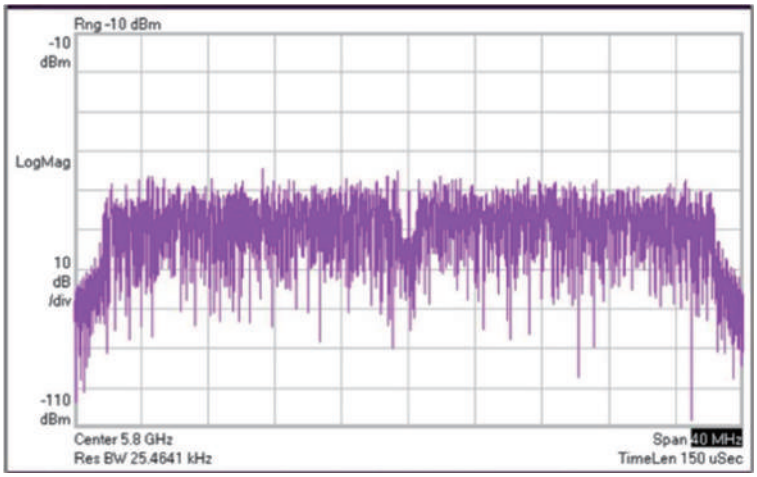
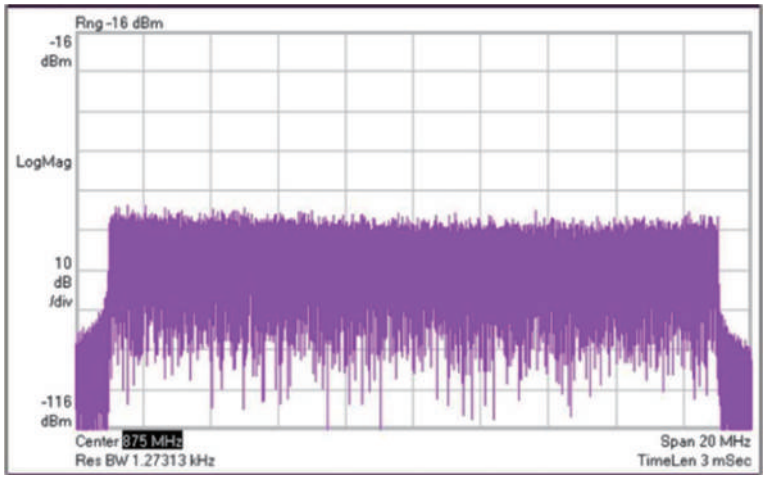


Figure 1. Multiservice RoF using WDM system experimental setup



a) WiFi 802.11ac Bandwidth:40MHz 64 QAM (2/3) modulation @ 5.8GHz



a) LTE FDD Bandwidth:20MHz 64 QAM modulation @ 875MHz

Figure 2. RF spectra for WiFi (a) and LTE (b)

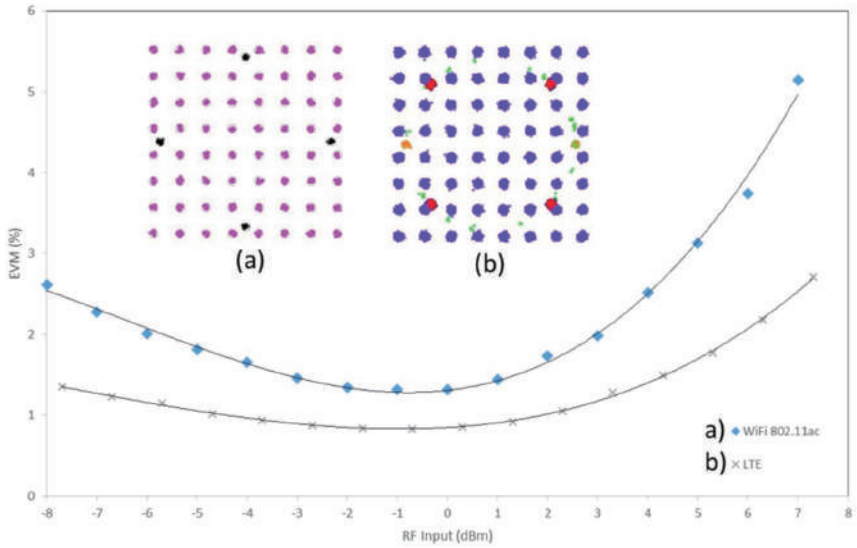


Figure 3. EVM vs. RF input power to MZM

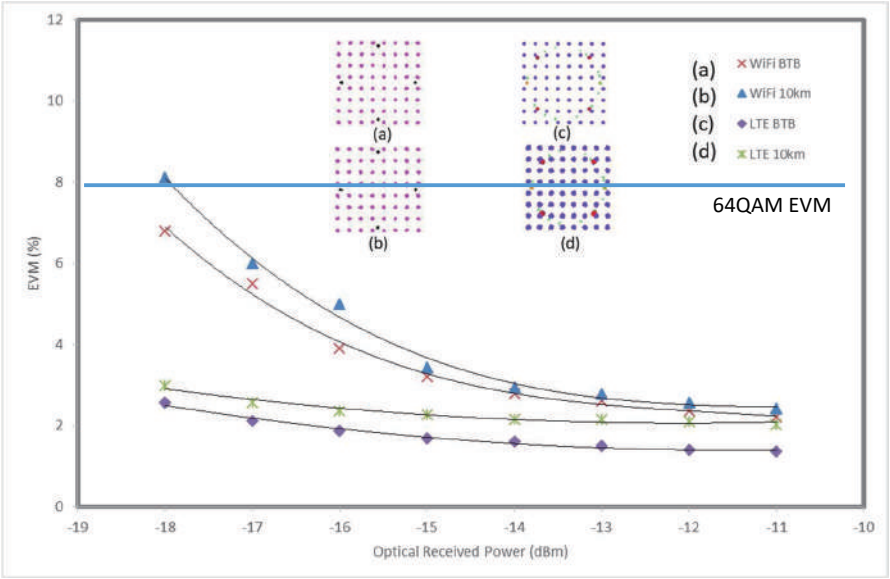


Figure 4. EVM vs. Optical received power

Research Trends of WDM QKD and Satellite QKD in Quantum Communication

Seungho Yoon
Department of Electrical and Electronic Engineering
Korea University
Seoul, Korea
seunghyoon@korea.ac.kr

Jun Heo
Department of Electrical and Electronic Engineering
Korea University
Seoul, Korea
junheo@korea.ac.kr

***Abstract*— In this paper, we show the latest trends in recent WDM-QKD and QKD-satellite research. Recently, the QKD technique is used alone by QKD technology, but it is also merged with the WDM technique for application with the actual communication network. In addition, research on merging QKD technology and satellite communication for long-distance communication is being conducted. This paper analyzes how the QKD technique is being studied in these WDM and satellite techniques.**

Keywords—Quantum key distribution(QKD), Satellite communication, Wavelength division multiplexing

I. INTRODUCTION

The quantum key distribution technique is a symmetric key distribution method that uses the physical properties of quantum mechanics to distribute the same key between the sender and the receiver. The quantum key distribution technique is a technique that allows the presence or absence of an eavesdropper to be checked by transmitting a single photon and allows the transmitter to share a secret key through a public channel. In general, when the sender(Alice) encrypts information to be transmitted to the signal, and transmits it to the receiver(Bob), the transmitter can confirm the presence of the eavesdropper if the eavesdropper(Eve) steals the encrypted information.

Quantum technology is classified into quantum communication, quantum computer, quantum sensor, etc. Quantum communication is emerging as a new paradigm technology in the field of information security, quantum computer is emerging as a nonlinear optimization problem solving, and quantum sensor is emerging as a new paradigm technology in the field of ultra-precise information detection. [1]

Quantum key distribution techniques go through a total of four steps. Quantum transmission, information reconciliation, privacy amplification, final key verification. To do this, two types of channels are used : a quantum channel and a public channel that assumes no loss or error. First, the transceiver divides the key using the quantum channel. As mentioned earlier, since it uses the physical properties of quantum mechanics, if the eavesdropper attempts to eavesdrop on the quantum channel, the transceiver can recognize it, ensuring security.[2]

The quantum bit error rate is calculated by disclosing some of the shared keys and checking that the quantum bit error rate does not exceed the threshold. If the threshold is not exceeded, confirm that there was no eavesdropper and starts the information reconstruction process to correct key errors caused by quantum channels. When the information reconciliation process is finished, the security amplification process is started to reduce the amount of information exposed to Eve in the process of reconciliation. Finally, the transmitter and the receiver check if both side shared the same key and whether the eavesdropper shared the key by disguising it as the sender or receiver.[3]

Wavelength division multiplexing(WDM) is an optical multiplexing method that transmits simultaneously through multiple wavelength bands to one optical fiber and input channel signals through different assigned channel wavelength bands. Since WDM allocates a different wavelength band for each input, it also divides the optical fiber low-loss wavelength band (especially 1550 nm) into several narrow channel wavelength bands. The WDM system process uses a multiplexer to combine and transmit signals, and the receiver's demultiplexer divides them and receives them. When a WDM system is used, if 100 kbps transmission and reception is possible with a specific single wavelength band, a higher transmission/reception speed can be achieved as well as 100 kbps if a 100 WDM system is used.

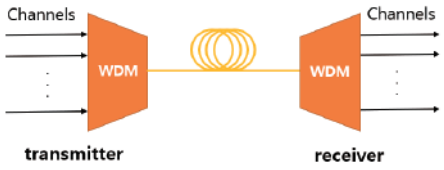


Fig. 1. basic concept of WDM

In order to implement free space communication technology, laser communication technology and PAT (Pointing, Acquisition, Tracking) technology are required. The laser communication technology is composed of an optical transmission and reception unit including a light source, a modulator, an optical system, an optical receiver, and an optical signal processing unit.

The PAT system can be configured using GPS (Global Positioning System), imaging device, beacon beam system, and QPD (Quadrant Photo Diode), Pan and Tilt Unit (PTU). The transmitter sequentially performs a pointing process toward the counterpart receiver, an acquisition process for receiving a tracking signal from the receiver, and a tracking process for maintaining the optical radio link through the counterpart terminal tracking. In order to construct a satellite-QKD system

base on PAT, a module for each part in manufactured by dividing it into FSM parts, sensor parts, and control parts that make up the QKD PAT part. The integrated structure should be designed in consideration of the optical signal path when the partial modules are arranged

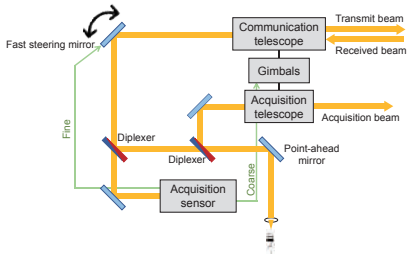


Fig. 2. PAT sub system

Therefore, laser communication technology requires the development of key technologies such as PAT technology to continuously maintain Line of Sight (LOS) between optical transmission and reception units, channel modeling technology to solve wavefront distortion caused by optical signal loss and atmospheric turbulence, optical system design technology for long-distance/wideband communication, and signal processing technology to secure robust wireless communication channel characteristics.[4]

II. WDM BASED QKD RESEARCH

In order for quantum cryptography to be used for commercial network, it is essential to present an interference control-based multiplexing technique between low quantum signals and high optical signals. High-power classical optical signals cause very threatening interference with low-power quantum signals, which leads to a fatal increase in quantum communication detection failures and error rates.

Interference phenomena in optical signals are being developed through improvement of quantum communication protocols, control of high-level optical signal power, and frequency multiplexing methods, but an integrated solution has not yet been developed. If a WDM-QKD system can be built, it is economical because there is no need to install additional lines dedicated to QKD in the existing backbone network. In addition, QKD-based security technology can improve security by fundamentally preventing external eavesdropping of the backbone network. Research on the implementation of Quantum key distribution is being actively conducted in various countries such as the US, China, EU, and Japan.

TABLE I. RESEARCH TRENDS OF QKD-WDM

Country	Transmission distance	Key rate
China	80km	1kbps
	66km	4.5kbps
	45km	-
EU	50/90km	507/7.6 kbps
	36.6km	1.906kbps
	25km	-
	51.5/95.5km	172.2/42 bps
Japan	10km	27.2kbps
US	50km	0.011kbps

A. *Research trends of QKD*

At the University of Geneva in Switzerland, researcher presented a quantum key distribution system with a 2.5 GHz repetition rate using a three-state time-bin protocol combined with a one-decoy approach. Taking advantage of superconducting single-photon detectors optimized for quantum key distribution and ultralow-loss fiber, researchers had distributed secret keys at a maximum distance of 421 km and obtain secret key rates of 6.5 bps over 405 km.[5]

In the United States, CV QKD was successfully implemented using local oscillator in 2015, and in 2016, succeeded in establishing a quantum communication backbone network from Columbus to Washington, D.C. Also in 2020, transmission of the entangled photon about 83.65 km between the Argon National Laboratory under DCE and the University of Chicago was successfully implemented.[6]

In China, the University of Science and Technology of China succeeded in distributing secret keys at a distance of 1,120 kilometers using entanglement based QKD in 2020. Also CV-QKD implementation using Ultra-low loss filter and TF QKD implementation using wavelength synchronization and time synchronization were successful in 2021.[7]

In Korea, Korea University succeeded in implementing a 25 km wired quantum communication system for the first time in the laboratory of a Korean university in 2018. Also SKT succeeded in transmitting a quantum encryption key in the 112 km round trip to Bundang, Yongin, and Suwon in 2017, and in 2018, KT succeeded in establishing the world's first multi-party (1:4) quantum encryption test network in each 10 km away.[8][9]

B. *Research trends of QKD-WDM*

In China, 16,64 QAM was used and the QKD receiver was tested using a fibre Bragg grating filter. This experiment was conducted in 2017. The transmission distance of this experiment was successful at 80 km and the transmission speed was 1.0 kbps. Also In 2018, the transmission experiment was successful by separating the wavelength band of the optical communication and the wavelength band of the quantum signal. [10]

In the UK, quantum cryptography and optical communication multiplexing techniques were successfully implemented in the same wavelength band by applying time filtering techniques in 2012. Through this, a transmission speed of 507/7.6 kbps was achieved at a distance of 50/90 km, respectively.[11] In addition, the DV-QKD was successfully tested for 121 km of field transmission by connecting several nodes in 2019.[12] In 2021, the 2.5 GHz DV-QKD WDM implementation experiment using the O-band was successfully implemented, with a transmission distance of 51.5/95.5 km at a transmission speed of 172.2/42 bps, respectively. [13]

In Japan, NEC/NICT joint research team developed a long-distance stable quantum cryptography system technology based

on optical fiber in 2013. [14] Also implementation using CV-QKD through 100 data channels and WDM was successful in 2019. [15]

In Korea, KT, a Korean telecommunications company, announced the development of unified quantum channel equipment that combines one data transmitting/receiving channel and one that transmits quantum signals in 2022. [16]

In the United States, a method of performing QKD between multiple transmitting units and multiple receiving units at a wavelength of about 1550 nm was proposed as a patent in 2013 through WDM link. [17] In addition, in 2011, a method and system for quantum key distribution through multi-user WDM networks were patented for wavelength routing. [18] In 2012, the temporary-filtering method was first proposed to increase the QKD transmission distance and key transmission rate when implementing the existing optical fiber system QKD. [19] Also In 2013, first proposed a QKD network method capable of WDM entanglement photon-based reconstruction and registered a patent [20]

III. SATELLITE BASED QKD RESEARCH

Wireless QKD is used to implement the satellite QKD. Therefore, many countries are conducting research on the merger with satellite QKD using wireless QKD. In order to implement satellite QKD technology, PAT and satellite payload technology are basically important. PAT technology refers to pointing, acquisition, tracking technology. PAT is a satellite protocol established to synchronize a ground station with an orbiting satellite

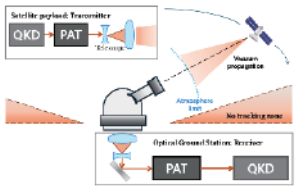


Fig. 3. Conceptual diagram of QKD system and ground station mounted in low-orbit satellites

Another important technology for satellite QKD is the payload technology. For satellite QKD design, the payload PAT module design must be carried out. In this process, the payload pointing error budget is analyzed (Field of uncertainty, FOU), beam acquisition/tracking(PAT), detector signal-to-noise ratio(SNR) is also analyzed. After that, the module design of the payload is carried out through the analysis of the beam divergence angle in each Coarse/Fine pointing step.

TABLE II. RESEARCH TRENDS OF SATELLITE QKD AND WIRELESS QKD

Country	Transmission distance	QBER	Key rate
Japan	-	3.7%	-
Germany	20km	4.8%	145bps
Canada	600km	6.5~8%	-
India	-	5%	35kbps
China	1200km	1%	14kbps
	4600km	-	-
Korea	2km	5%	1kbps
	100m	200kbps	200kbps

A. Research Trends of Satellite and Wireless QKD

In Japan, a demonstration test of quantum communication between a satellite-to-ground station was conducted between a low-orbit satellite orbiting an altitude of 628 km in NICT and a ground station built at NICT's Koganay headquarters in 2016. SOCRATES, a satellite used, is an ultra-small satellite with a size of about 50 cm3 and a mass of about 50 kg. [21]

In Germany, QKD experiments between aircraft moving at similar angular speeds and ground were successful. [22]

In Canada, a QKD experiment with a moving receiver similar to the angular velocity of a satellite at an altitude of 600 km was successful. [23]

In India, the BBM92 protocol was implemented using a flat light entangled double photon source (PEBS) based on the double pumping type 2 SPDC process in a Sagnac loop [24]

In China, succeeded in experimenting with uplink quantum teleportation and downlink entanglement using the low-orbit satellite trust node QKD in 2017. [25] In addition, in 2021, quantum cryptography communication was successful by combining optical fiber links on the ground and quantum satellites. [26]

Canada Announces Plans to Develop Experimental QKD Payloads to Test on Quantum Cryptography and Science Satellite QEYSSat [21]

In Korea, technology has been developed to enable semi-automatically ultra-precise orientation of (KT) transmission and reception devices, and quantum signals have been successfully transmitted in the 2 km round trip section of Cheongpyeong Lake in Gapyeong (2023). In addition, in 2018, (ETRI) succeeded in transmitting a secret key 200 kbps in the 100 m section of wireless QKD (ETRI). [21]

IV. DISCUSSION

Recently, QKD technology is not used alone, but research is being conducted in combination with other techniques such as WDM and satellite communication technology. Especially, when it come to WDM-QKD system, high power classical signals cause very threatening interference with low-power quantum signals, which leads to a fatal increase in quantum communication detection failures and error rates. The interference phenomenon that occurs in optical signals is being developed through improvement of quantum communication protocols, control of high-level optical signal power, and frequency multiplexing. However an integrated solution has not yet been developed, and these studies must be actively conducted.

Satellite QKD research is also being conducted with various techniques. Unlike wireless QKDs operating on the ground, the satellite QKD transmitting/receiving units are limited in weight

and size to be mounted on the satellite. Since the size of the optical unit that determines the QKD performance is limited, a technology that must overcome the limited conditions is required by advancing PAT technology and quantum signal transmission/reception technology. Therefore, these studies are being conducted in many countries. Additionally, components and design structures that have passed heat-resistant and radiation-resistant verification must be employed in order to operate normally in the environment at low orbit, and seismic motion design that can withstand the vibrations that occur when a satellite is launched is essential. It can be confirmed that these studies are being focused on various countries.

In the future, WDM and satellite techniques will serve as techniques that reinforce the advantages of the QKD technique and offset the disadvantages. It can be used for long-distance communication, communication security functions in disaster situations, and establishment of global QKD network

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00242396).

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2024-2021-0-01810) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2022-0-00463, Development of a quantum repeater in optical fiber networks for quantum internet)

REFERENCES

[1] C.H. Bennett and G. Brassard, in Proceedings IEEE International Conference on Computers, Systems and Signal Processing (Bangalore) IEEE, New York, 1984, pp. 175 – 179.

[2] Yamamoto, Y, et al. "Preparation, measurement and information capacity of optical quantum states." *Reviews of Modern Physics* 58.4 (1986): 1001.

[3] Jain, Nitin, et al. "Practical continuous-variable quantum key distribution with composable security." Nature communications 13.1 (2022): 4740.

[4] 허영순. “[광학 특집 특별 기고] 우주 레이저 통신 기술 동향” 『헬로 T』. July 25, 2023

[5] Boaron, Alberto, et al. "Secure quantum key distribution over 421 km of optical fiber." *Physical review letters* 121.19 (2018): 190502.

[6] Qi, Bing, et al. "Generating the local oscillator “locally” in continuous-variable quantum key distribution based on coherent detection." *Physical Review X* 5.4 (2015): 041009.

[7] Yin, Juan, et al. "Entanglement-based secure quantum cryptography over 1,120 kilometres." *Nature* 582.7813 (2020): 501-505.

[8] 김유성. “SKT, 양자암호통신 거리한계 극복.. 112km 시연 성공” 『이데일리』. June 19, 2017.

[9] 백지영. “10km 까지 전송 무선양자암호통신 시대 활짝… KT, 보안 패러다임 바꾼다” 『디지털데일리』. October 12, 2023.

[10] Mao, Yingqiu, et al. "Integrating quantum key distribution with classical communications in backbone fiber network." Optics express 26.5 (2018): 6010-6020.

[11] Patel, K. A., et al. "Coexistence of high-bit-rate quantum key distribution and data on optical fiber." Physical Review X 2.4 (2012): 041010.

[12] Wonfor, Adrian, et al. "Field trial of multi-node, coherent-one-way quantum key distribution with encrypted 5×100g dwdm transmission system." 45th European Conference on Optical Communication (ECOC 2019.IET, 2019.

[13] Grünenfelder, Fadri, et al. "The limits of multiplexing quantum and classical channels: Case study of a 2.5 GHz discrete variable quantum distribution system." Applied Physics Letters 119.12 (2021)

[14] Miki, Shigehito. "Superconductivity Web21."

[15] Eriksson, Tobias A., et al. "Wavelength division multiplexing of continuous variable quantum key distribution and 18.3 Tbit/s data channels." Communications Physics 2.1 (2019): 9.

[16] 박지성. “KT, 양자암호통신 양자채널 일원화 장비 개발…비용 절감, 통신효율 제고” 『전자신문』. November 7, 2022

[17] Luo, Y., Kam, TC.,(2011). System and methods for quantum key distribution over WDM links. US8050566B2. United States Patent

[18] Khalid, MAM., Ali, N., Witjaksono,G., (2011). Switchable integrated quantum key distribution system. WO2012044149A1. Patent Cooperation Treaty

[19] Patel, K. A., et al. "Coexistence of high-bit-rate quantum key distribution and data on optical fiber." Physical Review X 2.4 (2012): 041010

[20] Peiliang, GAO., 2013. Optical Isolator, Shutter, Variable Optical Attenuator and Modulator Device. US20090180776A1. United States

[21] 2023 양자정보기술 백서

[22] Nauerth, Sebastian, et al. "Air-to-ground quantum communication." Nature Photonics 7.5 (2013): 382-386.

[23] Bourgoin, Jean-Philippe, et al. "Free-space quantum key distribution to a moving receiver." Optics express 23.26 (2015): 33437-33447

[24] Chatterjee, Sourav, et al. "Polarization bases composition towards advancements in satellite-based QKD without active feedback." Communications Physics 6.1 (2023): 116.

[25] Liao, Sheng-Kai, et al. "Satellite-to-ground quantum key distribution." Nature 549.7670 (2017): 43-47

[26] Chen, Yu-Ao, et al. "An integrated space-to-ground quantum communication networkover 4,600 kilometres." Nature 589.7841 (2021): 214-219

Simulation Of Radar Cross Section of Foreign Object Debris Using High Frequency Simulation Software

MF Hakimi
Malaysia-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
hakimi99@graduate.utm.my

S Ambran
Malaysia-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
sumiaty@utm.my

SM Idrus
Faculty of Eletrical Engineering
Universiti Teknologi Malaysia
Johor Bharu, Johor, Malaysia
sevia@utm.my

PN Ja'afar
Malaysia-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
nadiahsyamimi20@gmail.com

***Abstract*—Safety term in aviation world is the most important thing. Foreign Object Debris (FOD) is the common thing that can be found on the airport runway around the world. Any type of FOD might cause an accident to the airplane. According to the Australian Transport Safety Bureau, on 27 November 2018, the nose wheel of Piper PA-31 was punctured caused by FOD on the runways. Effective ways to detect any hazardous on the runway is needed to ensure the safe and smooth airplane operations. The use of radar in FOD by reading the value of the radar cross section (RCS) continues to be developed over time. However, the use of radar in FOD detection is not effective without sufficient information because of the many types of FOD and also the many possible angles for each FOD which will result in various RCS values. Therefore, the focus of this project is simulation of edge light in different angles using Ansys High Frequency Structure Simulator at frequency of 93.1 GHz. The 12 different angles in the simulation will result in different surface areas reflecting the radar signal, and will also produce different RCS values. The wider surface area will give larger RCS value, and vice versa. Comparison of RCS values in simulation with experiments is also included in this project, where there is a large difference in values between experiment and simulation. This is due to distance, azimuth, and also the FOD environment. However, both have the same RCS value pattern. This project can be used as the basis information of detection FOD on the runway. This will improve the FOD detection system for aviation security systems.**

Keywords—foreign object debris, ANSYS HFSS, millimeterwave radar, radar cross section

I. INTRODUCTION

Nowadays, air transportation is one of the most time-efficient ways to travel from one place to another compared to other modes of transportation. The improvement in safety is even more impressive when considering the increase in air traffic. Airplane accidents can occur in various ways, one of which is due to Foreign Object Debris (FOD) on airport runways. FOD incurs millions of dollars in damage each year, particularly affecting airplanes during takeoff, landing, and taxiing.

In March 2015, accident investigators in New Zealand reported that damage to the flying controls of a 737 flight was due to a trapped cleaning cloth. Over the years, many researchers have developed systems to detect FOD, one of which involves using radar systems. The principle of radar is to detect, locate, track, and recognize objects of various kinds at considerable distances. Generally, the radar transmits electromagnetic energy toward FOD objects and measures the echoes returned from the objects. The values that bounce back to the receiver are called Radar Cross Section (RCS). Different RCS values depend on the material, size, angle, and other factors that might influence the RCS value.

The software used in this research provides comprehensive simulation capabilities, offering a vast array of tools for different engineering disciplines. One notable feature is the High Frequency Structure Simulator (HFSS). This research aims to provide a solution for measuring the RCS value of objects on airport runways using the software. By comparing the values obtained from the software with those from the radar, it can help create a safer environment in the aviation world.

II. METHODOLOGY

This section focuses on the design, simulation, and validation of results through experiments involving FOD to obtain the RCS values for edge lights. To achieve the project's objectives, the RCS needs to be measured from different angles of the edge light. The simulation results will then be compared with experimental results. During the design stage, the material is set to Perfect Electric Conductor (PEC), the standard for metallic materials. The object dimensions are as follows: a cylinder with a diameter of 150 mm and a height of 300 mm, a lower cone with a diameter of 200 mm, and an upper cone with a diameter of 180 mm. These dimensions are similar to the actual size of the edge lights used on runways. In Ansys HFSS, several parameters need to be determined to achieve the desired results, as shown in Table 1.

Table 1. Parameters used in HFSS simulation

Parameter	Description
Material	Perfet Electric Conductor (PEC)
Vector Input Format	Spherical (X=0, Y=0, Z=0)
I Wave Phi	Start = 0; Stop = 360, Step = 30
I Wave Theta	Start = 90; Stop = 90
Eo Vector	Phi = 1; Theta;0
Hybrid Region	PO Region
Frequency	93.1 GHz

Ultimately, the RCS values obtained from the simulations will be compared with those from the experimental work. The experiment was conducted at Kuala Lumpur International Airport (KLIA), exposing the setup to a real environment. The Radar Antenna Unit (RAU) served as the main system, with the edge lights on the runway as the targets. During the experiment, nine edge lights were placed on the runway, and measurements were taken at a 90-degree angle from the RAU. The angles varied from 0 to 360 degrees in 30-degree increments. The experimental setup described above is intended to serve as a comparison to the ANSYS HFSS simulation findings. The results from both the simulation and the experiment will be compared and used as the basis for future discussions.

III. RESULT AND DISCUSSION

The results obtained will be displayed and discussed. The measurement of the RCS value carried out by the Ansys HFSS software is the main objective of this project. Key parameters that need attention include the materials, frequency, and position of the FOD. This is to achieve the desired results, which will be compared with the RCS values obtained from the experiment conducted at a frequency of 93.1 GHz. The results of the ANSYS HFSS simulation are shown in Figure 1 and Table 2 below.

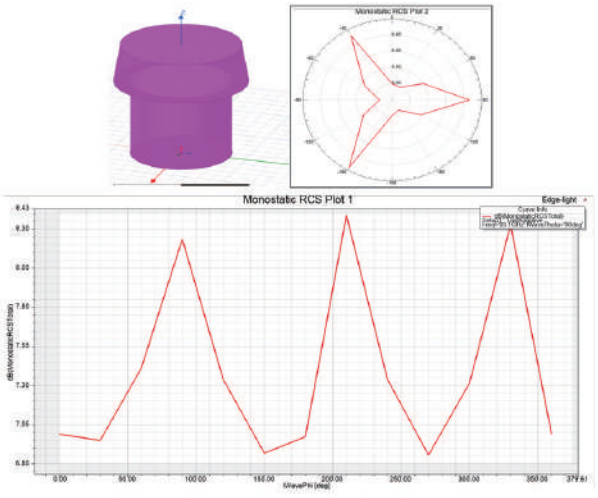


Figure 1. Experimental data for edge light

Table 2. ANSYS Simulation Result

Angle.	Radar Cross Section value (dbsm).
0°	6.990839
30°	6.951028
60°	7.410847
90°	8.233726
120°	7.339089
150°	6.869178
180°	6.974418
210°	8.386086
240°	7.340496
270°	6.857294
300°	7.317203
330°	8.325110
360°	6.990839

For each angle, the material is fixed to PEC, as this material is essential for obtaining valid RCS values. The type of material significantly affects the calculation of the RCS value. The cone and cylinder components must be unified using Boolean expressions before running the simulation; otherwise, errors will occur. For the hybrid region, the Physical Optic (PO) region is used instead of the Integral Equation (IE) region due to computer memory limitations. The frequency is maintained at 93.1 GHz throughout the simulation. A larger surface area results in a higher RCS value, while a smaller surface area results in a lower RCS value.

The simulation results are compared with experimental results conducted on the runway at Kuala Lumpur International Airport (KLIA). Nine edge lights were placed on the runway, and the angles varied from 0 to 360 degrees. However, for the simulation, only edge light number 9 was chosen to plot the data, as the data pattern is the same for each edge light, as shown in Figure 2.

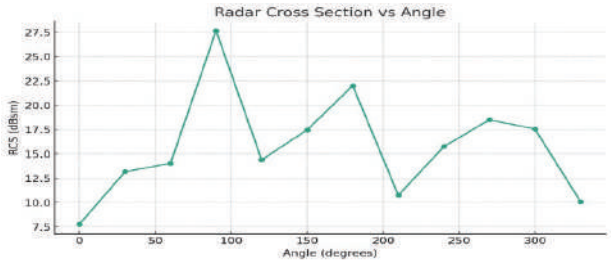


Figure 2. Experimental data for edge light

Based on Figure 2, the values can be transferred into Table 3, showing the values in decibels per square meter (dBsm). The graph patterns for the experimental and simulation results are quite similar, both showing three peaks of RCS. This indicates that the RCS values in the simulation are valid, even though they differ from the experimental values. In summary, the simulation runs as expected. Different angles of the edge light yield different RCS values. Environmental conditions can also influence the RCS value, but the graph pattern remains similar. The simulation results can be beneficial for aviation security systems to improve the performance of FOD detection systems.

Table 3. Experimental result

Angle.	Radar Cross Section value (dbsm).
0°	7.76dbsm
30°	13.15dbsm
60°	14.03dbsm
90°	27.65dbsm
120°	14.39dbsm
150°	17.46dbsm
180°	22.01dbsm
210°	10.74dbsm
240°	15.74dbsm
270°	18.50dbsm
300°	17.55dbsm
330°	10.08dbsm
360°	7.76dbsm

IV. CONCLUSION

As a conclusion, this project aims to determine the RCS value of the edge light from different angles. The values obtained from the simulation will be compared with those from the radar to minimize errors. Although there is a significant difference between the simulation and experimental values, the graph shows the same pattern. This project will enhance airport security systems in detecting any FOD on the runway. The same technique can be applied to other related FOD, significantly impacting aviation safety. As discussed in the previous chapter, this safety system can be improved. Integrating radar with Artificial Intelligence is one potential method for detecting FOD. Future research on FOD detection can utilize the technique proposed in this project. Thus, the innovation of this project can be refined and improved over time.

ACKNOWLEDGMENT

This research was funded by Universiti Teknologi Malaysia under the research grant vot no. 23H41.

REFERENCES

[1] Chauhan, T., Goyal, C., Kumari, D., & Thakur, A. K. (2020). A review on foreign object debris/damage (FOD) and its effects on aviation industry. Materials Today: Proceedings, 33, 4336-4339.

[2] Nori, H. (2020). A review of hijack events by airline employees, with a particular reference to Pilot Suicide Using an Aircraft (PSUA): an analysis of the triggering factors and current mitigation.

[3] Rajyalakshmi, P., & Raju, G. S. N. (2011). Characteristics of radar cross section with different objects. International Journal of Electronics and Communication Engineering, 4(2), 205-216.

[4] Stolarski, T., Nakasone, Y., & Yoshimoto, S. (2018). Engineering analysis with ANSYS software. Oxford, United Kingdom: Butterworth-Heinemann.

[5] Ja'afar, P. N. ., Idrus, S. M., Ambran, S., Yusri, N. A. ., Mohamed, N. ., Hamzah, A. ., Shibagaki, N., Kashima, K. ., & Hayashi, T. . (2023). Metallic Cylinder Reflected Power Measurement For 93.1GHz Frequency Modulated Continuous Wave Radar Calibration. International Journal of Integrated Engineering, 15(3), 187-192.

[6] Yusri, N., Idrus, S., Mohamed, N., Ambran, S., Iqbal, F., & Kanno, A. et al. (2021). Calibration of 93.1GHz FOD Detection Radar on Airport Runway using Trihedral Corner Reflector. 2021 8Th International Conference on Electrical Engineering, Computer Science, and Informatics (EECSI).

[7] P. D. L. Beasley, G. Binns, R. D. Hodgesl, and R. J. Badley. (2004). Tarsier- A millimetre wave radar for airport runway debris detection. pp. 261–264.

Performance Evaluation of Sampling Clock Offset Compensation in OFDM VLC System

Trio Adiono[*†], Erwin Setiawan[†], Michael Jonathan[†], Rahmat Mulyawan[*†], Nana Sutisna[*†], and Infall Syafalni[*†]
[*]School of Electrical Engineering and Informatics, Bandung Institute of Technology, Indonesia
[†]University Center of Excellence on Microelectronics, Bandung Institute of Technology, Indonesia
Email: {tadiono, erwin.s, michael.j, rahmat.mulyawan, nsutisna, infall}@itb.ac.id

***Abstract*—In this paper, we present simulation and FPGA verification results of the sampling offset effect in a (VLC) visible light communication (VLC) orthogonal frequency division multiplexing (OFDM) system. We model the sampling clock offset effect, which consists of frequency offset (SFO) and phase offset (SPO). These effects cause rotation in the constellation, thus increasing the BER. We found that the SFO can be compensated using the SFO compensator, while the SPO can be compensated using the MMSE equalizer. Our simulation results show that the SFO compensator can compensate for the for the SFO effect from -100 to 100 ppm with EVM values below 10% and the MMSE equalizer can compensate for the for the SPO effect from -0.5 to 0.5 with EVM values below 10%.**

***Index Terms*—OFDM, VLC, sampling offset, SFO, SPO**

I. Introduction

Visible light communication (VLC) is wireless communication that is carried out through the medium of the visible light spectrum. It is often called Li-Fi [1], [2]. One of the advantages of a VLC is its cost efficiency because it uses an LED transmitter that can also be used as lighting [3]. VLC is intended to be a complementary RF technology in the future of wireless communications.

Many VLC experiments are usually carried out under ideal conditions. Sampling offset is a problem that is usually ideally conditioned in lab experiments [4]. As done by [5] and [6], real-time transmission with the FPGA reaching a data rate of 2.5 Gbps and 5 Gbps is done with 2 boards each for TX and RX but using one common clock module for TX and RX. In real product conditions, this is not possible because the TX and RX are separated by a distance, so the TX and RX clocks must be independent.

There are many solutions to the sampling offset problem, as in [4], [7], [8]. However, these works only focus on the simulation level and are not implemented in real-time on the FPGA. Therefore, our contribution to this paper is the simulation and development of hardware modules implemented in FPGAs for sampling offset compensation.

The rest of the paper is organized as follows. The system model is presented in Section II. The sampling offset compensation is presented in Section III. Result is described in Section IV. Section V draws the conclusion.

II. Sampling Offset Modeling

Let's say $x(t)$ is the continuous OFDM signal at the transmitter and $y[n]$ is the signal at the receiver after ADC sampling. Then the $y[n]$ can be expressed by (1),

$$y[n] = x((n+\phi)T_{adc}) \tag{1}$$

where n is sampling index, ϕ is sampling phase offset, and T_{adc} is sampling period at the receiver's ADC. Under the ideal condition, the value of ϕ is 0 and T_{adc} is equal to T_{dac}, where T_{dac} is the sampling period at the transmitter's DAC.

The ideal condition can't occur in practical conditions because the oscillators for DAC and ADC clocks are independent, so the initial sampling instant between DAC and ADC can be misaligned [7]. They also have a non-zero tolerance, so the frequency sampling is not exactly the same [8].

Let's say the period difference between DAC's and ADC's clocks is $T_{dac}\Delta$, then the frequency offset would be $\Delta f = 1/T_{dac}\Delta$. Then, the sampled signal $y[n]$ can be rewritten as (2).

$$y[n] = x((n+\phi)T_{dac}(1+\Delta)) \tag{2}$$

The Δf is the tolerance of the crystal oscillator, which is usually expressed in units of ppm. Many wireless communication standards require the system to operate within a tolerance of 20–25 ppm [8].

III. Sampling Offset Compensation

A. Proposed System Block Diagram

The SFO effect can be compensated by using the SFO compensator as proposed in [8], and the SPO effect can be compensated using the frequency equalizer [7]. Fig. 1 shows our proposed system block diagram. We put the SFO module before channel estimation and equalization because our system uses a long preamble for channel estimation, which must also be compensated for by the SFO effect.

For the channel estimation, we use a long preamble for training and 8 pilots in every data symbol. The preamble and pilot pattern are based on the Wi-Max 802.16d standard with a few modifications. These modifications include the addition of pilots and Hermitian symmetry to the long preamble. For the equalization process, we use minimum mean squared error equalization (MMSE) equalization.

B. Sampling Frequency Offset Algorithm

The SFO estimation and compensation algorithm that we used is from this work [8]. To estimate the SFO, there are eight pilot subcarriers on every OFDM symbol. On the RX,

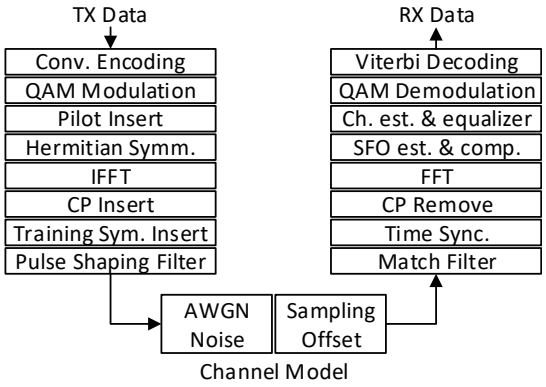


Fig. 1. System block diagram of OFDM VLC system.

from the received pilots and the known TX pilots, the $\hat{\Phi}$ can be calculated as

$$\hat{\Phi}_{p_i} = angle\left(\frac{Y_{p_i}}{X_{p_i}}\right),\ p_i \in Pilot,\ i = [1, 2, ..., 8] \quad (3)$$

where X_{p_i} and Y_{p_i} are the transmitted and received pilots, respectively. After that, the slope S_n needs to be calculated from the Φ defined as

$$S_n = \frac{\sum_{i=1}^{8} p_i.\hat{\Phi}_{p_i}}{\sum_{i=1}^{8} p_i^2} \quad (4)$$

where $p_i = -88, -63, -38, -13, 13, 38, 63, 88$ denotes the pilot subcarier index. Finally, in the compensation process, every IQ data subcarrier is rotated by the value $-iS_n$, as in 5.

$$Y^{'}[n] = Y[n].e^{-jiS_n} \quad (5)$$

The hardware module design, implementation, and optimization are presented in [9] and [10].

C. Sampling Phase Offset Algorithm

As justified in the work [7], the SPO can be compensated using a frequency equalizer. Therefore, we use channel estimation and equalization for the OFDM system to compensate the SPO effect.

The OFDM system uses the same eight pilots for channel estimation. The channel response is estimated by using a preamble symbol and pilots on every data symbol. the channel response is defined as

$$H(k) = \frac{Y(k)}{X(k)} \quad (6)$$

where $Y(k)$ is the received symbol, and $X(k)$ is the transmitted symbol. This process is done both for the preamble and the pilot. The response from the pilot updates the response from the preamble.

The MMSE equalization method used in our OFDM system is defined in (7). The noise variance σ_n^2 is obtained from the null subcarrier of the symbols.

$$Y_{equ}(k) = \frac{Y(k)}{|H(k)|^2 + \sigma_n^2}.(H(k))^* \quad (7)$$

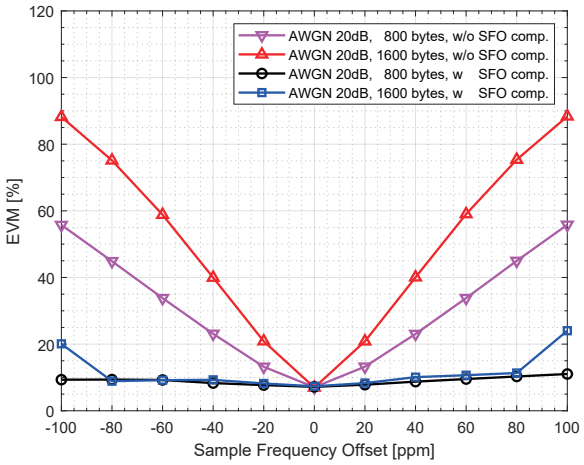


Fig. 2. EVM compared to SFO before and after compensation.

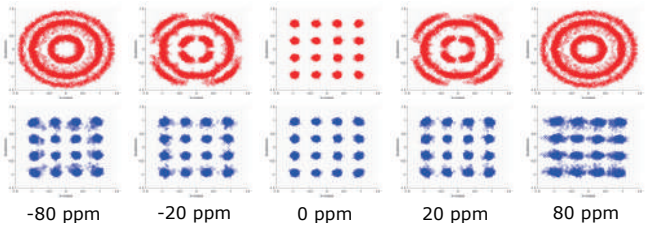


Fig. 3. Comparison of constellations before and after SFO compensation from -80 to +80 ppm.

IV. Result

A. SFO Compensation Performance

The SFO modeling and compensation are simulated in MATLAB, and the result is shown in Fig. 2. The result shows that the longer the frame (800 bytes vs. 1600 bytes), the worse the EVM. The SFO compensator can reduce the EVM to below 10% from -100 to +100 ppm SFO effect. The constellation of the system before and after SFO compensation is shown in Fig. 3.

B. SPO Compensation Performance

The SPO modeling and its compensation using the MMSE equalizer are simulated in MATLAB, and the result is shown in Fig. 4. The result shows that the frame length (800 bytes vs. 1600 bytes) has no effect on the EVM. The MMSE equalizer can reduce the EVM to below 10% from -0.5 to +0.5 SPO effect. The constellation of the system before and after SPO compensation is shown in Fig. 5.

C. BER Performance

We simulate the BER performance of our OFDM VLC system under an SFO value of 10 ppm and an SPO of 0.12, as shown in Fig. 6. We chose 10 ppm because this is the tolerance of the crystal oscillator on our target FPGA board. We chose 0.12 because this is the worst condition of EVM based on the result in Fig. 4.

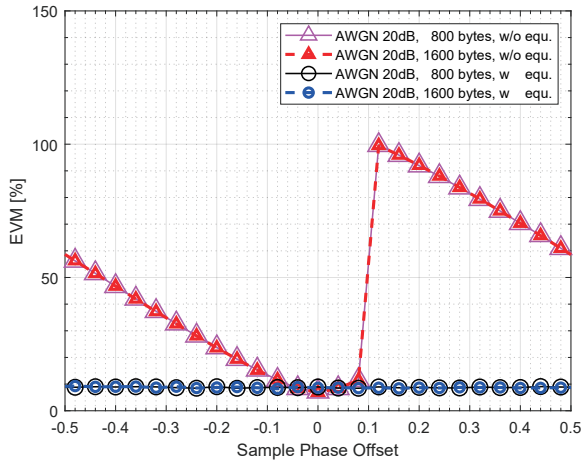


Fig. 4. EVM compared to SPO before and after compensation.

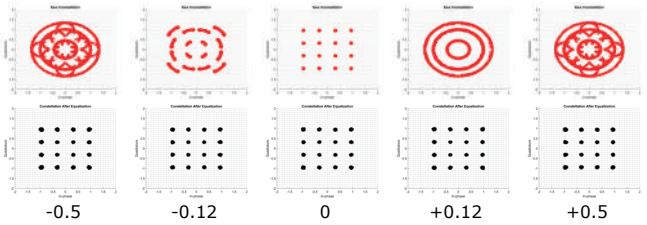


Fig. 5. Comparison of constellations before and after SPO compensation from -0.5 to +0.5.

D. FPGA Verification

The SFO and SPO modules have been implemented and verified on the Xilinx FPGA (Eclypse Z7 board [11] with Zmod AWG [12] and Zmod Scope [13] modules). Transmission is done using two FPGA boards with an independent clock. The setup for verification is shown in Fig. 7. We use an analog front-end that consists of an LED module, a bias-tee module from Mini Circuits [14], and an avalanche photodiode (APD) module from Hamamatsu [15].

The BER test result is shown in Table I. The BER without SFO and SPO compensation can't get below 10^{-2}, but after compensation, it can achieve error free condition except for 16-QAM 3/4.

TABLE I
FPGA BER PERFORMANCE

Modulation	BER without SFO compensation	BER with SFO compensation
QPSK 1/2	5.79e-2	0
16-QAM 1/2	3.19e-2	0
QPSK 3/4	4.08e-2	0
16-QAM 3/4	4.57e-2	1.24e-5

V. Conclusion

Our proposed hardware design can compensate for the for the SFO effect from -100 to 100 ppm with EVM values below 10% and the MMSE equalizer can compensate for the for the SPO effect from -0.5 to 0.5 with EVM values below 10%. FPGA verification shows that the SFO compensation can reduce the BER.

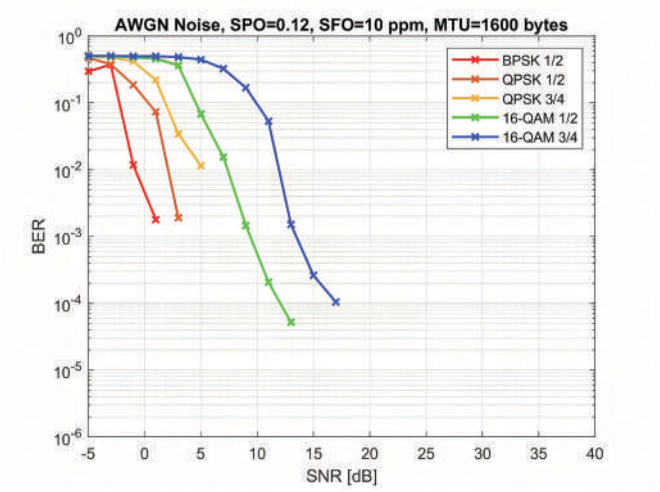


Fig. 6. BER performance of OFDM VLC system.



Fig. 7. A photograph of the FPGA verification.

References

[1] H. Haas, "Multi-Gigabit/s LiFi Networking for 6G," in *Proc. 2021 IEEE CPMT Symposium Japan (ICSJ)*, Japan, Nov. 2021, pp. 25–26.

[2] H. Abumarshoud, L. Mohjazi, O. A. Dobre, M. D. Renzo, M. A. Imran, and H. Haas, "LiFi Through Reconfigurable Intelligent Surfaces: A New Frontier for 6G?," *IEEE Vehicular Technology Magazine*, vol. 17, no. 1, pp. 37–46, Mar. 2022.

[3] H. Elgala, R. Mesleh, and H. Haas, "Indoor Broadcasting via White LEDs and OFDM," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 3, pp. 1127–1134, Aug. 2009.

[4] Y. Jiang, Y. Wang, P. Cao, M. Safari, J. Thompson, and H. Haas, "Robust and Low-Complexity Timing Synchronization for DCO-OFDM LiFi Systems," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 1, pp. 53–65, Jan. 2018.

[5] C. Wang, G. Li, Y. Ha, S. Han, and N. Chi, "A 2.5Gb/s Real-Time Visible-Light Communication System Based on Phosphorescent White LED," in *Proc. 2019 7th International Conference on Information, Communication and Networks (ICICN)*, China, Apr. 2019, pp. 140–145.

[6] M. Shi, C. Wang, G. Li, Y. Liu, K. Wang, and N. Chi, "A 5Gb/s 2 × 2 MIMO Real-time Visible Light Communication System based on silicon substrate LEDs," in *Proc. 2019 Global LIFI Congress (GLC)*, France, Jun. 2019.

[7] C. Chen et al., "Accurate Sampling Timing Acquisition for Baseband OFDM Power-line Communication in Non-Gaussian Noise," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1608-–1620, Apr. 2013.

[8] H. Shafiee, B. Nourani, and M. Khoshgard, "Estimation and Compensation of Frequency Offset in DAC/ADC Clocks in OFDM Systems," in *2004 IEEE International Conference on Communications*, France, June 2004, pp. 2397—2401.

[9] T. Adiono, E. Setiawan, M. Jonathan, R. Mulyawan, N. Sutisna, and I. Syafalni, ”FPGA Implementation of SFO for OFDM-based Network Enabled Li-Fi System,” unpublished.
[10] T. Adiono, E. Setiawan, M. Jonathan, R. Mulyawan, N. Sutisna, I. Syafalni, and W. O. Popoola, ”A Scalable Design of A Full-Stack Real-Time OFDM Baseband Processor for Network-Enabled VLC Systems,” unpublished.
[11] ”Eclypse Z7.” Digilent. https://digilent.com/reference/programmable-logic/eclypse-z7/start (accessed Mar. 7, 2024).
[12] ”Zmod AWG.” Digilent. https://digilent.com/reference/zmod/awg/start (accessed Mar. 7, 2024).
[13] ”Zmod Scope.” Digilent. https://digilent.com/reference/zmod/scope/start (accessed Mar. 7, 2024).
[14] ”ZFBT-4R2GW+” Mini-Circuits. https://www.minicircuits.com/pdfs/ZFBT-4R2GW+.pdf (accessed Mar. 7, 2024).
[15] ”C12702” Hamamatsu. https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRARY/ssd/c12702series_kacc1214e.pdf (accessed Mar. 7, 2024).

Comparison of YOLO-NAS and Roboflow 3.0 for Local Firearms Detection and Recognition

Cosmas Ifeanyi Nwakanma [1], Ikenna Uzoma Ajere [2], Ebuka Chinaechetam Nkoro [3], Judith Nkechinyere Njoku [3], Love Allen Chijioke Ahakonye [1], Dong-Seong Kim [3]

[1] ICT Convergence Research Center, *Kumoh National Institute of Technology*, Gumi, South Korea

[2] Information Technology, *Federal University of Technology Owerri, Nigeria*,

[3] IT-Convergence Engineering, *Kumoh National Institute of Technology*, Gumi, South Korea

(cosmas.ifeanyi; ebuka; judithnjoku24; loveahakonye; dskim)@kumoh.ac.kr

***Abstract*—Mass shootings, terrorism, and small firearm trafficking account for the rise in causes of death and grave injuries in most countries. However, one of the critical security concerns is the detection, screening, and recognition of small and locally handcrafted firearms. Deep learning though has played promising roles, more research options are still available. This work compared two recent object detection models - YOLO-NAS and Roboflow 3.0. In addition, we demonstrated the potential for collecting gun datasets that recognized the local guns to increase the robustness of the performance evaluation of deep learning models for gun detection. Results show that the hybrid adoption of YOLO-NAS and Roboflow 3.0 offers a potential for accurate, efficient, and real-time firearm detection and recognition for security and screening.**

***Index Terms*—Firearms, Deep Learning, Military Informatics**

I. Introduction

There is a growing concern over the increase in mass shootings, kidnappings, injuries, and deaths from the use of guns (licensed and trafficked). The United States, for instance, witnessed about 604 shootings in 2023 [1]. The situation is also worrisome in developing nations like Nigeria where there is a rise in insecurity and armed conflicts due to the ugly cases of the consequences of unrestricted use of local firearms [2]. In the 2023 Plateaue State massacre, there were 200 deaths, 500 were injured, and there was significant loss due to property damage. Terrorism and religious extremism are some of the reasons behind these mass shootings and executions [3]. One common denominator in these areas is the popularity of craft-handed local break-action shotguns and pistols four times cheaper than their industrially produced counterparts [4]. Thus, there is a need for a system to locate guns and weapons in public to ensure safety norms.

Hence, developing a firearm and weapon detection system is imperative to ensure public safety. Deep learning is a promising solution to detecting guns critical in security, surveillance, and content customization [5], [6]. The complexities in gun detection within intricate environments arise from clutter and the diverse geometries of guns. Thus, researchers have attempted fine-tuning and optimizing deep learning models with appreciable results. the You-Look-only-Once-Neural Architecture Search (YOLO-NAS) [7] and Roboflow 3.0 are chief among the recent solutions in this regard.

In this work, we address this problem in two-fold. First, we propose a novel dataset that captures the local, unconventional, and trafficked firearms using images from crime-related news, videos, and internet photos. Secondly, we implemented and compared the performances of the YOLO-NAS and Roboflow 3.0 taking note of the gains and limitations of each approach. The real-time firearm detection system helps bolster manual surveillance efforts along national borders, identifying potential security threats. Additionally, its deployment on wireless cameras and drones [6] facilitates the swift detection of suspicious weaponry, thereby enhancing efforts to apprehend fugitives within forested areas [3]. The general findings serve as a watershed to helping researchers in security, military, and firearm monitoring, enabling more efficient firearm localization and identification in diverse scenarios [5].

Although deep-learning-based firearm detection research has attracted attention, the study into unconventional, small handcrafted local firearm research is still in its infancy. In addition, the YOLO-NAS [7] is a new object detection model with superior performance to other YOLO versions. This work is arguably the first attempt to investigate the comparative performance of YOLO-NAS and Roboflow 3.0 [1] for gun detection and classification system research.

II. Related Works

Yadav [3] focused on firearm classification, achieving 92.723% accuracy with pre-training and 95.01% without pre-training using ResNet-34, and 95.011% accuracy with pre-trained ResNet-50 and 91.26% without pre-training. GunY-olo [8] improved rifle number detection by fusing authentic rifle images with existing digit images for training, achieving enhanced detection and recognition capabilities using rifle images and numerical datasets.

Xu et al. [9] proposed an attention mechanism-based CNN (AM-CNN) model to detect and localize guns and other prohibited items in X-ray images of airport security systems.

[1] https://roboflow.com/

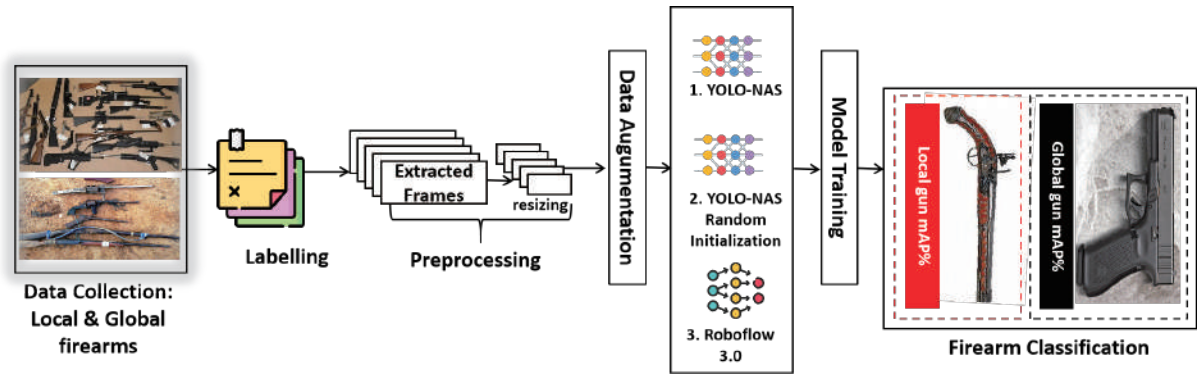


Fig. 1. Showing the proposed hybrid firearm detection methodology: integrating local & global approaches for diverse firearm detection.

The AM-CNN model was trained with data from the GDX-ray image dataset. The recognition and localization roles were played by CNN and AM respectively. The result shows that the proposed AM-CNN model is stable and efficient for single and multi-target detection, with a recognition and location accuracy of 98.3%, and 73.5% respectively for guns.

Mahmood et al. [5] proposed an attention mechanism capable of distinguishing humans from firearms by utilizing an adaptive average pooling layer of deep learning. The attention mechanism yielded an inspiring result motivating our work. However, the difference lies in the need for local firearms not captured in the datasets and the need for explainability.

Although deep-learning-based firearm detection research has attracted attention, the study into unconventional, small handcrafted local firearm research is still in its infancy. In addition, the YOLO-NAS [7] is a new object detection model with superior performance to other YOLO versions. This is the first attempt to deploy YOLO-NAS for local firearm detection and classification system research. The YOLO-NAS was compared with the Roboflow 3.0 [2] to widen the research in this area. This work presents a deep-learning real-time local firearms detection system custom-designed to detect and recognize local, handcrafted firearms not captured in existing works.

III. System Model and Design Methodology

A. System Architecture

Fig. 1 represents the collection of local firearms integrated with conventional and globally recognized firearms to have a unique dataset. This dataset was used to train and validate the performance of YOLO-NAS and Roboflow 3.0. Most existing works relied on publicly available datasets such as the Hands Guns and Phones (HGP) Dataset [10], and Gun Detection in Video Dataset [11]. However, this work involved the independent collection and processing of about 600 datasets of guns cutting across various countries and accommodating both licensed and trafficked guns. These firearm images, especially the local ones, were captured from images of local robbery attacks, shooting practices, and near-life situations representing the posture of local guns in Nigeria from the news and internet.

The adopted dataset captures the local, unconventional, and trafficked firearms using images from crime-related news, videos, and internet photos. In addition, we leverage data augmentation incorporating local firearms from Nigeria such as the **Beretta M12**, **Beryl M762**, **OBJ-006**, **Beretta BM59**, and deadly local handcrafted firearms [2], [4].

B. Implementation

The YOLO-NAS is deployed using the Super-Gradients library, a computer vision framework based on Deci's PyTorch implementation [7], [12]. See Medaramatla et al [12] for the comprehensive architecture of the YOLO-NAS. The YOLO-NAS was released in 2023 by Deci-AI to compensate for the trade-off between accuracy and latency [7]. Deci-AI's optimization engine known as AutoNAC served as a game changer making it possible to maximize the use of hardware while achieving faster deep learning acceleration. Pretraining was done on Roboflow based on knowledge distillation and distribution focal loss [7].

When combined with YOLO-NAS [7], Roboflow 3.0 [3] enhances firearm detection capabilities significantly. It offers streamlined dataset management, enabling efficient organization and annotation of firearm images. Advanced data augmentation techniques in Roboflow 3.0 improve model generalization and robustness, leading to higher detection accuracy. Roboflow 3.0's enhanced model training capabilities and performance evaluation tools further optimize the YOLO-NAS model, resulting in superior firearm detection performance with increased speed and accuracy [12].

The following steps were taken to implement the three scenarios: YOLO-NAS base model, YOLO-NAS with random initialization, and Roboflow 3.0. Data collection - Dataset annotation - Data preprocessing - Data Augmentation - Model training - Model testing.

The total data collected were 494 images Split into 70% (346 images), 20% (99 images), and 10% (49 images) representing the training, validation, and testing data respectively.

[2]https://roboflow.com/

[3]https://roboflow.com/

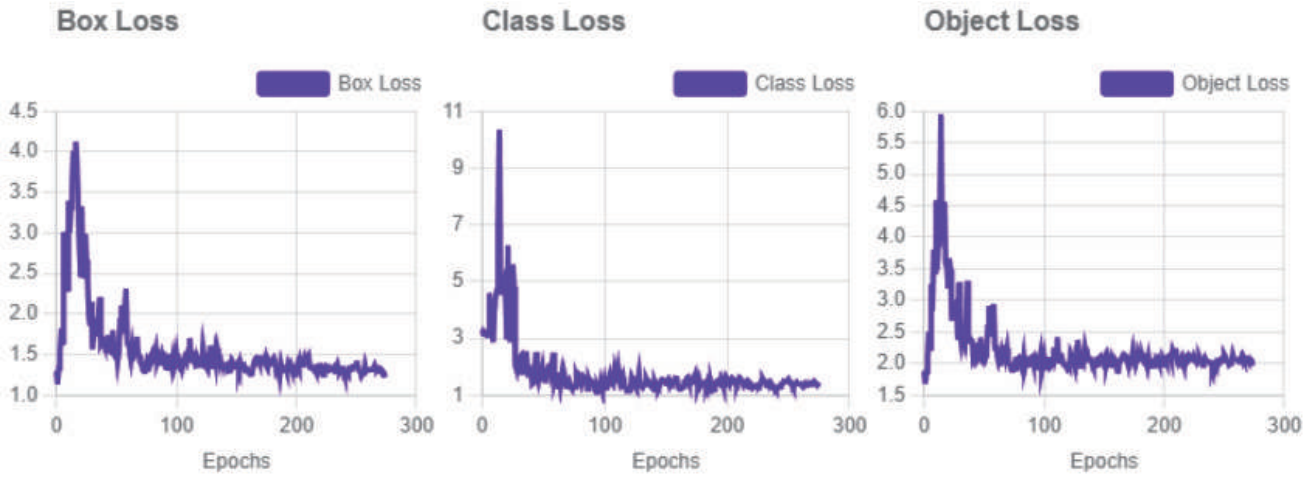


Fig. 2. Showing the losses of the best-performed model Roboflow 3.0. Box Loss, Class Loss, and Object Loss respectively.

These datasets were preprocessed using the grayscale, adaptive equalization contrast auto-adjust, and auto-orient. Data augmentation was also done to increase the dataset to 1186 (multiplied by 3). All testing was done on the baseline YOLO-NAS, Roboflow 3.0, and YOLO-NAS - Random initialization. Dataset after augmentation: 1177 comprising Training [1029] 87%, Validation [99]8%, and Test [49] 4%.

C. Attention Mechanism

The attention mechanism provides a means to select the input features relevant to the current task. It is used in DL because it allows DL models to pay more attention to useful features, thereby conserving storage and computing power and improving the accuracy and efficiency of DL models [13]. By assigning different weights to the hidden layer units of the DL model, the attention mechanism enables the hidden layer to pay more attention to the critical information. Computationally, AM is derived by: a). computing the attention information on the input information and b). computing the context vector according to the attention information. This is enumerated in Algorithm 1 for YOLO-NAS. The same process was implemented for Roboflow 3.0.

Algorithm 1 Attention Mechanism in YOLO-NAS

1: **Input:** Image X
2: $F \leftarrow \text{CNN}(X)$ {Extract feature map}
3: $Z \leftarrow \text{ReLU}(W_f \cdot F)$ {Apply ReLU activation}
4: $A \leftarrow \text{softmax}(W_a \cdot Z)$ {Compute attention map}
5: $F_w \leftarrow F \odot A$ {Weighted feature map}
6: **Output:** Predictions YOLO-NAs(F_w) {Object detection}

Where: X: Input image, F: Feature map, Z: Intermediate feature representation, A: Attention map, F_w: Weighted feature map, CNN: Convolutional Neural Network, YOLO-NAs: YOLO-Neural Architecture , W_f: Weight matrix for feature transformation, W_a: Weight matrix for attention map transformation , $\odot$: Element-wise multiplication , ReLU: Rectified Linear Unit activation function and softmax: Softmax function for normalization

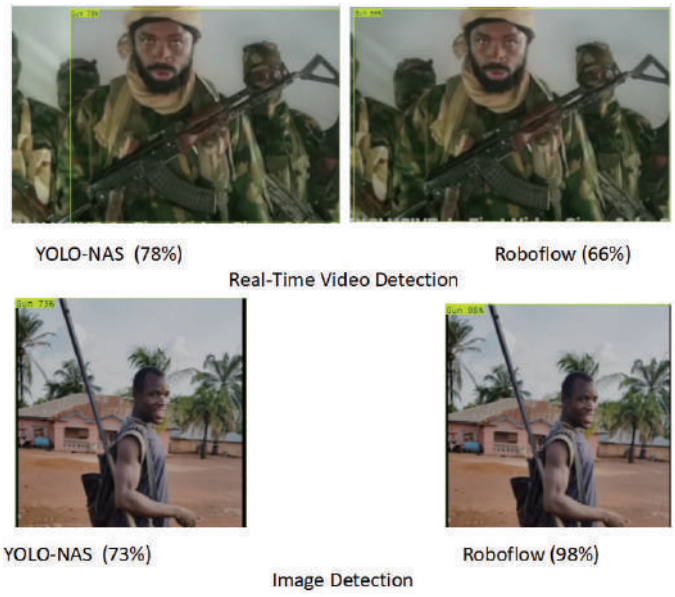


Fig. 3. Showing a Nigerian Local firearm used for the test, the superior performance of the Roboflow 3.0 and YOLO-NAS base model. It also shows Showing the Validation of YOLO-NAS and Roboflow 3.0. when tested with a real-time video of the BOKO HARAM and firearm detection of a local Nigerian firearm carried by a man.

IV. Performance Evaluation

Table I shows the summary of the performance of the three models. Deploying the Roboflow 3.0 and YOLO-NAS promises to be faster, more accurate, and more efficient for firearm detection and recognition. Table I outlines an overview of the performance of the investigated models, While Fig. 2 shows the negligible loss values of the Roboflow 3.0 model. Fig. 3 shows the comparative performance of the YOLO-NAS (78%) and Roboflow 3.0 (66%) when validated using the Boko Haram image from a video clip. In other scenarios,

the Roboflow showed 98% accuracy compared to the YOLO-NAS value of 73% for local firearm images. The mAP of the two models is presented in Fig. 4. Roboflow displayed an accuracy of 98% when deployed for image detection but had a low performance of 66% for real-time video detection. On the other hand, the YOLO-NAS had consistent performance of 73% and 78% for image and video-based gun detection.

TABLE I
PERFORMANCE COMPARISON OF THE THREE MODELS INVESTIGATED

Performance Metrics	YOLO-NAS	Roboflow 3.0	YOLO-NAS-R
Training mAP (%)	72.6	77.5	54.7
Training Precision (%)	81.3	90.4	74.5
Training Recall (%)	69.6	72.8	48.8
Validation Average Precision (%)	73	77	55
Testing Average Precision (%)	68	81	58

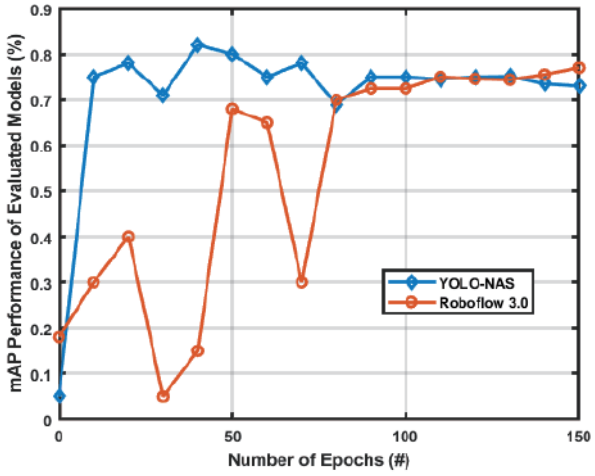


Fig. 4. The mAP Comparison between the YOLO-NAS and Roboflow 3.0 over the first 150 Epochs.

Fig. 4 presents the Mean Average Precision (mAP) results. The mAP is often preferred over accuracy in computer vision tasks because it provides a more comprehensive evaluation, especially in scenarios where there are multiple classes or when the distribution of classes is imbalanced [14], [15].

V. CONCLUSION

This work compared the novel YOLO-NAS and Roboflow 3.0 deep learning models for gun detection leveraging the uniquely collected datasets that captured diverse guns especially local and handcrafted guns which account for high mass shootings and terrorism in countries. Results show that the hybrid adoption of YOLO-NAS and Roboflow 3.0 offers a potential for accurate, efficient, and real-time firearm detection and recognition for security and screening. One limitation of this work is the delimitation of YOLO and Roboflow frameworks without consideration for other competing deep learning architectures. An area for future investigation lies in advancing the integration of holistic firearm regulation, improvement in security perceptions, and real-time response to security-related distress calls.

ACKNOWLEDGMENT

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government(MSIT) (IITP-2024-2020-0-01612, 50%) and Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2018R1A6A1A03024003, 50%)

REFERENCES

[1] P. Lu, D. Chen, Y. Li, X. Wang, and S. Yu, "Agent-based model of mass campus shooting: Comparing hiding and moving of civilians," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 3, pp. 994–1003, 2023.

[2] F. Ogwueleka, "Information communication technology, cybersecurity and small arms in africa," in *The Palgrave Handbook of Small Arms and Conflicts in Africa*, U. A. Tar and C. P. Onwurah, Eds. Palgrave Macmillan, Cham, 2021, p. 647–677. [Online]. Available: https://doi.org/10.1007/978-3-030-62183-4_31

[3] K. Yadav, R. Punia, and N. Raghava, "Towards the law enforcement of firearms," in *2020 IEEE 17th India Council International Conference (INDICON)*, 2020, pp. 1–6. [Online]. Available: https://doi.org/10.1109/INDICON49873.2020.9342202

[4] M. Nowak and A. Gsell, "Handmade and deadly craft production of small arms in nigeria," *Small Arms Survey: Briefing Paper*, pp. 1–20, June 2018. [Online]. Available: https://www.smallarmssurvey.org/resource/handmade-and-deadly-craft-production-small-arms-nigeria

[5] A. Mahmood, A. Basit, M. A. Munir, and M. Ali, "Detection and localization of firearm carriers in complex scenes for improved safety measures," *IEEE Transactions on Computational Social Systems*, pp. 1–11, 2023.

[6] S. O. Ajakwe, V. U. Ihekoronye, D.-S. Kim, and J. M. Lee, "Dronet: Multi-tasking framework for real-time industrial facility aerial surveillance and safety," *Drones*, vol. 6, no. 2, 2022.

[7] E. Khvedchenya and H. Sahota, "Yolo-nas by deci achieves state-of-the-art performance on object detection using neural architecture search," in *Algorithms*. Deci, May 3, 2023. [Online]. Available: https://deci.ai/blog/yolo-nas-object-detection-foundation-model/

[8] H. Kwon and S. Lee, "Novel rifle number recognition based on improved yolo in military environment," *Computers, Materials & Continua*, vol. 78, no. 1, pp. 249–263, 2024.

[9] M. Xu, H. Zhang, and J. Yang, "Prohibited item detection in airport x-ray security images via attention mechanism based cnn," in *Pattern Recognition and Computer Vision: First Chinese Conference, PRCV 2018, Guangzhou, China, November 23-26, 2018, Proceedings, Part II 1*. Springer, 2018, pp. 429–439.

[10] M. A. Duran-Vega, M. Gonzalez-Mendoza, L. Chang, and C. D. Suarez-Ramirez, "Tyolov5: A temporal yolov5 detector based on quasi-recurrent neural networks for real-time handgun detection in video," *arXiv*, pp. 1–16, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2111.08867

[11] D. Qi, W. Tan, Z. Liu, Q. Yao, and J. Liu, "A dataset and system for real-time gun detection in surveillance video using deep learning," *arXiv*, no. 2105.01058, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2105.01058

[12] S. C. Medaramatla, C. V. Samhitha, S. D. Pande, and S. R. Vinta, "Detection of hand bone fractures in x-ray images using hybrid yolo nas," *IEEE Access*, vol. 12, pp. 57 661–57 673, 2024.

[13] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, September 2021.

[14] J. N. Njoku, G. O. Anyanwu, I. S. Igboanusi, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "State-of-the-art object detectors for vehicle, pedestrian, and traffic sign detection for smart parking systems," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 1585–1590.

[15] S. O. Ajakwe, V. U. Ihekoronye, D.-S. Kim, and J.-M. Lee, "Alien: Assisted learning invasive encroachment neutralization for secured drone transportation system," *Sensors*, vol. 23, no. 3, 2023.

Design and Implementation of Simulator for Large-scaled Naval Combat Systems

Jae-Woo Kim
ICT Convergence Research Center
Kumoh National Institute of Technology
Gumi, Republic of Korea
jaewookim@kumoh.ac.kr

Gi-Hyeob Kwon
ICT Convergence Research Center
Kumoh National Institute of Technology
Gumi, Republic of Korea
navkwon@kumoh.ac.kr

Dong-Seong Kim
Dept. of IT Convergence Engineering
Kumoh National Institute of Technology
Gumi, Republic of Korea
dskim@kumoh.ac.kr

Won-Jae Ryu
ICT Convergence Research Center
Kumoh National Institute of Technology
Gumi, Republic of Korea
wj0828@kumoh.ac.kr

Jae-Ho Lee
Naval Systems R&D Center
Hanwha Systems
Gumi, Republic of Korea
jh8028@hanwha.com

Jae-Hak Yu
Naval Systems R&D Center
Hanwha Systems
Gumi, Republic of Korea
jahak@hanwha.com

***Abstract*—This paper proposes and implements a simulator design method for analyzing the performance of large-scaled naval combat systems. Through the simulator, it is possible to analyze the performance of the naval combat system network and ensure its suitability before actual network configuration. The proposed simulator is a comprehensive tool that not only tests the performance of the network devices constituting the naval combat system but also considers the performance of processing units. Additionally, it takes into account characteristic network traffic used in naval combat networks. Therefore, this paper derives requirements for simulating naval combat system and designs and implements a comprehensive simulator that meets these requirements. The simulator analyzes network performance and parameterizes the resources of each processing unit in the naval combat system to reflect the performance results. The proposed comprehensive simulator enables optimization of network configuration for naval combat systems and facilitates stability assessment, among other functionalities.**

***Index Terms*—Naval Combat Systems, Performance Simulation, Network Test**

I. Introduction

Naval combat system comprises various sensors, armament systems, command and control systems, etc., within a naval ship. These systems are integrated to control and manage all necessary information for efficient mission execution, including tactical situation assessment, command decisions, threat evaluation, weapon allocation, and engagement[1-2]. The integration of combat and tactical networks, along with the growing scale of combat system networks, is leading to increasingly larger networks. Testing such large-scale networks in reality is challenging. Therefore, pre-performance analysis through simulation becomes crucial for rapid prototype design, reproducibility, stability, education, and other perspectives [3]. Moreover, to deploy new equipment in next-generation naval ships, it is necessary to design resources efficiently, considering their potential impact on network performance. Hence, there is a need for simulation that supports parameterizing resources for optimal design. Existing network performance evaluation tools include OPNET, OMNET++, NS3, etc. However, these existing tools are not specialized for naval combat networks, do not reflect node resources in the results, and require a amount of time for large-scale testing. Additionally, some tools require paid licenses and do not have publicly available source code [4].

In this paper, we identify the requirements of naval combat system for performance evaluation of networks and processing units and propose a design methodology for a comprehensive simulator for performance evaluation. Additionally, we implemented a simulator according to the proposed design technique.

II. System design

A. Naval Combat System Test Requirements

Naval combat systems consist of large-scaled systems, including command, control, and armed systems, sensors, weapons, and data channels, and they incorporate a variety of subsystems tailored to missions and characteristics. Information exchange among the subsystems of the naval combat system occurs through networks. A simulation tool is necessary to conduct preliminary simulations for designing the network of the naval combat system. Furthermore, when designing a comprehensive simulation tool, it is necessary to reflect the performance evaluation requirements of the naval combat system. Through a survey of defense industry, we have defined the following simulation requirements for evaluating the performance of the naval combat system's network and processing units.

- Parameterize the specifications of network devices and processing units in the combat system.
- Parameterize the Layer 3 or Layer 3 data throughput of nodes and processing units in the naval combat system.
- Evaluate the traffic of Data Distribution Service(DDS) middleware used in the naval combat system.

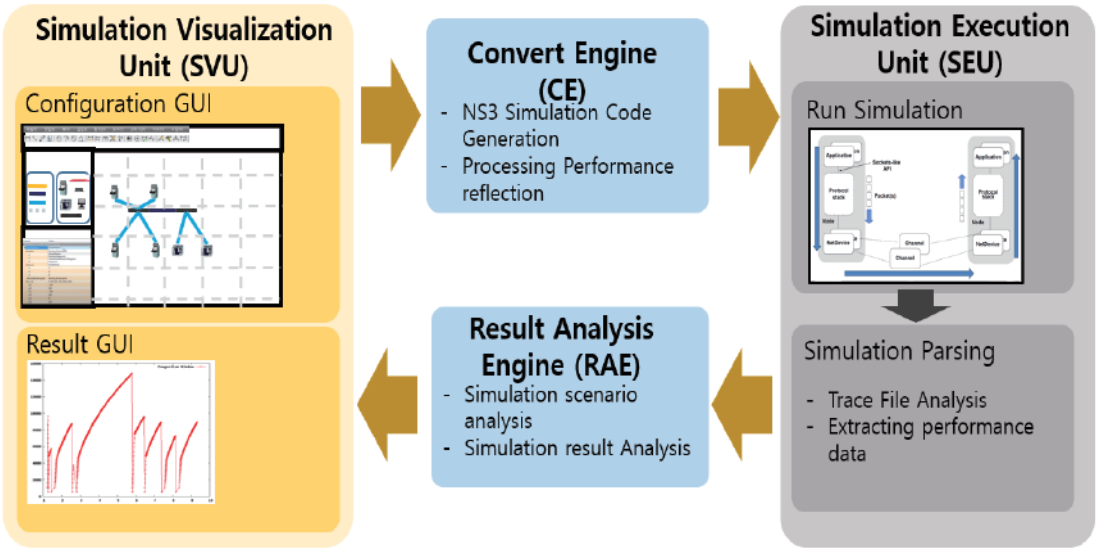


Fig. 1. Purposed Simulator System Design

- Quantitatively derive overall performance analysis of the naval combat system and define quantitative evaluation criteria.
- Measure end-to-end data transfer and latency of nodes in the naval combat system network.
- Measure the load distribution of each node when performing data transfer and processing in the naval combat system.
- Monitoring the progress of the simulation for the combat system should be possible.
- The evaluation simulator for naval combat systems should provide an interface that allows intuitive confirmation of network system scenarios and performance analysis results, considering user convenience.

The requirements of the evaluation simulator for naval combat systems include specification parameterization, evaluation of data throughput and middleware traffic, quantitative analysis of system performance, measurement of data transmission and latency, monitoring of simulation progress, and provision of an intuitive interface for scenario verification and analysis. In particular, simulations capable of simulating DDS-related data are needed for middleware traffic evaluation. In particular, DDS-based traffic tests, a real-time message-oriented standard, are required to evaluate middleware traffic used in military networks [5].

B. The Comprehensive Simulator System Design

Figure 1 depicts the block diagram of the proposed comprehensive simulator design for analyzing the performance of naval combat systems. The comprehensive simulator consists of a simulation visualization module unit (SVU), a simulation execution unit (SEU), as well as a Conversion Engine (CE) and Result Analysis Engine (RAE). The SVU constructs the simulation network. To parameterize the capabilities of processing units, the specifications of processing units are designed to be reflected in the simulation configuration. Additionally, the SVU allows setting scenarios for simulation operation. When setting up a simulation scenario, the comprehensive simulator can be configured by selecting the types of network layers 2 and 3. Furthermore, to reflect the characteristics of DDS middleware traffic in the simulation, SVU enables distinguishing and setting multicast and unicast scenarios. SVU also includes the part to visualize simulation results through graphs or values. CE generates code to perform network simulation based on the simulation structure and scenarios configured in SVU. The comprehensive simulator utilizes NS3 for network simulation. NS3 is developed based on C++, and simulation implementations require users to write code in C++ [6]. While NS3 offers various advantages, it lacks user-friendliness and requires knowledge of code implementation. In this paper, to improve the user-friendliness of NS3, SVU is used to easily configure the simulation structure, and based on this configuration, CE converts it into code for NS3 simulation. Additionally, CE is designed to generate code that reflects the specified parameters of the processing units in the naval combat system, ensuring that the simulation results are influenced accordingly.

The SEU performs simulations using the NS3 engine using generated code. NS3 outputs a trace file as a result of the simulation. The trace file contains information about all packets transmitted and received over the network. Therefore, the size of the Trace file increases with the simulation structure and time, along with the increase in information. For simulation result analysis, the SEU performs simulation parsing to extract meaningful result information from the Trace file obtained through simulation execution. The RAE retrieves simulation result data from the parsed Trace file and analyzes the simulation results by referring to the simulation scenarios. After analysis, the simulation results are passed to the SVU for graphical output. RAE derives the load distribution of network nodes, calculates transmission delays, and measures network

jitter to meet the performance analysis requirements of naval combat system. SVU is implemented to visualize these result items in an easily understandable manner.

C. Implementation

In this paper, we implemented a comprehensive simulator according to the proposed design method. We used the NS3 network simulation tool for evaluating the network performance of the simulator. SVU was implemented using PySide6, a GUI library based on the Python language [7]. PySide6 operates on both Windows and Linux and has the advantage of being freely available for commercial use. Furthermore, CE and other engines were also developed using the Python language, providing the advantage of being implementable with the same code on both Windows and Linux platforms.

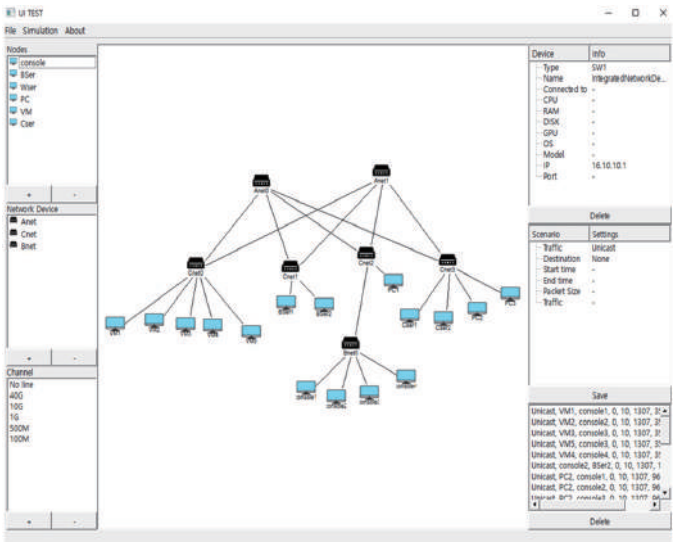


Fig. 2. Simulation Configuration Screen

III. Implementation Results

In this paper, to validate the implementation results, we defined a prototype of the naval combat system network and conducted performance evaluations using the simulator we developed. Figure 2 provides an example of the naval combat system network configuration. Initially, information such as nodes and switches to be deployed can be inputted by selecting the windows on the left. After inputting the information, related switches, nodes, etc., can be placed on the central screen, and connections between the placed nodes can be established by selecting link capacity and other parameters on the GUI. When configuring a node, the specifications of the processing unit are parameterized by setting the properties of the node. The properties of the set node affect the simulation results. Figure 3 shows the specification parameters that can be set when configuring the node. Parameterized properties include CPU, RAM memory, graphics card (GPU), the type and size of hard disk.

Once the network configuration is completed, it is possible to select individual nodes and input transmission scenarios. Inputs include Packet Size, Transmission Traffic, Transmission Time, and selection of Unicast/Multicast. Upon completing all inputs and saving, relevant transmission details are added as simulation scenarios at the bottom right.

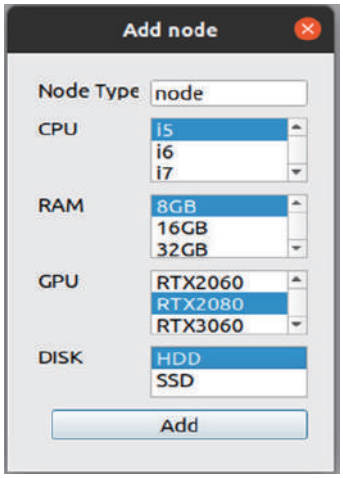


Fig. 3. Node specifications setting window

```
#include <iostream>
#include <fstream>
#include "ns3/flow-monitor.h"
#include "ns3/flow-monitor-helper.h"
#include "ns3/core-module.h"
#include "ns3/network-module.h"
#include "ns3/applications-module.h"
#include "ns3/bridge-module.h"
#include "ns3/csma-module.h"
#include "ns3/internet-module.h"
#include "ns3/point-to-point-module.h"
#include "ns3/netanim-module.h"
using namespace ns3;
NS_LOG_COMPONENT_DEFINE ("HanwhaNetworkSimulation");
int main (int argc, char *argv[])
{
  uint16_t udpEchoPort = 9;
  uint16_t multicastPort = 8;
  Config::SetDefault ("ns3::Ipv4GlobalRouting::RandomEcmpRouting",
  InternetStackHelper internet;
  Ipv4AddressHelper ipv4;
  PointToPointHelper p2p;

  Ptr<Node> VM1=CreateObject<Node> ();
  Ptr<Node> VM2=CreateObject<Node> ();
  Ptr<Node> VM3=CreateObject<Node> ();
  Ptr<Node> VM4=CreateObject<Node> ();
  Ptr<Node> VM5=CreateObject<Node> ();
  Ptr<Node> PC1=CreateObject<Node> ();
  Ptr<Node> PC2=CreateObject<Node> ();
  Ptr<Node> PC3=CreateObject<Node> ();
  Ptr<Node> Cser1=CreateObject<Node> ();
  Ptr<Node> Cser2=CreateObject<Node> ();
  Ptr<Node> BSer1=CreateObject<Node> ();
  Ptr<Node> BSer2=CreateObject<Node> ();
  Ptr<Node> console1=CreateObject<Node> ();
```

Fig. 4. Generated Simulation Code

Based on the added simulation scenarios, when simulation-related settings in the menu window are opened and executed, NS3 code is generated as shown in Figure 4, and NS3 is executed accordingly. Once NS3 is executed, and the simulation is completed, it generates Trace files as shown in Figure 5. The NS3 tracing data is built on the concepts of independent tracing sources and tracing sinks, and a uniform mechanism for connecting sources to sinks. Trace data are entities that can signal events that happen in a simulation and provide access to interesting underlying data. For example, a trace source could indicate when a packet is received by a net device and provide access to the packet contents for interested trace sinks [8]. The Trace files are parsed and saved as result files, which can be loaded and displayed in the simulation result window as shown in Figure 6. The outputted

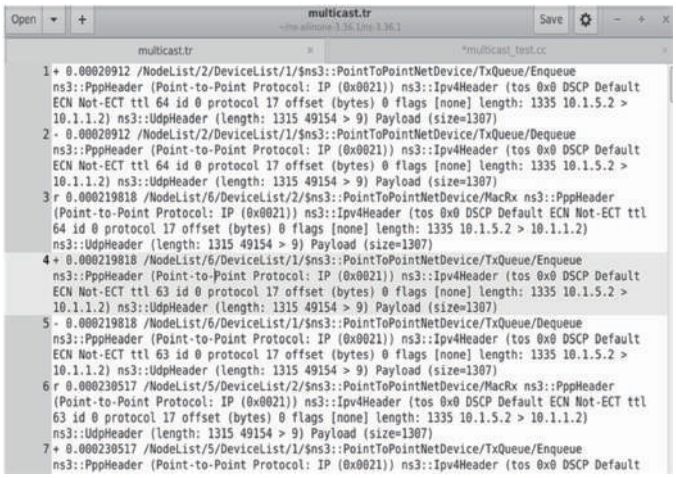


Fig. 5. Trace File Contents

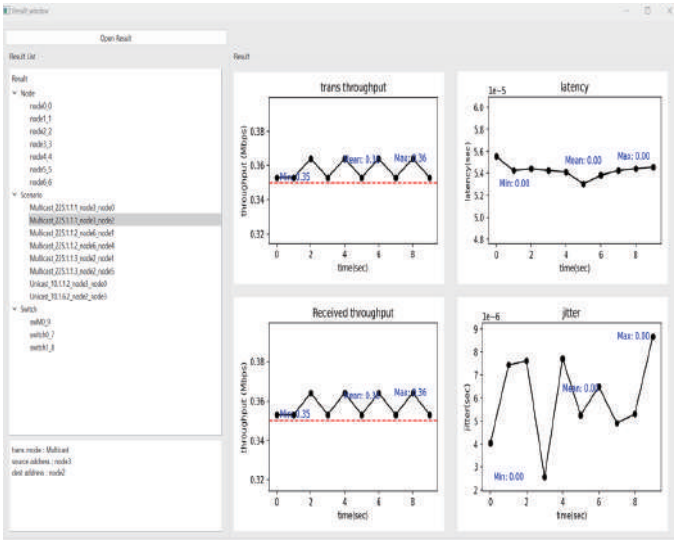


Fig. 6. Simulation Result

result values include Throughput, Goodput, Jitter, E2E delay, etc., which are organized by individual nodes and scenarios, allowing for visualization of changes over time. Throughput and Goodput are items extracted based on nodes and are processed by fetching length and payload values from trace files, respectively. Additionally, in the Trace file, operations entering the node's queue, leaving the queue, and arriving at the node are inherently categorized using symbols +, -, r. This allows for easy separation and organization of transmission and received data for the Throughput and Goodput of nodes. Latency and Jitter are items extracted based on scenarios, where outputs with matching Source IP and Destination IP in the Trace file are considered part of the same scenario. Subsequently, using the time values of lines printed together with the respective output results, Latency and Jitter are calculated.

IV. Conclusions

This paper presents the design and implementation of a simulator for evaluating the performan ce of naval combat system networks and processing units. Requirements for testing naval combat system were derived to guide the simulator design, and the implementation satisfied these requirements. Utilizing the implemented comprehensive simulator, simulations were conducted by configuring prototype networks and scenarios for naval combat system networks. The proper functioning of the simulator was verified. This study not only facilitates network design and performance analysis by providing simulation capabilities for optimizing the structure of next-generation naval combat system networks and processing unit specifications but also has educational applications. Additionally, the developed tool serves as a foundation that can be expanded for application in various network environments and scenarios in the future. In future research, we plan to enhance the tool's versatility and usability by expanding support for diverse network environments such as wireless networks, improving the user interface, and implementing additional features.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2018R1A6A1A03024003,2022R1I1A1A01071058)
and This research was supported by Hanwha Systems Co., Ltd.

References

[1] Chi-Sun, and Jin-Hyang Ahn. "A Study of the Standard Interface Architecture of Naval Combat Management System." Journal of The Korea Society of Computer and Information, vol. 26, pp. 147-154, Jan. 2021
[2] Seongwon Oh,"An Integrated Architecture for Control and Monitoring Systmes on Naval Surface Combatants" Journal of The KIMST, vol. 21,No. 1 pp. 103-114, Jan. 2018
[3] Jae-Woo Kim, Won-Jae Ryu, Jae-Min Lee, and Dong-Seong Kim. "Analysis of Network Simulation Tools for Military Network Design and Analysis" Communications of the Korean Institute of Information Scientists and Engineers, vol. 43(3), pp. 40-50, Mar. 2023
[4] E. Weingartner, H. vom Lehn and K. Wehrle, "A Performance Comparison of Recent Network Simulators," 2009 IEEE International Conference on Communications, Dresden, Germany, 2009, pp. 1-5,
[5] Cha Jung-hyuk, Kim Hyeong-jin, Lee Jae-min, and Dong-Seong Kim, " Message-oriented middleware research trends for smart defense distributed systems ", Journal of Information Science, vol.41, no.3, pp.32-39, Mar. 2023.
[6] George F Riley and Thomas R Henderson. The NS3 network simulator. In Modeling and tools for network simulation, pp 15–34. Springer, 2010
[7] Willman, Joshua M. "Creating GUIs with Qt Designer." Beginning PyQt: A Hands-on Approach to GUI Programming with PyQt6. Berkeley, CA: Apress, 2022. 217-258.
[8] Ns-3 Tutorials: Using the Tracing System. https://www.nsnam.org/docs/release/3.9/tutorial/tutorial_23.html

Visualization Scheme for Naval Combat System Network Simulator

Seung-Min Lee, Neul-Som Lee, Min-Hui Jang, Jae-Min Lee, Tae-Soo Jun, Dong-Seong Kim*
Dept. of IT Convergence Eng.
Kumoh National Institute of Technology
Gumi, Republic of Korea
{tmdals6164, alwaysom2, jmhee, ljmpaul, taesoo.jun, dskim*}@kumoh.ac.kr

***Abstract*—The initial naval combat system was designed to meet requirements such as navy performance, safety, survivability, and scalability. However, the next generation naval combat system requires design considerations beyond these requirements, including increased system efficiency, reduced system weight, improved operability, and enhanced security, among others. Therefore, it is difficult to meet these additional requirements with the initial design alone. Furthermore, sensor nodes in the distributed control network exchange large amounts of data in real-time, placing a burden on CPU processing capabilities and causing performance degradation and other issues. To address this, research on network performance analysis and optimization schemes is needed. While there are various network simulation tools available, NS-3, which offers advantages in data collection, visualization, and analysis of simulation results, is suitable for validating naval network designs. However, NS-3 has complex initial setup requirements and lacks visualization tools. Therefore, in this paper, we design a naval combat system scheme based on 3D network simulation. Through a 3D virtual environment, we can design a network model similar-to the real environment and provide a 3D-based GUI for effectively validating simulation results.**

Keywords—Naval Combat System, System efficiency, System weight reduction, Network performance analysis, 3D network simulation.

I. INTRODUCTION

The initial naval combat system was designed with a focus on meeting the requirements of the navy, such as performance, safety, survivability, and scalability. However, for the next generation naval combat system, design considerations go beyond fulfilling existing requirements; they also include enhancing system efficiency, reducing system weight, improving operability, and enhancing security, among others. Therefore, relying solely on the design of the initial naval combat system makes it challenging to meet these additional requirements.[1] Furthermore, various sensor nodes installed in the distributed control network of the naval combat system exchange real-time updates and large volumes of data with each information processing unit. This generates network traffic and places a burden on the processing capabilities of the CPUs mounted on multifunction consoles within the naval combat system. Consequently, this leads to issues such as performance degradation, maintenance-related problems, noise within the combat information center, and heat generation.[2] With the increasing need to address these issues by developing a new network structure and optimization schemes for naval combat systems, research on simulators and schemes for analyzing network performance is actively underway.[3] To conduct optimized network performance analysis, selecting appropriate tools from various network simulation tools already developed, which align with the requirements and characteristics of naval combat systems, and configuring the network simulation environment accordingly are crucial.

Therefore, in this paper, NS-3 was utilized to evaluate the network structure of naval combat systems. Additionally, a 3D-based visualization scheme is proposed by integrating it with the Unreal Engine. The proposed scheme utilizes a 3D virtual environment to design a network model similar to the real system environment and provides a GUI for visualizing simulation results. By doing so, we aim to ensure the reliability of this scheme while enhancing usability and intuitiveness, thereby establishing a concise and effective naval network development and performance analysis environment.

II. REQUIREMENTS AND TOOLS ANALYSIS FOR THE PROPOSED SCHEME

A. Requirements for Naval Combat System

To ensure high reliability and stability, the naval combat system must be designed to operate reliably in various scenarios, including external shocks and potential errors that may occur during operations. The naval combat system consists of various subsystems, with multiple nodes and servers distributed within each subsystem. The distributed nodes enable the integration and control of various sensors and equipment in complex naval combat systems, allowing for efficient management and access to large-scale data. However, the increase in the number of nodes brings about network complexity, leading to issues such as network congestion and data transmission delays. To address these challenges, it is necessary to support real-time data transmission for prompt and accurate responses to unexpected situations during operations. Additionally, enhancing the efficiency of data communication is crucial, especially in naval combat systems that require real-time processing of large-scale data.[4],[5] To meet these requirements, utilizing a DDS (Data Distribution Service) communication architecture is suitable for optimizing the complex communication structure of large-scale distributed systems. This approach ensures real-time performance, high reliability, stability, thereby satisfying the demands of the naval combat system.

B. Requirements for Network Simulation Visualization

For modeling and simulating network environments with user convenience in mind, visualization is a crucial element. Particularly, in 3D-based network simulation environments, realistic visualization offers the advantage of visually understanding network interactions. Requirements in such environments should be defined considering the accuracy, real-time performance, and user experience of network simulation. To simulate network operations accurately in real environments, realistic modeling such as network topology,

placement of equipment and sensors, etc., is essential in 3D-based network simulation environments. As network situations dynamically change according to real scenarios, it's crucial to reflect these changes in real-time to provide users with experiences similar to real environments during network operation simulation. Providing an interface for users to easily set up and execute network simulations is necessary, enabling users to manipulate the simulation environment easily and analyze results. Additionally, tools for visualization and data analysis are required to analyze network results and data.

To meet these requirements, integrating Unreal Engine with NS-3 could be a viable solution. While Unreal Engine was initially used for game development, its functionality has expanded over time and is now widely utilized across various fields. One of the reasons why many users prefer it is because Unreal Engine provides a Real-Time 3D creation platform, making it easier and faster to experience virtual environments.[6] Moreover, it offers high compatibility with NS-3 as it supports development in C++. By building a system that visualizes real-time network simulation results from NS-3 through Unreal Engine, it is possible to achieve real-time performance monitoring and visualization simultaneously.

C. Analysis of Network Simulation Tools

Network simulation tools are categorized into two methods depending on the network simulation approach: virtualization and discrete event simulation. Virtualization simulation method first creates virtual machines on top of the operating system kernel using virtualization technology, and then conducts simulation by configuring one network node for each virtual machine. The discrete event simulation method, on the other hand, involves programming network nodes for modeling and defining events occurring at each node to perform simulation.[7] Therefore, to establish a flexible network simulation environment, it is necessary to utilize the discrete event simulation method, which involves adjusting various scenarios and detailed settings through programming. There are various discrete event simulation tools such as OPNET (Optimized Network Engineering Tools), OMNet++, and NS-3. First, OPNET Modeler is a commercial software for network simulation, primarily used in military and defense industries. It provides a comprehensive development environment for specifying network standards, simulation, and performance analysis. OPNET offers various performance metrics for analyzing network performance and provides a GUI for configuring simulations and visualizing results. However, it requires high licensing costs, may have performance limitations for large-scale simulations, and demands significant computing resources.[7],[8] Next, OMNet++ is an open-source network simulator with a modular architecture containing a large library of various network models and simulation components. It offers high flexibility and scalability for modeling and simulating network requirements. However, it has a steep learning curve for initial setup, complex customization, and demands understanding of modeling, simulation, and analysis tools.[7],[9] NS-3 is a discrete event network simulator primarily used in education and research, providing an open environment for sharing software since it's open-source. It offers open and extensible network simulation, developed as an upgraded version of NS-2 with advantages such as reduced memory usage and faster simulation runtime. Each module is implemented in C++ except for OTcl, providing good scalability and allowing users to modularize and add required functionalities. Additionally, being a GPL open-source program, NS-3 allows free usage without separate licensing. However, being C++ code-based, it lacks intuitiveness, leading to user difficulties in some cases. Furthermore, visualization tools for displaying simulation results, such as NetAnim and PyViz, are still relatively limited.[7],[10]

Therefore, among various discrete event simulation tools, this paper chose NS-3 as the scheme to design for evaluating and optimizing the network structure of naval combat systems.

III. Proposed Scheme

A. Explanation of Overall Structural Design

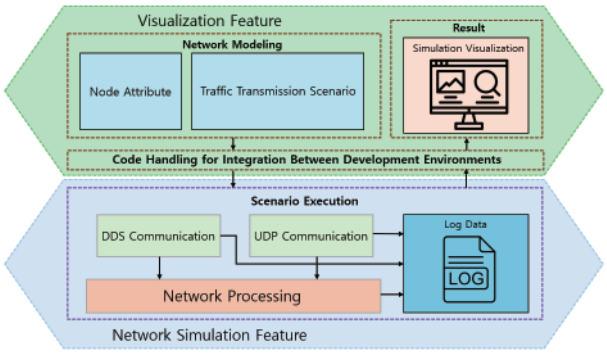


Fig. 1. The overall structure of the 3D-based network simulation scheme.

The overall structure of the 3D-based network simulation visualization scheme proposed in this paper for the naval combat system is illustrated in Fig. 1. The proposed architecture consists of visualization capabilities and network simulation functionality. Network modeling is a phase in forming the network topology of a naval battle system, where node properties and traffic transmission scenarios are configured in a GUI with visualization capabilities. Once the network modeling is complete, simulations are conducted based on the configured node properties and traffic transmission scenarios, utilizing NS-3 for code processing. Scenario execution is the stage where simulations, conducted after code processing, take place. Simulations progress through network processing via DDS communication and general UDP communication. Upon completion of the simulation, the generated log data undergoes code processing for visualization through the GUI. Subsequently, results are generated, including visualization elements such as tables, graphs, and animations, based on analyzed log data to evaluate the performance of the network.

B. Design of NS-3 and Unreal Engine Integration Method

The following explains the method of integrating simulation result data generated by NS-3 with Unreal Engine for 3D visualization. Both NS-3 and Unreal Engine support development based on C++. The C++ language provides high flexibility and extensibility, and since the same language is used to integrate the two development environments, complex processes such as syntax conversion or structural changes are not necessary. Efficient integration between NS-3 and Unreal Engine was achieved through parsing. There are two representative types of parsing: XML parsing and JSON parsing. JSON is more flexible and simpler than XML, can quickly analyze syntax, and has lowe5r data redundancy, so JSON parsing is utilized.[11] Here's an algorithm for integrating NS-3 with the Unreal Engine. First, create the data to be placed, such as nodes and switches, using buttons, and

then arrange them on the screen via drag-and-drop. After completing the network topology configuration, select the transmission status of each node and generate result data through network simulation in NS-3. Since NS-3 is based on C++, use the JSON for Modern C++ library to convert the generated data into JSON format. The converted JSON data is transmitted to Unreal Engine via DDS communication and UDP communication. In Unreal Engine, the received JSON data is converted into objects using TJSONReader and FJsonObject. TJSONReader is a tool for reading JSON data, and FJsonObject is an object for storing and manipulating JSON data. The parsed JSON data is then mapped to 3D objects in Unreal Engine. Through this process, network simulation data generated in NS-3 is visually mapped to 3D objects in Unreal Engine.

This algorithm describes the process of integrating NS-3 with the Unreal Engine. It involves collecting NS-3 simulation data, converting it into JSON format, and transmitting it to the Unreal Engine for mapping onto 3D objects. To generate NS-3 data, the NS-3 simulation needs to be executed, and simulation data needs to be collected. The run NS-3 Simulation command executes network scenarios and progresses the simulation for data generation. Subsequently, the collected simulation data is stored and returned via data = collectSimulationData(). To facilitate integration with the Unreal Engine, data collected from NS-3 simulation must be converted into JSON format. Utilizing createEmptyJSONObject(), an empty JSON object is initialized, and collected data is iterated through using a for loop to add each item to the JSON object. Each item comprises key-value pairs, which are appended to the JSON object using item.key for keys and item.value for corresponding values. Once all data is added, the JSON object is converted to a string using jsonObject.toString(), resulting in JSON data returned as jsonString. JSON data is transmitted using DDS or UDP communication methods. For DDS communication, DDS is initialized, and a ddsSubscriber is created. Subsequently, a ddsReader is generated via the Subscriber to receive JSON strings from DDS. Alternatively, if UDP communication is preferred, UDP is initialized, and a udpSocket is created to receive JSON strings using the generated socket. Upon receiving JSON strings, the ParseJSONData procedure parses them into JSON objects. To parse the input JSON string, a JSONReader is instantiated, which reads the JSON string, parses it into a JSONObject. If parsing is successful, the parsed JSON object is returned. Otherwise, an exception is raised with the message "Failed to parse JSON". Parsed JSON data is mapped to 3D objects in the Unreal Engine using the MapJSONDataTo3DObject procedure. The actor variable represents 3D objects in the Unreal Engine. To retrieve values from specific fields of the JSON object, x, y, and z coordinate values are extracted, indicating the position of the 3D object. Subsequently, the extracted x, y, and z coordinate values are used to set the location of the Unreal Engine's actor using the setLocation function.

Algorithm 1 Integrating NS-3 and Unreal Engine in C++

```
Input: NodeName, NodeType, SwitchName, SwitchType,
       DataSize, Communication Method, DataType,
       SendNode, ReceiveNode, LinkCapacity, LinkColor
Output: Graph, Throughput, Goodput

procedure GENERATENS-3DATA
    run NS-3 simulation
    data = collectSimulationData()
    return data
end procedure
procedure CONVERTDATATOJSON(data)
    jsonObject = createEmptyJSONObject()
    for all item in data do
        jsonObject.add(item.key, item.value)
    end for
    jsonString = jsonObject.toString()
    return jsonString
end procedure
procedure SENDJSONDATA(jsonString, method)
    if method == "DDS" then
        initializeDDS()
        ddsPublisher = createDDSPublisher()
        ddsWriter = createDDSWriter(ddsPublisher)
        ddsWriter.write(jsonString)
    else
        initializeUDP()
        udpSocket = createUDPSocket()
        udpSocket.send(jsonString)
    end if
end procedure
procedure RECEIVEJSONDATA(method)
    if method == DDS then
        initializeDDSReceiver()
        ddsSubscriber = createDDSSubscriber()
        ddsReader = createDDSReader(ddsSubscriber)
        jsonString = ddsReader.read()
    else
        initializeUDPReceiver()
        udpSocket = createUDPReceiverSocket()
        jsonString = udpSocket.receive()
    end if
    return jsonString
end procedure
procedure PARSEJSONDATA(jsonString)
    jsonReader = createJSONReader(jsonString)
    jsonObject = createEmptyFJsonObject()
    if jsonReader.read(jsonObject) then
        return jsonObject
    else
        throw Error("Failed to parse JSON")
    end if
end procedure
```

C. *Network Topology GUI Design*

The GUI design aims to simulate the complex network structures of naval combat systems in a realistic and interactive environment. This enables the evaluation of network performance to meet the unique requirements of naval combat systems, such as real-time data processing, high reliability, and operational efficiency. Therefore, this simulator is crucial for designing, testing, and optimizing naval combat systems to meet modern naval operations' rigorous standards. The visualization features proposed in this paper include creating network topology, processing integration code, and outputting simulation results. Network components such as nodes, links, and scenario information

undergo network modeling using the network topology creation GUI and are then passed to integration code processing in the form of classes defined by variables. The integration code processing utilizes these classes to construct NS-3 simulation execution code. Subsequently, NS-3 is executed to perform the simulation, and the log data generated by NS-3 is analyzed. The integration code processing refines this data through the functionality that outputs the simulation results and visualizes it via the simulation result output GUI.

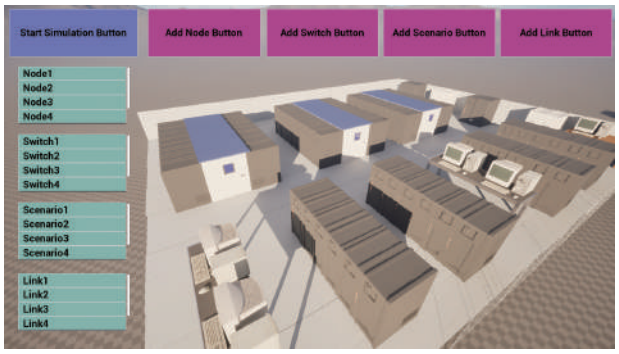


Fig. 2. Graphical User Interface (GUI) for Network Topology Creation.

Fig. 2. shows the design of a GUI that allows for the configuration of a network model in a 3D form by placing network components such as nodes and links. The Network Modeling View enables the construction of network topology in a virtual 3D space by dragging and dropping 3D models of network components onto the screen using various buttons. The Add Node Button allows users to create a node model by selecting the graphic type of the node and specifying detailed information such as the node's name and type. The Add Switch Button lets users generate a switch model by selecting the graphic type of the switch and setting details like the switch's name and type. The Add Scenario Button enables the creation of a desired scenario for experimentation, including communication method, sending node, receiving node, data transfer size, and data type. Lastly, the Add Link Button allows the creation of a link model by specifying the link's capacity, color, and dash style based on the capacity. Each List Panel displays a list of nodes, switches, scenarios, and links that have been created and applied in the Network Modeling View. The Start Simulation Button is linked to a function that analyzes the performance of the constructed network model. Users complete the network modeling through the node, switch, scenario, and link buttons and panels, and can easily proceed with network simulation by pressing the Start Simulation Button. By integrating NS-3 and Unreal Engine to provide a visualized GUI, users can more intuitively configure network topologies and visually analyze simulation results. This integration reduces user inconvenience caused by NS-3's complex setup and lack of visualization tools. Users can construct network models in a 3D environment and understand network interactions in real-time, reflecting dynamic changes to simulate actual network operations. This enables quicker network performance analysis and problem-solving. The simulation results are presented in tables, graphs, and animations, simplifying complex data analysis and improving result interpretation accuracy.

IV. Conclusion

In this paper, we designed a 3D-based visualization scheme through integration between NS-3 and Unreal Engine for evaluating the structure of naval combat system networks. NS-3 offers high scalability and flexibility, reflecting the characteristics of real networks to produce realistic simulation results. However, the lack of visualization tools compromises intuitiveness. Therefore, by integrating with the Unreal Engine, we addressed the lack of intuitiveness in NS-3's C++-based simulation and provided a visualized GUI. Through this, users can easily configure network topologies and efficiently analyze the results generated after simulation. As a future research direction, we plan to build a digital twin to make network simulation results even more realistic, and integrate it with the 3D-based visualization scheme implemented in this paper. This will enhance the accuracy and reliability of network simulation.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 33.3%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33.3%) and by the MSIT, Korea, under the ICAN support program(IITP-2024-00156394, 33.3%) supervised by the IITP.

References

[1] S.-M. Kwon and S.-M. Jung, “Virtualization based high efficiency naval combat management systemdesign and performance analysis,” Journal of The Korea Society of Computer and Information, Vol. 23, pp. 9-15, Nov. 2018.

[2] D.-K. Kim, D.-J. Min and D.-S. Kim, “Design and Implementation of Naval Combat System Using Virtualization Technology,” 2016 KICS Winter Conference, pp. 177-178, Jan. 2016.

[3] M.-H. Jang, J.-H. Lee, S.-H. Lee, J.-M. Lee and D.-S. Kim, “Design and Implementation of Simulatorfor Performance Evaluation of Naval Combat Network,”2023 KICS Summer Conference, pp. 1,272-1,273, June. 2023.

[4] G.-S. Kim, M.-H. Jang, J.-Y. Ryu, B.-I. Jung, J.-M. Lee and D.-S. Kim, “Edge Computing Based Framework for Next Generation Naval Ship Combat Systems,” The Journal of Korean Institute of Communications and Information Sciences, Vol. 47, pp. 2,078-2,085, Dec. 2022.

[5] C.-J. Choi, J.-W. Lee and D.-S. Kim, “Design and Implementation of Transmission Redundancy Schemefor Real-time Network of Naval Combat System,” Autumn Annual Conference of IEIE, 2018, pp. 298-300, Nov. 2018.

[6] M.-J. Oh and Y.-D. Seo, “A Study on the Creation Virtual Space using 2D Images based on Unreal Engine,” The Korean Institute of Broadcast and Media Engineers, pp. 175-176, Nov. 2022.

[7] J.-W. Kim, W.-J. Ryu, J.-M. Lee and D.-S. Kim, “An Analysis of the Network Simulation Tools for Military Network Design”, Communications of the Korean Institute of Information Scientists and Engineers, pp. 40-50, Mar. 2023.

[8] C. Smera and J. Sandeep, "Networks Simulation: Research Based Implementation using Tools and Approaches," 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), pp. 1-7, Oct. 2022.

[9] X. Chang, "Network simulations with OPNET," WSC'99. 1999 Winter Simulation Conference Proceedings. 'Simulation - A Bridge to the Future' (Cat. No.99CH37038), pp. 307-314, Dec. 1999.

[10] D.-B. Kim, S.-H. Lee and Y.-B. Ko, “NetVis: A New Visualization Tool for the ns-3 based Wireless Network Simulation,” Journal of KISS : Information Networking, Vol. 38, pp. 89-94, April. 2011.

[11] C. Sun, X. Zeng, C. Sun, H. Si and Y. Li, "Research and Application of Data Exchange based on JSON," 2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), p. 349-355, April. 2020

TwinMil: Semantic Segmentation-based Digital Twin Framework for Military Surveillance

Judith Nkechinyere Njoku *, Cosmas Ifeanyi Nwakanma †, and Dong-Seong Kim *
*IT Convergence Engineering, † ICT Convergence Research Center, Kumoh National Institute of Technology, Korea.

***Abstract*—This study proposes a novel framework that combines instance segmentation and digital twin technology to improve military surveillance. Integrating precise object identification with dynamic environmental modeling addresses critical surveillance challenges. This study presents early results of a digital twin system showcasing a military base with military personnel and an instance segmentation model for detection. This integration enables enhanced asset recognition and decision-making support, substantially advancing surveillance capabilities.**

***Index Terms*—AI, military, surveillance, visualization, digital twin**

I. Introduction

Surveillance is crucial for ensuring security and obtaining strategic advantages in military operations. Surveillance operations entail the detection and identification of entities within a designated facility [1]. Conventional surveillance systems, while effective to some extent, often struggle with the complexities of dynamic and densely populated environments. In these settings, distinguishing and tracking military assets or individuals among civilians presents significant challenges [2]–[4]. The accelerated development of artificial intelligence (AI) technologies, particularly semantic segmentation, and digital twin technology [5], has opened new avenues for improving the accuracy and reliability of surveillance systems [6]. Semantic segmentation, which entails the pixel-wise classification of images, enables the precise identification of objects within a scene [6], [7]. This capability renders it an indispensable tool for military surveillance applications. Moreover, semantic segmentation's usefulness extends beyond mere identification. By incorporating it within a digital twin framework—a virtual replica of a physical environment that mirrors real-time data and operations—its potential is amplified [5].

This paper proposes a novel strategy that merges the detailed insights provided by semantic segmentation with the comprehensive situational awareness made possible by digital twin technology. This synergy aims to enhance the accuracy of identifying military entities in surveillance footage and improve decision-making processes via real-time, data visualization and scenario simulation. By offering a framework for integrating these elements, this study seeks to address the pressing needs of contemporary military surveillance, providing solutions that cater to both immediate identification tasks and broader strategic planning and response mechanisms.

II. Related works and Research Justification

A. Semantic segmentation in surveillance

Recent advancements in semantic segmentation, particularly those utilizing convolutional neural networks (CNNs), have significantly improved object recognition accuracy in diverse environments. The work by Long et al. on fully convolutional networks (FCNs) for pixel-wise segmentation marks a pivotal advancement in this domain, enabling more refined scene understanding crucial for surveillance applications [8]. Additionally, the DeepLab series by Chen et al. introduced improvements with atrous convolution, enhancing the model's ability to capture finer details, which is particularly beneficial for identifying intricate patterns in military uniforms within surveillance imagery [9]. A study by Morocho et al. applied dilated convolutions and a modified encoder-decoder architecture to recognize the patterns of soldiers' uniforms [6]. This solution is particularly relevant when distinguishing authentic military personnel in a crowded setting.

B. Digital twins in defense applications

The concept of digital twins, initially popularized in manufacturing and aerospace, has found its way into defense applications, offering a new paradigm for operational planning and training. Studies by Tao et al. have outlined the foundational principles of digital twins, emphasizing their potential for real-time monitoring and simulation [10]. In military contexts, the adoption of digital twin technology, as discussed by Rosen et al., extends to strategic scenario simulations and equipment maintenance, highlighting the technology's versatility beyond traditional industrial applications [11]. A preliminary study by [12] explored and highlighted the potential of digital twins in military applications. Similarly, Mendi et al. [13] also reviewed the potential of digital twins in the military field. The digital twin holds much potential for surveillance, personnel training, and operational planning.

III. Methodology

This section presents the conceptual framework for the digital twin tailored for military surveillance, as illustrated in Fig. 1.

A. Digital twin framework for military surveillance

1) Design principles: The digital twin framework presented in this subsection was conceived to create a virtual replica of surveillance environments, capturing both static and dynamic elements with high fidelity. This involves:

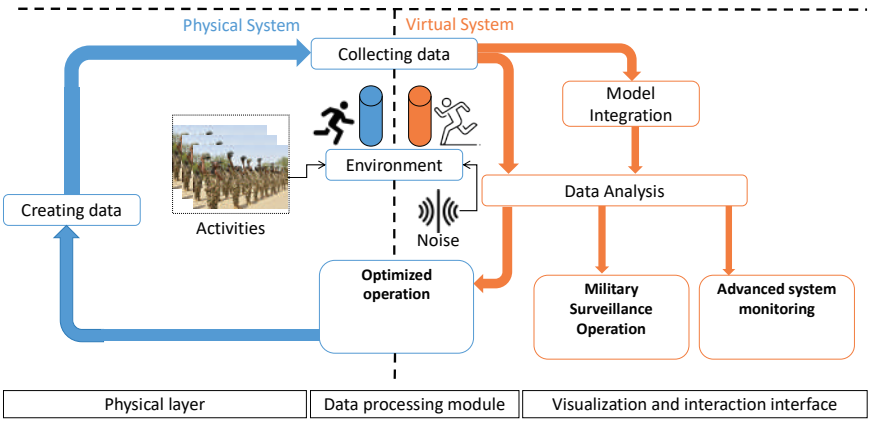


Fig. 1. Conceptual Framework of Digital Twin for military surveillance

- Data integration: Diverse data sources, including geographical information, architectural layouts, and real-time camera sensor data, can be used to construct a comprehensive digital model.
- Real-time synchronization: Implementing methods that enable continuous data exchange between the physical surveillance domain and its virtual counterpart, ensuring that the twin accurately reflects the current state.

2) Framework: The proposed framework of the digital twin is modular and comprises several interconnected components, including a physical layer, a data processing module, and a visualization and interaction interface.

- **Physical layer**: The primary objective of the physical layer is to collect all necessary data for developing a highly detailed digital twin model. Sensors, cameras and other IoT devices are deployed in the surveillance area to provide data to this layer.
- **Data processing module**: This module aims to process, and analyze the continuous stream of data flowing from the physical environment to the digital twin, using this data to update the twin in real time and extract valuable insights. Semantic segmentation algorithms are applied in this layer for detecting unusual activities or predicting potential security threats based on historical data.
- **Visualization and interaction interface**: The goal of this layer is to present a very detailed and accurate 3D virtual model that mirrors the physical surveillance area, including all relevant terrain, infrastructure, and dynamic entities.

The digital twin environment was developed using **Blender** and *Unity* development environments. The military personnel model was based on real military clothing and can be credited to *sketchfab.com*.

B. Semantic segmentation solution

1) Dataset preparation: The data employed in this study was sourced from a random search on the internet, using search terms such as **military**, **military uniform**, **soldier**, and **army**. Ultimately, a small sample size of 62 images was collected. These images were annotated using the segmentation annotation tool in *Roboflow*, as shown in Fig. 4. The dataset was preprocessed by applying Auto-Orient, cropping 25% - 75% off the vertical and horizontal regions, and a stretch of 640×640. Some data augmentation techniques were applied to increase the dataset by a scale of three, as illustrated in Table I. Thus, 129 images were used for training, 13 for validation, and 6 for model testing.

TABLE I
AUGMENTATION PARAMETERS

Augmentation	Parameter
Outputs per training example	3
Flip	Horizontal
Rotation	Between -15° and +15°
Shear	±10° Horizontal, ±10° Vertical
Hue	Between -15° and +15°
Brightness	Between -15% and +15%
Exposure	Between -10% and +10%
Blur	Up to 2.5px
Noise	Up to 0.1% of pixels

2) Model development: Instance segmentation is an innovative approach in computer vision that identifies objects within an image and highlights the precise pixel boundaries of each instance. This technique is valuable, especially where understanding the context and contours of individual items within the same class is essential.

In this research endeavor, the Yolov5 instance segmentation model was implemented. The Yolov5 model represents an enhancement to the detection architecture. In addition to the standard object detection head, there is a small, fully connected neural network (FCN) called ProtoNet. ProtoNet generates prototype masks for the segmentation model. It bears similarities to the FCN utilized for semantic segmentation tasks. ProtoNet is a three-layer network that comprises 2D convolutions with SiLU activation functions. Fig. 2 depicts the ProtoNet architecture and is included in the final instance

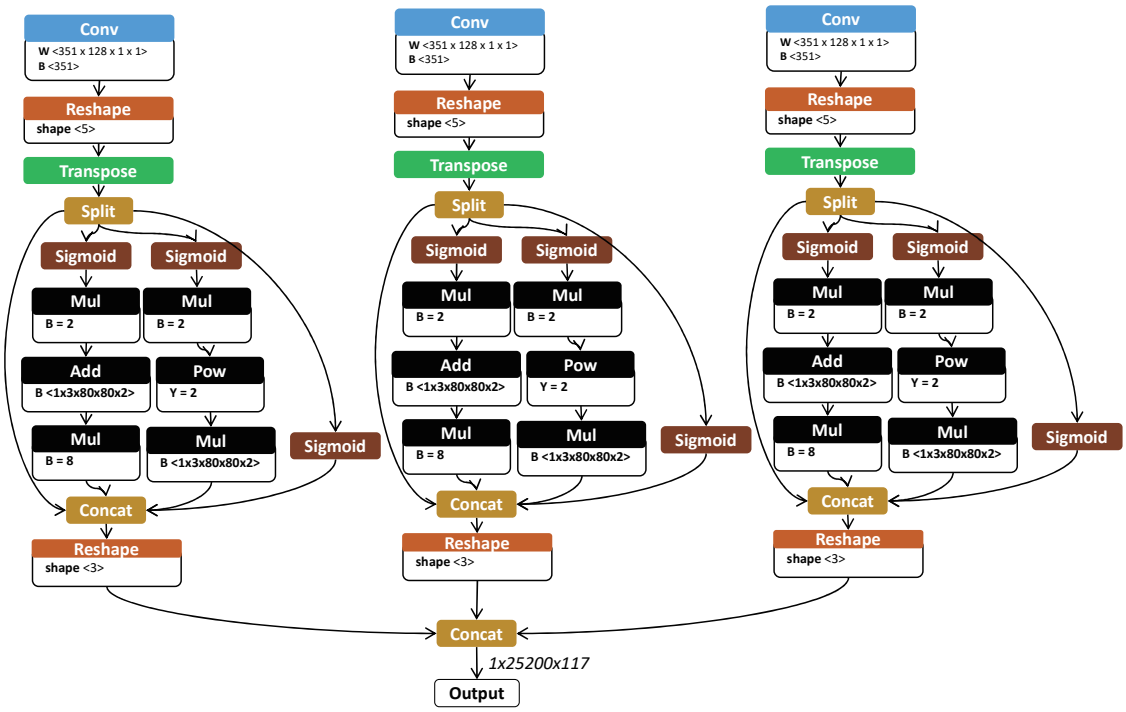


Fig. 2. YOLOv5 ProtoNet Architecture for Instance Segmentation

segmentation head in conjunction with the final 2D convolutional layers. The final convolution detection heads have 351 channels.

Three key performance metrics were employed in evaluating the model: mAP, precision, and recall. The mAP equates to the average of the average precision metric across all classes in the model; precision assesses the accuracy of the model's predictions, while recall measures the proportion of relevant labels that were effectively identified.

IV. Results and Discussion

Table II displays the YoloV5 instance segmentation model results. The metrics provided include mean Average Precision (mAP), precision, and recall for two model variants. The primary distinction between these two models is that they use different data augmentation techniques. Model version 1 utilizes all available augmentation techniques, while Model version 2 does not include blur and noise techniques. As

TABLE II
Performance Metrics Comparison Between Model Versions

Metrics	Model version 1	Model version 2
mAP	44.7%	51.8%
Precision	1.2%	71.3%
Recall	96.0%	56.0%



Fig. 3. Digital Twin environment illustrating the Military base (left), internal environment with a Soldier patrolling (center), and a close-up of a soldier (right)



Fig. 4. Sample results of instance segmentation using YOLOv5

observed, adding augmentation techniques has impacted the model's performance negatively. Fig. 5 also depicts the loss plot results for Model version 2. Fig. 3 represents the digital twin environment developed to highlight a military base with soldiers on patrol. The segmentation solution is designed to operate within this environment and accurately detect military personnel based on their uniforms.

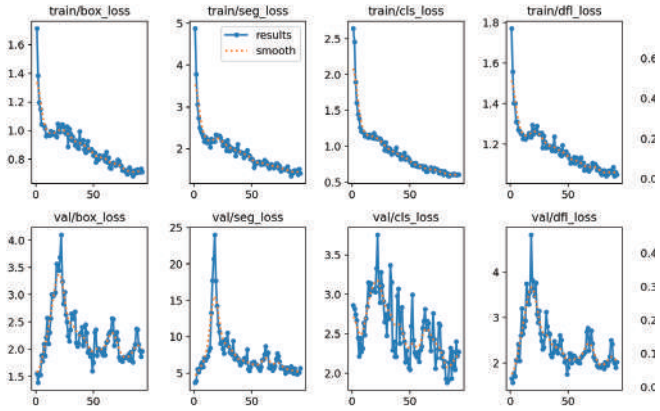


Fig. 5. Loss Plots of training results

V. Conclusion

This paper proposes a digital twin-aided surveillance framework for detecting military personnel in crowded scenarios, presenting the design principles and framework. An instance segmentation model based on Yolov5 is presented for detecting military personnel in the digital twin environment. Future research will focus on developing a more integrated system and testing the segmentation model within the digital twin environment.

Acknowledgments

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 50%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 50%)

References

[1] S. Band, R. Javeri, V. Kale, A. Morope, and S. Rudrawar, "Military surveillance system based on iot," in *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, 2022, pp. 1–6.

[2] S.-H. Liao, B.-L. Sun, and R.-Y. Wang, "A knowledge-based architecture for planning military intelligence, surveillance, and reconnaissance," *Space Policy*, vol. 19, no. 3, pp. 191–202, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0265964603000201

[3] J. N. Njoku, G. O. Anyanwu, I. S. Igboanusi, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "State-of-the-art object detectors for vehicle, pedestrian, and traffic sign detection for smart parking systems," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 1585–1590.

[4] J. N. Njoku, G. C. Amaizu, J.-M. Lee, and D.-S. Kim, "Real-time deep learning-based scene recognition model for metaverse applications," in *Proceedings of the KICS Winter Conference, Pyeongchang, South Korea*, 2022, pp. 195–198.

[5] J. N. Njoku, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Model comparison and selection for battery digital twin development using pybamm," in *The 33rd Joint Conference on Communication and Information (JCCI), Yeosu, South Korea*, 2023, pp. 1–2.

[6] M. E. Morocho-Cayamcela and W. Lim, "Pattern recognition of soldier uniforms with dilated convolutions and a modified encoder-decoder neural network architecture," *Applied Artificial Intelligence*, vol. 35, pp. 476 – 487, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:233302851

[7] S. B. Nemade and S. P. Sonavane, "Semantic segmentation using gsaunet," *ICT Express*, vol. 9, no. 1, pp. 1–7, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405959522001400

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[9] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[10] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9, pp. 3563–3576, Feb 2018. [Online]. Available: https://doi.org/10.1007/s00170-017-0233-1

[11] R. Rosen, G. von Wichert, G. Lo, and K. D. Bettenhausen, "About the importance of autonomy and digital twins for the future of manufacturing," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 567–572, 2015, 15th IFAC Symposium onInformation Control Problems inManufacturing. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896315003808

[12] S. Li, Q. Yang, J. Xing, and S. Yuan, "Preliminary study on the application of digital twin in military engineering and equipment," in *2020 Chinese Automation Congress (CAC)*, 2020, pp. 7249–7255.

[13] A. F. Mendi, T. Erol, and D. Doğan, "Digital twin in the military field," *IEEE Internet Computing*, vol. 26, no. 5, pp. 33–40, 2022.

Modular Open System Approach (MOSA) for Defense Unmanned Systems

Min-Seon Lee, Jae-Min Lee, Tae-Soo Jun, Dong-Seong Kim*
Dept. of IT Convergence Eng.
Kumoh National Institute of Technology
Gumi, Republic of Korea
{20236117, ljmpaul, taesoo.jun, dskim*}@kumoh.ac.kr

***Abstract*— The U.S. defense unmanned systems face numerous limitations due to inefficiencies such as technological redundancies resulting from the development of various small-scale and diverse unmanned systems, as well as difficulties in maintaining operations due to high-performance requirements for achieving complex missions. In response, the U.S. Department of Defense has introduced and mandated the implementation of the Modular Open System Approach (MOSA) to achieve significant cost savings, schedule reductions, rapid adoption of new technologies, technological upgrades, interoperability between systems, and mission integration, among other benefits. Recently, the South Korean Ministry of National Defense has been pursuing a policy of defense unmanned system standardization and modularization, known as the Korea-tailored Modular Open System Approach (K-MOSA). Therefore, this study aims to explore the research trends for the advancement of South Korea's defense unmanned systems through an analysis of the Modular Open System Approach (MOSA) in the United States.**

Keywords— Unmanned defense system, MOSA, K-MOSA

I. INTRODUCTION

The United States Department of Defense, the world's largest defense organization, annually allocates approximately $842 billion for technology acquisition. However, the U.S. defense unmanned systems face numerous limitations such as inefficiencies resulting from the development of various small-scale and diverse unmanned systems, as well as challenges in maintaining operations due to high-performance requirements for achieving complex missions [1]. To address these challenges, the U.S. Department of Defense introduced and mandated the implementation of the Modular Open System Approach (MOSA) through the Better Buying Power 3.0 guidelines in 2015. Additionally, the Department of Defense Tri Services memo, released in 2019, explicitly states that as many modular open system approach (MOSA) principles as possible should be included in all requirements, programming, and development activities for future weapon system upgrades and new start development programs. MOSA, previously known as Modular Open System Architecture, is a business and technical strategy for developing new systems or modernizing existing systems. By applying MOSA, substantial cost savings, schedule reductions, rapid adoption of new technologies, technological upgrades, interoperability between systems, and mission integration, among other benefits, can be achieved. Thus, MOSA is a prioritized method for implementing open systems within the U.S. Department of Defense (DoD) and is mandated by U.S. law[2].

South Korea faces increasing uncertainties in the global order and growing complex security threats. As an alternative to manpower resources amid the spread of a value for human life and decreasing military personnel, unmanned systems are gaining prominence[3]. In 2022, the government initiated a concept study on intelligent unmanned and manned hybrid combat systems and further developed defense unmanned system development plans. To enhance these efforts, the Ministry of National Defense drafted the Korea-tailored Modular Open System Approach (K-MOSA) policy in December 2023. As of May 2024, the Advanced Power Planning Division of the Ministry of National Defense is in the early stages of conducting research to implement and establish the foundation for the K-MOSA policy, aiming to enhance interoperability, reduce costs, and achieve rapid electrification of defense unmanned systems. By referencing successful outcomes from various weapon systems achieved through MOSA in the United States, the South Korean Ministry of National Defense can minimize trial and error and introduce effective strategies. Moreover, adhering to MOSA standards can ensure international interoperability of South Korea's weapon systems, enabling them to be effectively utilized in joint operations. Against this backdrop, the South Korean Ministry of National Defense emphasizes the necessity of analyzing and researching the application process of the U.S. MOSA policy for the defense unmanned system standardization and modularization (K-MOSA). This paper introduces the U.S. Modular Open System Approach (MOSA) and examines the technical characteristics and application cases of standards such as FACE (Future Airborne Capability Environment) and SOSA (Sensor Open Systems Architecture) that are implementing MOSA. Based on this, it presents the research trends of applying U.S. MOSA for the standardization and modularization of South Korea's defense unmanned systems.

II. OVERVIEW OF MOSA

MOSA stands for Modular Open Systems Approach, a strategic standardization plan pursued by the United States Department of Defense (DoD). It demands that the thousands of products and systems acquired by the military sector annually operate more productively together. MOSA emphasizes the ability to integrate the latest technologies and products while being able to advance systems and subsystems affordably and continuously. In practice, the U.S. Department of Defense is investing significant efforts into applying the MOSA methodology to defense acquisition programs. DoDD 5000.1

specifies that to be managed using a systems engineering approach aimed at optimizing the performance of future weapon systems and minimizing total ownership costs, the adoption of the Modular Open Systems Approach (MOSA) should be maximized. This approach can be applied to all weapon systems and has significant implications, particularly for unmanned systems that can adapt to changes in social environments such as population decline and shifts in the nature of warfare. South Korea's defense unmanned systems are mainly composed of various platforms performing missions such as surveillance, reconnaissance, strike, communication relay, electronic warfare, and transportation. However, South Korea's unmanned systems suffer from duplication in platform and technology development, as they are pursued independently by each military branch (Army, Navy, Air Force) without reflecting the rapid technological advancements of the Fourth Industrial Revolution and without a centralized control tower function or roadmap. Moreover, reliance on specific suppliers results in high costs, and if the production and supply of these suppliers are disrupted, it becomes difficult to replace parts, leading to shortened equipment lifespan and increased operational maintenance costs[4]. Given that unmanned system platforms are operated in diverse environments and situations, high levels of interoperability and rapid technological upgrades are essential[5]. These requirements align with the goals of MOSA, such as promoting system integration, mitigating risks associated with single-source system lifespan, cost savings in maintenance, and interoperability. Therefore, applying MOSA to South Korea's defense unmanned systems is deemed appropriate, and the benefits that can be gained are outlined in Table 1.

TABLE. 1. BENEFITS OF MOSA (MODULAR OPEN SYSTEM APPROACH)

Benefit	Description
Cost savings	Optimizes development costs by reusing modular components and leveraging multiple vendors
Schedule reduction and rapid deployment of new technology	Fast integration of new technologies or features with common standards and modular architecture
Opportunities for technical upgrades and refresh	Periodic replacement and upgrade of some technologies in the system without replacing the entire system
Interoperability	Integrate open architecture, modularize components and systems, allowing components to flexibly adapt and scale systems through a variety of vendors

Through the application of MOSA, various fields such as communication systems, military vehicles, weapon systems, aircraft, and maritime systems can enhance interoperability and achieve cost savings. Furthermore, these modularized systems provide flexibility to adapt to potential future changes, thereby extending the lifespan of systems and improving performance. Therefore, MOSA is expected to play a crucial role in developing and implementing innovative technologies and products in the military sector.

III. SYSTEM DESIGN METHODOLOGY THROUGH MOSA

To successfully apply MOSA (Modular Open Systems Approach), systems must be designed to adhere to five key requirements. The requirements are as shown in Figure 1. Firstly, it is necessary to establish an enabled environment. Program

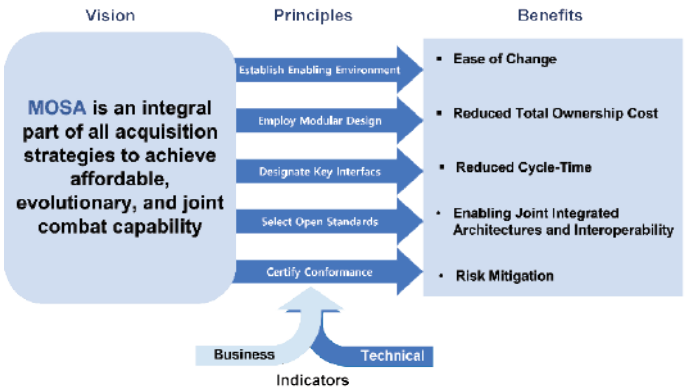


Fig.1. Overview of MOSA (Modular Open Systems Approach)

managers (PMs) must establish support requirements, business practices, technology development, acquisition, testing and evaluation, and product support strategies necessary for the effective development of open systems. Secondly, modular design must be adopted through consistency, encapsulation, independence, and modularization. While simple systems are relatively easy to design and acquire, more complex systems exhibit multiple levels of granularity. Current defense-scale unmanned systems are characterized by increasing components and complex interdependencies. Therefore, designing these systems in a modular fashion, decomposing them into loosely coupled subsystems, facilitates system development, maintenance, modification, and upgrade. Thirdly, interfaces should be grouped into major and minor interfaces. Classifying elements such as technically stable modules versus unstable modules, high-reliability modules versus low-reliability modules, and modules conveying less impactful interoperability versus those conveying critical interoperability information maximizes lifecycle. This enhances compatibility between systems and optimizes interactions between modules. Fourthly, open standards must be utilized. Interface standards must specify the physical, functional, and operational relationships among various elements, including hardware and software, and well-defined, mature, widely used, and readily available interface standards make designs more flexible and adaptable. Therefore, considering open standards, de facto standards, government standards, and proprietary standards in that order is common practice, although in some cases, proprietary standards may be more appropriate. Finally, certification and testing plans for compliance with internal and external open interfaces of system and component modules are required, and thorough compliance testing ensures verification and validation.

These requirements lay the foundation for the successful implementation of MOSA, maximizing the efficiency of system development and maintenance. The application of MOSA acts as a crucial element contributing to the achievement of defense strategic objectives, extending beyond merely a technical approach.

IV. ANALYSIS OF MOSA APPLIED STANDARDS

Unmanned systems in Korea play a pivotal role in the defense sector[6], and standardization is crucial for their efficient operation[7]. Among the key concepts to maximize this efficiency is MOSA, which continues to be applied in the U.S.

defense sector. Understanding the concept and system design methods of MOSA, followed by analyzing cases where it has been applied, can provide concrete evidence of its benefits and offer strategic insights into its application in Korea's defense unmanned system standardization and modularization policy (K-MOSA).In January 2019, a memo titled "Modular Open Systems Approach for our Weapon Systems is a Warfighting Imperative," issued by the secretaries of the U.S. Air Force, Army, and Navy, expressed concerns about the lack of interoperability among defense department systems. To address this concern, various standards were recommended. Key standards mentioned in this memo include Open Mission Systems (OMS) / Universal Command and Control Interface (UCI), Sensor Open Systems Architecture (SOSA), Future Airborne Capability Environment (FACE), and Vehicular Integration for Command, Control, Communications, and Computers (C4) C4ISR/Electronic Warfare (EW) Interoperability (VICTORY). These standards emphasize modularity and openness, contributing to the enhancement of unmanned system performance and flexibility. To align with Korea's defense unmanned system standardization and modularization policy (K-MOSA), analyzing the application of MOSA in the U.S. can facilitate the adoption of effective strategies in weapon system development, reducing trial and error and ensuring international interoperability[8].

This paper focuses primarily on the cases of the FACE and SOSA standards among various standards. FACE adheres most closely to MOSA principles among the recommended standards and is recognized as an essential technical standard for U.S. military operations. Similarly, the SOSA standard was initially developed as part of the FACE Consortium and integrates specifications for electrical and mechanical interfaces constituting SOSA sensor elements, applicable across all domains.

4.1 FACE (Future Airborne Capability Environment)

The Future Airborne Capability Environment (FACE) provides standards for interoperability and portability among airborne electronic equipment systems. Historically, the United States has produced manned aircraft for single-mission purposes through fixed suppliers, but modifying these systems entails long lead times and high costs. This is no longer suitable in the new era transitioning to more complex unmanned systems where software and system complexity is heightened. In response, FACE facilitates the reuse of software components in the airborne operational system, reducing development costs and enhancing operational capabilities. The FACE architecture, as depicted in Figure 2, is structured in a layered approach, allowing for the rapid replacement of software components as needed. It consists of five major segments: the Operating System Segment (OSS), Input/Output Services Segment (IOSS), Platform-Specific Services Segment (PSSS), Transport Services Segment (TSS), and Portable Components Segment (PCS). These segments are designed to promote problem isolation and establish a product line approach for reusable software functions. FACE provides third-party tools for seamless development (Software Life Cycle, Data Modeling, Conformance Support) and divides each segment into a modular structure to apply the architecture effectively. Furthermore, it distinguishes between major interfaces using open standards in OSS, IOSS, and TSS,

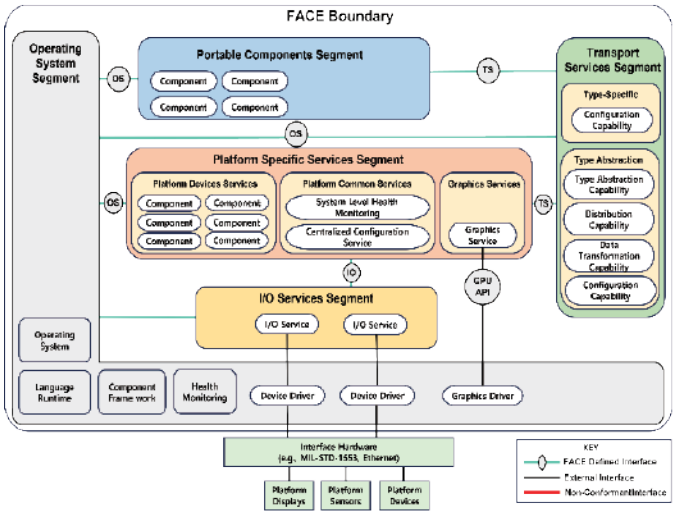


Fig.2. FACE (Future Airborne Capability Environment) Integrated Approach

as well as proprietary standards in PSSS and PCS. FACE already utilizes the widely adopted ARINC 66 standard and distributes Excel files outlining compliance program (CTS) operations and essential requirements in each segment. Thus, FACE effectively applies the principles of MOSA to enhance the efficiency of airborne operational systems and ensures the ability to flexibly adapt to the introduction of new airborne electronic equipment.

In Korea, research on hardware and software standards is underway to enhance the performance and reduce costs of aircraft electronic equipment [9]. Furthermore, research is also being conducted on proposing unmanned aircraft systems based on the Future Airborne Capability Environment (FACE) to address interoperability, portability, and modularity issues [10]. As a result, the Korean defense unmanned system can utilize a modularized structure through the application of FACE, enabling cost and time savings through the reuse of software components. Additionally, integrating standardized interfaces for various equipment enhances interoperability between systems and improves flexibility.

4.2 SOSA(Sensor Open Systems Architecture)

The Sensor Open Systems Architecture (SOSA) establishes modular entity architectures and standards for configuring sensors, including electronic-optical/infrared (EO/IR), electronic warfare (EW), radar, and signal information modes. SOSA aims to efficiently enhance the U.S. Department of Defense C5ISR (Command, Control, Computers, Communications, Cyber, Intelligence, Surveillance, and Reconnaissance) capabilities by integrating rapid and cost-effective sensor capabilities across various platforms, thereby reducing development costs and rapid field deployment times. SOSA is a collaborative effort among the Army, Navy, Air Force, other government agencies, industry, and academia to define best practices and standards for open modular sensor systems. It provides solutions based on best practices in system, software, hardware, electrical, and mechanical engineering and principles of Modular Open Systems Approach (MOSA). Furthermore, the SOSA architecture adopts modular design, defining each component individually and resolving requirements for integrated sensor function sets quickly and

inexpensively through standardized interfaces, enabling cross-platform functionality enhancement.

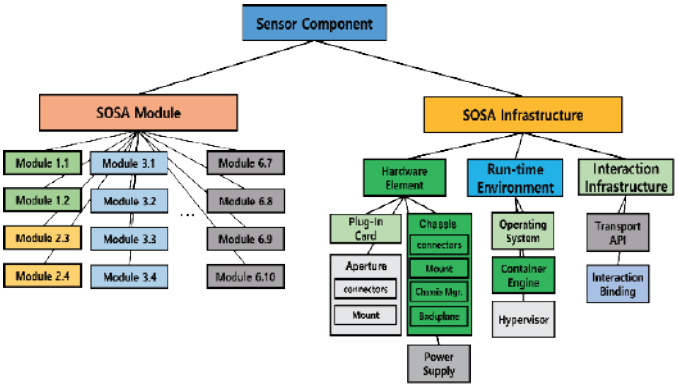


Fig.3. SOSA(Sensor Open Systems Architecture) Taxonomy

As depicted in Figure 3, SOSA sensor components are categorized into logical building blocks: SOSA modules, implementing specific functions and interfaces, and SOSA infrastructure providing platforms to execute SOSA modules. The SOSA architecture facilitates modularity and upgradeability by providing APIs for sensor interaction and standard interaction infrastructure, and products meeting SOSA compliance must undergo certification through testing by SOSA validation authorities. Thus, SOSA effectively applies the principles of MOSA to increase interoperability with sensor technologies acquired by the U.S. Department of Defense, provide cost-effective embedded solutions, and offer the advantages of integrating, repairing, or replacing commercial off-the-shelf (COTS) sensor technologies easily and quickly.

The Korean defense unmanned systems are implemented with a variety of equipment and sensors, leading to increasing complexity and connectivity in the software needed for communication between heterogeneous sensors and equipment[11]. To address these issues, applying the Sensor Open Systems Architecture (SOSA) can enhance interoperability, reuse, and rapid technology adoption by ensuring seamless integration of various sensors and electronic systems.

V. Conclusion and Future Works

This paper examines the increasing demand for interoperability enhancement, cost-effectiveness, and rapid empowerment of unmanned systems, leading to the growing necessity of utilizing the Korea-tailored Modular Open System Approach (K-MOSA) policy for improving the efficiency of defense unmanned systems. The reasons why MOSA is gaining attention in the unmanned systems field and an analysis of the trends in the application process of MOSA policies in the United States are explored. The concept and system design methodology of MOSA, as well as examples of standards following MOSA, are introduced.

Prior to the application of standardized open systems, limitations such as low interoperability, duplication of development, and increased maintenance costs existed due to the absence of system modularization and component standardization in weapon systems. However, with the active promotion of the Korean Ministry of National Defense's unmanned systems standardization and modularization policy, efforts are being concentrated on overcoming these limitations. The Korean Ministry of National Defense plans to establish common platform standards through research on the adequacy review and optimization of 16 representative platforms. In particular, research and development based on MOSA by the U.S. Department of Defense are actively underway, which is expected to draw attention to research on modular open system approaches for unmanned systems, enabling the Korean Ministry of National Defense to minimize trial and error and adopt effective strategies.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 33.3%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33.3%) and by the MSIT, Korea, under the ICAN support program(IITP-2024-00156394, 33.3%) supervised by the IITP

References

[1] D.-S. Park, and K.-W. Oh , “Study on Development of Korean Unmanned Systems through Analysis of U.S. Unmanned Systems Policy,” Journal of Aerospace System Engineering, vol. 15, no. 3, pp. 65-70, Jun. 2021.

[2] P. Zimmerman, "Considerations and examples of a modular open systems approach in defense systems," The Journal of Defense Modeling and Simulation : Applications, Methodology, Technology, vol. 16, no. 4, pp. 373-388, Apr. 2019.

[3] I.-R. LEE, “A research on the survey & analysis of the level of defense science & technology and its implications for Unmanned Ground System,” Journal of the Korea Academia-Industrial cooperation Society, vol. 25, no. 3, pp. 475-482, Mar. 2024.

[4] T.-Y. Lee, J.-D. Lee, D.-S. Son, M.-H. Jang, J.-M. Lee, and D.-S. Kim, “Design and Implementation of Inspection Equipment for KDX Hull Sonar Pre-Amplification PCB,” 2023 KICS Winter Conference, pp. 424-425, Feb. 2023.

[5] D.-S. Kim, “Establishment of an environment and suggestions for Expansion of National Defense and Logistics ICT convergence technology,” The Journal of The Korean Institute of Communication Sciences, vol. 41, no. 2, pp. 3-11, Jan. 2024.

[6] S.-H. Park, and S.-P. Namgoong, "A Study on the Future Army's Development of Combined Combat System with Manned and Unmanned Forces ," The Journal of the Convergence on Culture Technology (JCCT), vol. 9, no. 2, pp. 295-299, Mar. 2023.

[7] D.-H. Han, J.-Y. Hwang, J.-H. Bae, and J.-H. Lee, “Design of Open System Architecture and Data Model for Upgrading Ground based Weapon Syste,” The Journal of Korean Institute of Information Technology, vol. 19, no. 12, pp. 151-161, Dec. 2021.

[8] S.-W. Baik, H.-Y. Moon, S.-S. Kim, S.-H. Kim, Y.-J. Jeon, and J.-H. Kim, “Open Architecture-based Advanced Serialization Approach for LTVs,” Journal of the Military Operations Research Society of Korea, vol. 49, no. 3, pp. 30-44, Dec. 2023.

[9] S.-H. Kim, J.-W. Kim, and D.-S. Kim, “Energy Consumption Analysis of Beamforming and Cooperative Schemes for Aircraft Wireless Sensor Networks,” Applied Sciences, vol. 10, no. 12, pp. 4374-4391, Jun. 2020.

[10] H.-C. Jo, and K.-Y. Park, “Development and Verification Methodology for Small Civil Unmanned Aerial Vehicle System based on Open System Architecture,” JOURNAL OF PLATFORM TECHNOLOGY, vol. 8, no. 2, pp. 32–43, Jun. 2020.

[11] J.-W. Lee, H.-J. Kim, J.-M. Lee, T.-S. Jun, and D.-S. Kim, “Blockchain-Based Data Sharing Scheme to Enhance Reliability and Security for Naval Combat System,” The Journal of Korean Institute of Communications and Information Sciences, vol. 47, no. 6, pp. 809-817, Jun. 2022.

Deep Learning-Based Classification of Laguerre-Gaussian OAM Beams in Atmospheric Turbulence and Thermal Convection

Young Jae Moon
Department of Infromation and Communications Engineering
Pukyong National University
Busan, Republic of Korea
8068joshua@pukyong.ac.kr

Yeon Ho Chung
Department of Information and Communications Engineering
Pukyong National University
Busan, Republic of Korea
yhchung@pknu.ac.kr

***Abstract*—This paper discusses a free-space optical communication system that uses a Laguerre-Gaussian beam carrying orbital angular momentum. When transmitted through the optical link, intensity and phase figures of the beam are distorted because of atmospheric turbulence and high-thermal heat haze. We analyze the distorted original images and demonstrate the regeneration process with the help of a convolutional neural network system. Based on these results, we aim to simulate more realistic transmission conditions and produce empirical outcomes by implementing various optical communication scenarios.**

***Index Terms*—Atmospheric turbulence, Convolutional neural network, Free-space optical communication, Heat haze, Orbital angular momentum**

I. Introduction

In the recent big-data era, there is a demand for immediate access to the data for application software that requires high data capacity. However, traditional radio frequency (RF) network technologies face difficulties meeting this requirement as they have limited bandwidth due to the few dominant paths in propagation [1]. In contrast, free-space optical (FSO) communication has emerged as a promising alternative with optical bands' high bandwidth and directivity [2].

Subsequently, communication techniques containing optical orbital angular momentum (OAM) have gained considerable attention in recent years. The photon improves the optical transmission efficiency rate; therefore, the OAM communication system theoretically has infinite data capacity and fulfills tera-bit per second data transmission [3].

However, if the OAM beam is affected by multiple layers of air, such as convective turbulence in the heat haze, the light beam would be scattered or reflected throughout the air [4] and distorted. In this paper, we introduce an orbital angular momentum image regeneration and mode detection technique against air-based convection and atmospheric turbulence.

The paper is organized as follows: Section 2 provides background information on OAM-based FSO communication and thermal air convection. Section 3 discusses the experimental design with the convolutional neural network (CNN) topology and expected budget. Section 4 covers the regeneration evaluation process. The conclusions are drawn in Section 5.

II. Background Information

A. OAM-Based FSO Communication

The topological charges of OAM served in the Laguerre-Gaussian (LG) beam are the exact answers to the cylinder-coordinated Huygens-Fresnel integral, due to its paraxial approximation. Therefore, most OAM communication uses LG beams. While the LG multi-moded beam propagates along the propagation axis, its electric field distribution is given by (1) [5]:

$$u(r,\phi,z) = U_1 U_2 \sqrt{\frac{2p!}{\pi(p+|m|)!}} \frac{1}{w(z)} [\frac{r\sqrt{2}}{w(z)}]^{|m|} [\frac{2r^2}{w^2(z)}] \quad (1)$$

where U_1 and U_2 are (2) and (3), respectively.

$$U_1 = \exp[\frac{-r^2}{w^2(z)}] \exp[\frac{-ikr^2 z}{2(z^2+z_R^2)}] L_p^m \quad (2)$$

$$U_2 = \exp[i(2p+|m|+1)\tan^{-1}(\frac{z}{z_R})] \exp(-im\phi) \quad (3)$$

where r and ϕ are the respective radial and angular coordinates, and $k = 2\pi/\lambda$ is the wave number in radians per meter, with wavelength λ. $w(z)$ is the beam waist size at distance z, while z_R is the Rayleigh range and L_p^m is the Laguerre polynomial with positive radial mode p and angular mode m, where they determine phase and intensity distribution.

When considering LG beams, the orthogonality and completeness of OAM allow the independence of multiple data streams. This orthogonality of different OAM beams with different mode numbers can be described along the integration between signal u_m and conjugate of another signal u_n (4) [6]:

$$u_m(r,\phi,z)u_n^*(r,\phi,z) \triangleq \int u_m(r,\phi,z)u_n^*(r,\phi,z) r dr d\phi \quad (4)$$

where the result is always zero when $m = n$ is ensured.

B. Thermal Air Convection

Turbulence from thermal air convection can distort an LG beam phase and intensity structure while propagating throughout the free space [7]. This turbulence includes the distortions

caused by the refraction and distortion of the mirage. The core formula of thermal convection is the unsolved Navier-Stokes equation with Newtonian fluid [8], given by

$$\frac{\partial}{\partial t}(\rho u) + \nabla \cdot (\rho u \otimes u + pI) = \nabla \cdot \tau + \rho g \tag{5}$$

where u is the flow velocity, t is time, $\frac{\partial}{\partial t}u$ is the material derivative, ρ is the density at a given point of the continuum, τ is the shear modulus of elasticity, I is the identity matrix, g is gravitational acceleration, and $\otimes$ is tensor multiplication.

Alternatively, we implement a distortion equation when LG beams pass through heat haze by using natural convection equation [9] with Newtonian fluid for experimental purposes

$$\frac{h \cdot L}{k} = C(\frac{\rho^2 g_c \beta \Delta T L^3}{\mu k})^n \tag{6}$$

In this formula, β is the thermal expansion coefficient, g_c is the gravitational acceleration, h is heat transfer coefficient, k is air thermal conductivity, L is plate length, ρ is air density, ΔT is air-surface temperature difference, C and n are empirical constants, and μ stands for dynamic air viscosity.

III. Simulation Setting

We first generate the distorted phase of the double-moded LG beam while propagating heat haze. The LG beam used for the simulation has a wavelength of 0.6238 um, $w(z)$ is 250 um, and $z = 0$. We generate original five OAM phases when the angular mode m ranges from 1 to 5, while the transverse mode p is 1.

Subsequently, we model a probabilistic model for the FSO communication link, shown in Fig. 1. Considering its properties with heat haze, we consider the optical link length to be 100 meters and the link attenuation to be 10 dB/km. Therefore, we expect the received optical power to be -60 dB in a realistic experiment [10], with the plate temperature set to 60 degrees Celsius. For the heat haze, β is $3.4 \cdot 10^{-3}$, g_c is $9.81 m/s^2$, k is 28.79 mW/m K, ρ is 1.06 kg/m^3, and μ is $18.97 \cdot 10^{-6} Ns/m^2$.

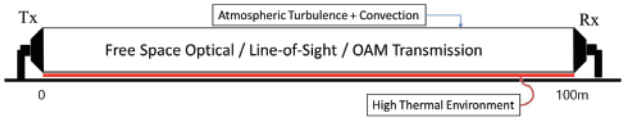


Fig. 1. Overall environment of the probabilistic model used in simulation.

As the settings for the optical communication link are finalized, we generate a convolutional neural network algorithm with 20 layers to regenerate distorted media against heat haze and atmospheric turbulence. The algorithm topology for regenerating distorted media is shown in Fig. 2.

IV. Evaluation

The evaluation process is performed as follows: First, we gain insights into the distortion performance of the beam across a thermally turbulent atmospheric transmission path with actual figures. Then, with a neural network algorithm,

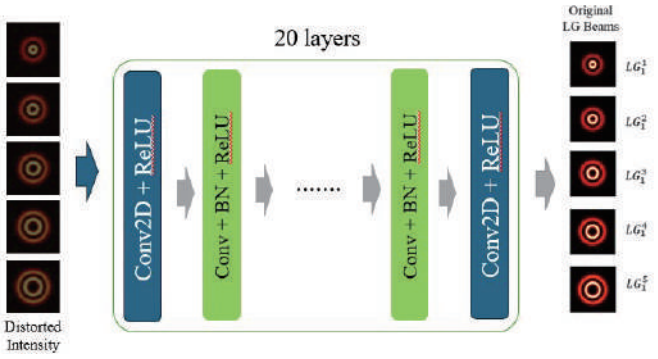


Fig. 2. Overall topology of regeneration CNN algorithm.

we evaluate the training loss of the system. After completing this analytical phase, we show the outcomes of the proposed reconstruction algorithm.

V. Conclusion

We have demonstrated a free-space optical communication system that employs a multiplexed LG-OAM beam. However, the intensity and phase images of the LG-OAM beam may be distorted in transmission due to atmospheric turbulence and inferior heat haze. To address this issue, we have utilized a convolutional neural network system with 20 layers. The outcome of this simulation allows us to produce more realistic results, such as an outdoor transmission system in the summer.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2023R1A2C2006860).

References

[1] F. Zhang, M. F. Bengtson, P. Kyösti, J. Kyröläinen and W. Fan, "Dynamic Sub-THZ Radio Channel Emulation: Principle, Challenges, and Experimental Validation," in IEEE Wireless Communications, vol. 31, no. 1, pp. 10-16, 2024.

[2] Walsh, S.M., Karpathakis, et al. "Demonstration of 100 Gbps coherent free-space optical communications at LEO tracking rates". Sci. Rep., 12, pp. 18345, 2022.

[3] W. Cheng, W. Zhang, H. Jing, S. Gao and H. Zhang, "Orbital angular momentum for wireless communications", IEEE Wireless Communications, vol. 26, no. 1, pp. 100-107, 2019.

[4] K. Rajasree, V. Vidyalal, P. Radhakrishnan, et al., "Use of mirage effect for the detection of phase transitions in liquid crystals", Materials Letters, vol.36, no. 1-4, pp. 76-80, 1998.

[5] Z. Lin, X. Chen, W. Qiu, and J. Pu, "Propagation Characteristics of High-Power Vortex Laguerre-Gaussian Laser Beams in Plasma",Appl. Sci. 2018, Volume 8, pp. 65, 2018.

[6] J. Wang, J. Y. Yang, I. M. Fazal, et al., "Terabit free-space data transmission employing orbital angular momentum multiplexing," Nat. Photonics, 6 ,7, pp. 488–496, 2012.

[7] Rasouli, Saifollah, Bagheri, Mohammad, "Resilience of radial carpet beams under propagation through indoor convective air turbulence". Journal of Optics. vol. 24, no. 1-14, pp. 075602, 2022

[8] Shiseng Wang, "Extensions to the Vavier-Stokes equations," Physics of Fluids, vol. 34, pp. 053106, 2022

[9] McAdams, W.H., "Heat Transmission", 3rd ed., McGraw-Hill,

[10] Willner, Alan, Xie, Guodong, et al., "Design challenges and guidelines for free-space optical communication links using orbital-angular-momentum multiplexing of multiple beams," Journal of Optics, 18, 2016.

Mathematical Detection Modeling for OAM-based MIMO Free Space Optical System

Merhawit Berhane Teklu
Artificial Intelligence Research Center
Pukyong National University, Busan 48513
Republic of Korea
Email: mberhane63@pknu.ac.kr

Yeon Ho Chung*
Department of Information and Communications Engineering
Pukyong National University, Busan 48513
Republic of Korea
Email: yhchung@pknu.ac.kr

***Abstract*—In free-space optical (FSO) communications, orbital angular momentum (OAM) multiplexing has recently gained considerable attention. The propagation of OAM modes through free space can be affected by atmospheric turbulence (AT) distortions. These distortions can cause intermodal crosstalk, power disparities between OAM modes, and mode-dependent losses (MDL), all of which can degrade system performance. In order to enhance the FSO system's performance, it is essential to implement strategies for mitigating AT. In this paper, we propose an improved reduced-search maximum likelihood (ML) detection method that provides near-ML performance while avoiding the computational complexity associated with ML detection.**

***Index Terms*—Orbital angular momentum (OAM), Free Space Optical (FSO), improved reduced-search ML (IRSML) detection**

I. Introduction

Orbital angular momentum (OAM) multiplexing has received considerable interest in free space optical (FSO) communications [1]. OAM modes is affected by atmospheric turbulence during propagation in free space, leading to intermodal crosstalk and power imbalances. An experiment was conducted to explore the impact of turbulence on intermodal crosstalk and power penalties in an OAM-based FSO multiplexing system [2]. Various OAM modes experience different channel losses, known as mode-dependent losses (MDLs), which contribute to performance degradation at the system level [3]. Thus, it is crucial to explore detection mechanisms for symbol recovery and enhance the overall system performance.

In this paper, we examine the impact of atmospheric turbulence, including MDL, on system performance. Our findings reveal that selecting specific OAM modes for multiplexing can decrease MDL, ultimately improving capacity. Symbols carried by OAM beams experience errors when transmitted through AT. While maximum likelihood (ML) detection is more effective than linear equalizers such as zero-forcing and minimum-mean square error equalizers in recovering the original signals, it requires significant computational resources. To address this issue, we propose an improved reduced-search ML detection method that attains near ML performance without the accompanying increase in complexity.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2023R1A2C2006860)
*Corresponding author

II. System model and Preliminaries

A. Orbital Angular Momentum Multiplexing

To implement OAM multiplexing, single and superpositions of orthogonal beams, an easy vortex beam Laguerre-Gaussian (LG) beam is used. Thus, the field distribution of the LG beams in cylindrical coordinate (r, ϕ, z) is given by [1]

$$u(r,\phi,z) = \sqrt{\frac{2p!}{\pi(p+|n|!)}}\frac{1}{w(z)}\left(\frac{r\sqrt{2}}{w(z)}\right)^{|n|} e^{\left(\frac{r}{w(z)}\right)^2} \times \mathbf{L}_P^{|n|}\left(\frac{2r^2}{w^2(z)}\right) e^{-i\frac{\pi r^2}{\lambda R(z)}} e^{i(|n|+2p+1)\Psi(z)} e^{-in\phi}, \tag{1}$$

where $w(z) = w_1\sqrt{1+\left(\frac{z}{z_R}\right)^2}$ is the beam radius at z where the Rayleigh range, $z_R = \frac{\pi w_1^2}{\lambda}$, where w_1 refers the beam Gaussian waist radius, and λ is the optical wavelength carrier; $\mathbf{L}_P^{|n|}(.)$ is the generalized Laguerre polynomial; $R(z) = z\left[1+\left(\frac{\pi w_1^2}{\lambda z}\right)\right]$; $(2p+|n|+1)\Psi(z)$ is the Gouy phase ($\Psi(z) = \tan^{-1}(\frac{z}{z_R})$); $e^{-in\phi}$ is the helical phase distribution, and p and n are the radial and angular mode numbers.

B. Optical turbulence model

The LG beams are generated using Spatial light modulators (SLMs). To simulate AT, random phase screens are positioned every 100 meters along the FSO channel. These screens are created using the power spectrum of refractive-index fluctuations derived from the modified atmospheric spectrum in [4]. In the FSO environment without AT, the vorticity of the propagating OAM ray is preserved and the OAM ray maintains orthogonality during propagation, which can be explained as

$$\int u_p(\boldsymbol{r},z)u_q^*(\boldsymbol{r},z)d\boldsymbol{r} = \begin{cases}1, & \text{if } p=q\\ 0, & \text{if } p\neq q,\end{cases} \tag{2}$$

where r is the radial position vector and $u_q^*(\boldsymbol{r},z)$ is the normalized field distribution of OAM mode of order q. Besides crosstalk, AT causes variations in losses among transmitted modes due to disruption in orthogonality, impacting the received signals and leading to an uneven SNR, ultimately affecting the system's performance.

C. OAM System Model

In a turbulent FSO channel, $M \times M$ MIMO can describe the transmission of M OAM beams as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{3}$$

where $\mathbf{y} \in^{M\times 1}$ is the received symbols vector; $\mathbf{x} \in^{M\times 1}$ is a vector of modulated symbols; $\mathbf{v} \in^{M\times 1}$ is the noise assumed to be additive white Gaussian; $\mathbf{H} \in^{M\times M}$ is the turbulent FSO channel, where each diagonal element h_{pp} stands for the amount of energy remaining in the OAM state p.

The OAM crosstalk because of turbulence is given by the channel inputs h_{pp} as

$$h_{pq} = \int u_p(\boldsymbol{r}, z) u_q^*(\boldsymbol{r}, z) d\boldsymbol{r}. \tag{4}$$

The crosstalk fading effect was shown in [5] to follow a Johnson SB distribution. The distribution in (4) is analytically intractable, making theoretical derivations infeasible. The demodulated signal, which is influenced by crosstalk among different OAM states, can be viewed as a linear combination of the initial signal, similar to a MIMO system.

III. Proposed Technique

A. Detection schemes

In order to recover the initial signals, we propose implementing MIMO detection technology at the recipient end of the OAM-based FSO communication system.

1) Linear equalizer MMSE: The linear equalizer MMSE is used because it suffers less from noise amplification than a ZF equalizer. The equalization is given by

$$\mathbf{r}_{\text{MMSE}} = \left(\mathbf{H}^H\mathbf{H} + \frac{\sigma_v^2}{\sigma_x^2}\mathbf{I}_M\right)^{-1}\mathbf{H}^H\mathbf{y} \tag{5}$$

where σ_v^2 and σ_x^2 represent the variances of the noise and transmit signal, consecutively.

2) Improved reduced-search ML (IRSML) detection: The IRSML algorithm performs better and is more flexible when compared with the reduced-search ML algorithm (RSML). Thus, the MMSE equalizer's output is processed by a hard decision unit and used to create a search set for the ML stage of the algorithm, which is defined by

$$\hat{\mathbf{x}}_{\text{IRSML}} = \underset{\mathbf{x}\in\mathbb{X}}{\text{argmin}} \left\|\mathbf{y} - \mathbf{H}\mathbf{x}\right\|^2, \tag{6}$$

where $\hat{\mathbf{x}}_{\text{IRSML}}$ is the estimated symbols vector and $\mathbb{X}$ is a subset of $\mathfrak{X}^M$ in which $\mathfrak{X}$ is the set of symbols in the constellation and $\mathbb{X}_1 \times ... \times \mathbb{X}_M$ corresponds to the Cartesian product of the tributary search set. In the search set that includes the ML solution, the algorithm gathers the constellation points of each tributary. The ML detector feeds the search set $\mathbb{X}$, where the maximum a posteriori (MAP) constructed the $\mathbb{X}$, which finds the hypothetical symbols $x_j, j = 1, ..., M$ whose probability $P(x_j/\mathbf{y}_{\text{MMSE}_j})$ is the highest. To achieve this, the Euclidean distance between the output of the MMSE equalizer and the detected symbols are determined by

$$\Omega_{\text{c,j}} = |\mathbf{y}_{\text{MMSE}_j - x_{c,j}}| \tag{7}$$

where $c = 1, ..., M$ is the index for all constellation points; $x_{c,j}$ is some constellation point whose index is c and its tributary is j. To arrange the Euclidean distances in (7) in ascending order, identify the pair c, j that meets

$$\underset{x_{\text{c,j}}}{\text{argmin}}\ \Omega_{\text{c,j}}. \tag{8}$$

This is added to the search set, $\mathbb{X}_j$, for the tributary j. Until the maximum number of hypersymbol candidates M_{max} is achieved or surpassed by adding another candidate, potential hypersymbols are continually added to the collection $\mathbb{X}_j$. Based on the required performance level and the allowed computing complexity, the maximum number of hypersymbol candidates M_{max} is chosen.

B. OAM Mode Selection

Mode-dependent loss is calculated based on the decibel ratio between the highest and lowest eigenvalues (Γ) of the channel matrix $\mathbf{H}$ according to [6] as given as

$$\text{MDL} = 10\log_{10}\left(\frac{\Gamma_{\text{max}}}{\Gamma_{\text{min}}}\right). \tag{9}$$

Assuming that there are $\mathcal{X}$ OAM modes available for multiplexing, our goal is to choose set $\mathcal{X}_{\text{p}}$ of OAM modes that performs the best in terms of error probability. This means that the best potential mode combination that satisfies the following requirements should be found as

$$\mathcal{X}_{\text{p}} = \underset{\mathcal{X}_{\text{sub}}\subset\mathcal{X}}{\text{argmin}}\ \text{MDL}, \tag{10}$$

where $\mathcal{X}_{\text{p}}$ spans all the possible subsets of $\mathcal{X}$.

IV. Conclusion

OAM modes in free space can be disrupted by turbulence, causing interference and power level variations, as well as MDL. This study applies MIMO detection technology to the OAM-based FSO communication system, considering the similarity between crosstalk and MDL caused by AT and the MIMO channel model. We propose an improved reduced-search ML detection method that achieves near ML performance without the added complexity. Our future work will include simulation results to validate the mathematical model for comparing the performance of detection mechanisms.

References

[1] A. E. W. et al, "Design challenges and guidelines for free-space optical communication links using orbital-angular-momentum multiplexing of multiple beams," *Journal of Optics*, vol. 18, no. 7, p. 074014, 2016.

[2] Y. R. et al., "Atmospheric turbulence effects on the performance of a free space optical link employing orbital angular momentum multiplexing," *Optics Letters*, vol. 38, no. 20, pp. 4062–4065, 2013.

[3] ——, "Atmospheric turbulence mitigation in an oam-based mimo free-space optical link using spatial diversity combined with mimo equalization," *Optics Letters*, vol. 41, no. 11, pp. 2406–2409, 2016.

[4] L. Andrews, "An analytical model for the refractive index power spectrum and its application to optical scintillations in the atmosphere," *Journal of Modern Optics*, vol. 39, no. 9, pp. 1849–1853, 1992.

[5] G. Funes, M. Vial, and J. A. Anguita, "Orbital-angular-momentum crosstalk and temporal fading in a terrestrial laser link using single-mode fiber coupling," *Opt. Express*, vol. 23, no. 18, pp. 23 133–23 142, 2015.

[6] E. Awwad, G. R.-B. Othman, and Y. Jaouën, "Space-time coding schemes for mdl-impaired mode-multiplexed fiber transmission systems," *Journal of Lightwave Technology*, vol. 33, no. 24, pp. 5084–5094, 2015.

Orbital Angular Momentum States Beamforming for Multiuser Optical Wireless Network

Siti H. Mohammad
Artificial Intelligence Research Institute
Pukyong National University
Busan, South Korea
Email: sitihasunah@gmail.com

Yeon H. Chung
Department of Information and Communication Engineering
Pukyong National University
Busan, South Korea
Email: yhchung@pknu.ac.kr

***Abstract*—Recognized because of its orthogonal characteristic which can provide high-capacity communication channel, orbital angular momentum (OAM) has become one of the promising technologies to enhance the Internet of Things (IoT) applications. As the orthogonality of the OAM beam provides big advantages in multiuser optical wireless network (OWC), it is the key interest that the higher capacity can be achieved further by beamforming method of the OAM states. In this paper, we proposed the derivation of ultraviolet (UV) channel impulse response considering the Laguerre-Gaussian beam with OAM states based beamforming to increase the channel capacity of the OAM-carrying UV communication system. It is shown via simulation that the proposed OAM states beamforming using UV channel impulse response achieves higher channel capacity and better error performance than the conventional single OAM state carrying UV beam.**

***Index Terms*—Orbital Angular Momentum, Laguerre-Gaussian beam, OAM beamforming, Optical Wireless Communication**

I. Introduction

One of the potential technologies that can enhance the performance of Internet of Things (IoT) applications is Optical Wireless Communication (OWC) network. OWC is an ideal convergence of an optical backhaul and wireless front end for a proficient IoT network due to its high capacity, flexibility, robustness, and cost-effectiveness [1]. The light beam in OWC operates in a broad license-free electromagnetic (EM) spectrum spanning the Infrared (IR), Visible Light (VL), and Ultraviolet (UV) bands [2]. Recently, the most popular implementation of OWC with high-speed data rate with multiple users is based on Laguerre-Gaussian (LG) beam. Due to its viability to create orthogonal and spatially distinct channel with multiple orbital angular momentum (OAM) modes, LG beam presents novel opportunities for future communication applications as it allows simultaneous data transmission from different users [3]. OAM has been used to increase the system capacity in free space communication [4], ultra-high spectral efficiency optical wireless communications [3] and real-world long-distance application scenario [5].

Even though the feasibility of OAM enabled the intense expansion of capacity supplied by optical communication system, there are many research challenges unresolved. For example, in different frequency ranges, electromagnetic wave carrying OAM beams experience different issues. Beams at higher frequencies are more likely affected with turbulence and crosstalk, while beams at lower frequencies have more divergence, limited distance and requires larger receiver apertures [6]. To address these concerns, OAM beamforming is one of the techniques used in high-capacity OWC networks. The concept of OAM beamforming is to superpose multiple OAM states within a single beam by modifying the amplitude, phase, and relative weights of each OAM state. Then, the resulting composite beam is shaped according to the desired spatial profile [7]–[9]. The beamformed OAM is less susceptible to interference compared to a conventional beam, with high tolerance to channel impairments and higher-quality transmission.

In this paper, we present a novel OAM based beamforming for indoor multiuser OWC using UV channel impulse response (UV-IR) coding for the first time as it is much simpler with low computational complexity. The proposed OAM beamforming method using IR code shows improved in bit error rate (BER) analysis and channel capacity for the UV links with beamformed OAM states compared to the conventional single OAM state beam carrying UV link without beamforming. The performance analysis from simulation also shows that higher OAM states interval beamformed together resulting in increasing signal-to-noise ratio (SNR), thus decreasing the BER and increase the channel capacity.

In this work, it should be noted that irrespective of the OAM mode value of LG beam, the presence of a beam only represents the bit "1". The OAM-carrying LG beam is considered as a data channel with bits "1" and "0" represented by the presence and absence of a beam, respectively. OAM multiplexing is not employed to transmit the data. The remainder of the paper is organized as follows. Section II describes the OAM signal model including the UV-IR coding, Section III presents the simulation setup on MATLAB, Section IV discusses the results and conclusions are drawn in Section V.

II. OAM Signal Model

The LG modes are a set of solutions to the paraxial wave equation in cylindrical coordinates. These modes have a helical wavefront and carry OAM. The system model of UV communication based on OAM in cylindrical coordinates is illustrated in Fig. 1 [10].

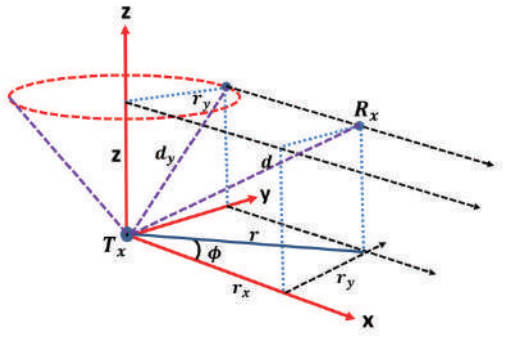


Fig. 1. OAM based UV communication link in cylindrical coordinates [10].

The reference plan for the sytem analysis is assumed to be the Cartesian xy-plane utilizing the cylindrical coordinate system (r, θ, z) with z = 0. r is the radial coordinate, represented by the radius of the cylindrical geometry formed, and calculated by $r = \sqrt{r_x^2 + r_y^2}$. r_x is the OAM beam dimension spreading along the x-axis, while r_y is the OAM beam widening along the y-axis.The transmitter T_x is assumed to be located at the coordinate (0, 0, 0). d represents the distance between the transmitter and the receiver located at the coordinates (r, θ, z), calculated by $d = \sqrt{r^2 + z^2}$.

The electric field of the LG beam is given by

$$\vec{E} = U_{l,p}(r, \theta, z)\hat{a}, \tag{1}$$

where $\hat{a}$ is a constant unit vector, and the term $U_{l,p}(r, \theta, z)$ is given by [11]

$$\begin{aligned} &U_{l,p}(r, \theta, z) \\ &= \alpha\sqrt{\frac{p!}{\pi(p+|l|)!}}\frac{1}{w(z)}\left(\frac{r\sqrt{2}}{w(z)}\right)^{|l|} L_p^{|l|}\left(\frac{2r^2}{w(z)^2}\right) \\ &\times \exp\left(-\frac{r^2}{w(z)^2}\right)\exp\left(-i\frac{z}{z_R}\left(\frac{r}{w(z)}\right)^2\right) \\ &\times \exp\left(-i(2p+|l|+1)\psi(z)\right) \\ &\times \exp(-il\theta)\exp\left(-ikz\right), \end{aligned} \tag{2}$$

where l is the azimuthal mode number (OAM state), p is the radial mode number, θ is the azimuthal angle between the reference direction on the chosen plane and the line from the origin to the projection of the receiver on the plane, and k is the wavenumber which is equal to $2\pi/\lambda$. z represents the propagation distance from the transmitter to the receiver at which the power of the beam drops by a factor of 2. $w(z)$ is the Gaussian beam width, which can be calculated by

$$w(z) = w_0\sqrt{1 + \left(\frac{z}{2R}\right)^2}, \tag{3}$$

where w_0 is the initial beam radius at $z = 0$. z_R is the Rayleigh range which is equal to $\frac{\pi w_0^2}{\lambda}$, where λ is the wavelength of the light source. $(2p + |l| + 1)\psi(z)$ is the Gouy phase shift with $\psi(z)$ equal to $\arctan\left(\frac{z}{z_R}\right)$. $L_p^{|l|}$ is the generalized Laguerre Polynomial of degree p and order l.

A. Channel Impulse Response

From [12], the refractive index of the channel, μ, and Rayleigh scattering coefficient, α_{sr}, are given by

$$\begin{aligned} &\mu \\ &= \left(8060.51 + \frac{2480990}{132.274 - \lambda_{\mu\text{m}}^{-2}} + \frac{17455.7}{39.32957 - \lambda_{\mu\text{m}}^{-2}}\right) \times 10^{-8} \end{aligned} \tag{4}$$

$$\alpha_{\text{sr}} = \frac{(24\pi^3(\mu^2-1)^2)}{(\lambda_{\mu\text{m}}^4 N(\mu^2+2)^2)} \times \frac{(6+3\rho_\mu)}{(6-7\rho_\mu)}, \tag{5}$$

where $\lambda_{\mu\text{m}}$ is the wavelength of the OAM-carrying UV beam in micrometers, N is the molecular number density and ρ_μ is the depolarization ratio, which is a function of the UV wavelength.

The atmospheric scattering coefficient, α_s, is defined by $\alpha_s = \alpha_{\text{sr}} + \alpha_{\text{sm}}$, where Mie scattering coefficient, α_{sm}, is expressed as [13]

$$\alpha_{\text{sm}} = \left(\frac{3.912}{v} - \alpha_{\text{sr}}(\lambda_{\text{nm}=550})\right) \times \left(\frac{\lambda_{\text{nm}}}{550}\right)^{0.585v^{1/3}}, \tag{6}$$

where λ_{nm} is the UV wavelength measured in nanometers and v is the atmospheric visibility in kilometers. The total coefficient, a_t, si defined by $\alpha_t = \alpha_s + \alpha_a$ [10]. Thus, the receiver gain, G, is defined by

$$G = \frac{A_r}{4\pi d_y^2}\exp(-\alpha_t d_y), \tag{7}$$

where A_r is the effective receiver area and d_y is the Eucladian distance of the transmitter to the point of receiver on the $y-z$ plane.

Then, the UV channel impulse response can be expressed as

$$h^l = \sqrt{G}\left(\frac{r}{r_y}\right)^{|l|}\exp\left[-\frac{(r^2 - r_y^2)}{w_z^2}\right]. \tag{8}$$

From this UV-IR equation, the OAM beam can be generated. The higher channel capacity with lower BER than conventional OAM beam can be achieved in OWC network by superposing many OAM states together with consecutives states.

III. Simulation Setup

A UV wavelength of 260 nm is used for the communication link in MATLAB simulation with a room dimension of 5 m × 10 m. α_{sr} and α_{sm} are obtained using (4), (5), and (6). α_{a} is set to 0.802 km^{-1} with v is set to 21 km [14]. The molecular number density N is equal to 2.4481×10^{25} m^{-3}, with depolarization ratio ρ_μ equal to 3.501×10^{-2} at λ equal to 260 nm [13]. Four UV links are served by OAM-carrying LG UV beam with OAM states, l, ranging from 5 to 30 and initial beam radius, w_0, equals to 0.5λ. One UV link is assigned with a conventional OAM beam, while another three UV links

are assigned with different sets of beamformed OAM states interval, Δl = 1, 2, and 3, respectively. The generated OAM beams for each UV link are shown in Fig. 3.

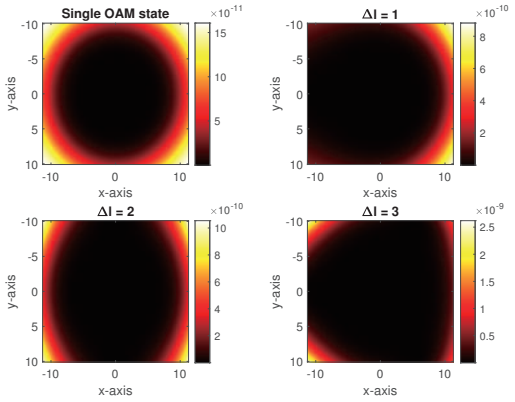


Fig. 2. OAM intensity profile for each UV link.

Each UV link has a ring-shaped intensity profile. From Fig. 2, it can be seen that the UV link with single conventional OAM beam has the lowest amplitude of intensity distribution, while the intensity distribution of the UV links with superposed beam shows higher amplitude. The larger the beamformed OAM states interval, the higher the intensity amplitude of the beam as high OAM states have higher energy due to narrower beam produced.

IV. Simulation Results and Discussion

Performance analysis of the UV-IR based OAM beamforming is visualized in terms of SNR, BER and channel capacity. As shown in Fig. 3, as the OAM state interval increases, the UV link with higher OAM state superposed with the initial OAM state has the highest SNR among all UV links. The performance of the network is enhanced better with larger OAM state interval due to narrower beam produced with higher superposed beam. The higher OAM states are superposed together, the SNR of the UV links increases, while BER decreases more to below effective coverage (3.8×10^{-3}) as shown in Fig. 4.

In term of capacity, the UV link with higher superposed OAM state interval has higher capacity over the UV link with lower superposed OAM state interval as shown in Fig. 5. As OAM states increases, the size of the outermost ring of LG beam increases with lower destructive interference, thus resulting in higher received energy at the receiver. This is also due to the larger OAM state value results in la larger beam size at the receiver, providing higher capacity as the received power increases.

This higher received energy increases the SNR, which then decreases the energy losses in the atmosphere and narrowing the beam, resulting in increasing capacity as shown in Fig. 6, and decreasing the BER. It also can be seen that the performance of the single OAM state carrying UV beam has the lowest capacity at the lowest SNR due to smaller beam size at the receiver.

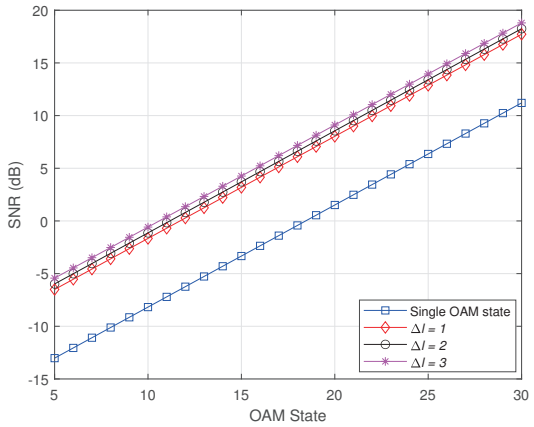


Fig. 3. SNR versus inreasing OAM states for different UV links.

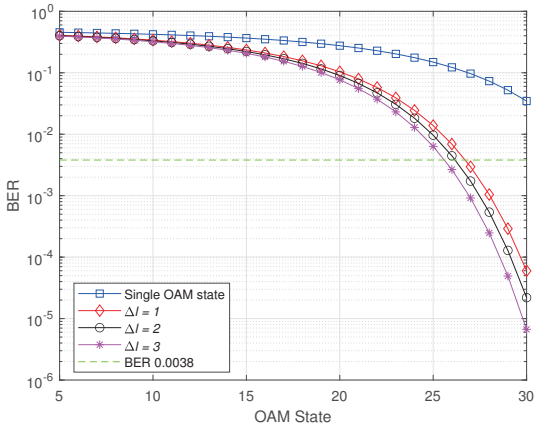


Fig. 4. BER versus inreasing OAM states for different UV links.

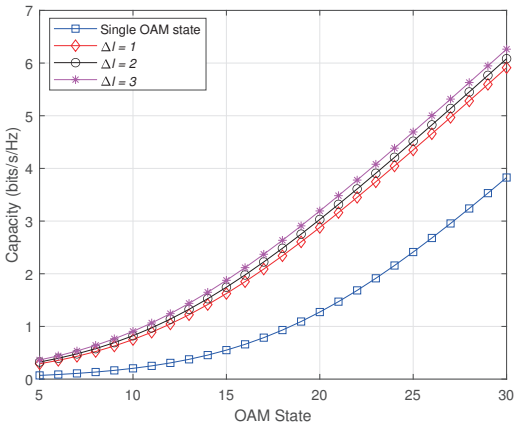


Fig. 5. Channel capacity versus inreasing OAM states for different UV links.

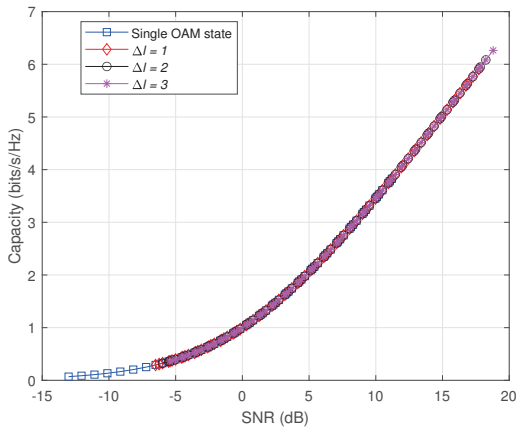


Fig. 6. Capacity versus SNR for different UV links.

V. Conclusion

In this paper, a multiuser optical wireless communication network model with OAM states beamforming is modelled using UV channel impulse response coding. The simplified UV-IR coding has been first derived for OAM beamforming in this work and the performance of the proposed network model have been analyzed. It is demonstrated that the beamformed OAM-carrying UV beam with higher superposed OAM states interval surpasses the conventional single OAM state carrying UV beam in terms of SNR, BER and channel capacity. More analysis on the UV-IR coding can be further studied in different scenario of OWC in the future.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2023R1A2C2006860).

References

[1] Mohammad, S. H., Zulkifli, N., Idrus, S. M., and Ramli, A. (2017). *Bandwidth Allocation Scheme with QoS Provisioning for Heterogeneous Optical and Wireless Networks.* Networks of the Future: Architectures, Technologies, and Implementations, 301-318.

[2] Hamza, A., and Tripp, T. (2023). *Optical wireless communication for the Internet of Things: Advances, challenges, and opportunities.* Authorea Preprints.

[3] Gibson, G., Courtial, J., Padgett, M. J., Vasnetsov, M., Pas'ko, V., Barnett, S. M., and Franke-Arnold, S. (2004). *Free-space information transfer using light beams carrying orbital angular momentum.* Optics express, 12(22), 5448-5456.

[4] Wang, J., Yang, J. Y., Fazal, I. M., Ahmed, N., Yan, Y., Huang, H., ... and Willner, A. E. (2012). *Terabit free-space data transmission employing orbital angular momentum multiplexing.* Nature photonics, 6(7), 488-496.

[5] Wang, J., Liu, J., Li, S., Zhao, Y., Du, J., and Zhu, L. (2021). *Orbital angular momentum and beyond in free-space optical communications.* Nanophotonics, 11(4), 645-680.

[6] Willner, A. E. (2021). *OAM light for communications.* Optics and Photonics News, 32(6), 34-41.

[7] Wang, Z., Zheng, S., Xiong, X., Zhu, Z., Chen, Y., Yu, X., Jin, X. and Zhang, X., 2020. Structure radio beam construction in azimuthal domain. IEEE Access, 8, pp.9395-9402.

[8] Kai, C., Feng, Z., Dedo, M.I., Huang, P., Guo, K., Shen, F., Gao, J. and Guo, Z., 2019. The performances of different OAM encoding systems. Optics Communications, 430, pp.151-157.

[9] Anguita, J.A., Herreros, J. and Djordjevic, I.B., 2014. Coherent multimode OAM superpositions for multidimensional modulation. IEEE Photonics Journal, 6(2), pp.1-11.

[10] Arya, S., and Chung, Y. H. (2019). *High-performance and high-capacity ultraviolet communication with orbital angular momentum.* IEEE Access, 7, 116734-116740.

[11] Djordjevic, I. B., Anguita, J. A., and Vasic, B. (2012). *Error-correction coded orbital-angular-momentum modulation for FSO channels affected by turbulence.* Journal of lightwave technology, 30(17), 2846-2852.

[12] Tomasi, C., Vitale, V., Petkov, B., Lupi, A., and Cacciari, A. (2005). *Improved algorithm for calculations of Rayleigh-scattering optical depth in standard atmospheres.* Applied optics, 44(16), 3320-3341.

[13] Patterson, E. M., and Gillespie, J. B. (1989). *Simplified ultraviolet and visible wavelength atmospheric propagation model.* Applied optics, 28(3), 425-429.

[14] Junge, D.M. and NA. (1977). *Non-line-of-sight electro-optic laser communications in the middle ultraviolet.* Doctoral dissertation, Naval Postgraduate School.

Precoding Scheme for Indoor Channel Modelling in OAM-based MIMO Communication System

Merhawit Berhane Teklu
Artificial Intelligence Research Center
Pukyong National University, Busan
Republic of Korea
Email: mberhane63@pknu.ac.kr

Yeon Ho Chung*
Department of Information and Communication Engineering
Pukyong National University, Busan
Republic of Korea
Email: yhchung@pknu.ac.kr

***Abstract*—Due to its low transceiver complexity, mode division multiplexing (MDM) using orbital angular momentum (OAM) has recently been explored as a new physical layer indoor wireless transmission technique. In practical applications, there are interferences between multiple users/pairs using OAM-based communication techniques. To mitigate the interferences and optimize the system performance, we mathematically formulate the OAM wireless channel using Laguerre-Gaussian (LG) beams, and then optimize the sum rate by designing the precoder matrix for each transmitter using leveraging of leakage concept.**

***Index Terms*—Orbital angular momentum (OAM), Channel Modeling, Sum rate, signal-to-leakage-noise ratio (SLNR)**

I. Introduction

Orbital angular momentum (OAM) is a mode division multiplexing (MDM) technique that has the potential for 6G optical communications. The authors [1], [2] conducted a study on mode selection criteria and assessed the capacity of various OAM mode combinations for OAM multiplexing systems. In [3], the study analyzed the combination of OAM with MDM (OAM-MDM) and conventional spatial multiplexing for a single transmitter and receiver pair. On the other hand, [4] investigated inter-pair interference in a multi-pair MDM system using mmwave OAM, where multiple transmitter and receiver pairs utilized the same multiple OAM modes.

This study proposes a precoding scheme for indoor OAM-based multiuser MIMO communication, aiming to enhance signal-to-leakage-and-noise ratio (SLNR) for all users concurrently by considering interference power generated by each user on others.

II. System model and Preliminaries

A. OAM System Model

Let us consider an indoor multiuser communication based on the OAM modes of multiplexing with N transmitting and N receiving antenna arrays. The transmit antenna at each transmitter can generate K OAM signals with different OAM modes, where $K \leq N$. So the set of OAM modes denotes $\mathcal{K} = [k_1, k_2, ..., k_a, ..., k_K]$ and the a^{th} OAM mode from each transmitter denotes l_a, where $a \in \{1, 2, ..., K\}$.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2023R1A2C2006860).

*Corresponding author

The received vector at the j^{th} receiver can be expressed by

$$\mathbf{y}_j = \mathbf{H}_{j,j}\mathbf{F}_i\mathbf{x}_i + \sum_{r\neq j}^{K} \mathbf{H}_{j,r}\mathbf{F}_r\mathbf{x}_r + \mathbf{v}_j, \quad \in \mathbb{C}^{N\times 1} \tag{1}$$

where $\mathbf{F}_i \in \mathbb{C}^{N\times K}$ and $\mathbf{x}_i \in \mathbb{C}^{K\times 1} = [x_{1i}, ..., x_{ai}, ..., x_{Ki}]^T$ are the precoding and transmitted data vectors from the i^{th} transmitter, respectively. Then, we normalize $\mathbf{x}_i$ and $\mathbf{F}_i$ as $\mathrm{E}[||\mathbf{x}_i||^2] = 1$ and $||\mathbf{F}_i||^2 = 1$, respectively. $\mathbf{H}_j \in \mathbb{C}^{N\times N} = [\mathbf{H}^l_{1j}, ..., \mathbf{H}^l_{aj}, ..., \mathbf{H}^l_{Nj}]^T$ is the channel response between the receiving and transmitting antennas with the OAM state l. $\mathbf{v}_j \in \mathbb{C}^{N\times 1} = [v_{1j}..., v_{aj}, ..., v_{Nj}]^T$ is the additive white Gaussian noise vector with variance σ_v^2.

B. OAM Channel Model

In cylindrical coordinates (r, ϕ, z), a transmitting antenna at the origin radiates Laguerre-Gaussian (LG) beams is given by

$$\begin{aligned} u(r,\phi,z) = \alpha\sqrt{\frac{p!}{\pi(p+|l|!)}}\frac{1}{w(z)}\left(\frac{r\sqrt{2}}{w(z)}\right)^{|l|} e^{\left(\frac{r}{w(z)}\right)^2} \\ \times \mathbf{L}_P^{|l|}\left(\frac{2r^2}{w^2(z)}\right) e^{-i\frac{\pi r^2}{\lambda R(z)}} e^{i(|l|+2p+1)\Psi(z)} e^{-il\phi}, \end{aligned} \tag{2}$$

where p is the radial index; $w(z) = w_l\sqrt{1+\left(\frac{z}{z_R}\right)^2}$ is the beam radius at z position where the Rayleigh distance, $z_R = \frac{\pi w_l^2}{\lambda}$, and the beam waist radius at $z = 0$, w_l, and $R(z) = z\left[1+\left(\frac{\pi w_1^2}{\lambda z}\right)\right]$; $\mathbf{L}_P^{|l|}(.)$ is the generalized Laguerre polynomial; $(2p+|l|+1)\Psi(z)$ refers the Gouy phase ($\Psi(z) = \tan^{-1}(\frac{z}{z_R})$); $e^{-il\phi}$ is the helical phase distribution. We assume the first each transmit antennas are positioned at origin. The position of the i^{th} transmit antenna is $(|i-j|\zeta, 0, 0)$, where ζ is the distance between adjacent transmit antennas. Then, we can define the relative radial position and transverse azimuthal angle from the i^{th} transmit antenna to the b^{th} receive antenna at the j^{th} receiver as r^b_{ji} and ϕ^b_{ji}, respectively and this applies for all receive antennas.

$$r^b_{ji} = \sqrt{(|i-j|\zeta)^2 + r^2_{max}(z, l_a)}, \tag{3}$$

$$\phi^b_{ji} = \begin{cases} \arctan\frac{r_{max}(z,l_a)}{|i-j|\zeta}, & j > i \\ \pi - \arctan\frac{r_{max}(z,l_a)}{|i-j|\zeta}, & j < i \\ \pi/2, & j = i, \end{cases} \tag{4}$$

$$h_{ji}^{ab} = \mathcal{A}\frac{\lambda}{4\pi d_{ii}^{a}}\left(\frac{r_{ji}^{a}}{r_{ii}^{b}}\right)^{l_a}\exp\left(-\frac{(r_{ji}^{a})^2-(r_{ii}^{b})^2}{w^2(z)}\right)^{l_a}\exp\left(-\frac{i\pi((r_{ji}^{a})^2-(r_{ii}^{b})^2)}{\lambda z\left(1+\frac{\pi w^2(z)}{\lambda z}\right)^2}\right)\exp(-ikd_{ii}^{a})\exp(-i\phi_{ji}^{b}l_a). \quad (10)$$

where $r_{max}(z, l_a)$ is the radius of the maximum energy strength circle region at the position of z with the function of z and l_a. With the obtained coordinates, the transmission distance between j^{th} transmitter and i^{th} receiver is given by

$$d_{ji}^{b} = \sqrt{z^2 + (|i-j|)^2\zeta^2 + r_{max}^2(z, l_a)}. \quad (5)$$

Based on the radius of the OAM circle region with the maximum energy strength on the plane z, all receiving antennas can be positioned in the following manner with each transmitter:

$$r_{max}(z, l_a) = \sqrt{\frac{|l_a|}{2}}w(z) = w_l\sqrt{\frac{|l_a|}{2}\left(1+\left(\frac{z}{z_R}\right)^2\right)}. \quad (6)$$

The OAM channel transfer from the i^{th} transmit antenna to the b^{th} receive antenna at the j^{th} receiver via the a^{th} OAM signal for $j \neq i$ can be expressed as

$$h_{ji}^{ab} = \mathcal{A}_{ji}^{ab}\frac{\lambda}{4\pi d_{ji}^{b}}e^{-ikd_{ji}^{b}}e^{-i\phi_{ji}^{ab}l_a}, \quad (7)$$

where $\mathcal{A}_{ji}^{ab}$ denotes the OAM channel gain coefficient and $k = \frac{2\pi}{\lambda}$ is the wave number. Similarly, the OAM channel transfer from the i^{th} transmit antenna to the a^{th} receive antenna at the i^{th} receiver via the a^{th} OAM signal, where $j = i$, is given by

$$h_{ii}^{aa} = \mathcal{A}\frac{\lambda}{4\pi d_{ii}^{a}}e^{-ikd_{ii}^{a}}e^{-i\phi_{ii}^{nn}l_a}, \quad (8)$$

where $\mathcal{A}$ represents desired OAM channel gain coefficient.

By recalling (2), when the OAM wave at state l is transmitted by the transmitting antenna j, the response u_{ii} at the receiving antenna i with $j = i$ is given by $u_{ii} = h_{ii}^{aa}x_{ai}$, where x_{mi} is the unit pulse input. Again, the response u_{ji} at the receiving antenna i with $j \neq i$ is given by $u_{ji} = h_{ji}^{ab}x_{ai}$. According to these channel responses, the proportional relationship [5] can be formulated

$$\frac{u_{ji}}{u_{ii}} = \frac{h_{ji}^{ab}}{h_{ii}^{aa}}. \quad (9)$$

To obtain the OAM channel matrix from the i^{th} transmitter to the j^{th} receiver, $\mathbf{H}_{ji}$, we can use the simplified expression in (10). We can obtain this expression by substituting (2), (7), and (8) into (9). This process can be applied for all j and i.

III. Performance Analysis

A. Proposed Technique

To alleviate the inter-mode interference among the OAM modes, we propose a precoding technique based on SLNR metric [6]. The optimal precoder, $\mathbf{F}_i^o$, at user i can be formulated as

$$\mathbf{F}_i^o = \arg\max_{\mathbf{F}_i\in\mathcal{C}^{N\times 1}}\frac{||\mathbf{H}_{ii}\mathbf{F}_i||^2}{N_j\sigma_v^2 + \sum_{r\neq j}^{K}||\bar{\mathbf{H}}_i\mathbf{F}_r||^2}. \quad (11)$$

where $\bar{\mathbf{H}}_i$ is extended OAM wireless channel without $\mathbf{H}_{ii}$. By eigenvalue decomposition in [6], (11) can be simplified.

B. Sum rate Optimization

The simplified maximum sum rate in bits/sec/Hz of the OAM-based MIMO communication system can be found by substituting the optimal precoder in (11) as

$$\mathrm{R} = \sum_{i=1}^{K} E[\log_2|\mathbf{I}_N + \gamma_i\mathbf{F}_i^{o,H}\mathbf{H}_{ii}^H\mathcal{Q}_I^{-1}\mathbf{H}_{ii}\mathbf{F}_i^o|] \quad (12)$$

where $\mathcal{Q}_I = \gamma_j\sum_{i\neq j}^{K}[\mathbf{H}_{ji}^H\mathbf{F}_j^{o,H}\mathbf{F}_j^o\mathbf{H}_{ji} + \mathbf{I}_N]$ and this is the covariance matrix of the interference plus noise; and γ represents the signal to noise ratio (SNR).

IV. Conclusion

There are interferences associated with OAM-based communication in practical applications. This paper aims to mathematically characterize the OAM wireless channel using LG beams and optimize the sum rate of the OAM interference channel by designing the precoder matrix for each transmitter based on the concept of leakage. By doing this, we can determine the maximum sum rates of indoor OAM-MIMO communication. In the future, we will compare the capacity of OAM-based MIMO systems to that of conventional MU-MIMO systems based on simulation results.

References

[1] R. Chen, J. Lin, B. Zhang, Y. Ding, and K. Xu, "Precoding-based downlink oam-mimo communications with rate splitting," *IEEE Transactions on Broadcasting*, vol. 69, no. 4, pp. 894–903, 2023.

[2] K. A. Opare, Y. Kuang, J. J. Kponyo, K. S. Nwizege, and Z. Enzhan, "The degrees of freedom in wireless line-of-sight oam multiplexing systems using a circular array of receiving antennas," in *2015 Fifth International Conference on Advanced Computing Communication Technologies*, 2015, pp. 608–613.

[3] Y. Ren, L. Li, G. Xie, Y. Yan, Y. Cao, H. Huang, N. Ahmed, Z. Zhao, P. Liao, C. Zhang, G. Caire, A. F. Molisch, M. Tur, and A. E. Willner, "Line-of-sight millimeter-wave communications using orbital angular momentum multiplexing combined with conventional spatial multiplexing," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3151–3161, 2017.

[4] W. Son, H. Lee, and B. C. Jung, "Capacity analysis of multi-pair orbital angular momentum interference networks," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–5.

[5] L. Wang, X. Ge, R. Zi, and C.-X. Wang, "Capacity analysis of orbital angular momentum wireless channels," *IEEE Access*, vol. 5, pp. 23 069–23 077, 2017.

[6] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user mimo channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1711–1721, 2007.

Beam Pointing for FSOC with Spatial Diversity

1st Hong-Seol Cha
School of Electrical Engineering
Korea University
Seoul, South Korea
chahongseol@korea.ac.kr

2nd Byungju Lim
Dept. of Electronic Engineering
Pukyong National University
Busan, South Korea
limbj@pknu.ac.kr

3rd Young-Chai Ko
School of Electrical Engineering
Korea University
Seoul, South Korea
koyc@korea.ac.kr

***Abstract*—Free space optical communications (FSOC) which adopt spatial diversity is actively underway to mitigate the effects of atmospheric turbulence and pointing error. Although beam pointing greatly affects the performance of FSOC, the beam pointing that maximize the performance have not been studied yet. In this paper, we analyze the convexity condition of the optimization problem for the center coordinates of the transmitted beam footprint maximizing the received signal-to-noise ratio (SNR) in a single-input multiple-output (SIMO) FSOC, which is affected by the atmospheric turbulence and pointing error.**

***Index Terms*—Free space optical communications (FSOC), spatial diversity, beam pointing**

I. Introduction

Free space optical communications (FSOC) is a promising technology, which offers numerous advantages such as high-speed data transmission, being license free, but there are several shortcomings, especially, atmospheric turbulence-induced fading and pointing error [1], [2]. To address these challenges, research aiming to enhance the link performance by adopting the spatial diversity is actively underway [3], [4] with transmitting the beams to the center of its paired receiver lens. However, due to the presence of channel effects, the center coordinates of the transmitted beam footprint maximizing the link performance may not be achieved by beam pointing to the center of its paired receiver lens.

II. Convexity Condition of Maximizing SNR

In a SIMO FSOC with M receiver lenses, after adopting the equal gain combining (EGC) technique at the receiver plane, the signal-to-noise ratio (SNR) for the diversity system is expressed as [3]

$$\gamma_{div} = \frac{|\sum_{i=1}^{M} \eta h_l h_{a_i} h_{p_i} P_t|^2}{MN_0/2}, \tag{1}$$

where η is the effective photo-electric conversion ratio, P_t is the transmit power. In (1), h_l is the path-loss factor, and h_{a_i} and h_{p_i} are the atmospheric turbulence-induced fading factor and the pointing error factor between a transmitter and i-th receiver lens, respectively.

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00260), Research on LEO Inter-Satellite Links.
(Corresponding author: Young-Chai Ko)

Since all terms in (1) are positive, the optimization problem for the center coordinates of the transmitted Gaussian beam footprint maximizing the received SNR can be simplified as

$$(\hat{x}_t, \hat{y}_t) = \arg \max_{(x_t, y_t)} \sum_{i=1}^{M} h_{a_i} \exp\left(-2|\mathbf{r}_{p_i}|^2 / w_{zeq}^2\right), \tag{2}$$

where $|\mathbf{r}_{p_i}|$ is the euclidean distance between the center of the beam footprint and the i-th receiver lens and w_{zeq} is the equivalent beamwidth.

Since the summation of convex functions is a convex, we only verify the convexity condition of the following function given as

$$f_i(x_t, y_t) = h_{a_i} \exp\left(-2|\mathbf{r}_{p_i}|^2 / w_{zeq}^2\right), \; i = 1, \ldots, M. \tag{3}$$

Then, to identify the convexity of $f_i(x_t, y_t)$, the Hessian matrix of (3) should be positive semi-definite (PSD). The determinant of the Hessian matrix can be calculated as

$$\det(\mathbf{H}_{f_i}) = A^2 w_{zeq}^2 \left(w_{zeq}^2 - 4|\mathbf{r}_{p_i}|^2\right), \tag{4}$$

where $A = -4h_{a_i} \exp\left(-2|\mathbf{r}_{p_i}|^2 / w_{zeq}^2\right) / w_{zeq}^4$. Note that since A always has negative value, (4) is always non-negative if $|\mathbf{r}_{p_i}| \leq w_{zeq}/2$, which is the convexity condition of (2).

III. Conclusion

In this paper, we proposed the tendency between received SNR and the center coordinates of the transmitted beam footprint in a SIMO FSOC. By identifying the convexity of the problem, it was confirmed that under specific conditions, beam pointing to the center of its paired receiver lens may not maximize the performance and through optimizing the beam pointing, we can further improve performance in FSOC adopting spatial diversity.

References

[1] M. A. Khalighi and M. Uysal, "Survey on free space optical communication: A communication theory perspective," *IEEE Commun. Surv. Tutor*, vol. 16, no. 4, pp. 2231–2258, 2014.
[2] A. K. Majumdar, *Advanced free space optics (FSO): a systems approach*. Springer, 2014, vol. 186.
[3] M. T. Dabiri, M. J. Saber, and S. M. S. Sadough, "On the performance of multiplexing FSO MIMO links in log-normal fading with pointing errors," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 9, no. 11, pp. 974–983, 2017.
[4] M. A. Amirabadi and V. Tabataba Vakili, "A new optimization problem in fso communication system," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1442–1445, July. 2018.

Transactions Time Synchronization for Time-Critical Blockchain

Ahmad Akmaluddin Mazlan[1], Azizul Azizan[1], Noor Hafizah Binti Hassan[1], Abdul Ghafar Bin Jaafar[1], Mohd Syahid Mohd Anuar[1], Mohamad Sofian Abu Talip[2]
[1]*Faculty of Artificial Intelligence (FAI), Universiti Teknologi Malaysia (UTM),* Kuala Lumpur, Malaysia
[2]*Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya (UM),* Kuala Lumpur, Malaysia

Corresponding author: Ahmad Akmaluddin Mazlan (e-mail: ahmadakmaluddin@graduate.utm.my)

***Abstract*— Time synchronisation is crucial in the field of time-critical applications. Although blockchain networks are chronologically ordered, Hyperledger Fabric (HLF) sequential processing during transaction validation causes time drift, impacting inaccurate transaction timestamps. These drifted transaction timestamps are exacerbated and decreased validity with increased number of nodes for time-critical applications. This research adopts and compares time synchronisation methods; (1) Tiny-Sync (TS) protocol and (2) Maximum Likelihood Estimator (MLE) to correct inaccurate timestamps for time-critical transactions, reflecting real-world time, for different node counts (two and three nodes). This research evaluates both methods through accuracy metrics. The results show in homogeneous setups that MLE outperforms TS with 6.68% higher MSE evaluation. However, R2 and TV show comparable result. This implies that both adopted methods can effectively synchronise and improve transaction timestamp accuracy, even in larger setups.**

Keywords—Time-Critical Blockchain, Hyperledger Fabric, Transaction Timestamps, Homogenous, Accuracy.

I. INTRODUCTION

Blockchain technology holds great promise for revolutionizing various industries such as finance, healthcare, and supply chain management by enhancing security, transparency, and efficiency [1]. However, several challenges hinder the realization of its full potential, with time synchronization being a critical concern [2]. Inaccurate timestamps can result in failed transactions, financial losses, and security vulnerabilities [3], [4], [5], highlighting the importance of precise time synchronization in blockchain networks. Timekeeping accuracy is essential for maintaining the correct order of transactions, particularly in time-sensitive applications like supply chain management, healthcare, and financial markets [6], where the sequencing of events is paramount for operational integrity and regulatory compliance.

This paper is structured as follows. Section II is about related works and explores time-critical blockchain and time synchronization. In Section III, research methodology, design, and implementation. Section IV discovers the results of the research questions. Section V addresses the findings of time synchronization for a time-critical blockchain system. Lastly, Section VI concludes our research and presents future directions.

II. RELATED WORKS

This paper is structured as follows. Section II is about related works and explores time-critical blockchain and time synchronization. Section III is about research methodology, design, and implementation. Section IV highlights the results, addresses the findings of time synchronization, and makes recommendations for a time-critical blockchain system. Lastly, Section V concludes our research with a summarisation and presents future directions.

A. Time-Critical Blockchain

Time-critical blockchain is a variant of blockchain technology specifically designed to handle time-sensitive transactions. The time-critical blockchain uses timestamping to record the exact time a transaction occurred [4], [7]. Timestamps in blockchain systems are critical for transaction validation, preventing double-spending, maintaining chronological order of blocks, and facilitating correct sequencing of transactions and consensus among network devices [8], [9], thereby ensuring the integrity and coherence of the distributed ledger. This timestamp verifies the transaction and ensures that it is added to the blockchain in the correct order [10].

In addition, using time synchronisation protocols in time-critical applications ensures that the network nodes agree on the correct time [11]. Synchronization protocol prevent time manipulation and ensure the network operates as intended [12]. Although blockchain networks are decentralized, the time synchronization method, such as the Network Time Protocol (NTP), is still centralised-based [13], [14]. This results in issues like time drift and inaccurate timestamps, which can significantly affect the operations within a network.

In blockchain systems, chain codes or smart contracts can be programmed for time-sensitive execution, becoming valid or invalid at specific dates, times, time windows, or recurring intervals. Like in recent study by [15] found additional complications related to these timestamp drifts where the timestamps of transactions could not effectively reference real-world time because the execution accuracy decreases markedly near interval boundaries. The interval boundaries can be vital for applications requiring time-specific actions, such as financial transactions or supply chain operations. Other than that, block timestamps in blockchain technology are not actual representations of world time, but references (with timestamps) indicating when a block was validated [16].

B. Time Synchronization In Blockchain System

Time synchronization in blockchain technology is crucial in ensuring all nodes or participants within a network maintain a unanimous and accurate understanding of time. When using a distributed and decentralised system, nodes rely on their internal clocks to keep track of time [17].

Trustless time synchronisation methods have leveraged time information in block headers to estimate true time with high accuracy, demonstrating global synchronisation despite the non-deterministic behaviour of blockchains [14]. This process is key to preserving the network's functionality and ensuring blockchain timestamps align with real-world scenarios and involves overcoming network latency and clock drift challenges. Accurate timestamp plays a pivotal role in

maintaining the coherence and integrity of distributed ledgers, particularly in applications such as IoT and smart contracts.

III. Research Methodology

The methodology will undergo the following five stages: experimental design, data collection, drift detection through TS and MLE, time correction, and evaluation metrics.

A. Experimental Design

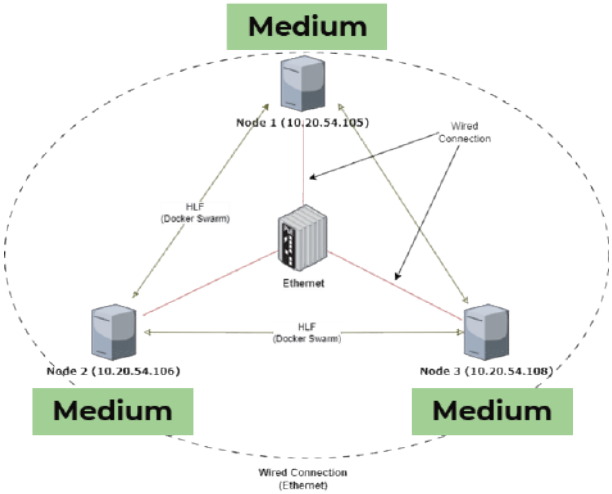


Fig. 1. Example of network topology based on HLF

Fig. 1 illustrates a diagram of a computer network topology, specifically highlighting a Hyperledger Fabric network connected through a Wireless Local Area Network (WLAN) that comprises three nodes and a router connecting all the nodes. All 3 nodes are represented as computers with same performance specifications, which are CPU 3.20 GHz, 8GB Memory, & 1TB storage.

This research employs two HLF nodes setup such as 2 nodes of homogeneous as the initial experiment. Meanwhile, Fig.1 shows the 3 nodes of homogeneous as the main experiment and adopting the transaction time synchronisation for this research.

B. Data Collection of Transactions Timestamp

Firstly, we recorded a total of 300 transactions from HLF and analysed the statistical distribution. We implemented the experimental design and executed the following capture script through chaincode.

```
String txTimestampString = stub.getTxTimestamp().toString(); // Get UNIX timestamp in milliseconds
```

C. Drift Detection

Drift detection played a huge role in the first part of the time synchronisation mechanism. The drift detection operation was designed based on TS and MLE. Drift detection serves as the transaction timestamp overshoots the offset limit and it will trigger the time correction operation.

Tiny-Sync (TS) provided optimal bound to serve as the offset limit. The optimal bound based on estimated offset (a_{13}) and drift (b_{13}). The three timestamps (t_o, t_b , t_r) collected from the reference node and client node, form a data point which effectively limits the possible values of parameters a_{12} and b_{12} in Eq. (1). Fig. 2 shows that the green line is the optimal bound, calculated through TS.

$$t_1(t) = a_{13}t_3(t) + b_{13} \quad (1)$$

Once $\boldsymbol{a_{12}}$ and $\boldsymbol{b_{12}}$ are estimated, node 1 (t_1) can always correct the reading of the local clock (using Eq. (1)) to have it match the readings of the clock at node 2 (t_2). The optimal bound serves as the offset limit to detect the time drift, thus invoking the time correction.

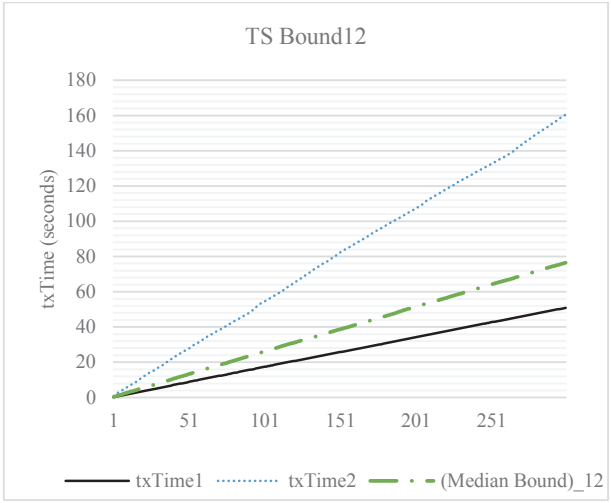


Fig. 2. Example of TS Median Bound as Drift Detection

Maximum Likelihood Estimator (MLE) provide the offset limit based on $\hat{\theta}_1 = \bar{X}$ (≈mean value). The MLE offset limit was calculated within the specific time windows based on the number of synchronisation frequency of TS.



Fig. 3. Example of MLE Offset Limit as Drift Detection

Fig. 3 shows that the orange line is the offset limit, calculated through MLE. The offset limit serves to detect the time drift, thus invoking the time correction.

D. Time Correction

Time correction is the second operation of the time synchronisation mechanism. The time correction will be invoked when time drift occurs through the drift detections of TS and MLE. Thus, the outputs will be the correction of the transaction timestamp and the sub-metric of complexity evaluation called synchronisation frequency.

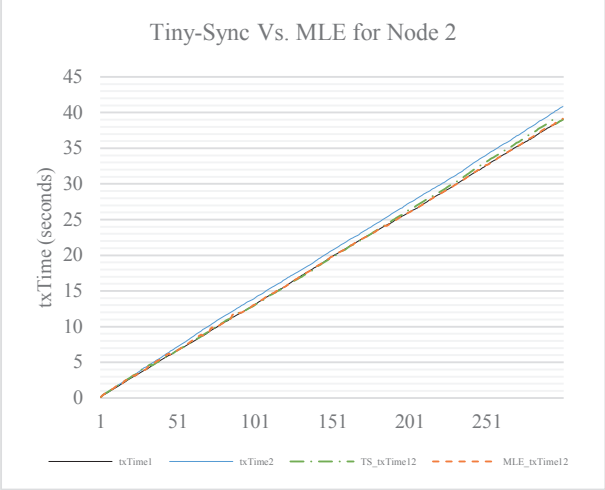


Fig. 4. Example of TS vs MLE Time Correction

Fig. 4 shows the example of time correction using TS and MLE for client node (Node 2) where the correction was tight as the differences of original txTime2 (client node) and txTime1 (reference node) was around 1 seconds. Thus, the aim of the time correction to decrease the drift of transaction timestamps and increase the time validity for the client node.

E. Evaluation Metrics

The fourth stage assesses the effectiveness of the time synchronization methods for time-critical blockchain. This stage examines the method's impacts on time synchronisation based on the accuracy and complexity evaluations. Accuracy evaluation metrics include R-squared (R^2) value, Mean Squared Error (MSE), and transactions validation (TV). This stage verifies the method's accuracy and complexity. Table III describes the aim of each evaluation metric.

TABLE I. EVALUATION METRICS

Metrics	***Aim***
R^2	To compare which value closer to 1
MSE	To compare which value closer to 0
TV	To compare which validity is higher

IV. RESULTS AND DISCUSSIONS

A. Time Distribution before Time Synchronisation

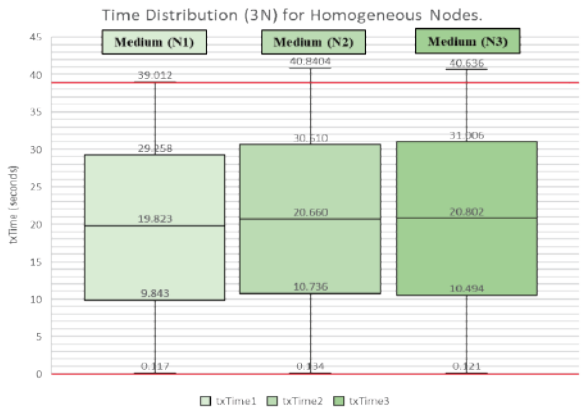


Fig. 5. Time Distribution before Time Synchronisation for Node 2 and Node 3

Fig. 5 shows the boxplot indicating that txTime2 and txTime3 operate efficiently comparing with the reference node, txTime1 under homogeneous setup. The time constraint interval is from 0.117 seconds to 39.012 seconds, where it indicates the successful transaction validation for the experiment.

TABLE II. TRANSACTION VALIDATION BEFORE TIME SYNC

Tx Range	***Validity***	***TxTime1***	***Homogeneous nodes***	
			TxTime2	***TxTime3***
1-300	Valid	300	286	287
Validity (%)	100.00	95.33	95.67	

Table II shows under 3N homogeneous setup, a slightly reduced overall validity rate of around 95% for txTime2 and txTime3.

Fig. 6 shows the time distribution comparison between the original timestamps and the corrected timestamps of node 2 and node 3 (TS and MLE) and places them in context with a baseline set (txTime1). Implementing TS and Maximum MLE has slightly improved transactions timestamp. Both methods have not only reduced the quartile values compared to the original timestamps but have also consistently maintained the minimum value of 0.134 sec for node 2 and 0.121 sec for node 3, demonstrating reliable performance at the lower end. Furthermore, both methods have decreased the maximum values compared to the original txTime2 and txTime3, indicating a reduction in longer transactions timestamp.

In summary, the Tiny-Sync (TS) and Maximum Likelihood Estimator (MLE) corrections have effectively reduced the transactions timestamp for txTime 2 and txTime3, bringing them closer to txTime1's performance. Across the board, TS and MLE shows a comparable difference in synchronisation, suggesting MLE only slightly higher in accuracy.

B. Accuracy Evaluations

Table IV shows the results of accuracy evaluations for 3 nodes in homogeneous setups.

TABLE III. 3N HOMOGENOUS SETUP

	R^2	***MSE***	***TV (%)***
Original	0.991	1.169	95.33
TS	0.999	0.086	99.67
MLE	1.000	0.008	99.33

For a homogenous setup, R^2 and TV highlight the negligible difference for both TS and MLE. However, MSE shows that MLE is better than TS with with value closer to 0. Based on the observed Mean-Squared Error (MSE) improvements, the Maximum Likelihood Estimator (MLE) demonstrates superior performance with a 99.32% improvement, which is 6.68% higher than Tiny-Sync's (TS) 92.64% improvement. Thus, the MSE result shows that accuracy of MLE is better than TS under the homogeneous setup.

C. Time Distribution after Time Synchronisation

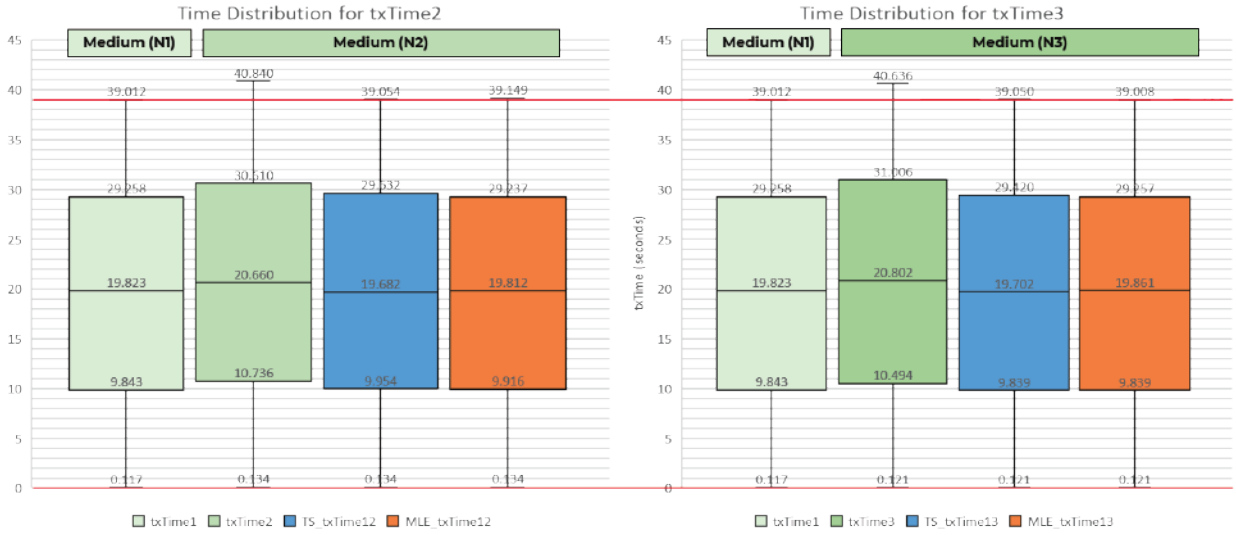


Fig. 6. Time Distribution after Time Synchronisation for Node 2 and Node 3

V. Conclusion

This research examines the impact of node count on transaction timestamp and validation, offering valuable insights into their accuracy through time synchronisation experiments. In a homogeneous setup, MLE outperforms TS based on MSE but R^2 and TV show comparable result. For MSE, MLE shows superior performance with 6.68% higher than TS improvement. Future research is recommended to include the complexity evaluations, explore different time synchronisation methods, the impact of network size on time sync, and the development of enhanced time synchronisation techniques for transactions timestamp. This research comprehensively analyses the Tiny-Sync and Maximum Likelihood Estimator (MLE) time synchronisation methods adopted for time-critical transactions in permissioned blockchains.

Acknowledgement

The authors would like to express their appreciation to the Faculty of Artificial Intelligence, Universiti Teknologi Malaysia for realizing and supporting this research work.

References

[1] C. Laroiya, D. Saxena, and C. Komalavalli, "Chapter 9 - Applications of Blockchain Technology," in Handbook of Research on Blockchain Technology, S. Krishnan, V. E. Balas, E. G. Julie, Y. H. Robinson, S. Balaji, and R. Kumar, Eds., Academic Press, 2020, pp. 213–243. DOI: 10.1016/B978-0-12-819816-2.00009-5.

[2] K. Fan et al., "Blockchain-based trust management for verifiable time synchronization service in IoT," Peer--Peer Netw. Appl., vol. 15, no. 2, pp. 1152–1162, Mar. 2022, DOI: 10.1007/s12083-021-01276-2.

[3] J. H. Mosakheil, "Security Threats Classification in Blockchains," Culminating Proj. Inf. Assur., May 2018, [online]. Available: https://repository.stcloudstate.edu/msia_etds/48

[4] G. Estevam, L. M. Palma, L. R. Silva, J. E. Martina, and M. Vigil, "Accurate and decentralized timestamping using smart contracts on the Ethereum blockchain," Inf. Process. Manag., vol. 58, no. 3, p. 102471, May 2021, DOI: 10.1016/j.ipm.2020.102471.

[5] S. Arunachalam, S. H J, G. Sivagurunathan, S. Das, D. Anand, and T. Raj M, "Cloud-based Decentralized Smart Healthcare for Patient Monitoring on Deep Learning," in 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), May 2023, pp. 459–466. DOI: 10.1109/ICAAIC56838.2023.10141120.

[6] P. Dutta, T.-M. Choi, S. Somani, and R. Butala, "Blockchain technology in supply chain operations: Applications, challenges and research opportunities," Transp. Res. Part E Logist. Transp. Rev., vol. 142, p. 102067, Oct. 2020, DOI: 10.1016/j.tre.2020.102067.

[7] D. Appelbaum and R. A. Nehmer, "Auditing Cloud-Based Blockchain Accounting Systems," J. Inf. Syst., vol. 34, no. 2, pp. 5–21, Jun. 2020, DOI: 10.2308/isys-52660.

[8] M. M. Akhtar, M. Z. Khan, M. A. Ahad, A. Noorwali, D. R. Rizvi, and C. Chakraborty, "Distributed ledger technology based robust access control and real-time synchronization for consumer electronics," PeerJ Comput. Sci., vol. 7, p. e566, Jun. 2021, DOI: 10.7717/peerj-cs.566.

[9] H. Takahashi and U. Lakhani, "Kernel Based Cache Effect on Blockchain Node Synchronization," in 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), Oct. 2020, pp. 784–786. DOI: 10.1109/GCCE50665.2020.9292007.

[10] L. Zhang and D. Kim, "A Peer-to-Peer Smart Food Delivery Platform Based on Smart Contract," Electronics, vol. 11, no. 12, Art. no. 12, Jan. 2022, DOI: 10.3390/electronics11121806.

[11] D. Petrov, S. Melnik, and T. Hämäläinen, "Distributed GNSS-based Time Synchronization and applications," in 2016 8th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT), Oct. 2016, pp. 130–134. DOI: 10.1109/ICUMT.2016.7765345.

[12] R. Peters, Ed., "Date and Time Manipulation," in Expert Shell Scripting, Berkeley, CA: Apress, 2009, pp. 19–24. DOI: 10.1007/978-1-4302-1842-5_3.

[13] D. L. Mills, "Internet time synchronization: the network time protocol," IEEE Trans. Commun., vol. 39, no. 10, pp. 1482–1493, Oct. 1991, DOI: 10.1109/26.103043.

[14] E. Regnath, N. Shivaraman, S. Shreejith, A. Easwaran, and S. Steinhorst, "Blockchain, what time is it? trustless datetime synchronization for iot," in 2020 International Conference on Omni-layer Intelligent Systems (COINS), IEEE, 2020, pp. 1–6. Accessed: Dec. 07, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9191420/?casa_token=2G_-AkYOKscAAAAA:GIPfvDhUY7fZpEmBfBb-kQtIilz_ygVfOjG8bwoYJ-JfwJyA9tLt3_f9dVlBQfyJu_UwsGySjCC3

[15] T. Eichinger and M. Ebermann, "Can We Effectively Use Smart Contracts to Stipulate Time Constraints?" in 2022 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS), Aug. 2022, pp. 11–18. DOI: 10.1109/DAPPS55202.2022.00010.

[16] J. Ladleif and M. Weske, "Time in Blockchain-Based Process Execution," in 2020 IEEE 24th International Enterprise Distributed Object Computing Conference (EDOC), Oct. 2020, pp. 217–226. DOI: 10.1109/EDOC49727.2020.00034.

[17] D. L. Mills, "Modelling and analysis of computer network clocks," Technical Report 92-5-2, Electrical Engineering Department, University of …, 1992.

Study on Trends in Blockchain-Based Energy Trading

Da-Jung Lee
ICT Convergence Research Center
Kumoh national institute of Technology
Gumi, Republic of Korea
djlee@kumoh.ac.kr

Ik-Hyun Kwon
Dept. of IT Convergence Engineering
Kumoh national institute of Technology
Gumi, Republic of Korea
inkwonaf@kumoh.ac.kr

Dong-Seong Kim
Dept. of IT Convergence Engineering
Kumoh national institute of Technology
Gumi, Republic of Korea
dskim@kumoh.ac.kr

***Abstract*—This study investigates cases where blockchain technology is applied in the energy trading sector, examining the potential value and prospects of blockchain technology in the energy trading market. Through previous research, it was found that blockchain-based energy trading offers advantages such as decentralization, transparency, security, trustworthiness of energy transactions, and efficient management with real-time monitoring. However, additional research is needed to address regulatory constraints and ensure the stability and scalability of practical industrial applications. Furthermore, to overcome inefficiencies in energy markets worldwide, blockchain-based distributed trading systems, including automated demand-supply contracts and incentivized compensation mechanisms, are expected to contribute to driving innovation in future energy markets.**

***Index Terms*—Blockchain, Blockchain-based Power Trading, Energy Trading System**

I. Introduction

Blockchain technology is being applied and utilized in the field of energy trading. It presents the possibility of resolving issues such as trust, security, and inefficiency caused by the existing centralized structure of power trading systems. Moreover, various empirical cases have already demonstrated the efficiency and transparency enhancements in power trading achieved through the application of blockchain technology. In 2016, Brooklyn Microgrid contributed to building a community involved in energy production and consumption by decentralizing the power network, enhancing transparency and efficiency in transactions. In 2017, Power Ledger supported energy trading in solar power plants and microgrids based on blockchain, enabling peer-to-peer transactions between energy producers and consumers and realizing intermediary-free transactions in the power market[1]. Recent trends in blockchain-based energy trading research include the development of smart contract technology, which automates and enhances the efficiency of power transactions, leading to increased speed and transparency of transactions. Integration research combining blockchain technology and artificial intelligence aims to optimize the management and operation of energy systems, thereby increasing energy system efficiency and enabling the prediction of energy consumption patterns. Standardization of blockchain in the energy sector is also a significant trend, aiming to enhance security, interoperability, and overall efficiency of energy systems by enabling broader adoption of blockchain technology. The objectives of this study are as follows: first, to review previous research on applying blockchain to power trading systems and explore the future development potential of blockchain-based power trading systems. Second, to reconfirm the potential and value of blockchain technology in power trading systems in major electricity markets and present the utility of blockchain for sustainable energy market innovation.

II. Related Works

A. Blockchain Technology Overview

Blockchain stores transaction records in blocks, forming a chain, and applies security technologies such as digital signatures, public keys, and hash functions to replicate and store them on numerous computers. Transaction records are provided to all users participating in the transaction, and each participant shares the transaction records whenever a transaction occurs, preventing data forgery or tampering. Smart contracts are programs that automatically execute when conditions are met, enabling transactions to be processed without intermediaries or mediators, and can be applied to energy trading, carbon emissions trading, and microgrid operations.

B. Blockchain-based electricity trading research

Summarizing the findings of these studies, we can identify the following advantages and limitations. First, blockchain-based power trading offers decentralization, transparency, and security. Second, it enhances efficiency through process automation and real-time monitoring, ensuring reliability by preventing fraudulent activities and data tampering in power trading systems. Third, it promotes environmentally friendly power systems by enabling real-time tracking of energy and facilitating the utilization of renewable energy sources. However, there are considerations in technical, legal, and economic aspects. Concrete cases of industrial application, along with research on stability and scalability, are needed to enhance system performance.

C. Electricity markets and issues

Examining the electricity markets and issues in major countries reveals the following: Current Status and Policies

TABLE I
BLOCKCHAIN-BASED ELECTRICITY TRADING PRECEDING STUDIES, RESEARCHER COMPILATION, SOURCE: [2-7]

Article	Author	Year	Content	Conclusion	Limitations
Blockchain Integration and Its Impact on Renewable Energy	John Doe et al.	2024	Examines blockchain's role in decentralized power dispatching, certificate trading, and energy management. Integration can reduce costs by 20%, increase renewable utilization by 15%, and enhance efficiency by 25%.	Blockchain improves transparency, security, and automation in energy management.	Scalability issues may limit transaction speed and privacy concerns could increase data protection costs by 10%.
Blockchain-Based Peer-to-Peer Energy Trading System Using Python	Alan Turing et al.	2024	Discusses a system using Ethereum and smart contracts, reducing transaction costs by 35% and increasing automation efficiency by 50%.	Enhances direct energy trading between prosumers and consumers, reducing reliance on central providers.	Security vulnerabilities could increase risk mitigation costs by 20%, and robust smart contract development may add 30% to project time.
Research on Distributed Energy Consensus Mechanism Based on Blockchain in Virtual Power Plant	Liu et al.	2022	Analyzes VPP composed of DERs, improving transparency and efficiency. The new consensus mechanism reduces collaboration time while maintaining fault tolerance.	Enhances DER coordination and security in VPP scenarios.	Needs more real-world validation, potentially increasing pilot costs by 20%, and integration with existing systems may add 25% to total costs.
Blockchain-based Energy Trading in Electric-Vehicle-Enabled Microgrids	Zhang et al.	2023	Proposes a scheme for energy trading between EVs and critical loads, enhancing energy management. Demonstrates a 30% improvement in transaction efficiency and a 25% cost reduction.	Directly connects prosumers and consumers, improving energy management.	Scalability issues could increase infrastructure costs by 15%, and integration with existing systems may require 20% additional development time.
V2GNet: Robust Blockchain-Based Energy Trading Method and Vehicle-to-Grid Coordination	Kumar et al.	2023	Develops a robust trading method and coordination for V2G systems, improving efficiency by 35% and reducing transaction costs by 20%	Facilitates efficient energy management between EVs and the grid.	Requires extensive infrastructure upgrades, potentially increasing initial costs by 30%, and technology upgrades may add 25% to maintenance costs.
Blockchain-Based Clean and Green Energy Trading System	Saibal et al.	2023	Discusses a system using private Ethereum blockchain for green energy trading, improving transaction transparency by 40% and reducing costs by 25%.	Promotes green energy trading with enhanced transparency and reduced costs.	Dependence on blockchain infrastructure could increase operational costs by 10%, and potential integration issues may add 20% to development time.

in Major Electricity Markets: Germany: Facing economic pressure on power companies due to energy overproduction and low prices, leading to energy wastage and decreased profits. In response, legislation on energy transition has been enacted to support renewable energy with subsidies and tax benefits, promoting sustainable energy production and usage. UK: Suffers from unstable power supply and customer service issues due to insufficient power network. Aims to enhance network efficiency and improve customer service through smart grid investment plans. France: Experiences environmental degradation and decreased energy efficiency due to a lack of technology and policies for sustainable energy production and usage. Implements energy efficiency programs and industrial policies to promote sustainability. Netherlands: Faces issues of market inefficiency and fairness due to regulatory gaps and lack of transparency in trading processes. Implements comprehensive policies to improve regulatory enforcement and introduce technological solutions for transparent trading. USA: Various states have different power market structures and regulatory regimes, leading to inconsistency and inefficiency in the electricity market. Recommendations include introducing consistent market regulations, promoting inter-market competition, expanding renewable energy, and increasing consumer participation. South Korea: High energy import dependency (95%) and regional disparities in power consumption and production lead to supply imbalances and market instability. Government implements policies to improve energy efficiency, increase renewable energy obligations and supply ratios, and promote electrification in the transportation sector for carbon neutrality. China: Faces instability in the strategic market structure due to excessive power consumption and inadequate supply of renewable energy. Complex market structure and intense competition hinder market efficiency. Implements multifaceted energy-saving and environmental protection plans to address excessive energy consumption and pollution[8].

III. CASE STUDY

Here are successful cases of blockchain-based energy trading platforms. WePower, an energy trading platform in Estonia, utilized blockchain technology in 2017 to connect energy suppliers who want to sell energy generated from solar and

TABLE II
RESEARCHER SUMMARY SOURCE [9-15]

Country	Status	Policy Name / Source	Solution
Germany	Overproduction and low prices	Energy Transition Law / European Commission, "Electricity Market Reform in the EU", 2020[9]	Policy and subsidy support for increasing energy efficiency
United Kingdom	Shortage in the power network and customer service issues[10]	Smart Grid Investment Plan/ National Grid Electricity System Operator, "Future Energy Scenarios", 2021	Introduction of policies to promote infrastructure investment and efficient management of power networks using smart grid and IoT technology
France	Sustainable energy production and use	Energy Efficiency Program/ European Environment Agency, "Renewable Energy in Europe", 2020[11]	Provision of subsidies and tax benefits for renewable energy, introduction of policies considering energy efficiency in construction and industrial sectors
Netherlands	Regulation of the power market and improvement of transaction transparency	Improvement of transaction transparency in power trading/ Authority for Consumers and Markets, "Electricity Market Monitoring Report", 2021[12]	Introduction of technological solutions to strengthen power market regulations and improve transaction transparency
United States	Increase in energy demand and associated load management issues	Smart Grid Implementation Plan/ U.S. Energy Information Administration, "Electric Power Monthly", 2019[13]	Introduction of smart grid and energy storage technologies, policy and technology development for power demand management
South Korea	Increase in renewable energy supply and supply management issues	Promotion Plan for Renewable Energy Supply/ Ministry of Trade, Industry and Energy, "Renewable Energy 3020 Plan", 2019[14]	Expansion of renewable energy generation facilities and subsidy support, policy and technology development for energy demand management and efficient energy use
China	Issues of energy consumption and environmental pollution	Energy Saving and Environmental Protection Plan/ Ministry of Trade, Industry and Energy, "Renewable Energy 3020 Plan", 2019[15]	Industrial restructuring for increased energy efficiency, introduction of policies for reducing energy consumption and promoting alternative energy use for environmental protection

TABLE III
BLOCKCHAIN-BASED POWER TRADING PROJECTS, SOURCE: 2020 COMMUNITY BLOCKCHAIN-BASED ENERGY TRADING TECHNOLOGY TRENDS

Project	Country	Type of Power Trading	Energy Source
Trasactive Grid L03 Energy	US	Peer-to-peer	Solar
SolarCoin	Multiple	Solar power generation companies and SolarCoin Foundation	Solar
Power Ledger	Australia	Peer-to-peer or energy company	Renewable energy
Grid Singularity	Austria	Solar power generation company	Solar
GE and Enedis	France	Peer-to-peer or corporate	Solar
TEPCO	Japan	Peer-to-peer	Solar
Alliander	Netherlands	Prosumer-to-prosumer	Solar
EnergoLabs	China	Peer-to-peer	Solar energy storage

wind power plants to consumers. The platform uses smart contracts to automate energy trading, records the entire transaction process transparently, and ensures transaction reliability. Additionally, microgrid projects in the United States operate independently or can be connected to traditional power grids, using blockchain technology to increase local energy self-sufficiency, enable energy sharing and trading, and promote the efficient utilization of renewable energy. Transactive Grid and LO3 Energy experimented with a blockchain-based peer-to-peer energy trading project on President Street in Brooklyn, New York. Five generating units with solar panels are located on one side of the street, while five demand units on the other side purchase surplus electricity. The generating units sell the produced energy to the connected demand units, and the feasibility of the trading system's interoperability with the existing power grid was confirmed using blockchain technology. Participants in the project have bidirectional smart meters that allow them to monitor their energy production and consumption in real-time. Furthermore, these smart meters are connected to the blockchain network, recording all transaction status so that anyone in the network can verify the information[16].

IV. BLOCKCHAIN-BASED ENERGY TRADING PROCESS

A. Blockchain-based energy trading process

The electricity trading process proceeds as follows: registration, sale conditions registration, purchase conditions registration, contract execution, transmission data recording, settlement, and initialization, as depicted in the above diagram. Upon user registration with the assumption of using coins, participants are granted participation rights in electricity trading and receive additional coins. Sellers and buyers register their respective trading conditions, which are recorded on the blockchain and executed. The electricity trading platform matches sale and purchase conditions to execute contracts, and

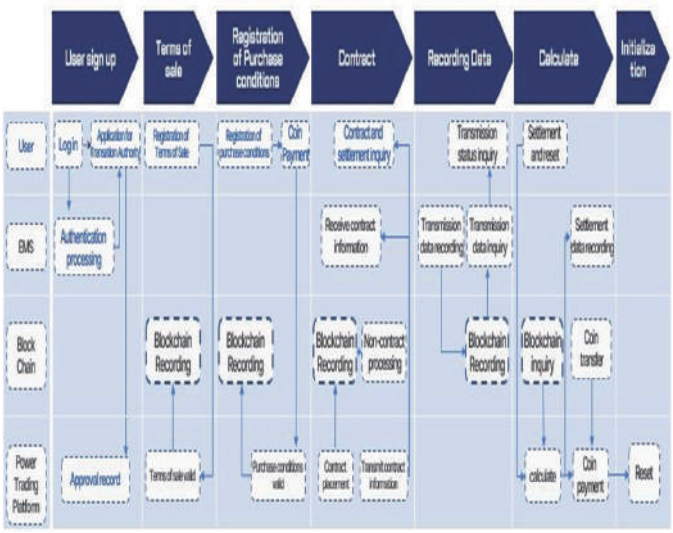


Fig. 1. Energy Trading Process, Source: Trends in Community-Based Blockchain Energy Trading Technology[17]

this information is recorded on the blockchain. Additionally, the contract details are provided to the EMS (Energy Management System) or users. Upon completion of settlement, the process is initialized.

V. Conclusions

The utility and potential value of blockchain-based electricity trading systems, as examined above, are as follows: First, they contribute to minimizing the intermediary role in electricity trading systems and enhancing the trustworthiness of transactions through decentralized systems. Second, by enhancing transparency and immutability of transaction records, blockchain-based electricity trading systems improve the transparency of the trading process and prevent tampering or forgery of transaction records. Third, by automating the trading process and reducing intermediaries' fees, blockchain-based electricity trading systems increase efficiency and contribute to cost savings. Fourth, the utilization of smart contracts enables the programming of conditions related to electricity trading, allowing for automated transaction execution. Last, blockchain technology enables the maintenance of a balanced supply and demand of energy and promotes the efficient utilization of renewable energy through distributed electricity trading networks. These advantages and technological values can contribute to the establishment of innovative and sustainable energy markets in the future. Furthermore, the impact of blockchain technology on sustainable energy market innovation includes enhancing transparency and reliability in energy trading, building distributed trading networks, supporting the integration and efficient operation of renewable energy, enabling automated transactions through smart contracts, and providing insights into the future direction of the energy market.

In the future direction of research, it is necessary to analyze how blockchain technology can contribute to innovation and sustainable development of the energy market through research on the advantages and disadvantages of more in-depth industrial application using blockchain technology, research on international standards to ensure the stability and scalability of the blockchain chain-based energy trading system, and research on how to optimize the operation and management of blockchain-based energy trading system using machine learning and deep learning technology.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 100)

References

[1] Power Ledger Team, 2017, ""Peer-to-Peer Energy Trading Using Blockchain: Power Ledger Case Study", Power Ledger Official Website and Research Publications, pp.1-3.
[2] John Doe et al,. 2024, "Blockchain Integration and Its Impact on Renewable Energy", MDPI, pp.102-120.
[3] Alan Turing et al., 2024, Blockchain-Based Peer-to-Peer Energy Trading System Using Python, SSRN, pp.34-56.
[4] Liu et al.,2022, Research on Distributed Energy Consensus Mechanism Based on Blockchain in Virtual Power Plant, MDPI, https://www.mdpi.com/79†Next Generation of Security and Privacy in IoT, Industry 4.0, 5G Systems and Beyond), pp.45-60.
[5] Zhang et al., 2023, Blockchain-based Energy Trading in Electric-Vehicle-Enabled Microgrids, ScienceDirect, pp.210-230.
[6] Kumar et al., 2023, V2GNet: Robust Blockchain-Based Energy Trading Method and Vehicle-to-Grid Coordination, IEEE Xplore, pp.178-194.
[7] Kim et al., 2023, A Blockchain based Framework for Secure and Decentralized Energy, IEEE Xplore, pp.99-115.
[8] Li Yuan, Sha Yu, Yishu Zhou,2018, "Analysis of China's power market structure and strategy"
[9] European Commission, "Electricity Market Reform in the EU", 2020, Electricity Market Reform in the EU.
[10] National Grid Electricity System Operator, 2021., "Future Energy Scenarios", Future Energy Scenarios
[11] European Environment Agency, "Renewable Energy in Europe", 2020.,Renewable Energy in Europe.
[12] Authority for Consumers and Markets, "Electricity Market Monitoring Report", 2021.,Electricity Market Monitoring Report.
[13] U.S. Energy Information Administration, "Electric Power Monthly", 2019, Electric Power Monthly.
[14] Ministry of Trade, Industry and Energy, "Renewable Energy 3020 Plan", 2019,Renewable Energy 3020 Plan.
[15] National Energy Administration, "13th Five-Year Plan for Energy Development",13th Five-Year Plan for Energy Development.
[16] Y.B, Moon, et al., blockchain Based Energy Trading Technology for the Community, 2020, ETRI, pp.97-99.
[17] Park Chan Kuk, 2016, "Study on the Feasibility of P2P Power Trading in South Korea". Ad-hoc Research pp.15-10, Korea Energy Economics Institute.

Enhanced Secure Communication in Maritime Tactical Networks Using Quantum Cryptography

Esmot Ara Tuli, Mohtasin Golam, Jae-Min Lee, and Dong-Seong Kim
Networked Systems Laboratory, Department of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea.
(esmot, golam248, ljmpaul, dskim)@kumoh.ac.kr

***Abstract*—The growing occurrence of cyber attacks requires advancing more complex encryption methods to protect sensitive communication in marine tactical networks. The study introduces Quantum DNA Partial Permutation (QDPP), a cryptographic scheme that combines quantum cryptography, DNA computing, and permutation-based methods to secure confidential communication in marine tactical networks. The protocol entails the interchange of two tables: a DNA Mapping table and a Rules table, which facilitate the sender and recipient's encoding and decoding of messages. The encryption procedure entails applying permutation rules from a pre-shared Rules database, selected according to a specific basis selection. The proposed approach uses the IBM Qiskit simulator, demonstrating its viability and capacity to enhance security in demanding underwater conditions.**

***Index Terms*—Maritime tactical networks, quantum cryptography, DNA mapping, and partial permutation encryption.**

I. Introduction

Maritime tactical networks (MTNs) are intricate systems created by combining many types of networks, including terrestrial broadband wireless networks, optical fiber networks, soldier-based personal area networks, and both wired and wireless local area networks [1]. The interconnected networks enable communication and data exchange between nodes, such as shipboard systems, aerial platforms, satellites, underwater sensors, submarines, naval combat systems, autonomous underwater vehicles, and other components. The marine industry is experiencing a growing trend towards decentralization, connectivity, and automation, increasing its vulnerability to cyber threats [2]. The World Economic Forum's Global Risks Report 2020 identifies cyber-attacks on maritime infrastructure as the fifth most significant threat [3]. An instance of a cyberattack is the Stuxnet worm, which targeted and caused disruption to the mechanical operations of an Iranian uranium enrichment facility [4]. The possibility of such an attack on ships and maritime infrastructure exists, which makes it necessary to secure communication in the MTNs.

Quantum cryptography is an innovative method for ensuring cybersecurity by utilizing the laws of quantum mechanics to create highly secure communication channels. Quantum cryptography uses the intrinsic characteristics of quantum particles, mainly photons, to protect data transfer, in contrast to classical encryption techniques [5]. The combination of quantum physics and encryption enables an unparalleled level of security in the transfer of data. The no-cloning theorem and quantum entanglement are two properties of quantum mechanics that make it impossible to intercept data [6]. Quantum cryptography provides a solution resistant to future advancements and potentially impervious to decryption even by the most powerful quantum computers. This system's high level of security makes it especially appealing for protecting susceptible data, such as financial transactions, classified national security information, and confidential medical records [7]. In MTNs, quantum cryptography can transform communication security in the quantum era as research continues to thrive.

Permutation-based cryptographic systems employ permutations to safeguard data by rearranging components according to a pre-established rule or key, rendering them less susceptible to unauthorized access [8]. Partial permutation techniques enhance the complexity of encryption algorithms, strengthening their resilience against cryptographic attacks. These methods effectively fight brute-force and differential attacks by increasing the complexity of recovering the original data without the key. Incorporating partial permutations into encryption algorithms enhances the scheme's ability to withstand cryptanalytic attacks. There are many modern uses for permutation-based methods in cryptography, such as quantum permutation cryptography, which is made to protect against attacks from quantum computers. In 2020, Kuang and Bettenburg presented the quantum permutation pad (QPP), a symmetric cryptographic method [9]. QPP uses linear algebraic representations of quantum gates, allowing it to be implemented in both quantum and classical systems.

DNA computing is an emerging discipline that harnesses the computational capabilities of DNA molecules, capitalizing on their extensive parallelism and dense storage capacity. This new approach shows promise in handling complex problems like extensive combinatorial issues, cryptography, encoding, and data manipulation beyond conventional digital technologies [10]. DNA-based encryption and steganography involve using DNA sequences to represent encrypted data. Adleman's research initiated the field of DNA computing, in which DNA molecules were employed to solve directed Hamiltonian path problems [11]. DNA coding techniques have been utilized for encryption, using the bio-inspired approach to perform

DNA permutation, addition, XOR operations, and encoding and decoding procedures. The double-helical structure of DNA is made up of the nucleotides adenine (A), guanine (G), cytosine (C), and thymine (T) arranged in different ways.

The study presents a novel encryption technique that combines quantum cryptography, DNA computing, and permutation methods to strengthen the security of marine tactical networks. This scheme offers a strong defense against complex cyber attacks by leveraging the distinct characteristics of each component to establish a comprehensive and resilient cryptographic solution. The integration is crucial to improving the security of the marine tactical network.

II. Proposed Scheme

This section extensively discusses the proposed scheme, including detailed explanations of each component utilized in this work. The QDPP encryption technique is a secure communication mechanism that begins with an initial step illustrated in Fig. 1. A secure quantum channel enables the transmission of two essential data tables: the DNA Mapping table and the Rules table. The DNA Mapping table is a reference, perhaps including a predetermined correlation between classical data bits and quantum representations, potentially encoded into photons. The Rules table prescribes the precise permutation operations employed during the encryption and decryption procedures. By sharing these tables, the sender and receiver have the essential resources to modify and comprehend quantum data. This initial interaction establishes the foundation for the following phases of encrypting and decrypting messages. During these stages, messages are altered and retrieved by utilizing shared information in the DNA Mapping and Rules tables. Below are the detailed explanations of the steps required for the encryption and decryption process:

A. Sharing the DNA Mapping and Rules Table

The technique involves the secure transmission of the DNA Mapping table and the Rules table between the transmitter and receiver using a secure quantum channel. The DNA Mapping table establishes a correlation between pairs of bits and specific nucleotides, while the Rules table delineates the permutation rules for encryption and decryption. This transaction lays the groundwork for subsequent procedures. The DNA Mapping table establishes a preset relationship between classical data bits and their quantum counterparts, while the Rules table specifies exact permutation techniques. The sender and receiver acquire a common understanding, which guarantees precise encoding and decoding of signals.

B. DNA Mapping

During this phase, the method transforms binary data into DNA sequences by associating each bit pair with a specific nucleotide. As an illustration, the bit pair '00' is transformed into the letter 'A', '01' into 'C', '10' into 'G', and '11' into 'T'. This process converts the complete binary message into a sequence of nucleotides. The sender then proceeds to the message encoding stage, where the classical binary message is transformed into a quantum representation for secure transmission. The translation relies on the DNA Mapping table, which functions as a Rosetta Stone, converting traditional bit sequences into their corresponding quantum equivalents. The binary message is systematically assigned to a specific nucleotide (A, C, G, or T) based on predetermined principles for each bit pair. This process efficiently converts the complete binary message into a series of nucleotides, generating a quantum representation of the original message. For example, the bit pair '00' could be consistently associated with the nucleotide 'A', '01' with 'C', and so forth. The DNA Mapping table is a vital connection between classical and quantum domains, facilitating encryption and ensuring message confidentiality.

C. Block Segmentation and Qubit Representation

Segmentation is dividing a nucleotide sequence into smaller, more manageable segments. The block size, denoted by κ, is determined by a modular operation represented as follows:

$$\kappa = N \mod i. \tag{1}$$

In this context, N denotes the length of the original binary message, whereas i represents a predefined block size measured in bits. This guarantees that each block contains an identical number of bits, which is essential for the encryption process since it facilitates the manipulation of the message.

Subsequently, every segmented data block is transformed into qubits, the fundamental components of quantum information. This process involves the use of Pauli operators. The Pauli operators represent rotations around specific axes of the Bloch sphere (X, Y, Z), providing a geometric representation of qubit states. The selection of specific options for this permutation (i.e., $\downarrow$, $\uparrow$, and $\rightarrow$) is of utmost importance and is pre-established in the Rules table, guaranteeing uniform implementation across the protocol. After the permutation-based encryption process, the DNA information blocks are sent along with the basis to the sender.

D. Partial Permutation Encryption and Decryption Process

The proposed technique employs permutation encryption, a process governed by the pre-shared Rules table. The encoded qubit blocks undergo operations to guarantee message secrecy. The selected basis determines the permutation rule, such as "ACGT" for the basis represented by the ($\downarrow$). For example, using the ($\downarrow$) basis could trigger the permutation rule "ACGT," which can be encoded in binary as "00011011." The rule governs the reorganization of qubits within each block, altering the qubit arrangement and making the original information indecipherable to unauthorized listeners.

Once the encryption process is complete, the modified qubit blocks and the permutation basis are securely delivered to the receiver. The recipient decrypts the message by utilizing the

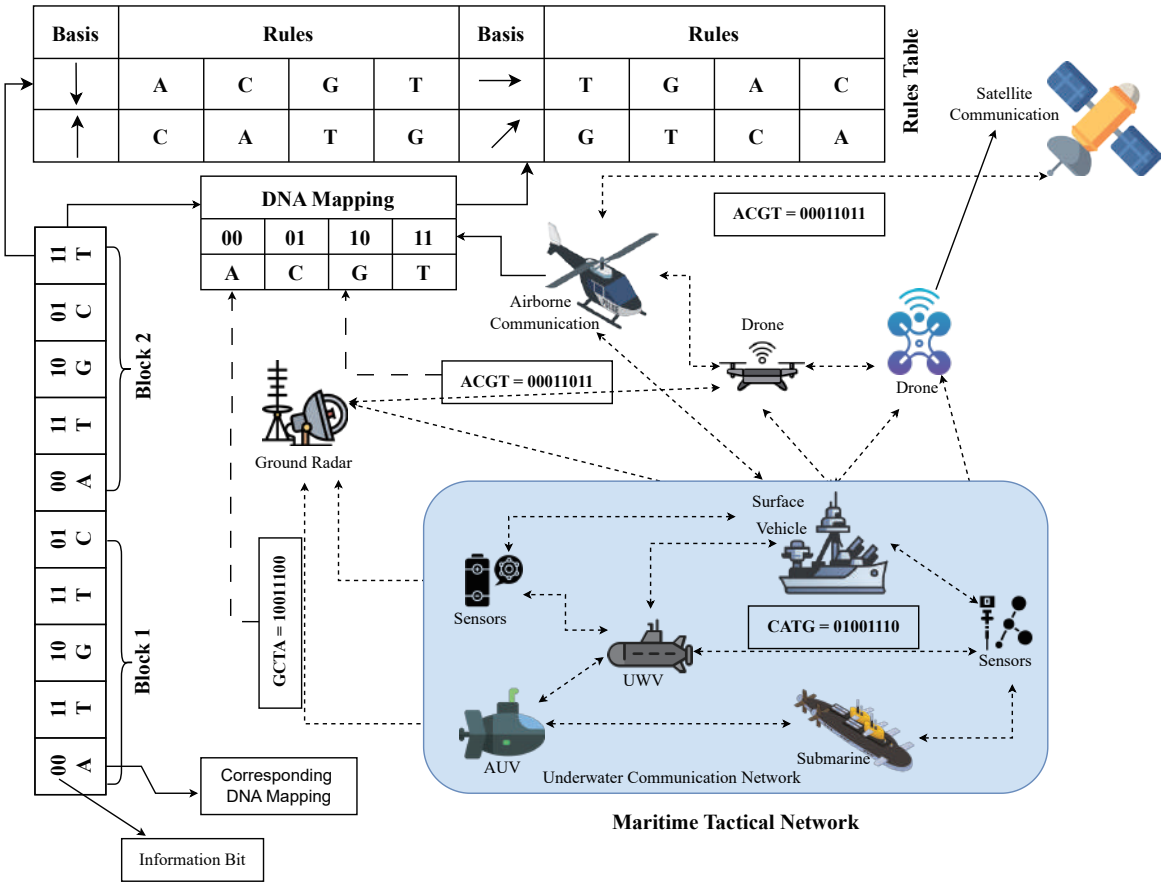


Fig. 1: An illustration of QDPP-based secure maritime communication

pre-established Rules table and applying the inverse permutation operation corresponding to the specific basis (i.e., the inverse of the "ACGT" rule for the ↓ basis). This procedure reverses the rearrangements during encryption, thus undoing the encryption and recovering the original DNA sequence. Subsequently, the recipient utilizes the DNA Mapping table to convert the obtained DNA sequence back into its conventional binary representation, retrieving the initial message. The sender and the receiver can easily change quantum data using the shared information in the Rules and DNA Mapping tables. This makes sure that communication is safe through the QDPP mechanism. Algorith 1 shows the QDPP-based secure communication protocol.

Algorithm 1 QDPP-based Secure Communication Protocol

Require: Shared DNA Mapping and Rules tables.
Ensure: Secure transmission of information.

1: Map each bit pair to a nucleotide using the DNA Mapping table.
2: Convert information to a bit string N.
3: Divide N into κ blocks, where $\kappa = N\%i$ and i is even.
4: Convert blocks to qubits using Pauli operators.
5: Encrypt qubits using permutation and the Rules table.
6: Send an encrypted message and a basis to the receiver.
7: Receiver decrypts using inverse permutation and Rules table.

III. Result Analysis

We utilized the IBM quantum computation framework, Qiskit, to simulate our work in Google Colaboratory using Python 3.9.7. As depicted in Figure 2(a), on the sender's side, the information bit is first converted into the corresponding nucleotide by mapping every two bits into one nucleotide, according to a predefined mapping table. These nucleotides are then grouped into blocks, with each block consisting of four nucleotides. Subsequently, a basis, which serves as the key for the permutation, is applied, and the blocks are processed through a quantum circuit created using Pauli gates (X, Y, and Z gates), as shown in Fig. 2(b) and Fig. 2(c). On the receiver's side, who already possesses the mapping table, the

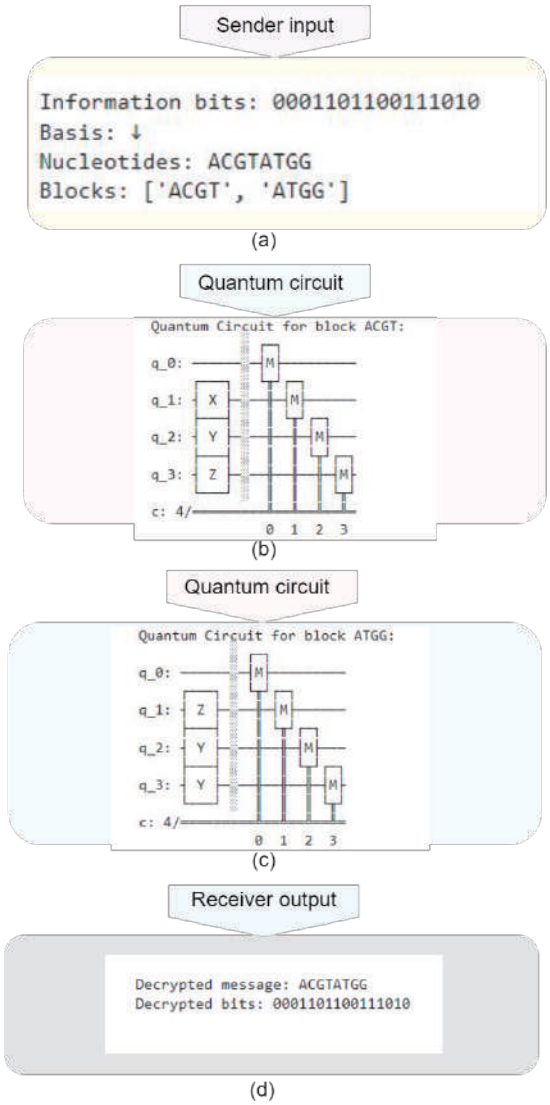


Fig. 2: Simulation results of encryption procedure in qiskit qasm simulator

information is decrypted using the basis information received from the message shows in Fig. 2(d). The encryption code can be found in github [12].

A. Security Analysis

The proposed method undergoes several levels of encryption. Initially, information is converted into corresponding nucleotides, with each nucleotide representing a pair of bits; however, this mapping can vary depending on the specific system requirements. Subsequently, these nucleotides are grouped into blocks. Our system considers each block to consist of four nucleotides, although this configuration can be adjusted. The blocks are then permuted using values derived from the rules that correspond to a specific basis. This multilayer encryption ensures that even if an attacker compromises one block, the information gained cannot be used to decipher other blocks. The continuous multi-layer encryption makes the information nearly impossible for an attacker to accurately guess or predict the correct combinations for the encrypted blocks.

IV. Conclusion

Maritime Tactical Networks (MTNs) are among the least studied network types. However, their financial and political significance is comparable to that of other networks. To enhance the security of communication, within MTNs, this paper introduces a quantum-based, multi-layer, bio-inspired encryption scheme known as Quantum DNA Partial Permutation (QDPP). By incorporating DNA coding into cryptographic processes, QDPP successfully encrypts messages and provides information-theoretic security against both classical and quantum attacks.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 50%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 50%)

References

[1] J. L. Burbank, P. F. Chimento, B. K. Haberman, and W. T. Kasch, "Key challenges of military tactical networking and the elusive promise of manet technology," *IEEE Communications Magazine*, vol. 44, no. 11, pp. 39–45, 2006.

[2] M. R. Subhan, M. F. Rahaman, M. Golam, D.-S. Kim, and T. Jun, "Elevating transparency in global maritime logistics through blockchain technology," , pp. 326–327, 2024.

[3] World Economic Forum, "Consequences of digital fragmentation," in *The Global Risks Report 2020*. World Economic Forum, 2020, pp. 63–63.

[4] R. L. Lendvay, "Shadows of stuxnet: Recommendations for us policy on critical infrastructure cyber defense derived from the stuxnet attack," Ph.D. dissertation, Monterey, California: Naval Postgraduate School, 2016.

[5] E. A. Tuli, M. Golam, J.-M. Lee, and D.-S. Kim, "Quantum superdense coding-based secure authentication for military metaverse."

[6] S. Subramani and S. K. Svn, "Review of security methods based on classical cryptography and quantum cryptography," *Cybernetics and Systems*, pp. 1–19, 2023.

[7] E. A. Tuli, J.-M. Lee, and D.-S. Kim, "Integration of quantum technologies into metaverse: Applications, potentials, and challenges," *IEEE Access*, vol. 12, pp. 29 995–30 019, 2024.

[8] D. Heuer, "On partial permutations and alternating sign matrices: Bijections and polytopes," Ph.D. dissertation, North Dakota State University, 2021.

[9] R. Kuang and M. Barbeau, "Quantum permutation pad for universal quantum-safe cryptography," *Quantum Information Processing*, vol. 21, no. 6, p. 211, 2022.

[10] J. Gao and T. Xie, "Dna computing in cryptography," in *Advances in Computers*. Elsevier, 2023, vol. 129, pp. 83–128.

[11] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *science*, vol. 266, no. 5187, pp. 1021–1024, 1994.

[12] E. A. Tuli, "Icmic-2024 quantum dna encoding," https://github.com/tulimoni/ICMIC-2024-Quantum-DNA-Encoding-, accessed: 2024-05-29.

Study on development of security requirements for CMVP security level1

Sun-Woo Lee[1,2]
[1]KEPCO KDN
[2]Chonnam University
Republic of Korea
hagler_sunwoo1@kdn.com

Tae-Hun Kim[1]
[1]KEPCO KDN
Republic of Korea
thkim_5117@kdn.com

Jun-Young Lee[1]
[1]KEPCO KDN
Republic of Koera
ljy.953386@kdn.com

Ieck-Chae Euom*[2]
[2]Chonnam University
Republic of Korea
iceuom@jnu.ac.kr

Abstract—**This study focuses on the application and evaluation of anti-tampering techniques within the Cryptographic Module Validation Program(CMVP) for software-based cryptographic modules. It recognizes the need for enhanced security measures against tampering attempts and proposes a new methodology to integrate anti-tampering requirements into cryptographic module validation. To this end, experiments were conducted on Windows, Linux and Embedded Linux platforms to select appropriate anti-tampering techniques for cryptographic modules and evaluate them from the perspective of CMVP security requirements. This contributes to strengthening the security of cryptographic modules and is expected to help establish new standards for enhancing the reliability of information security systems.**

Keywords—CMVP, Security Requirements, FIPS 140-2. FIPS 140-3, Finite State Model, State Transition Diagram

I. INTRODUCTION

Encryption technology plays a pivotal role in ensuring the confidentiality and integrity of information. One of the core components of this technology is the cryptographic module. Cryptographic modules perform encryption, decryption, authentication and integrity tasks in various information security systems. Therefore, the security of cryptographic modules directly impacts the reliability of the entire information security system.

Recognizing the importance, the national Institute of Standards and Technology(NIST) in the United States operates the Cryptographic Module Validation Program(CMVP), which defines and verifies the security requirements of cryptographic modules. To enhance and maintain the security of cryptographic modules, protective measures against unauthorized tampering are necessary. However, there are currently no security requirements for such protections.

This study presents a novel methodology for the integration and assessment of tamper-protection technologies in software-based cryptographic modules, aligned with the CMVP standards. It proposes an innovative framework to bolster the security of cryptographic modules, which, in turn, is expected to enhance the robustness of the overarching information security infrastructure. The core objective of this research is devise and refine criteria for the implementation and appraisal of tamper-resistant measures, thereby augmenting the dependability of information security systems. Specifically, his endeavor aims to facilitate the update of level 1 CMVP security requirements by elaboration on the development and evaluation methodologies for anti-tampering strategies within software cryptographic modules. The intent is fortify the security specifications for cryptographic modules, thereby elevating the integrity and reliability of the entire information security ecosystem.

II. BACKGROUND

A. Overview of the CMVP

The Cryptographic Module Validation Program was established by the National Institute of Standards and Technology(NIST) in the United States in 1994, defining the security requirements for cryptographic modules as outlined in the FIPS 140-1(Security Requirements for Cryptographic Modules) standard and its revisions, FIPS 140-2 and FIPS 140-3, issued in 2001. The program also defines the testing requirements for cryptographic modules as specified in the Derived Test Requirements(DTR) documents

B. The FIPS Series

The FIPS series serves as a critical benchmark for evaluating the security of cryptographic modules, presenting security requirements related to physical security, including tamper prevention and resistance. Those standards specify the level of protection that cryptographic modules must provide against physical attacks. The requirements are summarized in TABLE I.

TABLE I. FIPS's Tampering-related requirements

No	Category	Security Level
1	Operator Applied Security Mechanisms	Level 2 or higher
2	Tamper Evidence and Resistance Testing	Level 3 or higher

3	Requirements by Physical Security Level	Level 3 or higher
4	Hard Coating and Potting	Level 3 or higher

Security level 2 provides basic physical security, while Level3 and 4 require more enhanced security measures. For example, at Levels 3 and 4, cryptographic modules must be covered with a hard and opaque tamper-resistant coating or potting material. These security requirements ensure the maintenance of the physical security level of cryptographic modules but are not suitable for software-based cryptographic modules.

C. CMVP Security Requirements

In the CMVP, measures indirectly related to tampering protection include key zeroization, secure coding, obfuscation and integrity check. Key zeroization involves immediately removing and zeroing out variables such as encryption keys after their use. Secure coding prevents vulnerabilities such as buffer overflows at the programming stage. Obfuscation makes the source code difficult to analyze by obscuring it. Integrity verification is a techniques for ensuring that the code or data of a cryptographic module has not been altered. However, these measures are vulnerable to dynamic analysis through reverse engineering tools as shown in Figure 1.

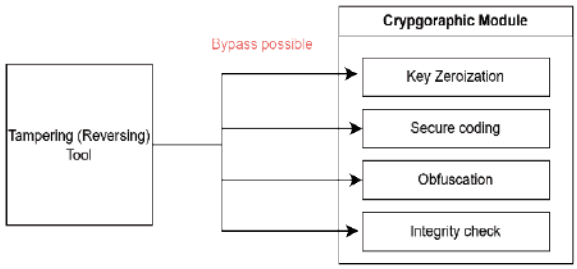


Figure 1. The relationship between CMVP requirements and tampering

III. ANTI-TAMPERING METHDOLOGIES

A. Development of Anti-Tampering Methodologies

Tampering, in the context of information security, carries the dictionary meaning of unauthorized alteration. It refers to the act of hackers illicitly modifying a program through reverse engineering, which is a form of attack aimed at bypassing or neutralizing the security mechanisms of software.

Tampering can be categorized into manual analysis and automatic analysis. Anti-tampering refers to the protective mechanisms designed to prevent unauthorized modification or reverse engineering of software, and it is broadly divided into thwarting manual analysis and thwarting automatic analysis. This study focuses on the former, specifically on code-level protection techniques that can be applied directly to cryptographic modules to obstruct manual analysis as shown in Figure 2.

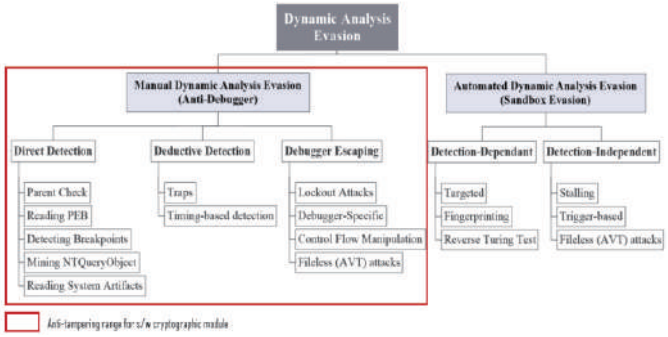


Figure 2. The Anti-tampering scope addressed in this study

Anti-tampering techniques are required to ensure security and stability. To achieve this, tamper-proofing techniques that have been recognized for their security among programs are applied. Additionally, considering multi-platform compatibility, only the techniques that have been validated for stability through experimentation are applied as shown in TABLE IIIII.

For Windows systems, the techniques to be applied are Mining NTQuery Object, Traps, Timing-based detection and Debugger Specific. The Mining NTQuery Object technique uilizes the NTQueryInformation system call to inquire about the debugger port information of the current process. By using the ProcessDebugPort information, it determines whether a debugger is connected and return this information to detect tampering attempts. The Traps technique uses INT3, INT0x2D, INT 0x41 and INT 1 to detect software interrupts. These interrupts cause exceptions when a debugger is active, allowing for the detection of tampering attemps. However, this is not applicable in a 64-bit environment for INT 2D and INT 0x41. Timing-Based detection involves the use of the RDTSC instruction to measure CPU cycle counts, detecting execution delays that typically accompany tampering attemps. It is important to note that this technique may not function correctly in environments beyond Windows Vista.

In Linux systems, the application of Mining NTQuery Object, Traps and Timing-Based detection is feasible. However, given that Linux lacks the Ntquery object, a similar effect is achieved by reading the /proc/self/status file and monitoring the TracePid value to detect tampering attempts. The Traps techniques is only applicable to INT 3. This method utilizes the SIGTRAP signal and the INT 3 command to detect tampering attempts.

TABLE III. results of Applying anti-tampering techniques

Tactic	Techniques	Windows		Linux	
		32	64	32	64
Mining NTQuery Object	ProcessDebugFlags	O	O	△	△
Traps	Interrupt 0x2D	O	×	×	×
	Interrupt 3	O	O	O	O

	Interrupt 0x41	O	×	×	×
Timing-Based detection	timeGetTime(), GetTickCount()	O	O	O	O
Debugger Specific	InputDebugString()	×	×	×	×
	SoftICE Interrupt 1	O	O	×	×

Anti-tampering features must be invoked during the self-tests prior to operation, as stipulated by the CMVP. In the event of a tampering attempt, the cryptographic module must transition to a critical error state. Furthermore, the cryptographic module is required to zeroize all critical security parameters and cease their use as shown in Figure 3.

```
uint32 KCMVP_Start()
{
...

#ifdef KCMVP_WIN
    if (DetectDebugger()) {

        KCMVP_mod_stat = KCMVP_MODULE_FATAL_ERROR;
        ret = KCMVP_MODULE_TAMPERED;
        printf("Start failed: %s\n", KCMVP_GetErrorMsg(ret));

        goto err;
    }

#endif
...

}
```

Figure 3. Methods of applying anti-tampering techniques in pre-operational self-test

Certainly, it may not be feasible to apply the aforementioned high-level anti-tampering techniques in all operational environments. In such cases, timing-based detection techniques, which are universally applicable, are employed exclusively.

B. Development of Anti-Tampering Test Methodology

Anti-tampering measures at the time of execution for cryptographic modules complement the existing security requirements of integrity verification and key zeroization by addressing their deficiencies. As shown in Figure 4, current security requirements for cryptographic modules cannot prevent variable monitoring due to tampering until key zeroization is performed. Furthermore, integrity checks, which do not detect tampering at the time of execution but rather the alteration of cryptographic module binary data, cannot prevent unauthorized access or parameter value modification through reverse engineering. However, the proposed application of anti-tampering measures can prevent tampering attempts using reverse engineering tools, thereby addressing the vulnerabilities in the existing security requirements.

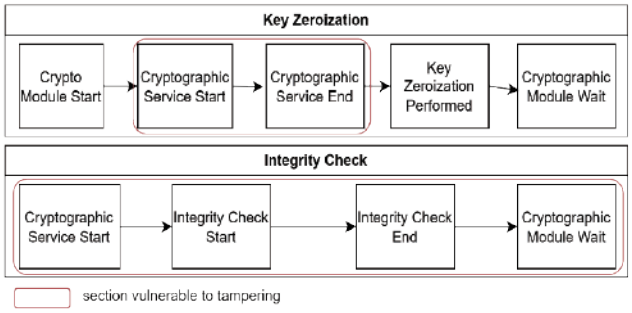


Figure 4. Tampering vulnerable sections in Cryptographic module

When implementing anti-tampering measures, it is imperative to maintain the finite state model, a pre-existing security requirement of the CMVP. This ensures that the cryptographic module transitions between specific states in a predictable and secure manner.

The testing for anti-tampering should verify whether the cryptographic module has maintained its finite state model. As shown in Figure 5, The anti-tampering features must necessarily be operational during the self-test processes that occur before the cryptographic module provides any cryptographic services. Thus, if a tampering attempt is detected during the self-test process, it must be verified that all subsequent operations of the cryptographic module are halted and that the module transitions into a critical error state. Moreover, the anti-tampering features should be accessible for invocation by the user at any time, as needed. This is to allow users the ability to verify the security state of the cryptographic module under specific circumstances. Testers must verify that these features can also be invoked following the pre-operational self-test.

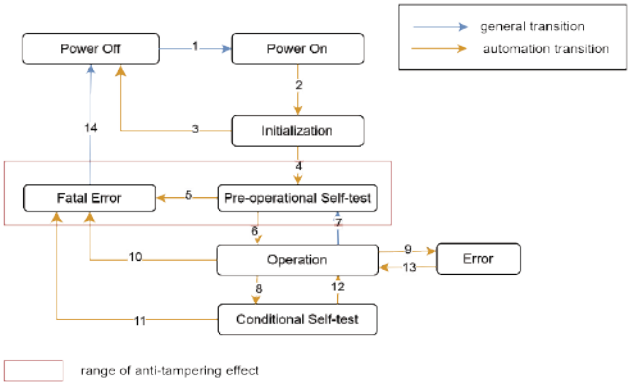


Figure 5. The scope of anti-tampering influence on the finite state transition

Tests to ensure the reliability of anti-tampering mechanisms must also be conducted. This is to ascertain whether the cryptographic module can guarantee normal operation across all the operational environments it intends to support.

IV. IMPLEMENTATION AND EVALUATION

This involves analyzing the results of the cryptographic services provided by the module in target operational environments. Both with and without the use of reverse

engineering tools. In cases of normal operation, where no tampering attempts are present, it must be confirmed that the success test data is output as expected, as shown in Figure 6.

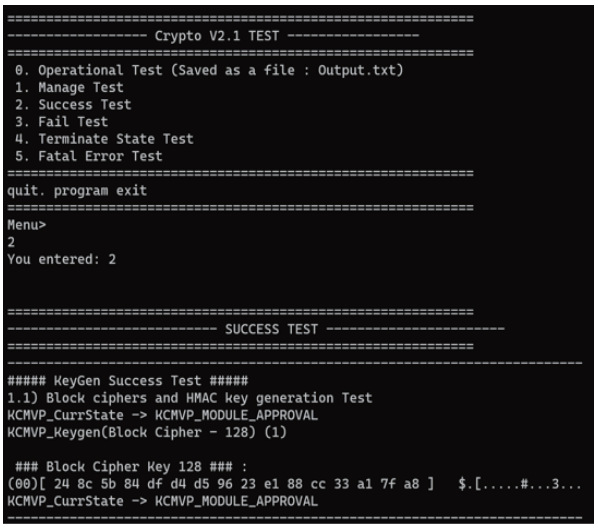


Figure 6. Output in the absence of a tampering attempt

However, in the event of a tampering attempt, it must be verified that the cryptographic module transitions into a critical error state, as shown in Figure 7.

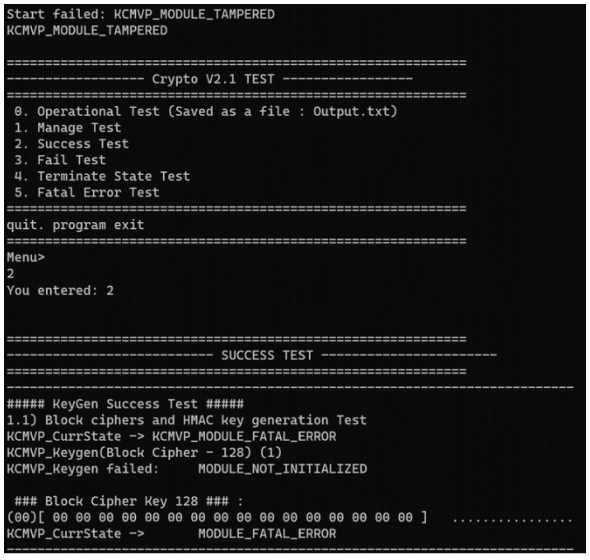


Figure 7 Result in the event of a tampering attempt

At this time, protections against tampering through reverse engineering must be implemented across all operational environments intended to be supported by the cryptographic module. Evidence through unit testing for all applied tampering protection measures must be provided, and for those operational environments where application was unavoidably not possible, details must be documented in the certification documents.

V. CONCLUSION

This research has established a distinctive methodology for the development and evaluation of anti-tampering techniques specifically for software-based CMVP cryptographic modules. This novel approach not only differentiates our work within the domain of cryptographic security but also underscores the urgent need for revising the security requirements of level 1 within the CMVP. The findings emphatically suggest that the integration of this methodology can play a crucial role in the upcoming revisions, providing a solid framework that could be readily adopted to enhance the security standards of cryptographic modules. By advocating for the incorporation of theses anti-tampering measures, this study reinforces the call for comprehensive update of CMVP level 1 security requirements, highlighting the practicality and necessity of applying our developed methodology to ensure the fortified security of cryptographic modules in an ever-evolving digital landscape.

REFERENCES

[1] FIPS 140-2, Security Requirement for Cryptographic Modules, NIST, 2002.12

[2] FIPS 140-3, Security Requirement for Cryptographic Modules, NIST,

[3] Dennis Distler and Charles Hornat. Malware analysis: An introduction.SANS Institute InfoSec Reading Room, pages 18{19, 2007.

[4] Afianian, A., Niksefat, S., Sadeghiyan, B., Baptiste, D. Malware Dynamic Analysis Evasion Techniques: A Survey. arXiv preprint arXiv:1811.01190 (2018).

[5] J. Baines, “Programming Linux Anti-Reversing Techniques” Dec. 2016.

[6] apriorit. Anti debugging protection techniques, 2016.

[7] P. Ferrie. The ultimate anti-debugging reference, 2011.

[8] Joshua Tully. An anti-reverse engineering guide, 9 Nov 2008.

[9] Xu Chen, Jon Andersen, Z. Morley Mao, Michael Bailey, Jose Nazario. Towards an understanding of antivirtualization and anti-debugging behavior in modern malware. In IEEE International Conference on Dependable Systems and Networks With FTCS and DCC. IEEE, 2008, 177-186.

[10] Nguyen Anh Quynh, Kuniyasu Suzaki. Virt-ice: Next-generation debugger for malware analysis. Black Hat USA (2010).

[11] Michael Chourdakis. Toggle Hardware Data/Read/Execute Breakpoints Programmatically. Received October 2020 from https://www.codeproject.com/Articles/28071/Toggle□hardware-data-read-execute-breakpoints-prog.

[12] Tyler Shields. Anti-debugging—A developers view. Veracode Inc., USA(2010).

[13] Kulchytskyi Oleg. Anti-Debug Protection Techniques: Implementation and Neutralization. Retrieved October 2020 from https://www.codeproject.com/Articles/1090943/Anti-Debug□Protection-Techniques-Implementation-an. https://anti□reversing.com/Downloads/Anti-Reversing/The_Ultimate_Anti-Reversing_Reference.pdf.

[14] Rodrigo Rubira Branco, Gabriel Negreira Barbosa, Pedro Drimel Neto. Scientific but not academical overview of malware anti-debugging, anti-disassembly and anti-vm technologies. Black Hat(2012).

[15] XPN.Windows Anti-Debug Techniques – OpenProcess Filtering. Retrieved October 2020 from https://blog.xpnsec.com/anti-debug-openprocess/.

[16] Afianian, A., Niksefat, S., Sadeghiyan, B., Baptiste, D. Malware Dynamic Analysis Evasion Techniques: A Survey. arXiv preprint arXiv:1811.01190 (2018).

[17] Microsoft msdn. CheckRemoteDebuggerPresent function. Retrieved October 2020 from https://docs.microsoft.com/en-us/windows/win32/api/debugapi/nf-debugapi□checkremotedebuggerpre

[18] J. Baines, “Programming Linux Anti-Reversing Techniques” Dec. 2016.

Exploring network security with MTD in Mininet

Gilhan Choi
Kongju National University
Cheonan 31080, Korea
choikilhan12@smail.kongju.ac.kr

Yonggang Kim*
Kongju National University
Cheonan 31080, Korea
ygkim@kongju.ac.kr

***Abstract*—This paper explores MTD in virtual network environments using Mininet. It covers the theoretical foundations of MTD, practical implementation in Mininet, and details on constructing virtual networks with IP hopping. Python code snippets are included for clarity and replication. In particular, the utilization of the virtualization platform Mininet highlights its crucial role in demonstrating the feasibility of applying dynamic defense strategies in virtual environments.**

I. Introduction

In recent years, several MTD strategies have been published which can be grouped into three main categories: (i) network-level MTD, (ii) host-level MTD, and (iii) application-level MTD [1]. This paper investigates the implementation of IP hopping, a network-level MTD technique, within virtual network environments utilizing Mininet. Virtual networks offer a flexible and scalable platform for testing and implementing innovative security strategies like MTD without impacting production systems. By leveraging Mininet's capabilities, researchers and practitioners can emulate complex network topologies and experiment with various defense mechanisms in a controlled environment.

The significance of this study lies in its contribution to advancing network security paradigms. By demonstrating how MTD can be effectively implemented within virtual networks, this research sheds light on the practical implications of dynamic defense strategies. Furthermore, the utilization of Mininet as a tool for constructing virtual networks underscores the importance of simulation-based approaches in evaluating and refining security solutions.

In the subsequent sections, we will delve into the theoretical foundations of IP Hopping, elucidate the process of setting up virtual networks using Mininet, and showcase the implementation of IP hopping via Python code snippets.

II. IP Hopping-Based Network-Level MTD Model

IP Hopping scheme provides a transformation between the original IP address (used for routing and other networking decisions) and the IP address used for MTD randomization. The process of converting from the original to the randomized IP address is called IP hopping, and the reverse process is called IP de-hopping. Both the source IP address and the destination IP address go through IP hopping and de-hopping processes [2]. This technology modifies the network state to increase the difficulty of attackers performing their attacks, demonstrating enhanced security through the table below.

TABLE I
Average Response Time for a Valid Client to Successfully Obtain a Service Provided by the Server Under Different States and Attack Rates (in Seconds) [3]

State	Attack Rate		
	0 times/s	100 times/s	1000 times/s
Normal State	0.025112	0.165025	0.639812
Hopping State	0.031997	0.049540	0.063001

An MTD system can be identified by the target system, goals, and MTD policies. If we represent the MTD system as Σ, then the $\Sigma = (\Gamma, G, P)$ where Γ, G, and P are the configurable system, a set of operational and security goals, and MTD policies [2]. The configurable system Γ can be defined as a tuple of $\Gamma = (S, \Lambda, \tau)$. The set of feasible configuration states of MTD and the set of possible actions is denoted as $S = \{s_1, s_2, \cdots, s_n\}$ and $\Lambda = \{\alpha_1, \alpha_2, \cdots, \alpha_n\}$, respectively. $\tau : S \times \Lambda \rightarrow S$ represents the state transition function. In this MTD model, if we denote π as a configuration parameter from the possible set of parameter type Π and the z is the unique assignment of value Z, then the configuration state $s_i \in S$ can be represented as $s = \pi \leftarrow z$ as follows [4]:

$$\overbrace{\underbrace{\text{IP address}}_{\text{configuration parameter}} = \underbrace{192.168.0.54}_{z}}^{\text{configuration state } s \text{ where } s=\pi\leftarrow z} \tag{1}$$

In this paper, we implement a Mininet environment for exploring the network security based on (1).

III. Mininet Environment Setup for MTD

1) *Downloading Oracle VM VirtualBox:* Mininet runs on Linux, so a virtual machine is needed to run Linux on a Windows computer. Oracle VM VirtualBox was downloaded for this purpose.
2) *Downloading Mininet VM image:* An image file is required to run Mininet. The Mininet image file was downloaded.
3) *Setting up Mininet Environment on VirtualBox:* The Mininet image (.ovf) was imported into VirtualBox via the "File" menu by selecting "Import Virtual System." Since the researcher uses Windows, two VMs were created.
4) *Configuring Mininet Network:* One VM was configured for networking. In the VM settings under the "Network" tab, "Adapter 2" was selected, "Enable Network

Adapter" was checked, and "Attached to" was set to "Host-only Adapter."

5) *Running Mininet:* After starting each VM, log in with the username and password "mininet." In the VM with network configuration, activate the host-only adapter using "sudo dhclient eth1" and check the IP address with "ifconfig." From another VM, use "ssh -X mininet@(eth1 IP address)" to complete the Mininet environment setup.

IV. Demonstration of IP Hopping with Python Code in Mininet Environment

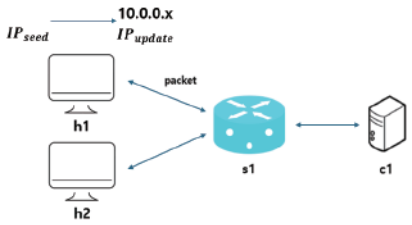


Fig. 1. Minimal Topology

The researcher aims to construct a virtual network with a minimal topology structure in the Mininet environment. Subsequently, they intend to demonstrate the IP hopping technique using Python code.

```
class Mtd Topo(Topo):
    def build(self):
        h1 = self.addHost('h1')
        h2 = self.addHost('h2')
        s1 = self.addSwitch('s1')
        self.addLink (h1, s1)
        self.addLink (h2, s1)
def change IP (host, newIP):
    host.setIP (newIP)
    info("changed IP address to {}\n".format(newIP))
def updateSwitchARP (switch, host, newIP):
    intf = host.defaultIntf()
    mac = intf.mac
    quietRun('arp -d {}-{}'.format(host.name, intf))
    quietRun('arp -s {} {} {}'.format(newIP, mac,
    host.name))
def pingTest (h1, h2, count=3):
    output = h2.cmd('ping -c {} {}'.format(count, h1
    .IP()))
    return '0% packet loss' in output
def main():
    topo = MtdTopo()
    net = Mininet (topo=topo, host=Host)
    net.start()
    h1 = net.get('h1')
    s1 = net.get('s1')
    h2 = net.get('h2')
    while True:
        new_ip_address = '10.0.0.{}'.format(random.
    randint(1, 254))
        if h2. IP() == new_ip_address:
            info('h1 and h2 IP are same! so change
    again\n')
            continue
        updateSwitchARP (s1, h1, new_ip_address)
        change IP(h1, new_ip_address)
        if pingTest (h1, h2):
            info("Host {} can communicate with host
    {}\n".format(h2.name, h1.name))
        else:
            info("Host {} cannot communicate with
    host {\n".format(h1. name, h2.name))
        time.sleep(30)
    net.stop()
if _name__ == '__main__':
    setLogLevel('info')
    main()
```

Listing 1. Demonstrating Simple IP Hopping in Minimal Topology using Python

1) *MtdTopo:* Inherits from Mininet's Topo class to create a custom topology with two hosts (h1, h2) and one switch (s1), each host connected to the switch.
2) *changeIP:* Changes the IP address of a given host to a new IP address.
3) *updateSwitchARP:* Updates the ARP cache of a switch for a given host and new IP address.
4) *pingTest:* Tests the network connection between two hosts by executing the ping command and checking for packet loss.
5) *main:* Runs an infinite loop where a host's IP address is changed to a randomly generated one, then tests the network connection and logs the results. The main function is called at the end, and the log level is set to run the program.

V. Conclusion

The implementation of dynamic defense strategies within virtual network environments plays a crucial role in enhancing network security. Utilizing virtualization platforms like Mininet to validate the effectiveness of dynamic defense strategies and demonstrate their applicability in real-world environments holds significant value. Network-level MTD techniques such as IP hopping prove to be effective in rendering attackers' efforts more challenging, thus strengthening overall security measures. Such research endeavors contribute to exploring innovative approaches in the field of network security and fostering safer and more resilient digital ecosystems. Future research efforts should focus on the development of more effective MTD strategies and explore their applicability in increasingly complex network environments.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2022-00166739).

References

[1] Morphisec Ltd, "Moving target defense: Common practices. website." Website, 2016. Available at: http://blog.morphisec.com/moving-target-defense-common-practices.

[2] S.-Y. Chang, Y. Park, and B. B. Ashok Babu, "Fast IP hopping randomization to secure hop-by-hop access in SDN," *IEEE Transactions on Network and Service Management*, vol. 16, no. 1, pp. 308–320, 2019.

[3] Y. Fan, G. Wu, K.-C. Li, and A. Castiglione, "Robust end hopping for secure satellite communication in moving target defense," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 16908–16916, 2022.

[4] J. Steinberger, B. Kuhnert, C. Dietz, L. Ball, A. Sperotto, H. Baier, A. Pras, and G. Dreo, "DDoS defense using MTD and SDN," in *IEEE/IFIP Network Operations and Management Symposium (NOMS)*, pp. 1–9, 2018.

3D Positioning Algorithm based on Decision Tree Classification Model in Indoor UWB Environment

Ho Chul Lee
Chief Executive Officer
Floware
Busan, Republic of Korea
calmtot@gmail.com

Dong Myung Lee
Prof., Dept. of Computer Engineering
Tongmyong University
Busan, Republic of Korea
dmlee@tu.ac.kr

Hee Min Park
Chief Executive Officer
Sky-Tech Co.
Daejeon, Republic of Korea
heemin35@nate.com

Gyoungbae Kim
Prof., Dept. of Software
Seowon University
Cheongju, Republic of Korea
gbkim@seowon.ac.kr

Abstract— In this paper, we propose a novel 3-dimensional (3D) positioning algorithm based on a decision tree classification model using heuristic depth searching and the median absolute deviation (MAD) cost function. Our algorithm enables accurate 3D positioning of tags, even in challenging blind spots commonly encountered in indoor ultra-wideband (UWB) environments, thereby enhancing the reliability of indoor positioning systems. To train our algorithm, we leverage a heuristic depth search method designed to handle limited distance measurement data, particularly in scenarios where trilateration is impractical. Additionally, we optimize the decision tree classification model using the MAD cost function to refine tag positioning, particularly in complex indoor settings.

Keywords- Ultra-Wideband, Decision Tree, Machine Learning, 3-Dimension, Indoor Positioning

I. Introduction

With advancements in sensors and wireless communication technology, the demand for indoor positioning application is steadily rising. However, an indoor positioning application faces challenges due to the susceptibility of sensors to surrounding radio wave interference, leading to errors in measurement data even when using trilateration. Interruptions in signal reception caused by obstacles can result in temporary tag location loss due to trilateration errors, creating blind spots in the user's system. Eliminating these errors is crucial for improving the accuracy of indoor positioning system.

In this paper, we propose a 3-dimensional (3D) positioning algorithm based on the decision tree classification model using heuristic depth searching and the median absolute deviation (MAD) cost function in order to improve the above problem. Our algorithm is designed to facilitate 3D positioning of tags even in blind states, which are common challenges in indoor UWB environment. This approach aims to enhance the reliability and accuracy of 3D indoor positioning algorithm.

II. Related Studies

UWB is a short-range wireless communication technology that utilizes high-frequency radio waves with precise spatial recognition and directional characteristics. It enables intelligent connectivity among various devices and finds applications ranging from remote payments to remote control location detection. The decision tree is a popular method in supervised machine learning for classification and regression tasks. It constructs a model by learning simple decision rules from collected data characteristics to predict the value of a target variable. The decision tree classification model organizes decision rules into a tree structure, facilitating classification and prediction tasks.

III. System Design

A. Design Philosophies

The research methodology employed in this paper is outlined as follows: Firstly, a comprehensive review of existing literature on 3D positioning algorithm based on decision tree classification model in indoor UWB environment was conducted. Next, a positioning model was developed by integrating UWB sensor data with a decision tree classification model.

The key aspect of model development involved selecting relevant feature points from UWB sensor data and training the decision tree model to predict the position of the 3D target object. Subsequently, the developed positioning algorithm underwent rigorous testing for accuracy and reliability using simulation data. Finally, the performance of the proposed positioning algorithm was evaluated in a real-world environment, comparing its accuracy and reliability against existing techniques.

B. System Architecture

The configuration diagram of proposed 3D positioning algorithm is depicted in Fig. 1. The system utilizes four UWB anchors, with location estimation tags measuring distances via signal exchange with these anchors. Measurements are taken at 100ms intervals, and the collected data is transmitted to the user's mobile phone. From the mobile phone, data is further transmitted to a cloud server via the network, where it is stored and utilized for creating a prediction model to facilitate location estimation. The model's accuracy can be enhanced through continuous learning processes [1, 5-6].

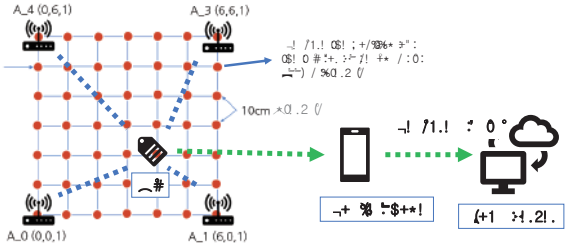


Fig. 1. Schematic diagram of proposed 3D positioning algorithm.

The tag measures the distance to each of the four anchors (labeled as anchors 1, 2, 3, and 4) and transmits this data to the

sink node. The use of four anchors allows the system to calculate the tag's position using data from any three anchors, even if one anchor becomes deactivated. The sink node collects all distance data between the anchors and the tag, transmitting this information to the monitoring server. The monitoring server then displays the tag's location in 3D coordinates [1].

C. Decision Tree Classification Model

The decision tree classification model is constructed by first completing the dataset using trilateration and then learning the measured data, which is subsequently divided into features (images) and labels, as illustrated in Fig. 2. The effectiveness of the generated model is evaluated for prediction accuracy using a dedicated test dataset. To enhance the accuracy of the decision tree classification model, we define a prediction model that combines the MAD and a cost function.

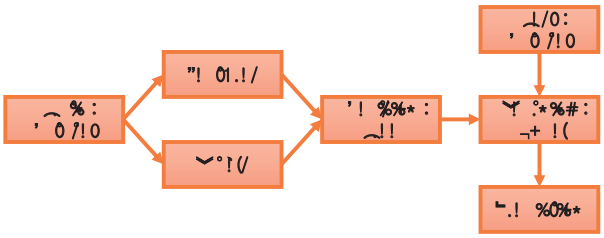


Fig. 2. A Process of decision tree classification model.

The primary cost function used in this model is the Gini Index described in Eq. (1). To enhance prediction accuracy, an improved prediction model was proposed by combining the Gini cost function (Eq. (1)) with Information Gain (IG), a feature evaluation function defined in Eq. (2). This combination aims to optimize the decision tree classification model for improved performance.

$$\sum_{k\neq k'} \hat{p}_{mk}\,\hat{p}_{mk'} = \sum_{k=1}^{k} \hat{p}_{mk}\,(1-\hat{p}_{mk}) \qquad (1)$$

From the four measured distance data points (Dist0, Dist1, Dist2, Dist3) obtained from the tag, the most influential distance data is identified using IG) calculated with Eq. (2) [2]. This process helps prioritize the relevant data for optimizing the decision tree classification model.

$$\mathrm{IG}(S,A) = Entropy(S) - \sum_{v\in value(A)} \frac{S_v}{S}\, Entropy(S_v) \qquad (2)$$

The proposed system utilizes a decision tree classification model that combines the Gini cost function and the median absolute deviation system [3]. The accuracy of this model can vary with tree depth, sometimes decreasing due to overfitting as the depth increases. To enhance accuracy, a heuristic search method is employed to determine the optimal tree depth.

Specifically, the model is iteratively trained with increasing tree depths. If a deeper tree results in lower prediction accuracy compared to the previous model, retraining occurs with a deeper tree depth. This iterative process continues until a decision tree classification model with optimal prediction accuracy is achieved [4].

IV. Experiments And Results

An experimental environment was established to evaluate the performance of the developed algorithm. The experiments were conducted indoors within a building with various structures. Four UWB anchor nodes were strategically placed at the vertices of a 10m × 10m ceiling. Tag nodes for positioning were positioned inside a rectangular area defined by the anchor nodes' vertices. The experimental setup utilized Qorvo DWM1001-DEV equipment for data collection and testing.

To collect data from the UWB equipment, embedded C code was developed and used to implement trilateration software using open-source tools. This software measures the distance values from each anchor and computes the 3D coordinates of the tag through trilateration. The measured data, along with the tag coordinates, are transmitted to a connected cloud server. The server stores the transmitted data on disk to facilitate the creation of a prediction model.

The tag coordinates were recorded in chronological order within 3D space. The coordinate estimation error of the UWB sensor in line-of-sight (LOS) conditions was approximately 10 cm, which was considered negligible for the experiment. During the experiment, the user carried the tag and followed a predefined movement path for about 30 minutes to collect coordinate data. To assess the performance of the proposed prediction model in 3D space, coordinate data was collected by moving the tag nodes to positions 0, 1, 2, and 3 meters along the z-axis at 30-second intervals.

V. Conclusion

In this paper, we propose a novel 3-dimensional (3D) positioning algorithm based on a decision tree classification model using heuristic depth searching and the MAD cost function. To evaluate the accuracy of the proposed algorithm, comparisons were made with existing decision tree models, random forest, and gradient boosting ensemble models. The experiment demonstrated that the proposed algorithm achieved a prediction accuracy of approximately 93%. Furthermore, the MAD classification model in the proposed algorithm reduced errors by about 2.9 m in the root mean square error (RMSE) index compared to tag coordinate predictions using the Kalman filter. The system presented in this paper is poised to advance the development of more reliable indoor location estimation applications, particularly by enabling 3D tag location estimation in blind conditions commonly encountered in indoor UWB environments.

References

[1] V. Maximov and O. Tabarovsky, “Survey of Accuracy Improvement Approaches for Tightly Coupled ToA/IMU Personal Indoor Navigation System,” International Conference on Indoor Positioning and Indoor Navigation (IPIN), Oct. 2013.

[2] F. S. Daníş, A. T. Cemgíl and C. Ersoy, “Adaptive Sequential Monte Carlo Filter for Indoor Positioning and Tracking With Bluetooth Low Energy Beacons,” IEEE Access, pp. 37022-37038. Mar. 2021.

[3] J. M. Quinteiro, D. S. Rodríguez, P. H. Morera and I. A. González, “A Low Complexity System Based on Multiple Weighted Decision Trees for Indoor Localization,” Sensors, vol. 15(6), pp. 14809-14829. Jun. 2015.

[4] Z. Deng, Y. Cao, P. Wang and B. Wang, “An Improved Heuristic Drift Elimination Method for Indoor Pedestrian Positioning,” Sensors, vol. 18(6), pp. 1-12, Jun. 2018.

[5] Ho Chul Lee, Woo Hyun Kim and Dong Myung Lee, “Indoor Positioning Algorithm Using a Decision Tree Model in UWB Environment,” Proceedings of the 2nd International Conference on Maritime IT Convergence (ICMIC 2023), Korea Information and Communications Society (KICS), pp. 89, Aug. 2023..

[6] Ho Chul Lee, “3D Indoor Positioning System Based On Decision Tree Classification Model Using Heuristic Depth Searching and MAD Cost Function,” Thesis for Degree of Doctor of Engineering, Dept. of Computers & Media Engineering, The Graduate School of Tongmyong University, pp. 1–42, Aug. 2023.

Smart Farm: Acquiring and Analyzing Environmental Data

Hoon-Seok Jang
IT Application Research Center
Korea Electronics Technology Institute
Jeonju, Republic of Korea
jhs0053@keti.re.kr

***Abstract*—Thanks to the advancement of information and communication technology (ICT), the development of Internet of Things (IoT)-based technologies is enhancing the quality of life and expanding its applications to diverse fields. In particular, smart farms have been gaining attention as a solution to the sustainability crisis faced by rural areas of Korea. A smart farm is a farm that optimizes the growing environment for crops to increase productivity and quality. If the relationships between environmental data in smart farms are analyzed, additional productivity enhancement and crop management will be possible. This paper proposes a method for analyzing the relationship between environmental data in a testbed for smart farm. To achieve this, nine data is initially acquired through RS-485 communication between the main and sensor boards, and stored in a database. Subsequently, the stored data is downloaded in Excel sheet format and subjected to analysis via histogram analysis, data chart analysis, and correlation coefficient heatmap analysis. This enables the confirmation of data distribution, trends in data changes, and correlations. We plan to conduct more detailed and accurate data analysis through the acquisition of long-term data in the future.**

Keywords—Smart Farm, Data Acquisition, Data Analysis, Environmental Data, Internet of Things (IoT)

I. Introduction

Thanks to the advancement of information and communication technology (ICT), the development of Internet of Things (IoT)-based technologies is enhancing the quality of life and expanding its applications to diverse fields. [1,2] In recent years, smart farms have attracted great attention in Korea as a solution to rural problems that are facing a sustainability crisis, such as aging farmers, population decline, decrease in production area, and stagnation of income, exports, and growth rates due to investment curtailment [3]. A smart farm is a farm that utilizes crop growth and environmental information to optimize growing conditions, and efficiently manages labor, energy, and nutrient usage through timely prescriptions to enhance both crop productivity and quality. In these smart farms, further productivity improvement and crop management will be possible if environmental data is analyzed to identify features and relationships [4]. Therefore, in this paper, we conducted a study to acquire environmental data and analyze the acquired data. This paper is organized as follows: Section 2 describes how we acquired the data, Section 3 explains how we analyzed the data, and Section 4 presents our conclusions.

II. Data Acquisition

Section 2 describes how to acquire the environmental data needed for analysis.

A. Establishment of a Smart Farm Testbed

To acquire data on the growth environment in smart farms, we constructed a testbed equipped with nine sensors measuring temperature, humidity, CO2, insolation, soil temperature, soil moisture, soil EC, Drainage EC, and Drainage pH. Figure 1 shows the smart farm testbed.



Fig. 1. Testbed for mart farm

B. Process of Acquiring Environmental Data

The process of acquiring environmental data is shown in Figure 2.

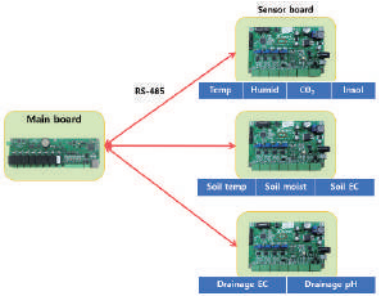


Fig. 2. Data acquisition process

It works by using RS-485 communication between the main board and the three sensor boards to send and receive data. The first sensor board is responsible for acquiring four types of data, including temperature, humidity, CO2, and insolation. The second sensor board is responsible for acquiring three types of data, including soil temperature, soil moisture, and soil EC. Finally, the last sensor board is responsible for acquiring two types of data, including Drainage EC and Drainage pH.

C. Establishment of Database for Smart Farm

The process of data storage is shown in Figure 3. The nine types of enviromental data obtained from the three sensor boards are transmitted from the main board to the web server via HTTP request. The web server stores the transmitted data in the database, and in this process, the success of data storage can be checked through the HTTP response sent by the web server.

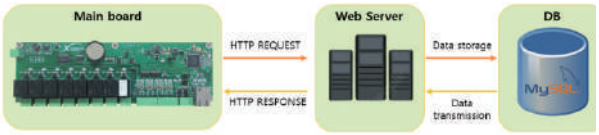


Fig. 3. The process of storing data in a database

D. Preprocessing of Environmental Data

Environmental data were obtained at 1-hour intervals from June 20 to November 16, 2022. The stored data were downloaded from the database into an Excel sheet, and after data preprocessing, about 300,000 data were acquired, excluding missing data.

III. Data Analysis

Section 3 describes how to analyze the environmental data. Histogram analysis, data chart analysis, and correlation heatmap analysis are used in this manuscript.

A. Histogram Analysis

To understand the distribution of the environmental data, we visualized it using a histogram. Figure 4 shows a histogram of the environmental data, including temperature, humidity, CO2, soil temperature, soil moisture, insolation, soil EC, Drainage EC, and Drainge pH arranged from top left.

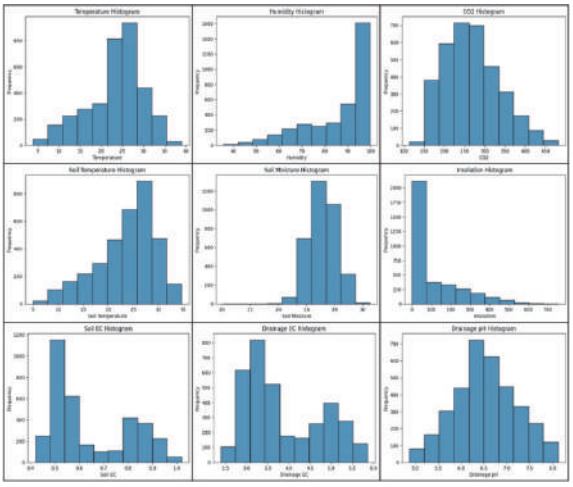


Fig. 4. Data histogram

When analyzing the data distribution, we observed that the temperature had a significant amount of data in the 20-30℃ range, while humidity had a large amount of data in the 80-100% range. The CO2 data was evenly distributed between 200 and 350 ppm, and similar to the temperature, the soil temperature data was heavily distributed between 20 and 30℃. Additionally, the soil moisture was found to be 26-28%, insolation ranged from 0-100 w/m^2, and soil EC was distributed in two ranges, 0.5-0.6dS/m and 0.8-0.9dS/m, and Drainage EC was also distributed in two ranges, 3.0-3.5dS/m and 4.5-5.5dS/m. Finally, the Drainage pH values ranged from 6.0-7.5pH. For further analysis, we created the data histograms for day and night, with the 06:00~18:00 time period when insolation data was frequently measured as day and the remaining time as night, as shown in Figure 5.

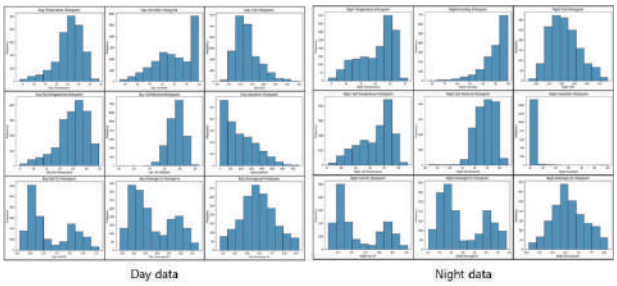


Fig. 5. Data histogram for day and night

It was observed that the temperature distribution showed a higher frequency in the range of 20~35℃ during the day and 20~30℃ at night. Humidity was the most distributed at 90~100% both during the day and night. CO2 had a large distribution of 150 to 250ppm during the day and an even distribution of 250 to 400ppm at night. Soil temperature was similar to the temperature and soil moisture was heavily distributed between 26~28% both during the day and at night. Insolation showed a distribution that converges at 0 w/m^2 during the day and a wide distribution of 0 to 700 w/m^2 at night. Soil EC, Drainage EC, and Drainage pH were distributed similarly to the histograms in the Figure 4, regardless of day or night.

B. Data Chart Analysis

After analyzing the data distribution using histograms, we separated the data into day and night and calculated the monthly average, median, minimum, and maximum values. These values were visualized on the chart to analyze the trends in the data over time. The chart used blue to represent the average, orange to represent the median, gray to represent the minimum, and yellow to represent the maximum values. After analyzing the temperature data, we found that the average and median values for both day and night remained stable from June to August, and then decreased from September to November, as shown in Figure 6.

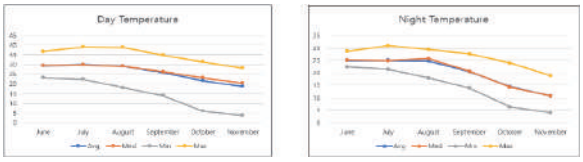


Fig. 6. Temperature data chart for day and night

When comparing the minimum and maximum values, we observed that the diurnal temperature range was greater during the day than at night. After analyzing the humidity data, we found that, unlike temperature data, the average and median values for both day and night remained consistent from month to month, as shown in Figure 7.

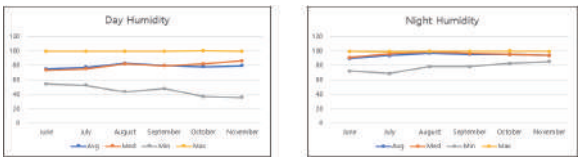


Fig. 7. Humidity data chart for day and night

The monthly average humidity during the day ranged from 74% to 83%, while the monthly average humidity during the night ranged from 89% to 97%. This indicates that the night had a higher average humidity than the day, and the range between minimum and maximum values also showed that the humidity during the night was more stable than during the day. After analyzing the CO2 data, we found that both day and night showed a gradual decrease in average and median values from June to November, as shown in Figure 8.

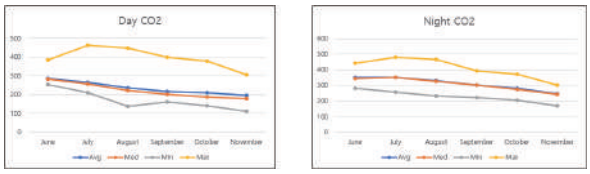


Fig. 8. CO2 data chart for day and night

Additionally, we observed that night had higher median and minimum values, as well as maintained higher CO2 concentrations compared to day. The maximum values were higher at night until August, but from September onwards, the day values were higher by a small margin. After analyzing the soil temperature data, we found that the data were similar to the temperature data, as shown in Figure 9.

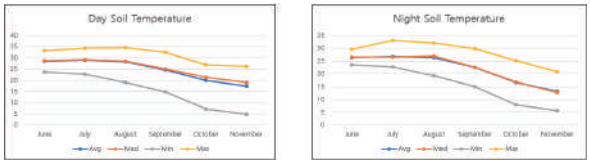


Fig. 9. Soil temperature data chart for day and night

The average values for both day and night stayed constant from June through August and then decreased from September through November. After analyzing the soil moisture data, we found that there was little difference in the average, median, minimum, and maximum values between day and night, except for September when there was little variability in the data, as shown in Figure 10.

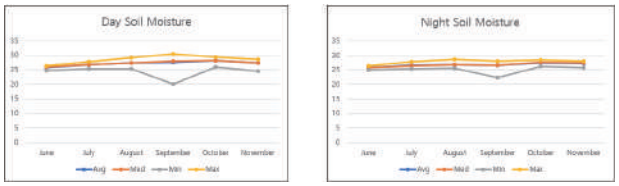


Fig. 10. Soil moisture data chart for day and night

We observed that the soil moisture levels were relatively stable throughout the day and night, with an average range of 25% to 28%. After analyzing the insolation data, we found that the data values converged to a much lower average value of 0w/m^2 at night than during the day, as shown in Figure 11. During the day, the data values increased from June to September and then decreased. The minimum value was zero in all months for both day and night, and the maximum value was 753w/m^2 in August for day and 60w/m^2 in July for night.

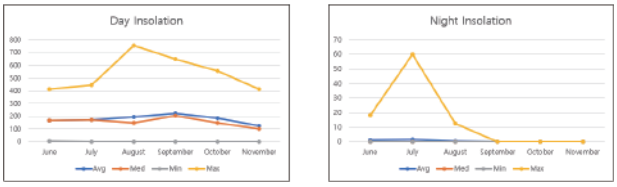


Fig. 11. Insolation data chart for day and night

After analyzing the soil EC data, we found that the average, median, minimum, and maximum values for both day and night were not significantly different, except for the maximum values in July and September, as shown in Figure 12.

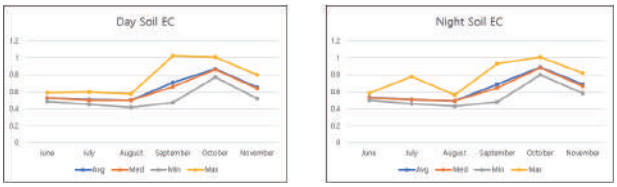


Fig. 12. Soil EC data chart for day and night

Average values during the day and night decreased from June to August, increased from September to October, and then decreased again in November. After analyzing the Drainage EC data, we found that the day and night charts show a similar pattern to the soil EC, as shown in Figure 13.

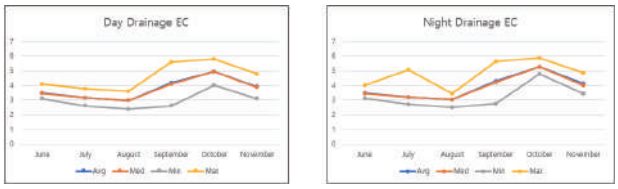


Fig. 13. Drainage EC data chart for day and night

Moreover, we found that the average, median, minimum, and maximum values of both day and night were not significantly different, except for the maximum value in July. The average values of day and night decreased from June to August, increased from September to October, and then decreased again in November. After analyzing the Drainage pH data, we found that, except for the minimum value in August, the average, median, minimum, and maximum values for both day and night were not significantly different, similar to the soil EC and Drainage EC, as shown in Figure 14.

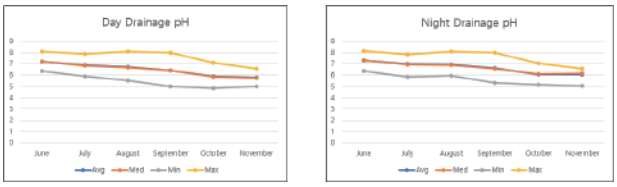


Fig. 14. Drainage pH data chart for day and night

We also observed a gradual decrease in the average value from June to November for both day and night.

C. Correlation Heatmap Analysis

After analyzing the data charts, we found that certain environmental data, such as temperature and soil temperature, soil EC, and Drainage EC, exhibited similar monthly changes. To analyze these relationships, we calculated the correlations of the data and drew the data heatmaps for day and night to provide a clear overview of the correlations. The Pearson correlation coefficient formula is shown in equation (1) [5].

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (1)$$

n represents the size of the data, x and y represent the data x and data y, respectively, and x_i and y_i represent the i^{th} value of x and y, respectively. The Pearson correlation coefficient values range between -1 and 1, with values closer to 1 indicating a strong correlation and values closer to 0 indicating little to no correlation between the two data sets. After analyzing the correlation coefficients for day, we found a very strong positive correlation between temperature and soil temperature, as well as between soil EC and Drainage EC, as shown in Figure 15.

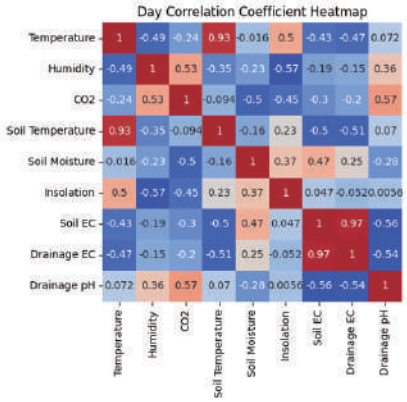


Fig. 15. Correlation coefficient heatmap for day

We observed a moderate positive correlation between temperature and insolation, humidity and CO2, and between CO2 and Drainage pH. In addition, we found moderate negative correlations between humidity and insolation, CO2 and soil moisture, soil temperature and soil EC, soil temperature and Drainage EC, soil EC and Drainage pH, and between Drainage EC and Drainage pH. The other data showed weak, very weak, or almost no positive or negative correlations. After analyzing the night correlation coefficients, we found a very strong positive correlation between temperature and soil temperature, as well as between soil EC and Drainage EC, similar to what was observed during the day, as shown in Figure 16. Additionally, we observed a strong negative correlation between temperature and soil EC. We found a moderate positive correlation between CO2 and Drainage pH, and moderate negative correlations between temperature and Drainage EC, soil temperature and Drainage EC, soil EC and Drainage pH, and between Drainage EC and Drainage pH. The other data showed weak, very weak, or almost no positive or negative correlations.

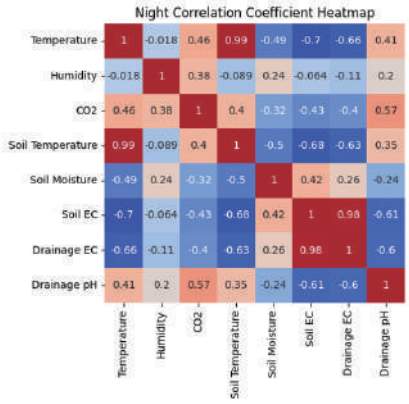


Fig. 16. Correlation coefficient heatmap for night

IV. Conclusion

In this paper, we first acquired nine types of data through RS-485 communication and stored them in a database to acquire the environmental data for smart farm. Then, we downloaded the data in Excel sheet format from the database and conducted data analysis. Through histogram analysis, data chart analysis, and correlation coefficient heatmap analysis, we were able to understand the data distribution, data changes for day and night, and the correlation between each data. Through this study, we were able to identify the correlation of environmental data. However, to acquire more accurate results, long-term data acquisition and analysis are required. Therefore, in future studies, we plan to conduct more detailed and accurate analysis through long-term data acquisition.

V. Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2021-0-00751, Development of multi-dimensional visualization digital twin framework technology for displaying visible and invisible information with lower than 0.5mm precision)

References

[1] S. Madakam, R. Ramaswamy, and S. Tripathi, “Internet of Things (IoT): A Literature Review,” Journal of Computer and Communications, vol. 3, no. 5, pp. 164–173, May 2015.

[2] N. H. Motlagh, M. Mohammadrezaei, J. Hunt, and B. Zakeri, “Internet of Things (IoT) and the Energy Sector,” energies, vol. 13, no. 2, pp. 494, January 2020.

[3] S. Kim, M. Lee, and C. Shin, “IoT-Based Strawberry Disease Prediction System for Smart Farming,” Sensors, vol. 18, no. 11, pp. 4051, November 2018.

[4] K. H. Son, H. S. Sim, J. K. Lee, and J. Lee, “Precise Sensing of Leaf Temperatures for Smart Farm Applications,” horticulturae, vol. 9, no. 4, pp. 518, April 2023.

[5] I. Jebli, F. Z. Belouadha, M. I. Kabbaj, and A. Tilioua, “Prediction of solar energy guided by pearson correlation using machine learning,” Energy, vol. 224, pp. 120109, June 2021.

Design of Performance Measurement System for Non-Face-to-Face Remote Workers

Min Goo Lee
Energy IT Convergence Research Center
Korea Electronics Technology Institute
Seongnam-si, Republic of Korea
emingoo@keti.re.kr

Yong Kuk Park
Energy IT Convergence Research Center
Korea Electronics Technology Institute
Seongnam-si, Republic of Korea
ykpark@keti.re.kr

Kyung Kwon Jung
Department of Electrical Engineering
Dongshin University
Naju-si, Republic of Korea
kkjung@dsu.ac.kr

***Abstract*— The COVID-19 pandemic triggered an unprecedented shift in work practices, propelling millions towards remote work arrangements. However, the companies are seeking methods for assessing the productivity and effectiveness of their non-face-to-face employees. This paper proposed the complexities of performance measurement in the remote work landscape, proposing a framework for evaluating employer efficacy in this new normal. A few indices were developed to reflect the performance measurement for remote workers who work from home or remote office is presented through the combined use of the fuzzy comprehensive evaluation and the entropy weight decision-making method. The fuzzy set approach was adapted to identify the membership degree of each index to various evaluation results, whereas the entropy weight method is used to acquire the sets of weighting factors. This method represented a unified one of the quantitative results obtained from the system data and the qualitative results based on the judgment of human resource domain experts. Finally, practical examples were served for demonstrating the developed approach.**

Keywords—non-face-to -face remote workers, work-from-home, fuzzy comprehensive evaluation, performance measurement

I. INTRODUCTION

The COVID-19 pandemic significantly altered the employment environment, propelling millions into remote work settings. While this shift offered potential benefits like improved work-life balance and cost savings, it presented novel challenges for employers evaluating and managing remote teams.

The onset of remote work has not only redefined the concept of the workplace but also necessitated the development of new metrics and methodologies for assessing worker performance, engagement, and well-being. The flexibility and autonomy offered by remote work, while advantageous in fostering a better work-life balance, also demand a reevaluation of how productivity and efficiency are measured and encouraged. As such, this paper seeks to explore and establish comprehensive criteria and best practices for the evaluation of remote workers, taking into account the diverse aspects of their work life, from performance metrics to the utilization of technology, self-management skills, and their overall well-being and job satisfaction[1-2].

Many companies have elected to maintain remote work policies, buoyed by research indicating improvements in productivity and work efficiency attributable to this mode of employment. This transition, while initially a response to global health concerns, has revealed the potential for remote work to enhance organizational performance and employee satisfaction.

However, the evaluation of performance in a remote work environment has emerged as a significant source of contention among workers. From the employer's perspective, there is a pressing need to monitor whether employees are effectively managing their tasks and utilizing their time efficiently. Conversely, employees have expressed concerns regarding the fairness and accuracy of these evaluations, highlighting a growing discord between the perceived benefits of remote work and the realities of its implementation.

In recent times, crowdsourcing through digital collaborative work platforms has garnered significant attention. Crowdsourcing, a portmanteau of "crowd" and "outsourcing," refers to the practice of opening up certain aspects of corporate activities to participation by consumers or the general public. This process involves soliciting knowledge from the public on problems that need solving, encouraging their participation in tasks, and offering appropriate rewards to those who contribute.

During the COVID-19 pandemic, the rapid spread of the coronavirus prompted the introduction of flexible working arrangements, such as staggered work hours, and remote working policies to minimize contact between workers while ensuring organizational continuity. This period saw an exponential growth in the crowdsourcing market as companies increasingly adopted online collaboration for remote work. The crowdsourcing industry is primarily composed of three key elements: platforms, requesters, and providers. Platforms serve as marketplaces, facilitating the connection between requesters and providers over the internet.

Requesters, often referred to as crowdsourcers, mainly focus on ensuring a dependable source of skilled remote workers who can execute tasks with high quality. On the other hand, providers, or remote workers, aim for a consistent job provision system that acknowledges their expertise and worth. The demands of the key stakeholders in the crowdsourcing industry can be simultaneously addressed by improving the mutual trust in the crowdsourcing system[3].

Based on these insights, the introduction of this paper aims to explore the significant impact of crowdsourcing in the context of the shift towards non-face-to-face work modalities accelerated by the COVID-19 pandemic. By examining the dynamics between platforms, requesters, and providers, this study seeks to understand how crowdsourcing has evolved as a crucial element of digital collaboration, highlighting the importance of trust and quality in these interactions. Through this exploration, the paper will contribute to the broader discourse on the future of work and the role of crowdsourcing in facilitating effective remote work environments.

TABLE I. INDICES FOR ASSESSING THE WFH WORKERS

First layer	Second layer	Third layer	Fourth layer
Assessing the WFH workers	Work attitude factor U_1	attendance compliance factor U_{11}	number of work start-time compliance U_{111}
			number of work end-time compliance U_{112}
			number of intensive work compliance U_{113}
			number of leave early U_{114}
			number of go out U_{115}
			number of agent program use U_{116}
		attendance violation factor U_{12}	number of absence without notice U_{121}
			number of lateness without notice U_{122}
			number of leave early without notice U_{123}
			number of go out without notice U_{124}
			number of leave work without notice U_{125}
			number of intensive work non-compliance U_{126}
			number of agent unuse U_{127}
		work sincerity factor U_{13}	number of work-related programs use U_{131}
			time of work-related programs use U_{132}
			number of work-related sites visit U_{133}
			number of work-related keyword use U_{134}
			total active-time of worker U_{135}
		work insincerity factor U_{14}	number of non-work-related programs use U_{141}
			time of non-work-related programs use U_{142}
			number of non-work-related sites visit U_{143}
			number of non-work-related keyword use U_{144}
			total idle-time of worker U_{145}
	Job ability factor U_2	work loyalty factor U_{21}	number of assignment task U_{211}
			number of assignment task completed U_{212}
			task execute time U_{213}
		job skill factor U_{22}	number of work-related programs tool U_{221}
			time of work-related programs hold U_{222}
		work performance factor U_{23}	number of deadlines compliance U_{231}
			number of deadlines early compliance U_{232}
			time of deadlines compliance U_{233}
			number of deadlines non-compliance U_{234}
	Partnership factor U_3	colleague reputation factor U_{31}	senior reputation score U_{311}
			colleague reputation score U_{312}
			number of complaints/accidents U_{313}
			number of praise U_{314}
			number of penalty U_{315}
	Security factor U_4	security compliance factor U_{41}	number of approved program installation U_{411}
			number of unauthorized programs installation U_{412}
			number of screen saver not working U_{413}
			computer password renewal cycle U_{414}
			agent password renewal cycle U_{415}
			installation status of security program U_{416}

This paper proposes a novel design for a performance measurement system specifically tailored for telecommuting workers. Traditional performance evaluation methods often struggle to accurately assess the output of remote employees due to the inherent challenges of measuring productivity and engagement in non-traditional work settings. To address this issue, the proposed system leverages the fuzzy comprehensive evaluation method, a multi-criteria decision-making technique adept at handling qualitative and imprecise data. With the user's consent, an information collection agent is installed on the worker's computer to automatically collect and transmit data to the server. In order to develop an evaluation platform, we constructed an FCM module using both quantitative worker information and qualitative evaluation information as a preliminary step. The usefulness of the proposed system was verified through testing.

II. THE EVALUATION INDEX SYSTEM

The proposed indicator system represents an evaluating indicator set in which all indicators are chosen according to the appraisal system of the pre-specified experts such as HR departments. The development of the evaluation system needs to observe technicality, rationality and feasibility. The implemented guideline of KPIs (Key Performance Indicators) have been referred to screen and optimize each target, including one first-level target as evaluation score, four second-level targets, nine third-level targets and forty three fourth-level targets. Table 1 is proposed Comprehensive evaluation indicator system.

III. FUZZY COMPREHENSIVE EVALUATION METHODS

A fuzzy comprehensive evaluation (FCE) is an artificial evaluation technique based on fuzzy mathematics, with real and accurate evaluation results that are extensively utilized in psychology and other fields. The fuzzy comprehensive evaluation is a process of integrating multiple independent variable scores into a dependent variable score. The traditional method is to determine the weight by expert scoring, which is subjective. We use other objective methods to obtain the weight distribution. The fuzzy comprehensive evaluation emerges as a powerful methodology for tackling inherent complexities and subjective interpretations. Leveraging the principles of fuzzy logic and set theory, this approach gracefully handles uncertainty and imprecision, often encountered when evaluating complex systems or phenomena. Through the construction of a meticulously designed evaluation index system and the subsequent assignment of

appropriate weights to individual indicators, this method empowers decision-makers with a nuanced understanding. Beyond the mere aggregation of quantitative data, fuzzy comprehensive evaluation seamlessly integrates both subjective and objective weights, yielding evaluations that are not only accurate but also reflect the inherent ambiguity often present in real-world scenarios. The versatility of this technique is attested to by its successful application in diverse fields, ranging from engineering education certification to creative dance teaching quality evaluation, demonstrating its capacity to enhance the objectivity, scientifically, and ultimately, the effectiveness of decision-making across a vast spectrum of disciplines[4-6].

The formal procedures of the general fuzzy evaluation model can be described by the following steps.

Step 1: Determining the set of evaluation factors

Evaluation factors can be defined according to the objectives of the product evaluation process. A set of n evaluation factors can be represented as a vector U = {u1, u2, ..., un}. For example, one can define U = {effectiveness, efficiency, user satisfaction} such that different measurements will be conducted to evaluate the product based on these three factors.

Step 2: Determining the set of appraisal grades

The appraisal set can be represented as a vector $V = \{v_1, v_2, \cdots, v_m\}$, in which m represents the number of levels in the appraisal. For example, if $m = 5$, the appraisal vector can be represented as V = {very poor, poor, medium, good, excellent} .

Step 3: Setting the fuzzy mapping matrix

The goal of the evaluation process is to provide a mapping from U to V. For a specific factor u_i the fuzzy mapping to the appraisal vector V can be represented by the vector $R_i = \{r_{i1}, r_{i2}, \cdots, r_{im}\}$, in which m represents the number of levels in the appraisal (see step 2), and r_k represents the fuzzy membership degree of appraisal factor i to grade k.

In general, the fuzzy appraisal matrix of all n factors can be derived and represented as a matrix R, such that if there are n factors and m levels of appraisal grades:

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{bmatrix} \tag{1}$$

In the above matrix notation for R, each row represents the set of appraisal membership degrees to the corresponding appraisal vector V for each evaluation factor u_i in the evaluation vector U.

Step 4: Determining the weight of each evaluation factor

To obtain a comprehensive usability evaluation, the relative importance of each evaluation factor on the overall grading of the product should be quantified. The weight vector can be represented by W, which can be formulated by the entropy weight method, as described in the next subsection.

$$H_i = -k \sum_{j=1}^{n} r_{ij} \ln r_{ij} \tag{2}$$

where, $k = 1/\ln n$. The entropy weight of the *i*th indicator is defined as follows:

$$w_i = \frac{1 - H_i}{m - \sum_{l=1}^{m} H_l} \tag{3}$$

Step 5: Getting the overall appraisal result

The overall appraisal result can be obtained by taking into the account the relative weights of each evaluation factor, such that a single vector with the same level of appraisal grades m can be represented by:

$$B = (b_1, b_2, \cdots, b_m) = W \circ R \tag{4}$$

Where '$\circ$' is a fuzzy composition operator, b_j could be operated by a number of possible models.

$$b_j = \bigvee_{h=1}^{k} (w_h \wedge r_{hj}) \tag{5}$$

Step 6: final comprehensive evaluation matrix

The final score can also be obtained by the evaluation grade scores mentioned above and the evaluation results vector calculated by (4). The evaluation grade scores could be written in a vector form as,

$$F = [1, 0.8, 0.6, 0.4, 0.2] \tag{6}$$

The evaluation result vector B obtained (4) needs to be normalized and could then be denoted as $\bar{B}$. The final score of the power grid structure is obtained as

$$T = \bar{B} \cdot F^T \tag{7}$$

IV. EXPREMENTS

In the traditional workplace setting, performance evaluations of employees have predominantly been conducted through face-to-face interactions, where evaluators and those being evaluated naturally share time and space. This arrangement has led to the development and application of in-person evaluation methodologies that considerably incorporate qualitative factors, centered around the physical premises of business operations. However, the recent societal shift towards remote and asynchronous work environments has highlighted the inadequacies of exclusively relying on conventional face-to-face performance appraisals in varied hybrid work contexts. As a result, there emerges a critical necessity to devise novel performance evaluation methodologies that uphold fairness, specifically designed to accommodate the distinctive characteristics of remote and asynchronous work settings. Such methodologies must also

consider the ephemeral nature of employment, collaboration, evaluation, and dissolution typical of crowdworker engagements, prompting a reevaluation of performance assessment practices to ensure they remain equitable and effective amidst the changing dynamics of the workplace.

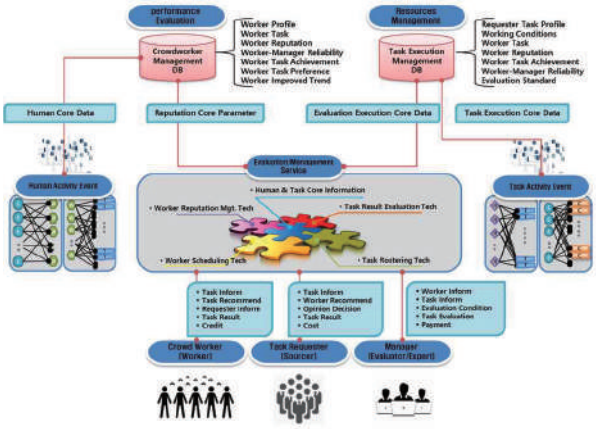
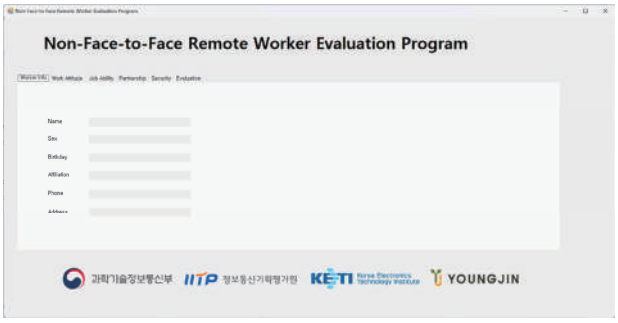
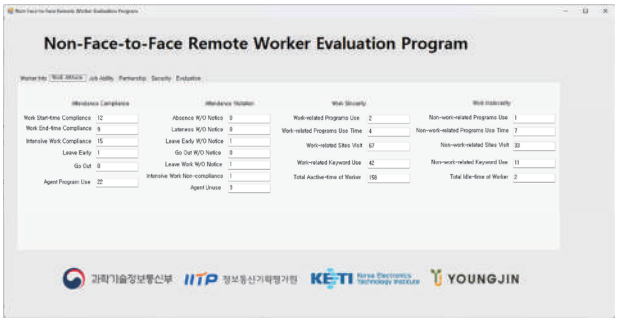


Fig. 1. Diagram of the Crowdsourcing Work Environment.

In Fig. 2, the screen depicted is developed from the proposed multi-attribute framework-based individual evaluation algorithm programmed in C#. It facilitates the input and storage of factors influencing individual evaluations, as well as the computation of individual evaluation outcomes.





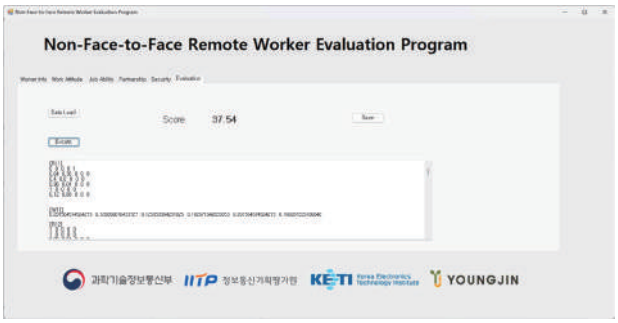


Fig. 2. Crowdworker Performance Evaluation Program.

V. Conclusions

This paper presents an evaluation index system for performance measurements of the work-from-home workers, and develops the evaluation method based on the fuzzy comprehensive evaluation and entropy weight method. This proposed method is a good combination of subjectivity and objectivity, and provides both qualitative and quantitative analysis. Through simulations and practical applications, the index system and the evaluation method presented in this paper have been proven to be reasonable and effective.

Acknowledgment

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.2022-0-00664, Development of non-face-to-face work platform that guarantees scalability based on the EX(Employee eXperience)).

References

[1] Kangoh Lee, “Working from home as an economic and social change: A review,” Labour Economics, Volume 85, 2023, 102462.

[2] Andrew Enaifoghe, Ntombizamakhwalosiziphiwe Zenzile, “The rapidly evolving situation of employee work-from-home productivity and the integration of ICT in Post-COVID-19 pandemic,” Scientific African, Volume 20, 2023, e01709.

[3] Danula Hettiachchi, Vassilis Kostakos, Jorge Goncalves, “A Survey on Task Assignment in Crowdsourcing,” ACM Computing Surveys, Volume 55, Issue 3, Article No.: 49, pp 1–35, 2022.

[4] Yuguo Qi, Fushuan Wen, Ke Wang, Li Li, S.N.Singh, “A fuzzy comprehensive evaluation and entropy weight decision-making based method for power network structure assessment,” International Journal of Engineering, Science and Technology, Vol. 2, No. 5, pp. 92-99, 2010.

[5] Jia Zhengyuan, Wang Chunmei, Huang Zhiwei, Zhang Gang, “Evaluation Research of Regional Power Grid Companies’ Operation Capacity based on Entropy Weight Fuzzy Comprehensive Model,” Procedia Engineering, Volume 15, pp. 4626-4630, 2011.

[6] Xiaowen Ding, Xiao Chong, Zhengfeng Bao, Ying Xue, Shanghong Zhang, “Fuzzy Comprehensive Assessment Method Based on the Entropy Weight Method and Its Application in the Water Environmental Safety Evaluation of the Heshangshan Drinking Water Source Area, Three Gorges Reservoir Area, China,” Water 9, no. 5: 329, 2017.

Forecasting Daily Demand in the Busan Metro Using LSTM-Based Encoder-Decoder

Sang-Hyun Hwang
Major of Information and Communications Engineering
Pukyong National University
Busan, South Korea
noul77@pukyong.ac.kr

Jun-Hyeong Park
Major of Information and Communications Engineering
Pukyong National University
Busan, South Korea
jason990820@gmail.com

Jeong-Won Seo
Major of Information and Communication Engineering
Pukyong National University
Busan, South Korea
sjw3570@naver.com

Pil-Goo Jeong
Line 1 Operations Office
Seomyeon Station
Busan Transportation Corporation
Busan, South Korea
vlfrn1830@naver.com

Jung-Tae Kim
Line 2 Operations Office
Hopo station
Busan Transportation Corporation
Busan, South Korea
busangull@humetro.busan.kr

Jun-Pyo Hong
Major of Information and Communications Engineering
Pukyong National University
Busan, South Korea
jp_hong@pknu.ac.kr

***Abstract*— In this study, we perform time series forecasting for the number of passengers. To achieve this, we use daily passenger data from the past ten years provided by the Busan transportation corporation and employ an LSTM (Long Short-Term Memory) based encoder-decoder model. In addition, we analyze the correlation coefficients between the number of passengers and various external independent variables, followed by constructing a time series forecasting model for the number of passengers. The experimental results show that the model performs best when trained with independent variables that have an absolute correlation coefficient of 0.25 or higher with the number of passengers. The proposed advanced metro demand forecasts can improve the cost efficiency of the Busan metro system by enabling appropriate train scheduling and budget planning.**

***Keywords*— Time Series Forecasting, LSTM, Deep Learning, Encoder-Decoder**

I. INTRODUCTION

Fares are the primary income source for Busan metro, accounting for 85.2% of its total operating revenue. To manage and increase this revenue efficiently, Busan transportation corporation has annually predicted and analyzed demand forecasting using multiple regression analysis. Although multiple regression analysis has been widely used for a statistical forecast, it has several significant limitations. Firstly, multiple regression analysis considers only linear relationships between variables, making it unsuitable for modeling the nonlinear patterns and complex dynamic changes in time series data [1]. Additionally, it fails to adequately reflect the autocorrelation of time series data, and ignoring these characteristics can significantly reduce the accuracy of predictions [2]. Therefore, other methods are needed to accurately model the complex patterns of time series data.

We study the forecasting of the monthly number of passengers for Busan metro using an LSTM (Long Short-Term Memory)-based encoder-decoder, a powerful architecture for analyzing time series data. Recently, this model has been utilized to predict PM10 concentrations based on real-time sensor data [3] and to analyze stock and shared bicycle data [4]. They have demonstrated that this model is a powerful tool capable of effectively capturing and analyzing the patterns and long-term dependencies of complex time series data.

Unlike previous studies that used fixed independent variables, this study is distinguished by performing modeling with varying independent variables and then comparing and analyzing the performance. Furthermore, it is significant in that it uses actual passenger data from the Busan transportation corporation, which was not addressed in previous studies, to verify the model's applicability and effectiveness.

II. MAIN BODY

A. Data Analysis and Preprocessing

For the time series forecasting of the number of passengers, this study uses a dataset of 3,652 days provided by the Busan transportation corporation from 2014 to 2023. The dataset is composed as Table I.

Table I. General Specification

Variable Type	Feature Name	Data Value	Measurement Interval
Dependent Variable	Number of Passengers	Numeric	Daily
Independent Variable	Holiday Status	Weekday (1), Saturday (2), Sunday and Public Holidays (3)	Daily
	Average Temperature	Numeric	Daily
	Precipitation	Numeric	Daily
	Economically Active Population	Numeric	Monthly
	Number of Registered Private Cars	Numeric	Monthly
	COVID-19 Impact	No Impact (0), Impact (1))	Daily

The data is preprocessed as Table II, depending on the feature type.

Table II. Processing by Feature Type

Feature Type	Feature Name	Description
Continuous Variables	Holiday Status, COVID-19 Impact	Applied Min-max scaling to normalize the data to a range between 0 and 1.
Categorical Variables	Number of Passengers, Average Temperature, Precipitation, Economically Active Population, Number of Registered Private Cars	Applied one-hot encoding for preprocessing.

The daily number of passengers from 2014 to 2023 are graphed as follows.

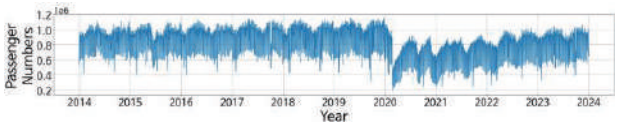


Fig. 1. Daily number of passengers from 2014 to 2023

Due to social distancing measures caused by COVID-19, the number of passengers from February 22, 2020, to April 1, 2022, show a different trend compared to the rest of the data. For this reason, the feature 'COVID-19 impact' is set to be 1 during this period. The correlation between the daily number of passengers (dependent variable) and other features (independent variables) is shown in Fig. 2.

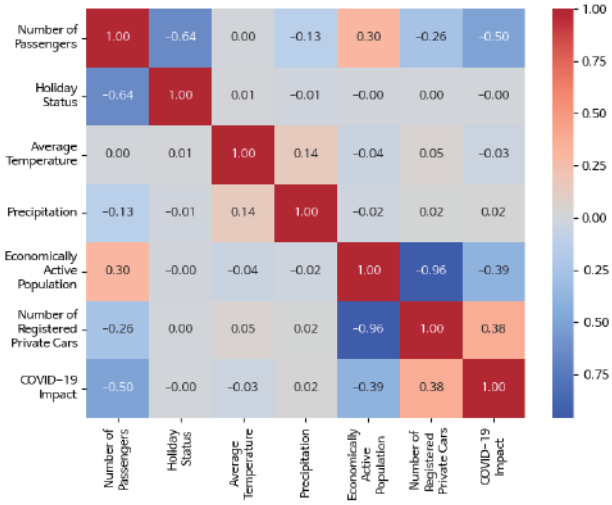


Fig. 2. Correlation matrix between daily passenger number and other features

The feature with the highest correlation with 'Number of Passengers' is 'Holiday Status', with a correlation coefficient of -0.64. This is followed by 'COVID-19 Impact' (-0.50), 'Economically Active Population' (0.30), 'Number of Registered Private Cars' (-0.26), 'Precipitation' (-0.13), and 'Average Temperature' (0.00).

From the given time series data, we constructed 2,769 training data samples, 308 validation data samples, and 157 test data samples using the sliding window method. The input window size is 180 days, and the output window size is 30 days.

B. Model Architecture

For predicting the number of passengers on the Busan metro, this study chose an LSTM-based encoder-decoder model. Fig. 3 shows the architecture of this model. The encoder extracts important features from the input sequence and passes the context vector to the decoder, which then uses this context vector to generate the output sequence. This structure can model the complex patterns of time series data more accurately and effectively solve long-term dependency issues, thereby improving prediction performance. In particular, LSTM overcomes the limitations of traditional Recurrent Neural Networks (RNNs) and excels at retaining information over long sequences. This advantage makes LSTM especially suitable for handling time series data, resulting in more accurate and reliable predictions [5-6].

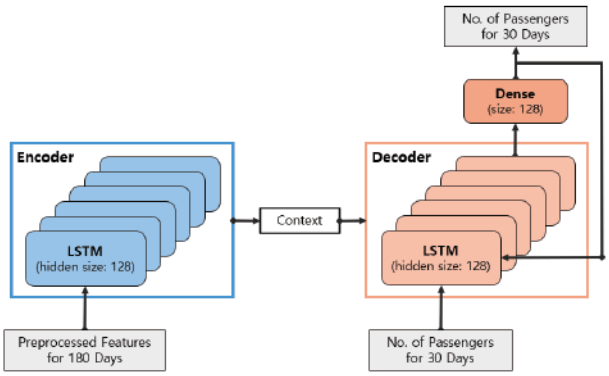


Fig. 3. Architecture of the LSTM encoder-decoder model

C. Model Training

We train the model with four different methods by varying the input features used during. The input features used for each method are summarized in Table III.

Table III. Input Features for Each Method

Method	Description
Method ①	Perform prediction using only dependent variables without considering independent variables. **Input Features: Number of Passengers**
Method ②	Use independent variables with a correlation coefficient greater than 0.5. **Input Features: Number of Passengers, Holiday Status, COVID-19 Impact**
Method ③	Use independent variables with a correlation coefficient greater than 0.25. **Input Features: Number of Passengers, Holiday Status, COVID-19 Impact, Economically Active Population, Number of Registered Private Cars**
Method ④	Use both dependent and independent variables. **Input Features: Number of Passengers, Holiday Status, COVID-19 Impact, Economically Active Population, Number of Registered Private Cars, Average Temperature, Precipitation**

The model is trained until the mean squared error (MSE) on the validation dataset converges. The Adam optimizer is adopted, and the final hidden state of the encoder was passed to the decoder as a context vector, with a linear layer added at the end of the LSTM to adjust the output. Google Colab

Pro+ and PyTorch framework are used. The Table IV shows the hyperparameters used during the training process.

Table IV. Hyperparameters Used for Training

	Method ①	Method ②	Method ③	Method ④
epoch	211	102	250	215
learning rate	0.0005			
batch size	64			
hidden size	128			
number of hidden layers	6			
teacher forcing rate	0.5			

D. Performance Evaluation

We use the MAPE(mean absolute percentage error) and the MASE (mean absolute scaled error) as the performance measures. MAPE measures the average percentage error between the predicted values and the actual values. MASE measures the error by considering the seasonality of the time series data through scaling. The prediction performance of the model is evaluated by comparing it with the performance of a simple random walk model to see how much the prediction model has improved. If the MASE value is less than 1, the model's predictions are considered better than a random walk. The test results are summarized in Table V.

Table V. Performance comparison

	MAPE	MASE
Method ①	13.20%±0.43%p	0.5451±0.0049
Method ②	12.64%±0.32%p	0.5096±0.0230
Method ③	11.31%±0.21%p	0.4775±0.0113
Method ④	11.40%±1.22%p	0.5237±0.1031

The test results show that the model built using Method ③ has a MAPE of 11.31% and a MASE of 0.4775, indicating the best performance. This is believed to be because Method ③ included independent variables with relatively low correlation coefficients, allowing the model to use more diverse information to more accurately predict the variability in the number of passengers. The added independent variables comprehensively reflect various external factors affecting the number of passengers in the complex system, contributing to improving the generalization capability of the prediction model.

However, in Method ④, the inclusion of independent variables with very low correlation coefficients results in performance degradations. This suggests the need for careful consideration in the selection of independent variables, emphasizing that efficient and effective variable selection is crucial for improving model performance.

Fig. 4 visualizes the result of passenger number prediction for one test data sample using the model trained with Method ③. The blue section represents the input period, the red section represents the period predicted by the model, and the light blue section represents the actual values of the predicted period.

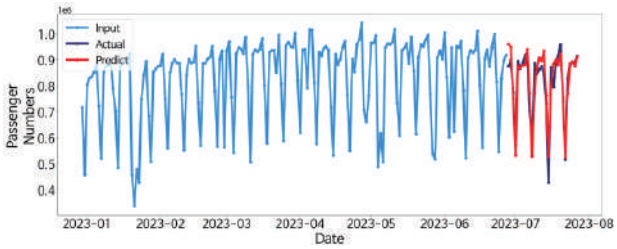


Fig. 4. Model prediction

The graph shows that while seasonality and periodicity are well predicted, some errors are observed in the specific values. In the future, it will be necessary to strengthen the model's generalization capability and improve prediction performance by including more data and various independent variables.

III. Conclusion

This study developed a passenger number prediction model using an LSTM-based encoder-decoder structure and analyzed its performance. Using actual passenger data from the Busan transportation corporation, the model was constructed and tested to verify its applicability and effectiveness in a real environment. The passenger prediction system utilizing the LSTM-based encoder-decoder model includes short-term fluctuations not captured by multiple regression models, enabling short-term financial planning and efficient resource allocation, which can be directly used for monthly supplementary budget formulation. Future research can enhance the model by incorporating more data and considering various external factors, continuously evaluating and improving its performance in real-world environments.

References

[1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, “Time Series Analysis: Forecasting and Control,” John Wiley & Sons, 2013.
[2] J. D. Hamilton, "Time Series Analysis,” Princeton University Press, 1994.
[3] S.-K. Kim and T.-I. Oh, “Real-time PM10 concentration prediction LSTM model based on IoT streaming sensor data,” Journal of the Korea Academia-Industrial Cooperation Society, vol. 19, no. 11, pp. 310–318, 2018.
[4] H. Zhang, S. Li, Y. Chen, J. Dai, and Y. Yi, “A novel encoder-decoder model for multivariate time series forecasting,” Computational Intelligence and Neuroscience, Article ID 5596676, 2022.
[5] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[6] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” Advances in Neural Information Processing Systems (NIPS), vol. 27, pp. 3104–3112, 2014.

Leveraging Data Mining and NLP for Service Readiness: The Case of Hotel Industry

Moner Alajeli Farj Abouoden
Faculty of Artificial Intelligence
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
monerabowden2@gmail.com

Siti Sophiayati Yuhaniz
Faculty of Artificial Intelligence
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
sophia@utm.my

***Abstract*— The rapid advancement of technology and the increasing availability of customer feedback through digital platforms have created new opportunities for evaluating service readiness across various industries. This study introduces a general-purpose framework utilizing data mining and natural language processing (NLP) techniques to assess service quality and operational efficiency. The framework is demonstrated through a case study focused on the hotel industry, where customer reviews from multiple web platforms are analyzed to extract insights into service readiness. Our approach involves systematically extracting aspect words from reviews, filtering relevant aspects, categorizing these aspects into predefined dimensions of service quality, and scoring their polarity to determine readiness levels. By leveraging these advanced techniques, we provide a comprehensive and data-driven assessment of hotel readiness, highlighting strengths and areas for improvement. The findings offer valuable insights for hotel management, enabling more informed decision-making to enhance service quality and customer satisfaction. Furthermore, the versatility of the proposed framework allows its application to other sectors such as retail, healthcare, and military, demonstrating its broad utility in assessing service readiness. This research underscores the potential of integrating data mining and NLP techniques to transform traditional service evaluation methods and drive improvements in various industries.**

Keywords—Data Mining, Natural Language Processing, Service Readiness, Hotel Readiness

I. INTRODUCTION

Understanding and meeting customer expectations is a complex challenge faced by various industries. Traditional methods of evaluating service quality, such as surveys and questionnaires, often fall short in capturing the dynamic and multifaceted nature of user experiences. The proliferation of digital platforms and the consequent explosion of customer feedback have necessitated the development of more sophisticated evaluation techniques.

This study introduces a general-purpose framework that leverages data mining and natural language processing (NLP) techniques to assess service readiness. This framework is designed to systematically analyze customer reviews from various web platforms, providing a nuanced understanding of service quality, user satisfaction, and operational efficiency. The versatility of this approach allows its application across diverse fields, including hospitality, retail, healthcare, and military.

To demonstrate the efficacy of this framework, we present a case study focused on the hotel industry. Hotels are an integral part of the hospitality sector, where service quality and customer satisfaction are paramount. By analyzing customer reviews, we aim to extract valuable insights into hotel readiness, encompassing various aspects such as service quality, cleanliness, amenities, and staff behavior.

The methodology involves extracting aspect words from reviews, filtering relevant aspects, categorizing these aspects into predefined dimensions, and scoring their polarity to determine readiness levels. This comprehensive approach provides hotel management with actionable data, enabling informed decision-making to enhance service standards and customer experiences.

The findings from the hotel industry case study highlight the framework's ability to identify strengths and areas for improvement, offering a detailed assessment of hotel readiness. Moreover, the general-purpose nature of the framework demonstrates its applicability to other sectors where similar challenges of service evaluation and improvement exist.

In summary, this research underscores the potential of integrating data mining and NLP techniques to revolutionize traditional service evaluation methods. By harnessing the power of big data, this framework offers a scalable and effective solution for assessing service readiness, driving enhancements in quality and customer satisfaction across various industries.

This paper is organized into the following sections: Literature Review, providing a comprehensive review of existing methodologies and their applications in various industries; Methodology, detailing the proposed general-purpose framework including data mining and NLP techniques; Case Study: Evaluating Hotel Readiness, applying the framework to assess hotel readiness with data collection, analysis, and findings; Results and Discussion, presenting and discussing the case study results, highlighting key insights and implications; Applicability to Other Fields, exploring potential applications of the framework in industries such as retail, healthcare, and military; and Conclusion, summarizing the research findings, contributions, and future research directions.

II. LITERATURE REVIEW

The integration of data mining and natural language processing (NLP) techniques in service readiness evaluation

has proven to be highly effective across various fields. The systematic reviews and studies discussed highlight the significant impact of these methodologies on enhancing customer satisfaction, service quality, and operational efficiency.

From service management to human resource management, and from the hotel industry to healthcare, text mining and NLP have been instrumental in extracting valuable insights from large datasets. The work of Kumar, Kar, and Ilavarasan [1] emphasizes the growing relevance of text mining in managing big data across digital platforms, providing a comprehensive research agenda for future studies. Piwowar-Sulej et al. [2] further illustrate the dynamic trends in HRM, underscoring the importance of text mining in identifying non-obvious research trends.

The application of these technologies in the hotel industry, as reviewed by Yu and Hu [3], showcases their utility in enhancing marketing strategies and organizational performance. Similarly, the systematic review by Beller et al. [4] on the use of text mining for study identification in systematic reviews highlights the efficiency and workload reduction benefits of these methods.

Moreover, the applications of text mining extend to educational data, financial sectors, and healthcare, as demonstrated by the studies of Bansal et al. [5], Bach et al. [6], and Liu and Zhang [7], respectively. These diverse applications emphasize the versatility and broad impact of text mining and NLP technologies.

Finally, the reviews by Wang et al. [8] and Chen et al. [9] provide a comprehensive overview of NLP's role in enhancing bibliometric analysis and its evolving status in various research domains. The practical challenges and solutions in applying text mining for systematic reviews, as discussed by Clark et al. [10], underline the continuous need for innovation and development in this field.

Collectively, these studies underscore the critical role of text mining and NLP in advancing service readiness evaluation and improving decision-making processes across different industries. The ongoing research and development in these technologies promise even greater enhancements in the future.

III. Methodology

This section provides a detailed description of the framework, which leverages data mining and natural language processing (NLP) techniques to evaluate service readiness across various fields. The methodology involves the following steps:

Data Collection: Customer reviews are gathered from multiple online platforms, ensuring a diverse and comprehensive dataset. This data forms the basis for the subsequent analysis.

Data Preprocessing: The collected data undergoes preprocessing, including cleaning (removing irrelevant information), tokenization (splitting text into words), and normalization (standardizing text to a uniform format).

Aspect Extraction: Key aspects of service quality are extracted from the reviews using NLP techniques. This involves identifying and categorizing relevant words and phrases that indicate different dimensions of service readiness, such as service quality, cleanliness, amenities, and staff behavior.

Sentiment Analysis: Sentiment analysis is performed on the extracted aspects to determine the polarity (positive, negative, or neutral) of the customer feedback. This step involves using advanced machine learning algorithms to analyze the sentiment expressed in the reviews.

Aspect Categorization: The aspects and their associated sentiments are categorized into predefined dimensions of service quality. This structured approach helps in systematically evaluating different facets of service readiness.

Scoring and Analysis: Each aspect is scored based on the sentiment analysis results. These scores provide a quantitative measure of service readiness, highlighting areas of strength and those requiring improvement.

Visualization and Reporting: The results are visualized using charts and graphs, making it easier to interpret and communicate the findings. A comprehensive report is generated to present the insights and recommendations.

IV. Case Study

To demonstrate the efficacy of the proposed framework, we apply it to assess hotel readiness using customer reviews from a selected dataset. The case study involves the following steps:

Data Collection: Reviews for 1000 hotels are sourced from Kaggle's Datafiniti Business Database [11]. The dataset includes reviews, ratings, and additional contextual information.

Data Preprocessing: The reviews are cleaned and standardized to remove noise and irrelevant data. This process ensures the accuracy of the subsequent analysis.

Aspect Extraction: NLP techniques are used to extract key aspects related to hotel service quality from the reviews. These aspects include service quality, amenities, cleanliness, staff behavior, and overall customer satisfaction.

Sentiment Analysis: Sentiment analysis is conducted on the extracted aspects to determine the overall sentiment expressed by customers. This step helps in identifying positive, negative, and neutral feedback.

Aspect Categorization: The extracted aspects are categorized into predefined dimensions of hotel service quality. This categorization provides a structured framework for assessing different facets of hotel readiness.

Scoring and Analysis: Each aspect is scored based on the sentiment analysis results, providing a quantitative measure of hotel readiness. These scores highlight the hotel's strengths and areas for improvement.

Visualization and Reporting: The results are visualized using charts and graphs, facilitating easy interpretation. A detailed report is generated to present the findings and actionable insights for hotel management.

V. Case Study: Evaluating Hotel Readiness

Context and Importance of the Hotel Industry

The hotel industry is a critical component of the hospitality sector, playing a pivotal role in providing accommodations and various services to travelers. The industry's success hinges

on delivering high-quality service and ensuring customer satisfaction, which are directly influenced by various factors such as cleanliness, amenities, staff behavior, and overall service quality. In an increasingly competitive market, understanding and improving these aspects are crucial for maintaining a positive reputation and achieving operational success.

Data Collection

Reviews for 1000 hotels were sourced from Kaggle's Datafiniti Business Database [11]. The dataset includes reviews, ratings, and additional contextual information, providing a rich source of data for analysis.

Data Preprocessing

The reviews were cleaned and standardized to remove noise and irrelevant data. This process involved removing special characters, stop words, and normalizing the text to ensure accuracy in the subsequent analysis.

Aspect Extraction

NLP techniques were used to extract key aspects related to hotel service quality from the reviews. These aspects included service quality, amenities, cleanliness, staff behavior, and overall customer satisfaction. The extraction process involved identifying relevant phrases and categorizing them into predefined dimensions of service quality.

Sentiment Analysis

Sentiment analysis was conducted on the extracted aspects to determine the overall sentiment expressed by customers. This step helped in identifying positive, negative, and neutral feedback, providing insights into customer perceptions and experiences.

Aspect Categorization

The extracted aspects were categorized into predefined dimensions of hotel service quality. This categorization provided a structured framework for assessing different facets of hotel readiness, allowing for a comprehensive evaluation of service quality.

Scoring and Analysis

Each aspect was scored based on the sentiment analysis results, providing a quantitative measure of hotel readiness. These scores highlighted the hotel's strengths and areas for improvement, offering actionable insights for hotel management.

Visualization and Reporting

The results were visualized using charts and graphs, facilitating easy interpretation. A detailed report was generated to present the findings and actionable insights for hotel management, enabling them to make informed decisions to enhance service standards and customer experiences.

VI. Results

Our findings comprise an assessment of the readiness level across various dimensions, including amenities, general aspects, location, meals, rooms, service, and staff. This comprehensive evaluation serves to offer hotel owners and management a nuanced understanding of service quality. By delving into each aspect, it illuminates both strengths and weaknesses, providing a detailed image of the hotel's performance and areas that may require improvement.

TABLE I. The Readiness for each aspect

Aspect	Compound score	Readiness Score	Readiness Level
amenities	0.23	3.46	Expected level of readiness
general	0.11	3.21	Not ready needs some work
location	0.18	3.35	Not ready needs some work
meal	0.28	3.75	Ready but needs few improvement
room	0.23	3.45	Expected level of readiness
service	0.32	3.64	Ready but needs few improvement
staff	-0.31	2.37	Not ready needs a lot of work

During the comprehensive readiness analysis, diverse aspects were scrutinized, each assigned a numerical readiness score and a corresponding readiness level. Amenities and rooms exhibited readiness scores of 3.46 and 3.45, respectively, placing them at an expected level of readiness. The meal aspect secured a score of 3.75, signifying readiness but with acknowledgement of areas that warrant improvement. Service closely followed with a score of 3.65, also falling within the ready category but indicating room for enhancement. Conversely, the general aspect and location scored 3.21 and 3.35, respectively, both falling short of readiness, necessitating attention and improvement. The staff aspect, with a notably lower score of 2.37, revealed a significant deficiency, categorizing it as unprepared and in dire need of substantial improvement. In essence, the analysis underscores a nuanced state of readiness, with some areas meeting expectations, others poised for improvement, and a critical need for targeted interventions, particularly in the staff domain, to elevate the overall readiness levels effectively.

VII. Applicability to Other Fields

The proposed framework's versatility allows its application to various industries beyond hospitality, including retail, healthcare, and military. For instance:

Retail: In the retail sector, customer reviews can be analyzed to assess product quality, customer service, and overall shopping experience. This helps retailers identify areas for improvement and enhance customer satisfaction.

Healthcare: In healthcare, patient feedback can be analyzed to evaluate service quality, treatment effectiveness, and patient satisfaction. This aids healthcare providers in improving patient care and operational efficiency.

Military: In the military, feedback from personnel can be analyzed to assess unit readiness, operational efficiency, and morale. This helps in identifying areas that require attention and improving overall performance.

By leveraging data mining and NLP techniques, organizations in these fields can gain valuable insights from customer feedback, enabling them to make informed decisions and enhance service readiness.

VIII. Conclusion

This research demonstrates the effectiveness of integrating data mining and NLP techniques in evaluating service readiness across various industries. The proposed framework provides a comprehensive and systematic approach to analyzing customer feedback, offering valuable insights into service quality and operational efficiency. The case study on hotel readiness illustrates the practical application of the framework, highlighting its ability to identify strengths and areas for improvement.

The findings contribute to the existing body of knowledge by showcasing the versatility and impact of text mining and NLP technologies. Future research can explore the integration of more advanced algorithms and the application of this framework to other domains. Continuous innovation and development in these technologies promise to further enhance service readiness evaluation, driving improvements in quality and customer satisfaction across different sectors.

References

[1] S. Kumar, A. K. Kar, and P. V. Ilavarasan, "Applications of text mining in services management: A systematic literature review," International Journal of Information Management Data Insights, vol. 1, no. 1, pp. 100008, Apr. 2021.

[2] K. Piwowar-Sulej, R. B. Cooper, and M. P. Sanders, "Research trends in human resource management. A text-mining-based literature review," Emerald Insight, vol. 27, pp. 123-136, Jul. 2021.

[3] T.-Y. Yu and D.-C. Hu, "Marketing management in the hotel industry: A systematic literature review by using text mining," Sustainability, vol. 14, no. 4, pp. 2344, Feb. 2022.

[4] M. Beller, J. Tatsioni, and L. P. Ioannidis, "Using text mining for study identification in systematic reviews: A systematic review of current approaches," Systematic Reviews, vol. 3, pp. 65, Dec. 2014.

[5] B. Bansal, A. Singh, and N. Sharma, "Sentiment analysis and opinion mining on educational data: A survey," International Journal of Educational Technology in Higher Education, vol. 19, no. 1, pp. 25-40, 2022.

[6] S. Bach, J. L. Dowdell, and P. Y. Johnson, "Comprehensive review of text-mining applications in finance," Journal of Financial Data Science, vol. 3, no. 1, pp. 23-45, 2019.

[7] Y. Liu and X. Zhang, "Applications of text mining in healthcare: A systematic review," Journal of Healthcare Informatics Research, vol. 5, no. 1, pp. 1-17, 2021.

[8] Y. Wang, Q. Lu, and W. Huang, "NLP-enhanced bibliometrics: An overview of applications and future directions," Journal of Informetrics, vol. 15, no. 2, pp. 100109, 2021.

[9] H. Chen, L. Xu, and Y. Li, "Vision, status, and research topics of natural language processing," Information Processing & Management, vol. 59, no. 3, pp. 102453, 2022.

[10] R. Clark, S. Haddow, and A. Ball, "Challenges in applying text mining for systematic reviews," Journal of Information Science, vol. 47, no. 1, pp. 78-95, 2021.

[11] "Hotel Reviews," Kaggle, Datafiniti's Business Database, [Online]. Available: https://www.kaggle.com/datafiniti/hotel-reviews.

Application of Big Data Analytics in Airside Operations To Improve Wildlife Hazard Management For The Case of Malaysia Airports

1st Tisiyanah Abd Baki
Airport Operations Division
Malaysia Airports (Sepang) Sdn Bhd
KLIA, Sepang, Selangor, Malaysia.
tisiyanah@malaysiaairports.com.my

2nd Badriyah Noordin
Aerodrome Safety & Standards
Malaysia Airports Holdings Berhad
KLIA, Sepang, Selangor, Malaysia.
badriyah@malaysiaairports.com.my

3rd Norliza Mohamed
Faculty of Artificial Intelligence
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
norlizam.kl@utm.my

4th Sevia Mahdaliza Idrus
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Johor Bahru, Johor, Malaysia
sevia@utm.my

5th Nor Azlina Mohd Isa
Technical Services Division
Malaysia Airports Holdings Berhad
KLIA, Sepang, Selangor, Malaysia.
azlinaisa@malaysiaairports.com.my

6th Syamil Azmin
Airport Operations Division
Malaysia Airports Sdn.Bhd
KLIA, Sepang, Selangor, Malaysia.
syamil@malaysiaairports.com.my

***Abstract*— Air travel provides vital links for the burgeoning global tourism industry. It is estimated that over half (58%) of all international tourists travel by air, so the aviation and tourism industries depend on each other for sustainable growth. In the aviation industry, big data plays a transformative role across various facets of aviation operations to ensure seamless traveling. Moving together with other countries towards Big Data Analytics, Malaysia Airports Holdings Berhad (MAHB) without exception also embarked in this journey. With many layers of operations, each layer has its own way of integrating with Big Data Analytics. For instance, by using web-based application, IoT, radar over fiber and many more, many system like Airport Collaborative Decision Making (A-CDM), MAVCOM Bus Performance Report, the biometric facial recognition system called EZPaz, self-bag drop facility called EZBagz, Digital Integrated Airside Services System (DIASS), the Foreign Object Debris Detection System (FODDS), Airside Wildlife Safety Data, the Airside Closure Coordination Platform, Customer Real-Time Feedback & Survey System (CRFSS) at the terminal and Operational Information System, Electronic Centralized Documents Management System (eCDMS), and many other operational systems are available at KL International Airport (KLIA). Airside Wildlife Hazard Management at KL International Airport (KLIA) in the perspective of Big Data Analytics will be discussed in this research.**

Keywords*— *Airside Operations, Big Data Analytics, Wildlife Hazard Management

I. INTRODUCTION

Big Data refers to the large and complex data sets that traditional data processing methods cannot handle effectively [1]. It is characterized by its volume, velocity, and variety, encompassing both structured and unstructured data. The term gained prominence in the 1990s due to the explosion of data generated by the internet and digital devices, necessitating the development of new technologies for data storage, processing, and analysis. Innovations such as Google's MapReduce and the Hadoop platform enabled large-scale data processing, while advancements like NoSQL databases, in-memory computing, and cloud computing further propelled the field. Big Data has become crucial across various industries, driven by the rise of machine learning and artificial intelligence, which require extensive data to train and refine models.

Big data has become the strategic resource of an organization and a country. Big data analytics is at the heart of any business, management, decision making, and socio-economic development. Big data analytics is revolutionizing businesses and different industries around the world [2][3].

In aviation industry, airports have always been one of the biggest contributors of big data to the aviation ecosystem. With the abundance of data available, big data analytics can help transform the airports to smart ones. It explores the value of applying intelligent big data analytics in an airport from an operations perspective and strategic differentiation perspective as per shown in Figure 1.

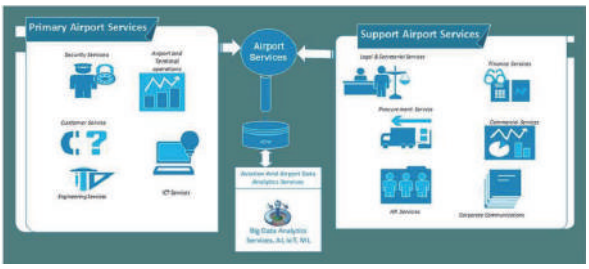


Fig. 1. Airport Services Model with Big Data analytics [3]

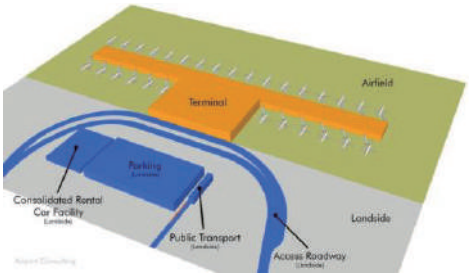


Fig. 2. Airport Operations Segments [4]

There are 3 main segments in airport operations namely landside, the terminal, and the airside/airfield [4]. Landside activities include the airport's passenger pickup and drop-off sites as well as local highways, railroads, and other intermodal modes of transportation. A successful wildlife management strategy must consider the surrounding environment of the airport in addition to the accessibility to the terminal area.

Check-in, bag drop, security, ground handling, airplane handling, luggage handling, parking, shopping areas, communications, and maintenance are all included in the terminal.

The airport's airfield, ramps, safety, and security are under the control of those who work in airside operations, sometimes known as airfield or airside operations. In a busy airport with thousands of aircraft movements in a day, safety of aircraft operations on airport runways, taxiways, and aprons is a main focus of this department. Managing many hazards related to aircraft operations is always a priority and one of it is wildlife hazards.

Wildlife management at airports involves strategies to minimize wildlife strikes particularly bird strikes, which pose safety risks to aircraft. This includes modifying habitats to deter wildlife, using harassment techniques to keep animals away from airside areas, and employing radar and bio-acoustic systems to monitor activity. Regular inspections and collaboration with wildlife experts help identify and address threats, ensuring safe aircraft operations and protecting both aviation personnel and wildlife [5].

As an airport operator, Malaysia Airports Holdings Berhad (MAHB) is implementing a comprehensive plan to leverage big data analytics to enhance its airport operations and passenger experiences [6]. This initiative involves collecting and analyzing extensive data from various sources such as flight schedules, passenger movements, and operational systems. By harnessing the power of big data analytics, MAHB aims to optimize resource allocation, streamline processes, and improve decision-making across its airports. Predictive analytics will enable proactive maintenance of airport infrastructure, minimizing downtime and enhancing reliability. Real-time data insights will be used to manage passenger flow effectively, leading to smoother and more efficient airport operations. Furthermore, MAHB plans to integrate big data analytics with advanced technologies like artificial intelligence (AI) and the Internet of Things (IoT) to create a smart airport environment. This holistic approach to big data analytics underscores MAHB's commitment to innovation, efficiency, and providing a world-class passenger experience across its airport network.

This paper is organized as follows: First, provide an overview of previous research related to big data analytic in an airport. Next, application of big data analytics in managing wildlife hazards in an airport. This paper discusses a real business scenario on the journey and plan in managing wildlife hazards at KL International Airport. We conclude this paper by highlighting key observations and offering recommendations for future research.

II. BIG DATA AND SMART AIRPORT

Big data and smart airports are closely intertwined, with big data serving as the backbone of smart airport operations. By collecting and analyzing vast amounts of data from various sources, such as sensors, flight information systems, passenger check-ins, and security systems, airports can gain valuable insights into their operations [7]. This data-driven approach allows for predictive maintenance, efficient resource allocation, and enhanced real-time decision-making. Smart airports leverage these insights to optimize everything from passenger flow and baggage handling to energy management and security measures [8]. The integration of big data with advanced technologies like IoT, AI, and machine learning transforms airports into intelligent ecosystems that offer improved efficiency, safety, and a superior passenger experience.

A. Continuous Airport Vehicle Asset Tracking System (CAVATS)

The big data analytics approach in the Continuous Airport Vehicle Asset Tracking System (CAVATS) at Dulles Airport involves using advanced data collection and analysis techniques to optimize the management and operation of airport vehicles [9]. Here are the key components of this approach:

1) **Data Collection**: CAVATS collects data from various sources including GPS tracking devices installed on vehicles, RFID tags, and sensors that monitor vehicle performance and environmental conditions.
2) **Real-Time Monitoring**: The system provides real-time tracking of all airport vehicles, allowing operators to monitor their locations, movements, and status continuously. This ensures that vehicles are where they need to be and are operating efficiently.
3) **Predictive Analytics**: By analyzing historical data on vehicle usage, maintenance records, and operational patterns, predictive analytics can forecast future maintenance needs and potential breakdowns. This proactive approach helps in scheduling maintenance activities before issues arise, reducing downtime and extending the lifespan of the vehicles.
4) **Optimization Algorithms**: Big data analytics applies optimization algorithms to manage the fleet efficiently. This includes route optimization to reduce travel time and fuel consumption, and efficient allocation of vehicles based on demand patterns and operational needs.
5) **Performance Metrics**: The system tracks key performance metrics such as fuel efficiency, idle times, and utilization rates. Analyzing these metrics helps identify areas for improvement and implement strategies to enhance overall efficiency.
6) **Automated Alerts and Reporting**: CAVATS generates automated alerts for maintenance needs, unusual activity, or deviations from planned routes. It also produces detailed reports on vehicle performance and operational efficiency, providing valuable insights for decision-making.
7) **Integration with Other Systems**: The data collected and analyzed by CAVATS is integrated with other airport management systems, such as scheduling and logistics platforms, to provide a holistic view of airport operations and facilitate coordinated efforts.

By leveraging big data analytics, CAVATS at Dulles Airport enhances the efficiency, reliability, and safety of vehicle operations, contributing to smoother airport logistics and better overall performance.

B. Willdife/Bird Hazard Management at Flughafen Berlin Brandenburg Airport.



Fig. 3. Bird Radar and Deterrence Devices at Flughafen Berlin Brandenburg

The big data analytics approach in the bird deterrence system at Flughafen Berlin Brandenburg (FBB) airport involves leveraging advanced technologies and real-time data to enhance the effectiveness of bird control measures [10]. This approach includes several key components:

1) **Data Collection**: The system collects extensive data from the MAX® avian radar, which tracks bird movements across the airport. Additional data is gathered from bird deterrence devices and environmental sensors.
2) **Real-Time Monitoring**: Using the Mobile Viewer software, bird controllers monitor bird activity in real-time. The software provides a live background map that visualizes bird movements, updating every second to ensure full situational awareness.
3) **Predictive Analytics**: The collected data is analyzed to identify patterns and predict high-risk areas for bird activity. This helps in anticipating potential threats and deploying deterrence measures proactively.
4) **Automated Alerts and Logging**: The system logs observations and generates visual alerts when bird activity increases in predefined zones. This enables quick responses to emerging threats.
5) **Remote Activation**: Bird controllers can remotely activate deterrence devices positioned around the airfield from their vehicles using tablets. This allows for immediate action to scare away birds and mitigate risks.
6) **Integrated Response**: The integration of radar data with deterrence systems ensures a coordinated response to bird threats. Controllers can assess the situation using the tablet interface and deploy the appropriate measures based on real-time data and predictive insights.

By applying big data analytics, FBB airport's bird deterrence system enhances the ability to detect, track, and mitigate avian threats efficiently, ensuring safer airport operations.

III. MALAYSIA AIRPORT'S WILDLIFE HAZARDS MANAGEMENT AT KL INTERNATIONAL AIRPORT

For the past few years, MAHB has introduced various initiatives related to Big Data Analytics throughout its airports especially KL International Airport (KLIA). KLIA has introduced many initiatives airport operations such as Airport Collaborative Decision Making (A-CDM), MAVCOM Bus Performance Report, the biometric facial recognition system called EZPaz, self-bag drop facility called EZBagz, STARDesk, AIS3 [11][12], Digital Integrated Airside Services System (DIASS), the Foreign Object Debris Detection System (FODDS) [13], Airside Wildlife Safety Data, the Airside Closure Coordination Platform, Customer Real-Time Feedback & Survey System (CRFSS) at the terminal and Operational Information System, Electronic Centralized Documents Management System (eCDMS), and many more.

In this research work, we are focusing on Wildlife Hazard Management at KL International Airport.

A. The Compliance of Wildlife Hazard Management

According to Civil Aviation Guidance Material 1400 (CAGM – 1400) issued by the Civil Aviation Authority of Malaysia (CAAM) which provide guidance for the Wildlife Hazard Management [14] at the aerodrome as follows: -

a) assignment of personnel (roles and responsibilities of personnel in the programme);
b) a process to report, collect and record data on wildlife strikes and observed wildlife;
c) a wildlife safety risk assessment method and procedure, including annual reviews;
d) a process of habitat and land management both on the airport and in its vicinity in order to reduce the attractiveness of the area to birds/wildlife;
e) a process for the expelling, deterring and removing of wildlife, including lethal means where appropriate;
f) a process for liaison or coordinating with internal and external stakeholders to ensure the airport operator is aware of developments that may contribute to creating additional bird hazards in the infrastructure, vegetation, land use and activities in the airport vicinity; and
g) a Training programme for the personnel involved in Wildlife Hazard Management; and
h) Evaluating of wildlife control programme.

Applying big data analytics to manage wildlife hazards in airports accordance to the compliance [12] involves several strategic steps:

1) **Data Collection**: Gather extensive data from multiple sources such as radar systems, surveillance cameras, and wildlife tracking sensors to monitor wildlife activity around the airport.
2) **Integration of Diverse Data Sources:** Combine data from various sources like environmental conditions, wildlife sightings, historical strike records, and habitat characteristics to create a comprehensive dataset.
3) **Predictive Analytics**: Use machine learning algorithms to analyze historical data and predict potential wildlife movements and strike risks based on patterns and trends. This helps in anticipating and mitigating risks before they occur.
4) **Real-Time Monitoring**: Implement real-time monitoring systems that use big data analytics to continuously analyze incoming data, providing live updates on wildlife activity and potential threats.
5) **Automated Alerts and Responses**: Develop automated alert systems that notify relevant personnel about immediate wildlife hazards, allowing for quick and coordinated responses to mitigate risks.

6) **Behavioral Insights**: Analyze data to understand the behavior and migration patterns of different wildlife species, enabling the development of targeted management strategies.
7) **Risk Mapping**: Create detailed risk maps highlighting areas of the airport most vulnerable to wildlife incursions, informing focused management and mitigation efforts.
8) **Evaluation and Adaptation**: Regularly assess the effectiveness of implemented strategies using big data analytics, adapting and refining approaches based on data-driven insights to continuously improve wildlife hazard management.

By leveraging big data analytics, airports can enhance their ability to predict, monitor, and respond to wildlife hazards, ensuring safer and more efficient airport operations.

B. Wildlife Hazard Management in KL International Airport

Malaysia Airports Holdings Berhad (MAHB) manages wildlife hazards through a comprehensive and data-driven approach designed to minimize the risk of wildlife strikes, ensuring the safety and efficiency of airport operations [15]. Practices by MAHB as per TABLE I: -

TABLE I WILDLIFE HAZARD MANAGEMENT AT KL INTERNATIONAL AIRPORT

Strategic Elements	Practices in KL International Airport
Data Collection	1) Data collection from observation and finding during inspection 2) Data stored manually in google sheet
Integration of Diverse Data Sources	Data collection from pilot report, regulator's report
Predictive Analytics	Data analysis by using MS Excel & Google Sheet
Real-Time Monitoring	No Real-Time Monitoring Monitoring by Safety Officer during aerodrome and wildlife inspection
Automated Alerts and Responses	1) No Automated Alert and Response 2) Alert and responses through communication medium such as publishment in AIP (Aeronautical Information Publication) and NOTAM (Notice To Airmen)
Behavioral Insights	Data analysis by using MS Excel & Google Sheet
Risk Mapping	1) Risk mapping by using MS Excel, Google Sheet and manually plot in map. 2) Mitigation action such as deterrent method is conducted manually by airside personnel.
Evaluation and Adaptation	Assessment based on reported cases and observation.

Various type of improvement can be implemented based on Big Data Analytics approach such as:-

1) Integration data with regulator, airlines and other stakeholders.
2) Usage of radar system, surveillance camera and sensors.
3) Application of AI in managing the data such as risk mapping, and mitigation measures and its assessment.
4) Usage of real-time monitoring system and device.
5) Integration with other potential radar or system own by other stakeholders in the same industry.

CONCLUSION

From our observation, improvement to the current Wildlife Hazard Management should be considered to ensure this hazard can be mitigated timely and at the same time improve airside operations management through Big Data Analytics approach at KL International Airport and for all other airports in Malaysia. Further study on the reliability of the systems shall be conducted to meet the compliance and the operational requirement.

ACKNOWLEDGMENT

This work was supported by the Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS/1/2020/TK0/UTM/01/6) and by Universiti Teknologi Malaysia institutional grant vot. 08G49. The authors would like to thank Malaysia Airport(Sepang) Sdn Bhd and Malaysia Airport Holding Berhad in the field Master Research study.

REFERENCES

[1] https://www.linkedin.com/pulse/what-big-data-introduction-history-types-examples-jobs-ram-narayan
[2] Blayney, P. J., & Sun, Z. (2019). Using Excel and Excel VBA for Preliminary Analysis in Big Data Research. In Z. Sun, & Z. Sun (Ed.), Managerial Perspectives on Intelligent Big Data Analytics (pp. 110-136). Hershey, IGI Global. doi:DOI: 10.4018/978-1-5225-7277-0.ch007
[3] Desmond Narongou, Zhaohao Sun Applying Intelligent Big Data Analytics in a Smart Airport Business: Value,Adoption, and Challenges (2022)
[4] ATSB 2010 ATSB Transport Safety Report (Australian Transport Safety Bureau), 2011
[5] https://www.mjinc.com/news/blog?p=the-importance-of-wildlife-management-at-airports
[6] https://www.malaysiaairports.com.my/sites/corporate/files/2024-05/MAHB%20AR2023%20%5BSingle%5D-compressed.pdf
[7] https://nexusintegra.io/smart-airports-the-digital-transformation-of-airports/
[8] https://www.researchgate.net/figure/A-model-for-enhancing-airport-services-with-big-data-analytics_fig3_368802510
[9] Matthew H. Jones, Stephen D. Patek, and Barbara E. Tawny, Continuous Airport Vehicle Asset Tracking System (2004)
[10] https://www.robinradar.com/bird-radar-flughafen-berlin-brandenburg-airport
[11] https://www.malaysiaairports.com.my/media-centre/news/airport-collaborative-decision-making-increase-operational-efficiency-klia
[12] https://www.malaysiaairports.com.my/sites/corporate/files/2024-05/MAHB%20AR2023%20%5BSingle%5D-compressed.pdf
[13] G. Fizza1, S.M. Idrus2*, F. Iqbal3, W.H.W. Hassan4, R. Ahmad5, Z. Meerah6, A. Kanno7, T. Kawanish8, N. Shibagaki9, K. Kashima10, Review on Foreign Object Debris Detection Technologies and Advancement for Airport Safety and Surveillance, (2021)
[14] https://www.caam.gov.my/wp-content/uploads/2021/12/CAGM-1400-Wildlife-Hazard-Management-Programme.pdf
[15] https://flightsafety.org/asw-article/it-for-wildlife-management/
[16] https://mahb.listedcompany.com/newsroom/MAHB_Annual_Report_2022_-_Part_2.pdf
[17] https://aviationbenefits.org/social-development/tourism-enabler/
[18] Thun, Sylvi; Kamsvåg, Pål F; Kløve, Birgit; Seim, Eva A; Torvatn, Hans Y, Industry 4.0: Whose Revolution? The Digitalization of Manufacturing Work Processes (2019)
[19] Pech, Martin; Vrchota, Jaroslav, The Product Customization Process in Relation to Industry 4.0 and Digitalization (2021)

Performance Analysis of Ultrasonic-Based Localization: Impact of Receiver Spacing and Array Distances

1st JaeJun Lim
Department of Electronic Engineering
Pusan National University
Busan, Republic of Korea
limlim258@pusan.ac.kr

2nd Jaehan Joo
Department of Electronic Engineering
Pusan National University
Busan, Republic of Korea
jhjoo2018@pusan.ac.kr

3rd Hunyoul Lee
Department of Electronic Engineering
Pusan National University
Busan, Republic of Korea
yeal96011@pusan.ac.kr

4th Dongjhin Noh
Department of Electronic Engineering
Pusan National University
Busan, Republic of Korea
rst007@pusan.ac.kr

5th Suk Chan Kim*
Department of Electronic Engineering
Pusan National University
Busan, Republic of Korea
sckim@pusan.ac.kr

***Abstract*—In this paper, we investigate the impact of transmitter and receiver placement on Active ultrasonic localization. We analyze the variation in localization performance based on the distance between transducers and receiver arrays, the spacing between receivers, and the arrangement of objects. Through simulation, we demonstrate and analyze the effect of transmitter and receiver placement on position estimation and Angle of Arrival (AOA) estimation when employing the Threshold Crossing algorithm.**

***Index Terms*—Ultrasonic, Threshold Crossing, Outdoor Localization,Time of Arrival,Angle of Arrival**

I. Introduction

Recent studies on location-based systems, crucial for autonomous vehicles and robotics, have highlighted the effectiveness of ultrasonic signals in enhancing sensor performance, particularly when traditional sensors like LiDAR and ToF cameras struggle with close-proximity object detection. [1] This paper focuses on optimizing transmitter-receiver arrangements in active localization systems, examining how their placement and spacing under the far-field assumption impact TDOA and AOA estimation accuracy, thereby affecting system performance. [2]

II. Related works

This study use an AOA estimation method using TDOA under the far-field assumption, employing the Threshold Crossing algorithm. This algorithm determines TOA and, subsequently, the distance between the object and sensor by analyzing the envelope of ultrasonic signals. [3] TDOA, crucial for AOA estimation, is derived by repeatedly measuring the rising time when the signal exceeds a predefined threshold. This measurement is averaged across all receivers to refine TDOA estimation.

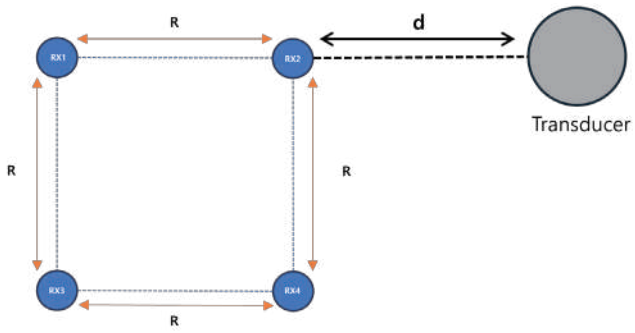


Fig. 1. Transducer and receiver arrangement

III. Method

In this section, we explain how the localization performance varies with the distance between the receiver array and the transducer from the Time of Arrival (TOA) perspective.

A. Distance Between Transducer and Receiver Array

The distance between the transducer and the receiver array affects the size of the sensor. A larger sensor size can lead to decreased portability and compatibility due to physical constraints. Therefore, as illustrated in Figure 1, we arrange the receiver array and the transducer to derive the overlapping section of the TOA of the ringing signal and the TOA of the reflected wave based on the distance d and R which represent interval of each receivers.

To account for the ringing effect, the distance between the transducer and each receiver is represented as follows. If the duration of the ultrasonic signal is denoted by T, to prevent

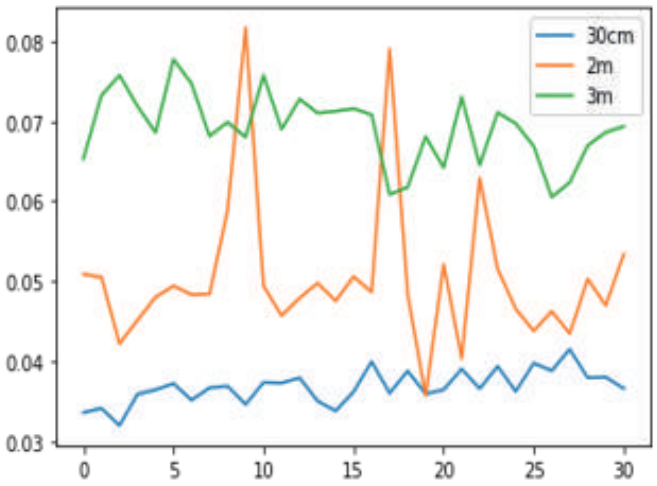


Fig. 2. RMSE by Transducer-Receiver Distance

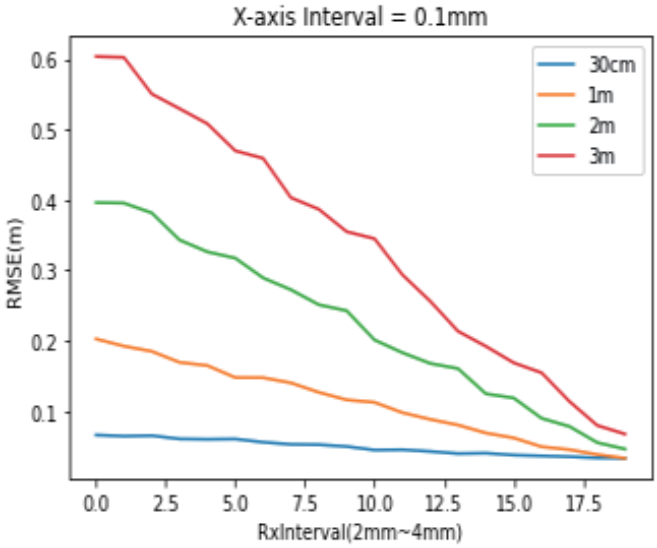


Fig. 3. RMSE by Receiver Interval

overlap between the reflected wave and the Ringing Effect, the following relationship must be satisfied:

$$T + \frac{d_i}{v_{sig}} < TOA_i \tag{1}$$

Through the above equation, we can see that the shorter the distance between the receiver array and the transducer, the stronger the ringing effect.

B. Receiver intreval

In Far-Field, AOA is approximated as follows:

$$\theta^{'} = arccos(\frac{v(\tau + n_\tau)}{R}) \tag{2}$$

It can be seen that the larger R becomes, the more robust it is against TDOA estimation errors.

IV. Result

This section deals with simulation. The verification of the algorithm was performed through simulation. Subsection A deals with the simulation configuration environment, and subsection B deals with the simulation results.

The transmitted signal is an ultrasonic signal and has a center frequency of 40kHz. The channel was created considering pathloss and awgn. The sampling rate is 2MHz. Additionally, the ringing effect was considered. [3]

The simulation was conducted by increasing the distance between the transducer and the receiver array from 0.005 m to 0.05 m at intervals of 0.0015 m, then placing objects at random angles, and repeating the simulation 50 times at one distance. The hoizontal and vertical angles were randomly placed between -40 and 40 degrees. In addition, in order to check the position estimation error depending on the distance between the object and the sensor, an experiment was conducted with the distance between the object and the sensor set to 30cm, 2m, and 3m.

Additionally, in order to analyze the effect of the receiver interval on location estimation, a simulation was performed after fixing the distance between the receiver and the transducer to 0.005m. After increasing the receiver interval from 0.002m to 0.004m at 0.0001m intervals, the object was placed at a random angle to estimate its position. In order to check the position estimation tendency according to the distance between the object and the sensor, Simulations were performed by increasing the distance at intervals of 30cm, 1m, 2m, and 3m.

V. Conclusion

In this paper, we analyzed the effect of the distance between the transducer and the receiver array and the distance between the receivers on localization performance. The effect of the distance between the transducer and the receiver array on location estimation performance did not have a significant effect unless the ultrasonic signal was very long. However, it was confirmed that receiver spacing has a significant impact on location estimation performance.

Acknowledgment

This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program) (00144500, Development of 3D ultrasonic sensor technology for vehicles based on meta structure) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea)

References

[1] Dorrington, A & Payne, Andrew & Cree, Michael. (2010). An evaluation of time-of-flight range cameras for close range metrology applications. 38.

[2] Luo X, Wang H, Yan S, Liu J, Zhong Y, Lan R. Ultrasonic localization method based on receiver array optimization schemes. International Journal of Distributed Sensor Networks. 2018;14(11).

[3] N. Rogel, D. Raphaeli and O. Bialer, "Time of Arrival and Angle of Arrival Estimation Algorithm in Dense Multipath," in IEEE Transactions on Signal Processing, vol. 69, pp. 5907-5919, 2021

Enhancing IoT connectivity: Implementation of Zigbee Direct for interoperability in multiprotocol network

1st Seonghyeon Park
Department of Electronic Engineering
Pusan National University
Busan, Republic of Korea
shpark24@pusan.ac.kr

2nd Jaehan Joo
Department of Electronic Engineering
Pusan National University
Busan, Republic of Korea
jhjoo2018@pusan.ac.kr

3rd Minseok Kim
Department of Electronic Engineering
Pusan National University
Busan, Republic of Korea
alstjr2650@pusan.ac.kr

4th Sukchan Kim*
Department of Electronic Engineering
Pusan National University
Busan, Republic of Korea
sckim@pusan.ac.kr

***Abstract*—In this Work-in-process paper, we investigate Zigbee Direct, wireless communication technology between Zigbee and Bluetooth Low Energy(BLE). Through simulation, we demonstrate that Zigbee and BLE devices can communicate through a Zigbee Direct device. Additionally, we show that a BLE device can monitor the status of a Zigbee Direct device.**

***Index Terms*—Zigbee Direct, multiprotocol, Zigbee, Bluetooth Low Energy, Generic Attribute profile**

I. Introduction

As the IoT market continues its rapid expansion, the demand for adaptable wireless communication technologies capable of functioning in diverse situations among IoT devices has become critically important. However, a significant challenge has emerged: interoperability issues among devices utilizing different communication protocols [1]. To handle this challenge, various multiprotocol solutions have been developed. We will talk about one of them, Zigbee Direct. In Section II, we introduce the Zigbee Direct with some descriptions of the terminology. In Section III and IV, we demonstrate the establishment of a network configuration in which Zigbee and BLE devices can directly control Zigbee Direct device with simulation.

II. Zigbee Direct

Zigbee Direct represents a novel solution designed to bridge Zigbee and BLE, facilitating seamless communication between devices operating on these protocols. Zigbee Direct aims to enable encrypted and authenticated access to Zigbee devices and networks from BLE devices. This enables smart devices and hubs to commission and control devices on a Zigbee network [2]. This network contains two types of devices, each serving unique functions and roles: the Zigbee Direct Device and the Zigbee Virtual Device.

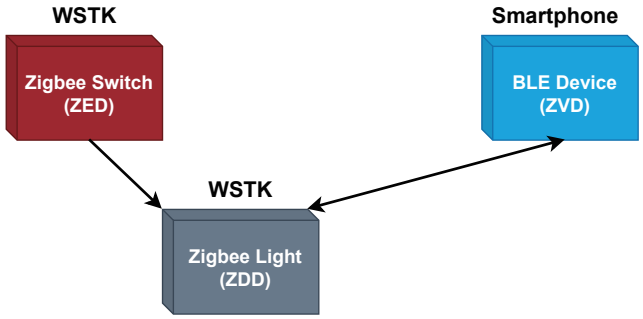


Fig. 1. Zigbee, BLE Device arrangement

A. Zigbee Direct Device (ZDD)

A Zigbee device including end device, router, coordinator, trust center that can run Zigbee and BLE stacks concurrently. It also facilitates data exchange between BLE devices and the Zigbee network [2].

B. Zigbee Virtual Device (ZVD)

Contrary to terminology, it is a BLE-based device and can operate within a Zigbee network as if it were a Zigbee device itself or perform certain roles, such as a trust center or a commissioning tool through a ZDD.

C. Zigbee Direct services

- Zigbee Direct Commissioning Service : A mandatory BLE Generic Attribute profile(GATT) service containing several characteristics that enables ZVD to perform operations such as configuration and management of the network through ZDD.

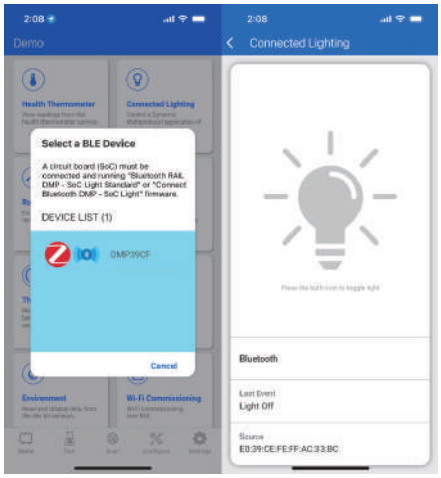


Fig. 2. Building connection with ZVD

- Zigbee Direct Tunnel Service : A BLE GATT optional service by ZDD that allows ZVD to exchange data with Zigbee network for performing certain role. Simply put, it provides a trusted link between the ZVD and ZDD, which cannot be utilized when the ZDD acts as an end device within the Zigbee network.

III. Simulation Setup

To form Zigbee Direct network, we used two EFR32MG21 WSTKs as Zigbee devices and smartphone as BLE device, shown in Fig. 1. Network configuration was managed with Simplicity Studio v5, and the EFR Connect application on the smartphone allowed for monitoring and controlling the ZDD's status, as depicted in Fig. 2. We configured the Zigbee Direct device to serve as the coordinator for the Zigbee network and as the scanner for the BLE network. The WSTKs were closely positioned with an RSSI of around -40dBm, while the smartphone was positioned at a moderate distance with an RSSI of about -60dBm.

IV. Simulation Result

We toggled the LED on the ZDD through the ZVD, and the process log in Simplicity Studio is shown in Fig. 3. Additionally, the LED was toggled using a standard Zigbee end device, a switch, and the changing state of the ZDD could be monitored from the ZVD. Fig. 4 shows that ZDD reports its status to the ZVD.

V. Conclusion

We have configured communication between Zigbee and BLE devices through Zigbee Direct, performing functions necessary for IoT systems. Unfortunately, the Zigbee Direct stack in EFR Connect application is not fully completed to enable the ZVD to perform all Zigbee Direct services, so we conducted the Tunnel service and basic commissioning services. Moving forward, the development of multiprotocols like Zigbee Direct will enable us to resolve interoperability issues between protocols.

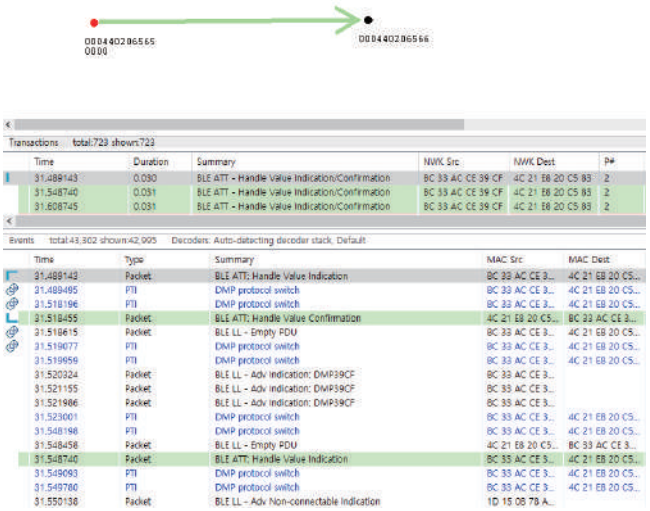


Fig. 3. Process logs of BLE toggle

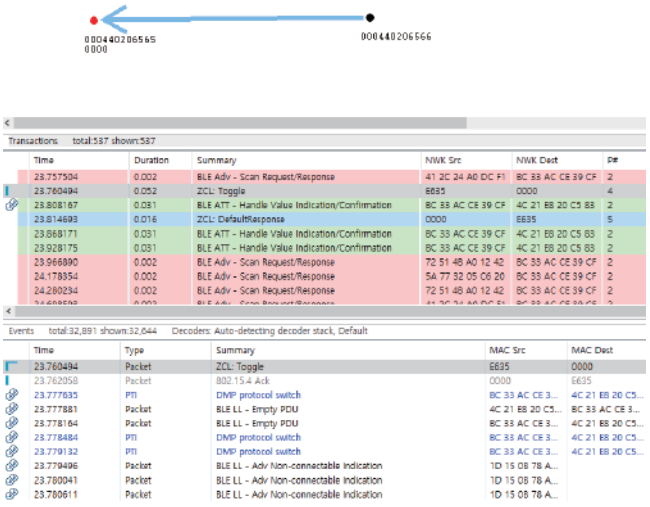


Fig. 4. Process logs of Zigbee toggle and BLE ATT update

Acknowledgment

This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program) (00144500, Development of 3D ultrasonic sensor technology for vehicles based on meta structure) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea) and also financially supported by Korea Ministry of Land, Infrastructure and Transport(MOLIT) as 「Innovative Talent Education Program for Smart City」.

References

[1] Polastre, J.; Hui, J.; Levis, P.; Zhao, J.; Culler, D.; Shenker, S.; Stoica, I. A Unifying Link Abstraction for Wireless Sensor Networks. In Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems, San Diego, CA, USA, 2–4 November 2005; Association for Computing Machinery: New York, NY, USA, 2005; pp. 76–89.

[2] Connectivity Standards Alliance, "Zigbee Direct Specification," 1.0, Document 20-27688-037, Nov. 30, 2022.

Covertness Analysis for the Covert Communications in Multi-antenna Two-way Relay Systems

Zhilin Fu*, Jihwan Moon†, *Member*, *IEEE*, Sangwon Hwang*‡, Sangmin Kim*
Di Zhang§, *Senior Member*, *IEEE* and Inkyu Lee*, *Fellow*, *IEEE*
*School of Electrical Engineering, Korea University, Seoul, South Korea
†Department of Mobile Convergence Engineering, Hanbat National University, Daejeon, South Korea
‡Mobile Experience Business, Samsung Electronics Co., Ltd., Suwon, South Korea
§School of Information Engineering, Zhengzhou University, Zhengzhou, China,
Email:*{celynnfu,tkddnjs3510,smgeem,inkyu}@korea.ac.kr, †anschino@staff.hanbat.ac.kr, §dr.di.zhang@ieee.org

***Abstract*—In this paper, the covert communications in multi-antenna two-way relay systems is studied. While the relay primarily assists in forwarding public messages between two source nodes, it also attempts to deliver covert information without being detected by an illegitimate node, who is often referred to as a Warden. The detection error probability (DEP) at the warden is a key metric for determining covertness, which is also essential in the subsequent design of covert communication strategies. We derive the closed-form solution of DEP and analyze the DEP performance in different scenarios.**

***Index Terms*—covert communications, multi-antenna two-way relay**

I. Introduction

The next-generation wireless communication systems provided outstanding network connectivity and lots of innovative applications [1]. Along with this, the communication security attacks have improved accordingly [2]. Traditional information security techniques, such as cryptography, have limitations in responding to security attacks [3]. Especially, some occasions require even stronger security, such as military operations, financial transactions, and the Internet of Things (IoT), where the very existence of communications links must be hidden from any external entities [4]. This calls for covert communications or low-probability-of-detection communications [5].

Considering the vast number of devices and the complex communication environments may disable the direct transmission link between two nodes, covert communication in the relay-assisted transmission has been explored [6]–[9]. However, most previous studies focused on single-antenna systems [6], [7]. For multi-antenna systems, the expression of DEP becomes more difficult to derive due to higher dimensions. [8] derived the DEP expression for the multi-antenna one-way decode-and-forward (DF) relay systems. [9] considered a two-way communication protocol with an intelligent reflecting surface system, and derived an approximate DEP at the warden based on the central limit theorem, and analyzed the effect of the system parameters on the covert link.

The multi-antenna two-way relay system in which the reception or forwarding of multiple source messages is simultaneously processed has received attention for its potential to double the spectral efficiency and high degree of freedom [10]. The first priority in investigating how such an integrated technology affects covert performance is the DEP analysis. In this paper, we analyze the covertness performance for the covert communications in a two-way amplify-and-forward (AF) multi-antenna relay system. The expression for DEP at the warden is derived, and its accuracy is verified through simulation experiments.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2021R1I1A3050126).

II. System Model

We consider a two-way relay network that consists of four nodes: two source nodes S_1 and S_2, an AF two-way relay R with M antennas, and a warden node W. We assume S_1 and S_2 simultaneously transmit their public message x_1 and x_2 to each other only through R. Meanwhile, the relay also tries to deliver covert messages while avoiding being detected by W . $\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_\mathrm{w}$ denote the Rayleigh distributed channel coefficients for S_1-to-R, S_2-to-R, and R-to-W, respectively. The channel state information (CSI) of $\mathbf{h}_1$, $\mathbf{h}_2$ and the channel distribution information (CDI) of $\mathbf{h}_\mathrm{w}$ is assumed be available. Each channel satisfies the channel reciprocity due to the time division duplex (TDD) mode [10].

During the first time slot, R received the signal from S_1 and S_2 which is denoted as $\mathbf{y}_\mathrm{r}$. In the second time slot, R opportunistically sends its own covert messages to S_1 and S_2 while broadcasting the amplified received signal. Denoting p_r as the power allocation ratio, $p_\mathrm{r}P_\mathrm{t}$ $(0 < p_\mathrm{r} \leq 1)$ is allocated to public message transmission while the remaining $(1-p_\mathrm{r})P_\mathrm{t}$ is assigned to covert messages transmission, where P_t is the total transmit power. The transmit signal at R is given by

$$\mathbf{r} = \mathbf{A}(\sqrt{p_\mathrm{r}}\mathbf{x}_\mathrm{s} + \sqrt{1-p_\mathrm{r}}\mathbf{x}_\mathrm{c}), \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{M\times M}$ is the precoder at R. Both the power of forwarding signal $\mathbf{x}_\mathrm{s} = \beta_\mathrm{m}\mathbf{y}_\mathrm{r}$ and the covert messages $\mathbf{x}_\mathrm{c}$ are one. The transmit power at R has $P_\mathrm{t} = \mathbb{E}\left[\|\mathbf{r}\|^2\right] = \mathrm{tr}(\mathbf{A}\mathbf{A}^H)$.

For the warden, a binary hypothesis test is employed to discriminate the existence of covert messages. The received signals at W under two hypotheses are written by

$$y_\mathrm{w} = \begin{cases} \sqrt{p_\mathrm{r}L_\mathrm{w}}\mathbf{h}_\mathrm{w}^T\mathbf{A}\mathbf{x}_\mathrm{s} + n_\mathrm{w}, & \text{for } \mathcal{H}_0, \\ \sqrt{L_\mathrm{w}}\mathbf{h}_\mathrm{w}^T\mathbf{A}\mathbf{x}_\mathrm{r} + n_\mathrm{w}, & \text{for } \mathcal{H}_1, \end{cases} \tag{2}$$

where $L_{\rm w}$ and $n_{\rm w}$ indicate the path-loss from R to W and the noise at W, respectively, and $\mathbf{x}_{\rm r}$ is defined as $\mathbf{x}_{\rm r} = \sqrt{p_{\rm r}}\mathbf{x}_{\rm s} + \sqrt{1-p_{\rm r}}\mathbf{x}_{\rm c}$ with power of one. In this test, the null hypothesis $\mathcal{H}_0$ is that R does not transmit covert messages, while the alternative hypothesis $\mathcal{H}_1$ represents that R conveys covert messages to two source nodes.

Assuming the warden detects the covert message based on the received signal power $Y_{\rm w}$ by a radiometer as [8]

$$Y_{\rm w} = \begin{cases} p_{\rm r} L_{\rm w} \left\|\mathbf{h}_{\rm w}^T \mathbf{A}\right\|^2 + \sigma_{\rm w}^2, & \text{for } \mathcal{H}_0, \\ L_{\rm w} \left\|\mathbf{h}_{\rm w}^T \mathbf{A}\right\|^2 + \sigma_{\rm w}^2, & \text{for } \mathcal{H}_1, \end{cases} \tag{3}$$

Then, the detection criterion at W is given based on a certain threshold τ as $Y_{\rm w} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \tau$.

Then, defining $Z \triangleq \left\|\mathbf{h}_{\rm w}^T \mathbf{A}\right\|^2$, the DEP, ξ, at W is given by

$$\xi = 1 - F_Z\left(\frac{\tau - \sigma_{\rm w}^2}{p_{\rm r} L_{\rm w}}\right) + F_Z\left(\frac{\tau - \sigma_{\rm w}^2}{L_{\rm w}}\right) \tag{4}$$

where $F_Z(z)$ indicates the cumulative distribution function (CDF) of the random variable Z. To compute (4), we need to determine the distribution of $\left\|\mathbf{h}_{\rm w}^T \mathbf{A}\right\|^2 = \mathbf{h}_{\rm w}^T \mathbf{A}\mathbf{A}^H \mathbf{h}_{\rm w}^*$.

Given that $\mathbf{A}\mathbf{A}^H$ is a positive semi-definite matrix, we apply eigenvalue decomposition to $\mathbf{A}\mathbf{A}^H$ as $\mathbf{A}\mathbf{A}^H = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$ where $\mathbf{Q}$ stands for an $M \times M$ unitary matrix and $\mathbf{\Lambda}$ is a diagonal matrix whose m-th diagonal element is the corresponding eigenvalue l_m^2. Accordingly, $\left\|\mathbf{h}_{\rm w}^T \mathbf{A}\right\|^2$ can be computed as

$$\left\|\mathbf{h}_{\rm w}^T \mathbf{A}\right\|^2 = \tilde{\mathbf{h}}^H \mathbf{\Lambda} \tilde{\mathbf{h}} = \sum_{i=1}^{K} l_i^2 \left|\tilde{h}_i\right|^2, \tag{5}$$

where $\tilde{\mathbf{h}} = \mathbf{Q}^H \mathbf{h}_{\rm w}^* = \left[\tilde{h}_1, ..., \tilde{h}_M\right]$, and $K = \text{rank}(\mathbf{A})$. As each element $h_{{\rm w},m}$ of $\mathbf{h}_{\rm w}$ is independent with each other and $h_{{\rm w},m} \sim \mathcal{CN}(0,1)$, which results in $\mathbb{E}\left[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\right] = \mathbf{I}_M$. Thus, we have

$$\mathbb{E}\left[\left\|\mathbf{h}_{\rm w}^T \mathbf{A}\right\|^2\right] = \mathbb{E}\left[\sum_{i=1}^{K} l_i^2 \left|\tilde{h}_i\right|^2\right] = \sum_{i=1}^{K} l_i^2, \tag{6}$$

where $K \leq 2$. As $\mathbb{E}\left[\left\|\mathbf{h}_{\rm w}^T \mathbf{A}\right\|^2\right] = \text{tr}\left(\mathbf{A}\mathbf{A}^H\right) = P_{\rm t}$, it follows $\sum_{i=1}^{K} l_i^2 = P_{\rm t}$.

The distribution of $\left\|\mathbf{h}_{\rm w}^T \mathbf{A}\right\|^2$ can be confirmed as a hypo-exponential distribution [11] with the CDF $F_Z(z) = 1 - \sum_{i=1}^{K}\left(\exp\left(-\frac{z}{l_i^2}\right)\prod_{j=1, j\neq i}^{K}\frac{l_i^2}{l_i^2 - l_j^2}\right)$. [1] Then, the DEP ξ for the second case is computed as

$$\xi = \begin{cases} 1 + \mathcal{C}_{\rm FA} - \mathcal{C}_{\rm MD}, \tau > \sigma_{\rm w}^2, \\ 1 \qquad\qquad\quad , \tau \leq \sigma_{\rm w}^2, \end{cases} \tag{7}$$

where

$$\mathcal{C}_{\rm FA} = \frac{l_1^2}{l_1^2 - l_2^2}\exp\left(-\frac{\tau - \sigma_{\rm w}^2}{l_1^2 p_{\rm r} L_{\rm w}}\right) + \frac{l_2^2}{l_2^2 - l_1^2}\exp\left(-\frac{\tau - \sigma_{\rm w}^2}{l_2^2 p_{\rm r} L_{\rm w}}\right),$$

$$\mathcal{C}_{\rm MD} = \frac{l_1^2}{l_1^2 - l_2^2}\exp\left(-\frac{\tau - \sigma_{\rm w}^2}{l_1^2 L_{\rm w}}\right) + \frac{l_2^2}{l_2^2 - l_1^2}\exp\left(-\frac{\tau - \sigma_{\rm w}^2}{l_2^2 L_{\rm w}}\right).$$

[1]The distribution of Z has two special cases: it is identical to the exponential distribution when $l_1^2 = P_{\rm t}$ and $l_2^2 = 0$, and equals the gamma distribution if $l_1^2 = l_2^2$.

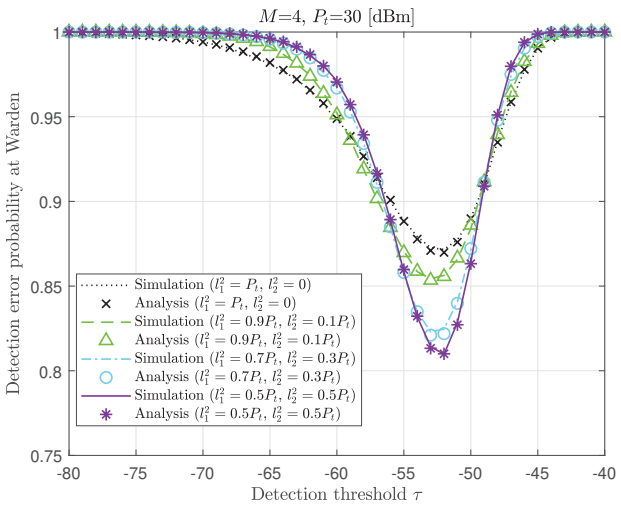


Fig. 1: The DEP with respect to τ.

III. Numerical Results and Conclusion

In this section, the effectiveness of the derived DEP expression is validated by numerical results. We compare the performance with different power allocation schemes. Fig. 1 presents the DEP ξ in terms of the threshold τ with all three cases. In the plot, the lines represent the simulation results through Monte Carlo experiment, while the marks indicate the analytic derivations. The accuracy of derivations in (7) is confirmed by comparing with the results of the experiments.

References

[1] E.-K. Hong, I. Lee, B. Shim, Y.-C. Ko, S.-H. Kim, S. Pack, K. Lee, S. Kim, J.-H. Kim, Y. Shin, Y. Kim, and H. Jung, "6G r amp;d vision: Requirements and candidate technologies," *J. of Commun. Netw.*, vol. 24, pp. 232–245, April 2022.

[2] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, pp. 1125–1142, Oct. 2017.

[3] F. Jameel, S. Wyne, G. Kaddoum, and T. Q. Duong, "A comprehensive survey on cooperative relaying and jamming strategies for physical layer security," *IEEE Commun. Surveys Tuts.*, vol. 21, pp. 2734–2771, Aug. 2019.

[4] G. Gui, M. Liu, F. Tang, N. Kato, and F. Adachi, "6G: Opening new horizons for integration of comfort, security, and intelligence," *IEEE Wirel. Commun.*, vol. 27, pp. 126–132, Oct. 2020.

[5] Z. Liu, J. Liu, Y. Zeng, and J. Ma, "Covert wireless communications in IoT systems: Hiding information in interference," *IEEE Wirel. Commun.*, vol. 25, pp. 46–52, Dec. 2018.

[6] C. Gao, B. Yang, X. Jiang, H. Inamura, and M. Fukushi, "Covert communication in relay-assisted IoT systems," *IEEE Internet Things J.*, vol. 8, pp. 6313–6323, April 2021.

[7] J. Moon, "Covert communications in a compress-and-forward relay system," *ICT Express*, Aug. 2023.

[8] L. Lv, Z. Li, H. Ding, N. Al-Dhahir, and J. Chen, "Achieving covert wireless communication with a multi-antenna relay," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 760–773, Feb. 2022.

[9] D. Deng, X. Li, S. Dang, M. C. Gursoy, and A. Nallanathan, "Covert communications in intelligent reflecting surface-assisted two-way relaying networks," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 12380–12385, Nov. 2022.

[10] K. Lee, N. Lee, and I. Lee, "Achievable degrees of freedom on MIMO two-way relay interference channels," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 1472–1480, April 2013.

[11] S. M. Ross, *Introduction to probability models (11th Edition)*. Academic press, 2020.

A Localization Algorithm using k-Nearest Neighbors in Indoor UWB Environment

Mookyung Jung
Senior Manager, Cloud Project Team
Goodmorning Information Technology Co., Ltd.
Busan, Republic of Korea
mkjung@goodmit.co.kr

Dong Myung Lee
Prof., Dept. of Computer Engineering
Tongmyong University
Busan, Republic of Korea
dmlee@tu.ac.kr

Sook Kyoung Cho
Chief Executive Officer
Nestors Co., Ltd.
Seoul, Republic of Korea
skyindb@naver.com

Kwangchul Jung
Prof., Research Institute of Information & Culture Technology
Dankook University
Yongin, Republic of Korea
jungkc6650@dankook.ac.kr

Abstract—This paper proposes a localization algorithm using k-nearest neighbors (k-NN) in an indoor ultra-wideband (UWB) environment. The experimental data under non-line of sight (NLOS) conditions and the actual coordinate performance with distance compensation using k-NN are analyzed. The proposed system utilizes learning data generated from experiments to determine line of sight (LOS) or NLOS states using k-NN, allowing for the calculation of standard deviation to account for distance value variations in LOS. This system aims to enhance reliability by reducing localization errors caused by NLOS and multipath effects typical in indoor positioning, thereby improving accuracy.

Keywords- Ultra-wideband, k-nearest neighbors, Line of Sight, Non-Line of Sight, Indoor Localization

I. INTRODUCTION

In indoor localization, accurately detecting the line of sight (LOS) or non-line of sight (NLOS) radio condition is crucial for achieving accurate localization, particularly in indoor ultra-wideband (UWB) environments [1-2]. In such condition, it would be very meaningful if NLOS state detection could be applied at low cost in a UWB environment.

Building on this background, this paper introduces a localization algorithm based on distance compensation using k-nearest neighbors (k-NN) in indoor UWB environments. The algorithm analyzes experimental data under NLOS conditions and evaluates the actual performance of distance compensation using k-NN. By conducting experiments to generate learning data, the algorithm distinguishes between LOS and NLOS conditions using k-NN. It then calculates the standard deviation of distance variations under LOS conditions and applies compensation for NLOS based on this deviation value [3-4].

II. RELATED STUDIES

In a UWB environment, the distance between an anchor and a tag is determined using two-way ranging (TWR) as shown in Fig. 1. T_{round_trip} represents the round trip time (RTT) for distance measurement, starting from when the anchor transmits the message until it receives the tag's response. T_{reply} indicates the tag's response time (processing time) [2]. The propagation delay time (T_p) between the two devices is calculated using Eq. (1), allowing the distance between them to be obtained using Eq. (2) (where c = 299,792,485 m/s).

$$T_p = \frac{t_{round_trip} - t_{reply}}{2} \quad (1)$$

$$\mathrm{d} = t_p \times c(propagation\ speed) \quad (2)$$

k-NN is a machine learning algorithm that classifies sample data points based on the k nearest data points in terms of distance. The classification performance of k-NN depends on choosing an appropriate value for k. The distance measurement method in k-NN typically uses the Euclidean distance between two data points, which can be calculated using Eq. (3). In this equation, i represents the maximum dimension of the data, and x_i and y_i denote the coordinates of the i^{th} dimension of data points x and y, respectively [4].

$$\mathrm{d}(\mathrm{x},\mathrm{y}) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2} \quad (3)$$

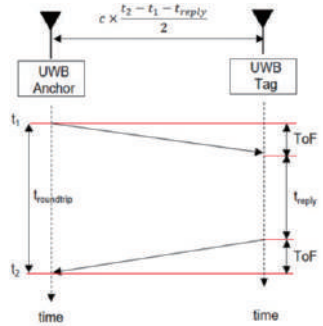


Fig. 1. Two way ranging concept.

III. SYSTEM DESIGN

The localization algorithm proposed in this paper, which utilizes distance compensation with k-NN in an indoor UWB environment, consists of three main steps as illustrated in Fig. 2. In the first step, distance data between UWB tags and anchors is collected. Next, in the second step, data is generated based on the change rate of the collected distance values between the UWB tag and anchor compared to the previous values. Finally, learning data for distinguishing between LOS and NLOS conditions is generated based on the calculated change rates [5].

In the third step, NLOS conditions are detected and determined using machine learning based on the change rate of distance values between tags and anchors within the learning dataset. In the final step, localization is performed on the distance values detected as NLOS. The localization

function adjusts the NLOS distance values based on the standard deviation of the change rate observed in LOS conditions. The performance of the corrected distance values is then analyzed in comparison to actual localization data [6].

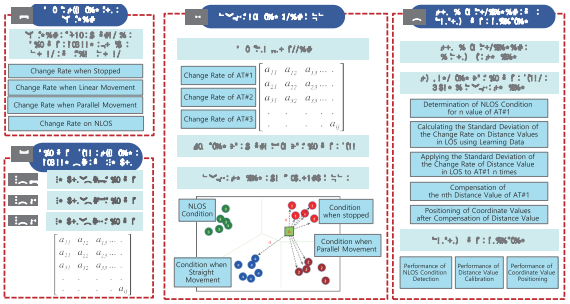


Fig. 2. Architecture of proposed localization algorithm.

IV. Experiments And Results

Experiments were conducted to evaluate the performance under LOS, NLOS, and multipath conditions using learning data of the propagation conditions between UWB tags and anchors. The UWB equipment utilized in the experiments was the DWM1001-DEV, manufactured by Qorvo.

The system configuration for indoor localization experiments in a UWB environment is illustrated in Fig. 3. The tag measures the distance to each anchor (1, 2, 3, and 4) using TWR and transmits these distance values to the anchors. All distance values between the anchors and the tag are transmitted to the monitoring server. The monitoring server processes these distance measurements to determine the location coordinates, which are then provided to the user.

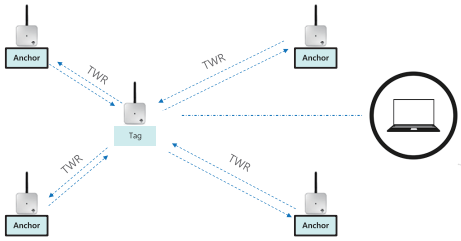


Fig. 3. System configuration for indoor localization experiments.

The learning data experiment was conducted in the following manner: First, distance values for LOS, NLOS, and multipath conditions between the UWB tag and Anchor 1, Anchor 2, Anchor 3, and Anchor 4 were measured. Data on the change rate between these distance values and the previously measured values were collected. Next, the change rates of the collected distances from the tags to anchors were classified into LOS or NLOS categories. Machine learning techniques were applied to each labeled dataset to detect LOS or NLOS conditions and assess the performance of condition discrimination. The discriminant performance of the machine learning models was verified using the Orange 3 machine learning data analysis tool. The data collection conditions for LOS and NLOS are illustrated in Fig. 4. The picture on the right of Figure 4 shows the NLOS environment, and the parts marked in black represent the pillar of the experimental space.

The experiment was conducted in an office and building hallway environment where actual NLOS and multipath conditions can occur. As a result of the learning data experiment, the proposed system demonstrated a reduction in root mean square error (RMSE) in both stationary and moving states. Specifically, the RMSE was reduced by 5.2 cm and 8.2 cm in the first stationary and moving states, respectively. Furthermore, it was observed that the RMSE decreased by 4.03 cm in the office environment and by 131 cm in the building hallway environment.

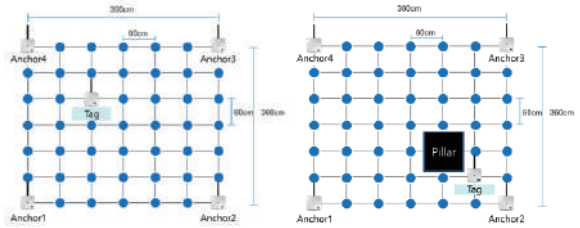


Fig. 4. Data collection conditions in LOS (Left) and NLOS (Right).

In addition, when comparing the NLOS precision of the proposed system with other studies, it was observed that the RMSE was reduced by a maximum of 82 cm and a minimum of 13 cm, while the sensitivity was improved by a maximum of 17% and a minimum of 7%. Finally, experimental results confirmed that the indoor localization system based on distance compensation using k-NN in the UWB environment, as proposed in this paper, effectively improves localization error accuracy in NLOS and multipath conditions commonly encountered in indoor localization.

V. Conclusion

In this paper, we proposed a localization algorithm based on distance compensation using k-NN in an indoor UWB environment. Through an analysis of the performance of support vector machine (SVM), decision tree, neural network, and k-NN algorithms, it was determined that k-NN exhibited the highest performance with an accuracy of 0.983. The experiments conducted included both learning data experiments and real-world experiments in office and building hallway environments where NLOS and multipath conditions could occur. Therefore, the system proposed in this paper is expected to be more reliable, reducing localization errors attributed to NLOS and multipath conditions commonly encountered in indoor localization and improving overall accuracy.

References

[1] H. Liu, H. Darabi, P. Banerjee, J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, issue: 6, pp. 1067–1080, Nov. 2007.

[2] S. Gezici,et. al., "Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks," IEEE Signal Processing Magazine, vol. 22, issue 4, pp. 70–84, Jul. 2005.

[3] T. V. Nguyen, Y. Jeong, H. Shin, M. Z Win, "Machine Learning for Wideband Localization," IEEE Journal on Selected Areas in Communications, vol. 33, issue 7, pp. 1357-1380, Jul. 2015.

[4] G. Guo, H. Wang, D. Bell, Y. Bi and K. Greer," KNN model-based approach in classification," OTM 2003: On The Move to Meaningful Internet Systems 2003, Springer, Berlin, Heidelberg, Germany, pp. 986-996, 2003.

[5] Mookyung Jung and Dong Myung Lee, "A Design of the k-NN based NLOS Recognition Algorithm in UWB Indoor Positioning Environment," Proceedings of the 33rd Joint Conference on Communications and Information (JCCI 2023), KICS, pp. 306-307, 26-28 Apr. 2023.

[6] Mookyung Jung, "Indoor Positioning System Based on Distance Compensation Using k-NN in UWB Environment," Thesis for Degree of Doctor of Engineering, Dept. of Computers & Media Engineering, The Graduate School of Tongmyong University, pp. 1–58, Aug. 2023.

An Indoor Localization Algorithm using Pose Estimation Method in Obstacle Environments for Pedestrians

Tae Wan Kim
Software Developer
Freelancer
Busan, Republic of Korea
wanykt@gmail.com

Dong Myung Lee
Prof., Dept. of Computer Engineering
Tongmyong University
Busan, Republic of Korea
dmlee@tu.ac.kr

Sook Kyoung Cho
Chief Executive Officer
Nestors Co., Ltd.
Seoul, Republic of Korea
skyindb@naver.com

Kwangchul Jung
Prof., Research Institute of Information & Culture Technology
Dankook University
Yongin, Republic of Korea
jungkc6650@dankook.ac.kr

Abstract—Most existing image-based methods have the disadvantage of low accuracy in environments with moving obstacles. In this paper, we propose an indoor localization algorithm named weight real-time model using pose estimation method for pedestrians (WRTMDet PEM) to improve this problem. The algorithm is designed to detect pedestrian objects and accurately estimate their locations based on real-time images in indoor environments with obstacles. Our experiments demonstrate the effectiveness of the proposed algorithm in estimating pedestrian locations under these challenging conditions. Additionally, we evaluate the algorithm's performance across various pose models in non-obstacle, virtual obstacle, and actual obstacle environments.

Keywords- Indoor Obstacle Environments, Pedestrian, Object Detection, Human Pose Estimation, Localization

I. INTRODUCTION

With the recent advancements in deep learning technology, there is a growing demand for algorithms capable of real-time pedestrian location estimation using CCTV images without the need for specialized hardware or equipment. Existing image-based methods rely on visual information to detect human objects and estimate pedestrian locations through pose estimation. However, these methods experience reduced accuracy in environments with moving obstacles. Sensor-based approaches face challenges such as sensor placement burdens and environmental factors affecting measurement accuracy. To address these issues, this paper proposes an indoor localization algorithm using pose estimation method in obstacle environment for pedestrians. By leveraging human pose estimation technology alongside image-based position estimation methods, this algorithm offers robust localization performance in various indoor settings.

II. RELATED STUDIES

Human pose estimation (HPE) techniques encompass RGB image/video-based and sensor-based approaches. The real-time models for object detection (RTMDet) framework was developed to enable efficient real-time object detection using the Faster R-CNN architecture [1]. RTMDet is designed for scalability across various object recognition tasks, including object segmentation and rotation-based detection.

RTMDet follows a structured process for object detection. Input images undergo preprocessing to match the model's input size before being fed into a convolutional neural network (CNN) backbone. The feature maps generated by the CNN are then processed by the path aggregation network-based feature pyramid network (PAFPN), which identifies feature points of objects, handles rotation detection, and supports instance segmentation. The final results are forwarded to the detection Heads for further analysis [2-4].

III. ALGORITHM DESIGN

A. Problem Analysis

The detection results of human objects using Faster R-CNN, Mask R-CNN, and RTMDet tiny models in a blind spot environment, along with the human pose estimation results using the ResNet50-FPN model, are illustrated in Fig. 1. The figures indicate that human object detection and pose estimation are typically confined to the image range.

However, the ankle joint was not detected for position estimation, leading to errors in pedestrian location estimation and reduced accuracy, as shown in Fig. 1. Therefore, accurate detection of the lower body region is crucial for precise location estimation based on image analysis.

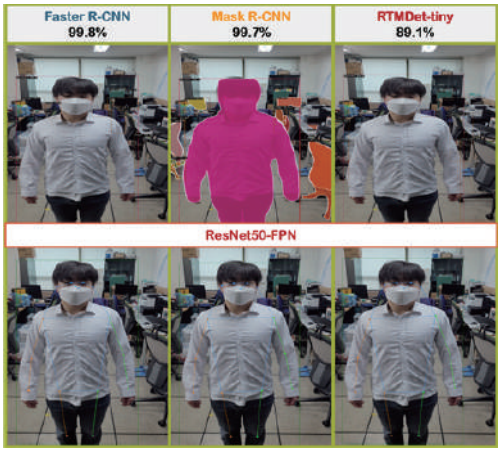


Fig. 1. Human object and human pose estimation results in blind spots.

B. Design Considerations

- To achieve accurate pedestrian location estimation in indoor obstacle environments, human object detection and posture estimation are crucial components.

- Information regarding a person's lower body is particularly important in indoor blind spots or obstacle-rich environments, as it significantly influences the accuracy of location estimation using image-based methods.
- The accuracy of human object detection and posture estimation tends to decrease in indoor obstacle environments. Therefore, developing algorithms that can effectively address these challenges is essential for improving pedestrian location estimation.
- Accurate analysis of location estimation results in indoor obstacle environments requires image distortion processing and map projection techniques.

C. Algorithm Architecture

The architecture of an proposed algorithm (WRTMDet PEM) is illustrated in Fig. 2, comprises the following key steps: preparation, WRTMDet execution, posture estimation processing, and location determination [5-6].

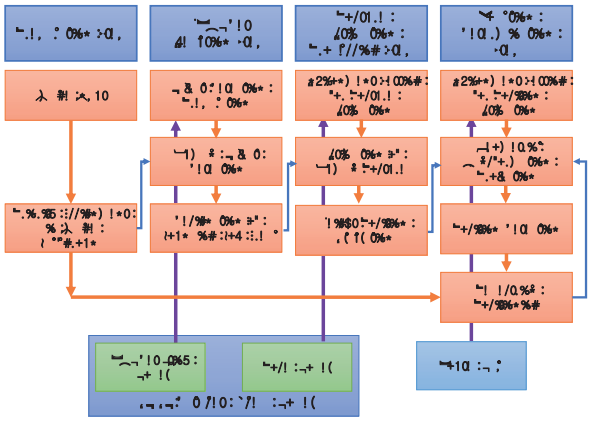


Fig. 2. Architecture of proposed algorithm (WRTMDet PEM).

D. WRTMDet Model

The weight real-time models for object detection (WRTMDet) model comprises object detection preparation, human object detection, and bounding box designation steps. For object detection preparation, the RTMDet-tiny model, a modified version of RTMDet, detects fast-moving pedestrian objects in real-time environments. The backbone divides the input image into segments and passes it to the neck, which includes Resize (R), Upscale (U), Merge (M), and Layer (L) operations. The neck processes the image through enlargement and merging before forwarding it to the Head, which applies a loss function to identify the bounding box area and detect human objects.

The proposed posture estimation function involves three main steps: environment setup for posture estimation, human pose estimation, and weighted position calculation. Once posture estimation is completed, the positioning function proceeds through environment setup for position estimation, geometric transformation projection, position detection, and pedestrian positioning. During the position estimation environment setup, image information is projected onto map coordinates. The original image is converted to a black and white format to detect and outline moving pedestrian areas, which are then mapped onto specific coordinates. In the geometric transformation projection step, coordinates of moving pedestrian areas in image information are projected onto map coordinates.

IV. Experiments And Results

The experimental results demonstrate significant improvements in the average root mean square error (RMSE) for the movement trajectory distance error of pedestrian objects using the proposed WRTMDet PEM algorithm compared to real-time models for object detection using B-Box (RTMDet B-Box) and predicted seamless human positioning (RTMDet PSHP) in various environments:

- In a non-obstacle movement environment, WRTMDet PEM improved RMSE by 69.42% over RTMDet B-Box and 39.04% over RTMDet PSHP.
- In a virtual obstacle movement environment, WRTMDet PEM showed performance gains of 65.94% over RTMDet B-Box and 29.82% over RTMDet PSHP.
- In an actual obstacle movement environment, WRTMDet PEM achieved improvements of 87.97% over RTMDet B-Box and 76.36% over RTMDet PSHP.

Furthermore, the average RMSE for the movement zone of the entire pose model was enhanced by 81.55% over RTMDet B-Box and 74.87% over RTMDet PSHP in a non-obstacle movement environment.

V. Conclusion

This paper proposed an indoor localization algorithm named weight real-time model using pose estimation method for pedestrians (WRTMDet PEM), designed specifically for detecting pedestrian objects in obstacle-rich environments. The algorithm's efficacy in estimating pedestrian locations under these conditions has been validated through experimentation. Evaluation metrics include the RMSE of pedestrian object trajectories and the average RMSE of error distances across the entire pose model's movement area. Performance evaluation involved comparative analyses using various pose models in non-obstacle, virtual obstacle, and actual obstacle environments.

References

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, Jun. 2017.

[2] C. Lyu, W. Zhang, H. Huang, Y. Zhou, Y. Wang, Y. Liu, S. Zhang, and K. Chen, "RTMDet: An Empirical Study of Designing Real-Time Object Detectors," arXiv preprint, arXiv:2212.07784, 2022.

[3] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, UniteSA, pp. 8759-8768, 2018.

[4] Y. Zheng, X. Zhang, R. Zhang, and D. Wang, "Gated Path Aggregation Feature Pyramid Network for Object Detection in Remote Sensing Images," Remote Sensing, vol. 14(18), pp. 1-13, Sep. 2022.

[5] Tae Wan Kim, Woo Hyun Kim and Dong Myung Lee, "Path Estimation Algorithm based on W-RTMDet HPE-WP in Indoor Real-time Obstacle Environment," Proceedings of the 2nd International Conference on Maritime IT Convergence (ICMIC 2023), Korea Information and Communications Society (KICS), pp. 90, Aug. 2023.

[6] Tae Wan Kim, "Pedestrian Localization System Using W-RTMDet and Pose Estimation Method in Indoor Obstacle Movement Environments," Thesis for Degree of Doctor of Engineering, Dept. of Computers & Media Engineering, The Graduate School of Tongmyong University, pp. 1–94, Aug. 2023.

Real-time sports pose estimation training based on skeleton simulation

Geon-Woo Kim
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
kgw20010916@gmail.com

Ji-Yoon Kweon
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
jiyxon03@naver.com

Ji-Soo Kim
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
ddui0@naver.com

Woo-Hyeon Kim
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
whkim712@naver.com

Jeong-Eun Kim
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
kjung1105@kyonggi.ac.kr

Kyungyong Chung
Dept. AI Computer Engineering
Kyonggi University
Suwon, South Korea
dragonhci@gmail.com

***Abstract*—In modern society, as awareness of health increases, the importance of exercise is emphasized. However, many people are unable to exercise adequately due to time and economic constraints. Furthermore, incorrect exercise postures can lead to injuries, making it essential to maintain accurate poses. This study aims to develop a program that precisely identifies users' exercise postures through pose estimation of exercise videos using YOLOv8-pose, and suggests personalized exercise methods based on this information.**

Keywords—real-time, pose estimation, healthcare, video similarity, web page, personal training

I. Introduction

In recent years, the importance of health has been increasingly recognized by modern individuals, making the need for exercise a significant issue. It is crucial to engage in physical activities under the guidance of professionals regularly. However, due to time and financial constraints, many people are unable to maintain this routine with expert assistance. Additionally, exercising with improper posture can lead to injuries and pose risks. Identifying and correcting one's own exercise mistakes can be challenging when training alone, which makes solo workouts prone to causing injuries. Consequently, exercise initiated for health benefits can, paradoxically, have detrimental effects on one's health.

To address this, there are programs that track exercise frequency, duration, and calories burned to provide related information. However, even those programs that claim to offer personalized information merely provide data based on an individual's height, weight, BMI, or InBody measurements. For effective exercise, physical characteristics are also crucial. Nonetheless, injuries during exercise often occur due to not considering the individual's unique physical capabilities or from performing exercises with incorrect postures.

The primary functions of commonly used exercise programs focus on providing set workout routines that consider body shape rather than physical capabilities, concentrating mainly on dispensing information. Consequently, these programs do not emphasize posture estimation or correction, making it difficult to reduce the risk of injuries. Additionally, the lack of customized feedback that considers an individual's physical abilities presents a challenge in effectively utilizing the provided exercise routines.

In this paper, we propose a program equipped with a personal training feature that aims to prevent injuries during exercise and provide methods tailored to the individual. This program is designed to offer customized exercise routines that take into account each user's unique physical capabilities and correct exercise postures, ensuring safer and more effective workouts.

The proposed program contributes by offering solutions to prevent injuries during exercise and providing maximized workout effects through support for personal training. It monitors movements around the joints—arms, legs, neck, and back—ensuring exercises are performed safely. By executing exercises in the correct posture, the program can maximize the effectiveness of the workout. It takes into account the varying physical conditions and exercise capabilities of individuals, allowing for precise analysis of personal exercise postures and coaching suitable methods and positions. The program suggested in this paper will also be beneficial for individuals with extreme obesity or underweight, enabling them to exercise effectively and safely. Furthermore, it can be used by professional athletes for efficient training in the absence of a trainer, based on accurate posture analysis. It differs from existing programs in that it compares exercise postures.

The advantage is that it can be easily followed by disabled individuals or the elderly. Based on an intuitive interface, it allows easy imitation of body movements, and immediate posture adjustments can be made by checking the motion descriptions described at the bottom of the web page. Additionally, because it allows real-time comparison between one's own exercise video and an expert's video, it has the advantage of being able to understand the accuracy of the movements.

The structure of this paper is as follows: Chapter 2 discusses existing methods of pose estimation. Chapter 3 describes the process of preprocessing exercise videos, real-time processing, backend integration, and the technology used for uploading to web pages. Finally, Chapter 4 concludes the paper.

II. Related Work

Existing exercise videos are often difficult for users to follow and do not enable them to perform exercises with accurate postures. Consequently, more advanced methods of exercise video analysis are being developed. In this section, we review recent methods of exercise video analysis and the operational principles of the technology proposed in this program. We also discuss how these methods are connected to the current research.

A. Analysis of Existing Exercise Videos

Traditional exercise support applications are used on smartphones and wearable devices. These apps track the user's

exercise frequency, duration, and calories burned, providing related information. However, because they do not focus on posture estimation or correction, there is a significant limitation in their actual usability. This results in a gap between simply tracking activity and facilitating effective, injury-free workouts.

In baseball, technologies like Rapsodo utilize pose estimation and tracking to provide valuable data to pitchers by monitoring their pitching form, speed, pitch type, and spin rate [1]. However, this technology is specialized for baseball and not applicable to other sports. The high costs associated with its use and the necessity for professional trainers also limit its accessibility to the general public, reducing its usability for amateur athletes and casual enthusiasts.

B. Basic Posture Estimation Techniques

C. Lugaresi et al. [2] proposed MediaPipe, a machine learning framework developed by Google, which offers features such as face tracking, hand tracking, and object recognition, along with the capability to track human poses. MediaPipe performs well in single pose estimation. However, it has limitations, such as failing to accurately analyze poses when people overlap or are obscured by objects.

H. C. Nguyen et al. [3] proposed a real-time pose estimation model using YOLOv5 as the backbone, which also performs 3D pose analysis, achieving high processing speeds and accuracy. However, this model tends to estimate stereotypical human poses and has limitations in its versatility, affecting its broader applicability.

In this study, we build upon existing exercise video analysis methods and pose estimation technologies to provide users with the correct way to exercise. This approach allows for the extraction of more diverse and objective information compared to traditional exercise assistance programs, exploring the potential to make exercise more effective. Therefore, this research is expected to offer a new exercise assistance program that can better serve modern individuals by enhancing their workout effectiveness and safety.

III. Real-time sports pose estimation training based on skeleton simulation

Videos that feature exercises without providing additional information can be challenging for users who seek precise instructions on proper technique. To address this, our research employs a three-stage approach using pose estimation to deliver accurate information from exercise videos [4]. In the first stage, reference exercise videos are preprocessed using the YOLOv8-pose model [5]. This preprocessing aims to enhance the videos' clarity and usefulness as instructional material. The second stage involves uploading these preprocessed videos to the web, making them accessible to a broad audience, and allowing users to reference and follow these enhanced exercise routines. Finally, the third stage involves real-time input of user exercise videos to estimate poses. By comparing these with the reference videos, the program measures similarity and provides feedback to the user on how to correct their posture to match the ideal demonstrated in the preprocessed videos. Fig. 1 illustrates the process of preprocessing exercise videos using YOLOv8-pose.

Fig. 1 describes the pose estimation step based on YOLOv8-pose. All these videos generally undergo a preprocessing stage. In this stage, images are resized to appropriate dimensions and undergo transformations such as color normalization. Subsequently, features are extracted and the pose of human objects is estimated. The main body parts of each individual are identified, and a fusion process connects all the key points. Finally, the model outputs the results of the pose estimation, including each individual's bounding box and confidence score.

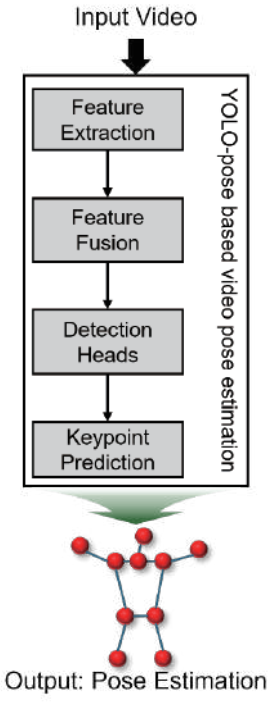


Fig. 1. YOLO-pose based pose estimation process.

A. Exercise Video Preprocessing

In this paper, the preprocessing step for exercise videos is a process that visualizes the movements of arms and legs to make it easy for users to follow existing exercise videos. The exercise video preprocessing step consists of three stages: dividing the exercise video into frames, pose estimation based on YOLOv8-pose, and video redefinition process.

The process of dividing exercise videos into frames is a step to prepare jpg-format files as input for the YOLOv8-pose model. Different datasets of exercise videos for various activities are downloaded and preprocessed at a frame rate of 1fps.

Afterward, the YOLOv8-pose model is used to extract the exercise posture of individuals from each frame. The preprocessing process using the YOLOv8-pose model is as follows: YOLOv8-pose is applied to all frames to find key points on the human shoulders, pelvis, knees, and ankles [6]. Then, pose estimation is performed by connecting all relevant key points.

Finally, the frames with pose estimation applied are redefined into a single mp4-format video. Fig. 2 depicts the analyzed exercise video through preprocessing techniques.

In Fig. 2, key points can be observed on the human skull, neck, shoulders, elbows, wrists, pelvis, knees, and ankles. Additionally, the key points are connected to divide the body into head, upper body, lower body, and waist. The approach assumes that the key points are connected based on the principle of continuity. However, when key points are occluded by other body parts, a method matching only two body parts simultaneously is employed. Finally, after applying two relaxation methods, the Hungarian algorithm is utilized to find the optimal matching solution [7].

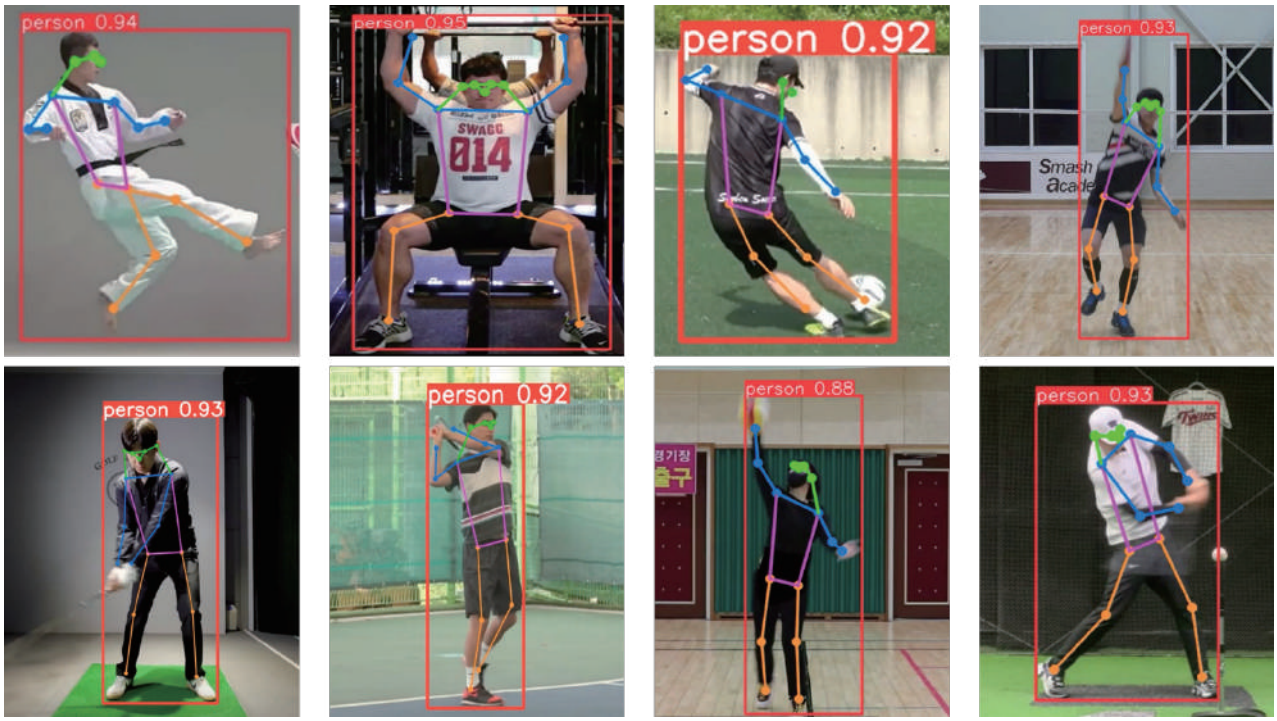


Fig 2. YOLOv8-pose based pose estimation results

$$C'_{ij} = C_{ij} - \min_{k}(C_{ik}) \quad (1)$$

$$C''_{ij} = C'_{ij} - \min_{k}(C'_{ik}) \quad (2)$$

Equation (1) and (2) are represent the computation process of the Hungarian algorithm. Let's generate matrix C as the cost matrix of $n \times n$, where c_{ij} represents the cost when each worker performs each task. (1) calculates the minimum value of each row, while (2) calculates the minimum value of each column. Finally, adjust the rows and columns to cover the zero values and repeat these steps until the optimal assignment of key points is found, where key points need to be connected. Fig. 3 represents the initial screen of the webpage.

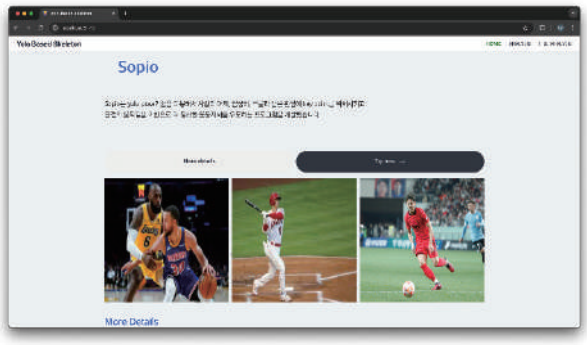


Fig. 3. Pre-processing video upload results

Fig. 3 shows the screen before entering the proposed webpage. By adding scenes of athletes exercising, it informs that it is a healthcare program and activates a start button to enable the execution of the exercise assistance program.

B. Upload Pre-processing Video Web Page

In Section 3.1, the preprocessed video is stored in Naver Cloud and then delivered to the front end through API integration on a URL basis. The received video from Naver Cloud is displayed on the corresponding exercise webpage along with exercise instructions and exposed on the webpage. Fig. 4 depicts a page for users to search for exercise videos.

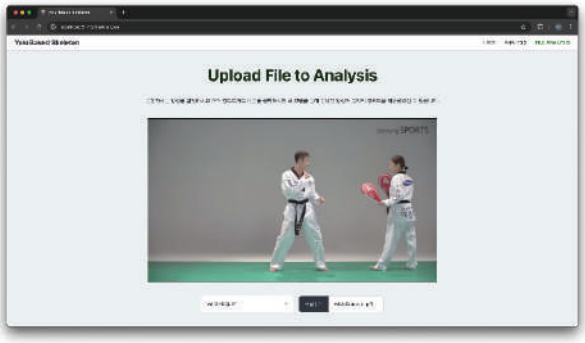


Fig. 4. Pre-processing video upload results

Fig. 4 shows information on the web page about which exercises will be performed along with the existing version of the video to be analyzed. When a user searches for the desired exercise at the bottom, the preprocessed video by the deep learning model is played, and it is designed to allow the user to click on various sports for convenience.

Additionally, it provides users with explanations of the movements in the video, aiding in more effective exercise. On the webpage, exercises are divided into ball sports and non-ball sports on the left side. By dividing the exercises, it increases accessibility for users to access the exercise videos of the program. Fig. 5 presnets the preprocessed video along with its description and exercise name displayed on the webpage.

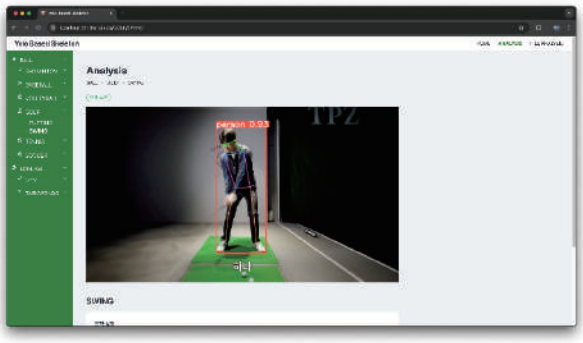


Fig. 5. Pre-processing video upload results

C. Analyzing Real-Time User Exercise Images

In this section, similar to Section 3.1, exercise videos are divided into frames to estimate poses. Using a webcam, the user's video is analyzed in real-time using YOLOv8-pose. Then, the similarity between the player and the user's actions is analyzed, and the similarity is visualized to provide the user with the progress of the exercise [8]. To upload the analyzed video to the webpage, frames with estimated poses are sent to the backend. Then, similar to Section 3.2, communication with the frontend is conducted via URLs to upload requested videos to the webpage. Fig. 6 illustrates a screen showing the user's exercise video analyzed in real-time and the similarity measure between the user's actions and the reference video [9].



Fig. 6. Real-time similarity analysis and pose estimation

In Fig. 6, as seen with the preprocessed video, real-time similarity of movements is provided to the user. This allows the user to see the cosine similarity-based visual similarity between their posture and the tracking results, which can also be accurately compared to reference footage [10]. This similarity result allows the user to modify their movements to extract high similarity. Consequently, it encourages users to perform efficient exercises with correct posture.

IV. Conclusion

This paper proposes a pose estimation-based exercise assistance healthcare program. The proposed method utilizes the YOLOv8-pose model to assist users in efficient exercise with real-time similarity processing. Through this research, it was confirmed that the developed exercise assistance healthcare program can efficiently analyze exercise video data and extract information. This implies the ability to provide users with high-quality feedback. This study will serve as a foundation for future research and applications in this field.

In future research, the following plans are scheduled. Firstly, instead of estimating poses for the entire exercise video, the aim is to track more key points to develop a more precise exercise assistance program. Providing specialized feedback for exercises requiring precise postures, such as joints in fingers, can offer users a higher level of feedback. Secondly, a program combining 3D pose estimation technology based on multi-angle shooting will be developed. Exercise is not merely an action performed in a 2D space but also in a 3D space. Therefore, by developing an exercise assistance program that encompasses both horizontal and vertical movements, users will be able to perform exercises with more accurate and refined postures. Exploring these avenues will enhance the versatility of this methodology.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation)"(2021-0-01393).

References

[1] S. K. Pal, A. Pramanik, J. Maiti and P. Mitra, "Deep learning in multi-object detection and tracking: state of the art", Applied Intelligence., vol. 51, pp. 6400-6429, 2021.

[2] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays and M. Grundmann, "Mediapipe: A framework for building perception pipelines", arXiv preprint arXiv:1906.08172, 2019.

[3] H. C. Nguyen, T. H. Nguyen, R. Scherer and V. H. Le, "Unified end-to-end YOLOv5-HR-TCM framework for automatic 2D/3D human pose estimation for real-time applications", Sensors, vol. 22, no. 14, pp. 5419, 2022.

[4] M. Andriluka, L. Pishchulin, P. Gehler and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis", In Proceedings of the IEEE Conference on computer Vision and Pattern Recognition (CVPR), 2014, pp. 3686-3693.

[5] D. Maji, S. Nagori, M. Mathew and D. Poddar, "Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 2637-2646.

[6] Y. Li, S. Zhang, Z. Wang, S. Yang, W. Yang, S. T. Xia and E. Zhou, "Tokenpose: Learning keypoint tokens for human pose estimation", In Proceedings of the IEEE/CVF International conference on computer vision (CVPR), 2021, pp. 11313-11322.

[7] E. Hamuda, B. Mc Ginley, M. Glavin and E. Jones, "Improved image processing-based crop detection using Kalman filtering and the Hungarian algorithm", Computers and electronics in agriculture, vol. 148, pp. 37-44, 2018.

[8] L. Yuan, T. Wang, X. Zhang, F. E. Tay, Z. Jie, W. Liu and J. Feng, "Central similarity quantization for efficient image and video retrieval", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2020, pp. 3083-3092.

[9] N. Pérez, F. S. Granda, D. Benítez, F. Grijalva and R. Lara, "Toward real-time volcano seismic events' classification: A new approach using mathematical morphology and similarity criteria", IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-13, 2021.

[10] Z. Lu, V. Rathod, R. Votel and J. Huang, "Retinatrack: Online single stage joint detection and tracking", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2020, pp. 14668-14678.

Data Placement Policy considering File System Input/Output Characteristics for Maritime IoT Platform

Sungmin Koo
Maritime ICT·Mobility Research Department
Korea Institue of Ocean Science & Technology(KIOST)
Busan, Republic of Korea
smkoo@kiost.ac.kr

Gunhee Choi
Dept. of Computer Engineering
Dankook University
Yongin, Republic of Korea
choi_gunhee@dankook.ac.kr

Jongmoo Choi
Dept. of Computer Engineering
Dankook University
Yongin, Republic of Korea
choijm@dankook.ac.kr

Juhyun Kim
Maritime ICT·Mobility Research Department
Korea Institue of Ocean Science & Technology(KIOST)
Busan, Republic of Korea
jhk@kiost.ac.kr

Seungjae Baek*
Maritime ICT·Mobility Research Department
Korea Institue of Ocean Science & Technology(KIOST)
Busan, Republic of Korea
baeksj@kiost.ac.kr

Abstract*— For an efficient maritime IoT platform, it is essential to consider the lifespan and performance of storage devices. IoT devices generally adopt flash memory as storage devices, which has many advantages over HDDs. In flash memory, the data placement policy highly influences the computer system performance and storage lifetime. In this paper, we propose a novel data placement policy considering file system input/output characteristics for maritime IoT platform. In terms of the file system, meta and journal data are hot data as they are more frequently accessed than user data. Using the file system's semantic information, the proposed scheme discerns hot data, including meta and journal data, and then places hot data into the same block. As our hot data classification is inferred from the immovable file system structure, our scheme is free from faulty classification. The proposed scheme is implemented with a QEMU-based flash memory simulator called FEMU, and demonstrated by Postmark benchmark. The results show that our scheme improves the storage performance by about 38.67% in EXT4 and by about 10.55% in XFS. Also, we demonstrated that the proposed scheme can lengthen the storage lifetime by about 9.1% in EXT4 and by about 3.9% in XFS.

Keywords—Flash memory, Meta data, Superblock, Journal, IoT

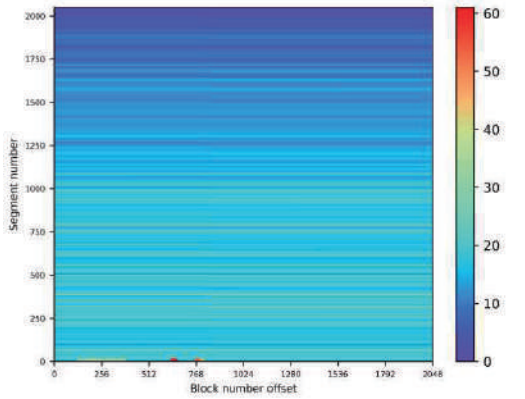
I. INTRODUCTION

Fundamental changes in the marine environment are accelerating due to ongoing marine pollution and climate change [1-3]. Obtaining long-term observation data is essential to respond to ocean changes that greatly impact human life. In harsh marine environments, remote sensing data acquisition is a great option to reduce hazards and costs. Recently, various research using Internet of Things (IoT) technology has been actively conducted to acquire marine data in real-time [4-9].

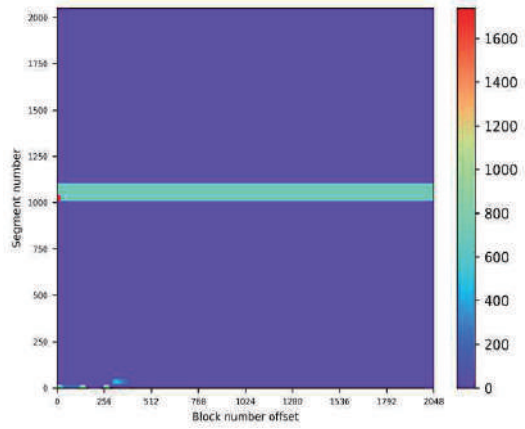
IoT devices generally adopt flash memory as storage because it is small, lightweight, low-power, and high-performance compared to HDDs. However, Triple Level Cell (TLC) and Quadruple Level Cell (QLC) based flash memory, widely adopted as a commercial product, have a short life expectancy due to their physical characteristics. Lengthening the storage lifetime is essential for a reliable IoT platform because replacing the storage device used by marine equipment would be both expensive and time-consuming. Additionally, Improving storage performance, which is the bottleneck of modern hierarchical computer systems, is crucial for real-time data acquisition.

Data are categorized into hot data with high access frequency and cold data with relatively low access frequency. According to the application's perspective, hot data classification is hard work because the data access frequency depends on the application's locality and can be changed during the application's execution. However, hot data can be managed more efficiently by considering the file system's semantic information. From a file system perspective, data can generally be divided into meta data, journal data, and user data. The meta and journal data are hot data because they are more frequently accessed than user data. Each file system has a unique semantic structure for data management such as file system layout and data management policy. Semantic structures of file systems lead to unique disk I/O patterns [10].

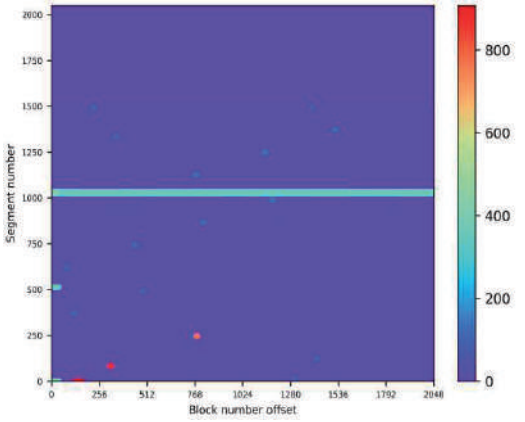
In this paper, we propose a novel data placement policy considering file system input/output characteristics for maritime IoT platforms. We first demonstrated that the file system's meta and journal data are hot data, and each file system had a unique I/O pattern because of its unique semantic structure. Second, we optimized the data placement policy considering the file system's semantic information. Our scheme places hot data, including meta and journal data, into



(a) exFAT access heatmap



(b) EXT4 access heatmap



(c) XFS access heatmap

Fig. 1. I/O patterns of file systems

the same block. Generally, large amounts of small-sized I/O requests are generated in IoT systems. When many small I/O operations occur, data access to the meta and journal area becomes more frequent. Accordingly, the proposed scheme greatly improves performance and life expectancy in IoT systems.

The proposed scheme was implemented with a QEMU-based flash memory simulator called FEMU [11, 12] and tested on exFAT, EXT4, and XFS using Postmark benchmark [13]. The results show that the proposed scheme improves write throughput. Specifically, the write throughput is

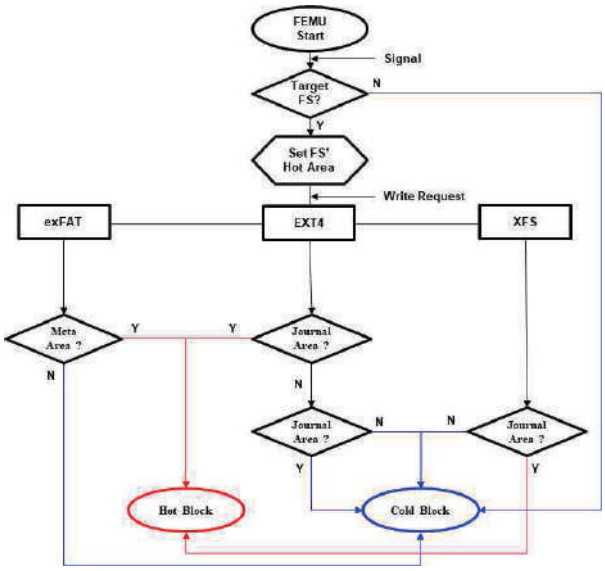


Fig. 2. Page placement procedure

improved by about 38.67% in EXT4 and by about 10.55% in XFS. Also, we demonstrated that the proposed scheme can lengthen the storage lifetime by about 9.1% in EXT4 and about 3.9% in XFS by decreasing the erase count.

II. Motivation

To analyze the I/O pattern of file systems, we conducted experiments with exFAT, EXT4, and XFS file systems using Postmark benchmark designed to simulate short-lived small files by creating files, executing transactions, and deleting files. For the experiment setup, we use Ubuntu 20.04 with Linux kernel 5.13.0 and 1TB Samsung 980 pro M.2 SSD. Using default options, we first created a file system on the 2G partition, and then we created 150 files of size 10MB and executed 10,000 transactions. I/O requests were traced by Blktrace [14], which traces all the I/O processes from the block I/O layer to storage. The file system access heatmap is shown in Figure 1. The x-axis denotes the block offset number, while the y-axis represents the segment (1,024 blocks) number. exFAT evenly distributes the write request over the entire storage space but frequently accesses the FAT and bitmap area. EXT4 shows a high access frequency to the journal and the fixed meta area managed by the flex group. The metadata of XFS is scattered all over the disk because XFS manages metadata using B+tree. The journal area of XFS is created in the middle of the disk.

III. Data Placement Policy considering File System Input/Output Characteristics

Figure 2 shows the proposed data placement procedure. First, the proposed scheme discerns the file system, meta area, and journal area. The file system, meta area, and journal area are determined through the superblock with detailed information about the file system. Then, when the write request arrives, our scheme decides whether data should be stored in a hot area based on the Logical Block Address (LBA) of the I/O request. If the LBA of the I/O request corresponds to a hot area, it is stored in the dedicated blocks. The dedicated blocks are not worn out faster than other blocks because of the wear-leveling algorithm. The page on which hot data are stored quickly becomes an invalid page. Accordingly, the proposed scheme increases the proportion of invalid pages in

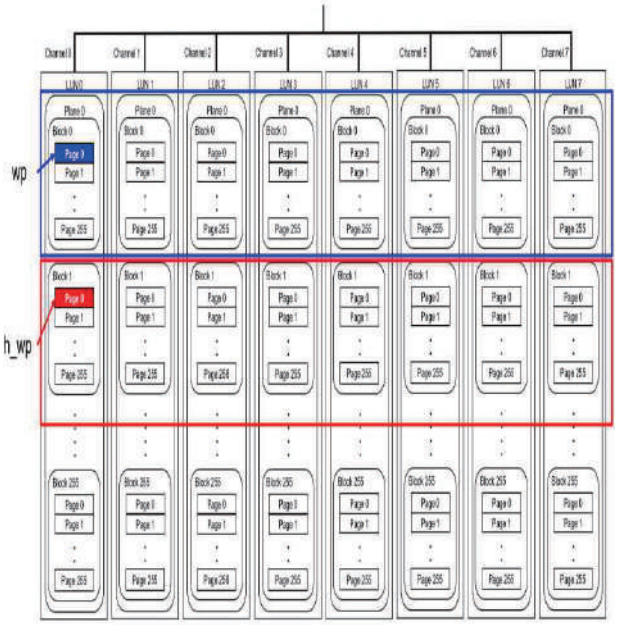


Fig. 3. Page allocation structure of FEMU

the block. Thus, the performance and the lifetime are increased by reducing the copy of valid pages in the victim block.

The proposed scheme is implemented with a QEMU-based simulator called FEMU [15, 16]. FEMU creates vSSD (Virtual SSD) in backend DRAM and imposes latency according to operation. FEMU supports multi-channel interleaving and manages pages in a line unit. The default vSSD structure of FEMU is as follows: it has 8 channels, a channel has 8 LUNs, a LUN has 256 blocks, and a block has 256 pages. The vSSD is composed of 256 lines for 4G bytes of space by default. A line of vSSD means a set of blocks with the same block sequence among blocks owned by LUNs. When a write is requested to vSSD, the pages inside the line pointed by the write pointer are allocated. A new line is allocated if all the pages within the line are used. For the implementation of the proposed scheme, we add an additional 'hot write pointer' as shown in Figure 3. When a write is requested to the vSSD, the Logical Page Number (LPN) of the request is checked. If the LPN of the write request is in the hot area, the page is allocated from the line of the hot write pointer. Specifically, in exFAT file system, pages are allocated from the hot write pointer if the LPN of the write request corresponds to the meta area. In EXT4 file system, pages are allocated from the hot write pointer if the LPN of the write request corresponds to the meta and journal area. In XFS file system, pages are allocated from the hot write pointer if the LPN of the write request corresponds to the journal area. Our policy incurs negligible runtime overhead of approximately 3.9ns, and therefore, it can be executed inside of the SSDs without host-side support.

IV. Evaluation

A. Experimental Setup

The host system consists of Intel(R) Core(TM) i9-9900K CPU @ 3.60GHz processor, 64GB DRAM, and Samsung SSD 980 PRO 1TB. We use 64-bit Linux 5.13 as a host operating system. We allocated 8 cores CPU, 4GB DRAM, and 500GB storage space with pre-allocated qcow2 format to

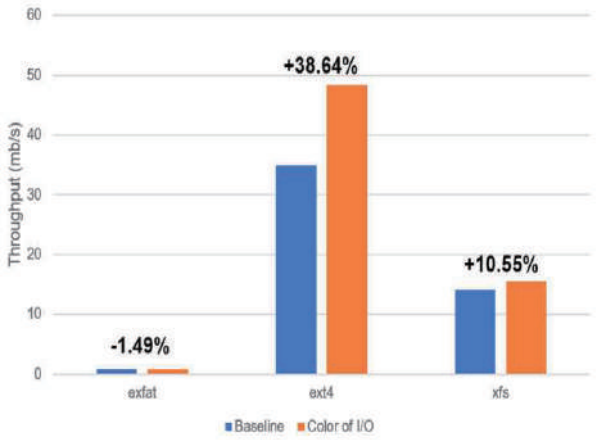


Fig. 4. Postmark benchmark result

the FEMU guest machine. For the experiment, we installed 64-bit Linux 5.13 on a guest machine and allocated 4GB DRAM to FEMU vSSD. To evaluate the proposed scheme, we chose Postmark benchmark that demonstrates system performance for short-lived small files because the IoT system creates a large number of small files.

B. Experimental Results

Figure 4 presents the performance evaluation results. We generated numerous I/O requests to specifically analyze the impact on performance by GC. In this experiment, we created 100,000 files of size 16KB and executed 1,000,000 transactions. The results show that the proposed scheme improves write throughput by about 38.67% in EXT4 and by about 10.55% in XFS. Our proposed scheme allocates hot data in the same block. Because hot data is updated frequently, only a small number of valid pages are left when the hot block is chosen as a victim. Accordingly, performance improves as fewer valid pages need to be copied. Additionally, the proposed scheme increases the proportion of blocks consisting of only cold data. Specifically, our scheme lengthens the storage lifetime by about 9.1% in EXT4 and about 3.9% in XFS by decreasing the erase count.

The experimental results show that EXT4 has a greater performance improvement than XFS. Both meta and journal areas can be discerned in EXT4 because the location of the meta area is fixed. However, XFS can discern only the journal area because the metadata of XFS is scattered all over the disk area. Conversely, exFAT file system did not benefit from our scheme. The exFAT, which is generally used as a file system for removable storage, does not use a journaling mechanism and has a small metadata area. Accordingly, it showed that the access frequency of user data is higher than in other file systems.

V. Conclusion

In this paper, we present a novel data placement policy considering file system input/output characteristics for maritime IoT platform. The contribution of this paper is as follows. First, we demonstrated that each file system has a unique I/O pattern, and meta and journal data are more frequently accessed than user data. Second, we confirmed that our scheme could improve the performance and lifetime of flash memory as hot data, including meta and journal data, are placed into the same block. As our hot data classification is inferred from the immovable file system structure, the

proposed scheme is free from faulty classification. Third, we showed that the policy incurs negligible runtime overhead of approximately 3.9ns, and therefore, it can be executed inside of the SSDs without host-side support.

The experimental results showed that the proposed scheme was an effective placement policy in EXT4 and XFS used as the main file system in the IoT platform based on Linux. Specifically, The results showed that the proposed scheme improved write throughput by about 38.67% in EXT4 and by about 10.55% in XFS. Also, we demonstrated that the proposed scheme can lengthen the storage lifetime by about 9.1% in EXT4 and by about 3.9% in XFS by decreasing the erase count.

ACKNOWLEDGMENT

This work was supported by research grants from Korea Institute of Ocean Science and Technology (PEA0202).

REFERENCES

[1] Tabari, H., "Climate change impact on flood and extreme precipitation increases with water availability" *Sci Rep* **10**, 13768, Aug. 2020.

[2] Mora, C., Dousset, B., Caldwell, I. et al., "Global risk of deadly heat", Nature Clim Change 7, 501–506, Jun. 2017.

[3] Doran, R., Ogunbode, C.A., Böhm, G. et al., "Exposure to and learning from the IPCC special report on 1.5 °C global warming, and public support for climate protests and mitigation policies", npj Clim. Action 2, 11, May. 2023.

[4] S. Geetha and S. Gouthami, "Internet of things enabled real time water quality monitoring system," Smart Water, vol. 2, no. 1, pp. 1-19, Jul. 2017.

[5] P. Salunke and J. Kate, "Advanced smart sensor interface in internet of things for water quality monitoring," in Proc. International Conference on Data Management, Analytics and Innovation (ICDMAI), pp. 298-302, Pune, India, Feb. 2017.

[6] B. Das and P. C. Jain, "Real-time water quality monitoring system using Internet of Things," in Proc. Int. Conf. on Computer, Communications and Electronics (Comptelix), pp. 78–82, 2017.

[7] C. Z. Myint, L. Gopal, and Y. L. Aung, "Reconfigurable smart water quality monitoring system in IoT environment," in Proc. IEEE/ACIS 16th Int. Conf. on Computer and Information Science(ICIS), pp. 435–440, 2017.

[8] ZHANG, Jinglin; DAI, Minghui; SU, Zhou. Task allocation with unmanned surface vehicles in smart ocean IoT. IEEE Internet of Things Journal, 2020, 7.10: 9702-9713.

[9] WANG, Chengcai, et al. Design, modeling, control, and experiments for a fish-robot-based IoT platform to enable smart ocean. IEEE Internet of Things Journal, 2021, 8.11: 9317-9329.

[10] Baek, S. J., Son, J. H., Song, Y. J., and Koo, S. M.. Color of I/O: I/O Characterization for Storage Systems. The 25th Korea Conference on Semiconductors, Kangwon, South Korea, February, 2018.

[11] LI, Huaicheng, et al, "The CASE of FEMU: Cheap, accurate, scalable and extensible flash emulator", In 16th USENIX Conference on File and Storage Technologies (FAST), pp. 83-90, February, 2018.

[12] FEMU. https://github.com/ucare-uchicago/FEMU

[13] Postmark. https://www.filesystems.org/docs/auto-pilot/Postmark.html

[14] Blktrace. https://github.com/sdsc/blktrace

[15] LI, Huaicheng, et al, "The CASE of FEMU: Cheap, accurate, scalable and extensible flash emulator", In 16th USENIX Conference on File and Storage Technologies (FAST), pp. 83-90, February, 2018.

[16] FEMU. https://github.com/ucare-uchicago/FEMU

Design of Smart Urban Environmental Monitoring System using Personal Mobility Devices

Kyung Kwon Jung
Dept. of Electric Engineering
Dongshin University
Naju-si, Republic of Korea
kkjung@dsu.ac.kr

Yeon Man Jeong
Dept.of Information and Telecommunication Engineering
Gangneung-Wonju National University
Wonju-si, Republic of Korea
ymjeong@gwnu.ac.kr

Byung Sang Choi
Dept. Biomedical Engineering
Korea Polytechnics Wonju-Campus
Wonju-si, Republic of Korea
cbs1235@kopo.ac.kr

Ho Soon Jang
Dept. Biomedical Engineering
Korea Polytechnics Wonju-Campus
Wonju-si, Republic of Korea
janghs@kopo.ac.kr

Yong Joong Kim
Dept. Biomedical Engineering
Korea Polytechnics Wonju-Campus
Wonju-si, Republic of Korea
kyj11111@kopo.ac.kr

***Abstract*— This paper presents a smart urban environmental monitoring system that utilizes personal mobility devices, specifically electric scooters and bikes, equipped with low-cost environmental IoT sensor modules. These modules comprise ESP32 microcontroller units, LTE-Cat.M1 device for real-time data communication, GPS for location tracking, and environmental sensors for measuring temperature, humidity, PM10, PM2.5, volatile organic compounds (VOCs), and nitrogen oxides (NOx). Data collected from these sensors are logged in real-time to server, including latitude, longitude, and sensor readings, enabling comprehensive urban environmental analysis. Furthermore, the use of Google Maps API enhances the system's ability to depict the routes taken by the e-scooters along with sensor readings. This feature offers a clearer understanding of the varying environmental conditions throughout distinct urban sectors.**

Keywords—urban environmental monitoring, IoT, personal mobility, air quality sensing

I. INTRODUCTION

Air pollution is a significant public health concern in many urban areas around the world. It can exacerbate respiratory illnesses, cardiovascular diseases, and other health problems. Traditional air quality monitoring systems rely on a network of fixed monitoring stations strategically placed throughout a city. These stations are typically expensive to install and maintain, requiring specialized equipment and dedicated infrastructure. Additionally, fixed stations have limitations in their spatial coverage. They are often concentrated in specific areas, leaving significant gaps in data collection, particularly in densely populated urban environments where air quality can vary greatly within short distances. This limited data makes it difficult for city planners and environmental agencies to get a complete picture of air quality variations across a city and develop effective pollution control strategies.

To address these limitations, researchers are exploring novel approaches for environmental monitoring that leverage advancements in sensor technology, low-power communication protocols, and the Internet of Things (IoT)[1].

This paper introduces a novel and cost-effective environmental monitoring system that utilizes personal mobility devices, specifically electric scooters (e-scooters) and electric bikes (e-bikes), as mobile sensor platforms. E-scooters and e-bikes are rapidly gaining popularity in urban transportation, offering a convenient, environmentally friendly, and increasingly ubiquitous mode of travel for short distances. The proposed method capitalizes on this growing trend by equipping e-scooters and e-bikes with low-cost sensor modules that collect real-time environmental data as they navigate through urban environments. These sensor modules can capture a range of environmental parameters, including temperature, humidity, particulate matter (PM), volatile organic compounds (VOCs) index, and nitrogen oxides (NOx) index.

II. SMART URBAN ENVIRONMENTAL MONITORING SYSTEM

Proposed platform is composed of three main physical elements: (i) the IoT sensor unit, (ii) the LPWAN(Low Power Wide Area Network) communication, and (iii) the backend server.

Fig.1 presents the architecture and the wireless technologies used. The core of the IoT sensor unit is ESP32 device, equipped with a number of environmental sensors and the LTE Cat-M1 interface. The LTE Cat-M1 is a low-power, wide-area network technology designed specifically for Internet of Things devices. The sensors monitor humidity, temperature, PM10, PM2.5, VOC index and NOx index. All hardware components are encapsulated in plastic encloser and deployed in personal mobile devices. The server stores the received data and generates map reports that can be accessed through the web page.

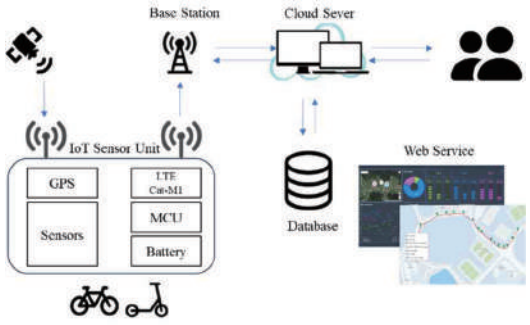
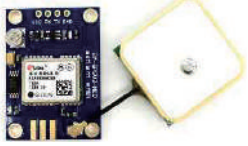


Fig. 1. System architecure.

A. IoT sensor unit

The IoT sensor unit was designed with components as shown in Table 1. The main MCU utilized is the ESP32, and it is connected to sensors via I2C for the measurement of temperature, humidity, particulate matter (PM10, PM2.5), VOC index, and NOx index. The GPS module receives latitude and longitude data, while the LTE Cat-M1 module facilitates the transmission of data to the server. The use of LTE Cat-M1 for communication ensures the reliable and rapid transmission of data. Data transmission utilizes the Vodafone Global IoT SIM with a 30MB/month plan[2-5].

TABLE I. TABLE TYPE STYLES

	ESP32-WROOM-32 Espressif Systems Xtensa® dualcore 32bit LX6 microprocessor
	SEN55 Sensirion.com Sensors: PM, RH, T, VOC, NOx
	GY-GPSV3-NEO NEO-7M GPS module (Ublox)
	LTE Cat-M1 CodeZoo.co.kr TYPE1SC chipset (Murata) Data package: 30MB/month

As shown in Fig. 2, the MCU, LTE Cat-M1 module, GPS, and multi-sensor are placed around the main board. Finally, a lithium power bank with 2,000 mAh of capacity has been attached to the unit and placed below the boards. In our experiments, we have obtained a battery life time of more than five days, working at 5V/3.3V and reporting data regularly. The dimensions of the whole set are 15x7x5 cm, and an adequate enclosure has been designed for its easy integration on-board a e-scooters.

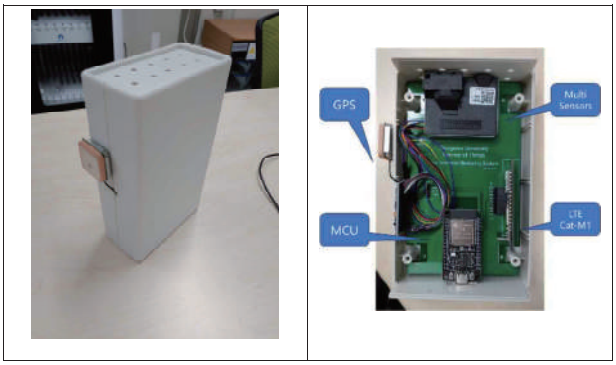


Fig. 2. IoT sensor unit.

B. Sever

The server system is composed of a database server and a web server. The data from each IoT sensor unit were collected into a database every 5 seconds, and then the data are processed to visualize the environmental data via web protocol.

The IoT sensor unit generates HTTP GET packets and injects them into the ThingSpeak cloud server[6]. ThingSpeak is an open IoT platform that allows aggregation, visualization, and analysis of live data streams in the cloud. Fig. 3 shows the data screen stored on the ThingSpeak server.

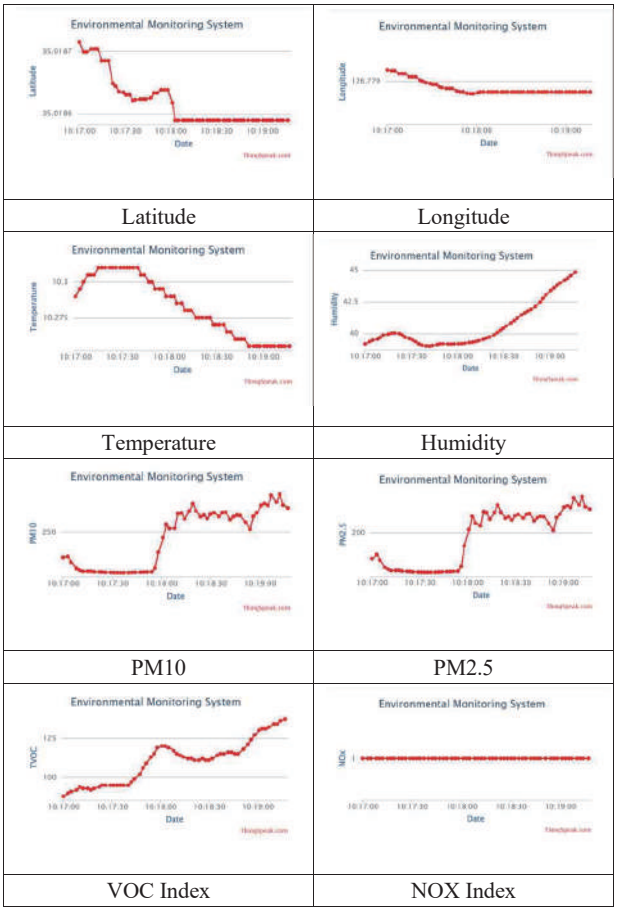


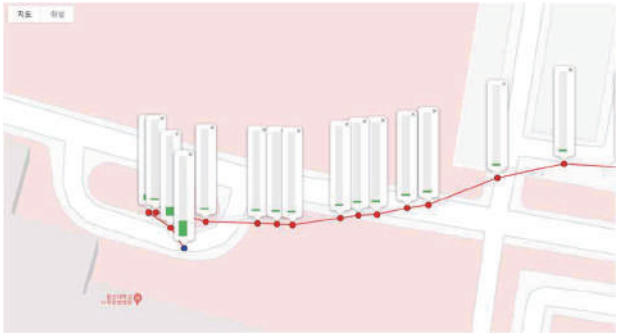
Fig. 3. Sensor data in ThingSpeak server.

The data stored on the cloud server was retrieved in JSON format and displayed on Google Maps API to show the location of personal mobility device and current sensor values, as illustrated in Fig. 4(a). Through the management page, it is possible to display multiple e-scooter, and selecting a stored marker displays the routing path and sensor values.

Fig. 4(b) presents an enhanced feature of the server system: heatmap visualization based on PM sensor data collected from personal mobility devices. This visualization provides a more comprehensive view of air quality variations along the travel route, allowing users to make informed decisions about air pollutions.



(a)



(b)

Fig. 4. Web based routing maps.

III. Experiments

To verify the operation of the proposed system, an IoT module was attached to an electric scooter, and data were collected while operating the scooter across the campus. The location information was stored on a server, and a separate website was developed to allow for the monitoring of operational data, as shown in Fig. 5.



Fig. 5. IoT sensor unit set-up in e-scooters.

The web page shown in Fig. 6 represents the route of the electric scooter during the experiments, with each marker clickable to present the measured sensor information.

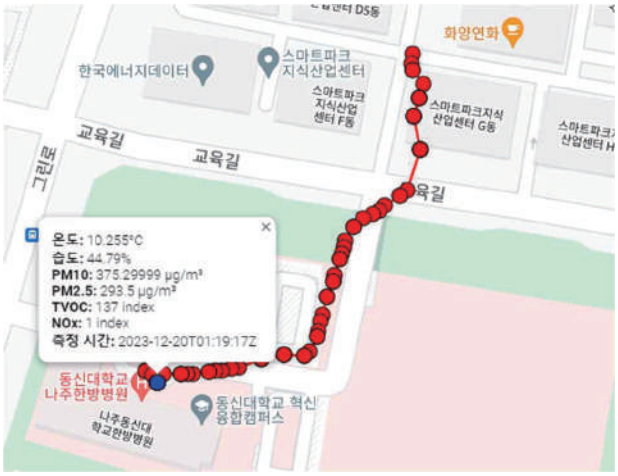


Fig. 6. Location routing of e-scooter.

IV. Conclusion

In this paper, a solution for Environmental awareness in urban environments using personal mobility devices with environmental sensors is presented.

Through LPWAN communications, compact urban vehicles become integrated with the smart city infrastructure. As compared to traditional approaches based on fixed stations, which involve costly deployments to cover wide areas, the application of IoT technologies into urban transportation emerges as a novel idea for advanced generation sensing.

Personal mobility devices enable the collection of detailed and real-time data on air pollution and weather conditions, improving accuracy when identifying unhealthy areas, as well as helping in decision making to reduce the air pollutants. The IoT sensor unit has been developed for this purpose for personal mobility devices. It includes low-power and wide-area IoT connectivity through LTE Cat-M1.

Acknowledgment

This research was supported by a grant (21163MFDS501) from Ministry of Food and Drug Safety in 2024.

References

[1] Waheb A. Jabbar, Thanasrii Subramaniam, Andre Emelio Ong, Mohd Iqmal Shu'Ib, Wenyan Wu, Mario A. de Oliveira, "LoRaWAN-Based IoT System Implementation for Long-Range Outdoor Air Quality Monitoring," Internet of Things, Volume 19, 100540, 2022.

[2] Espressif Systems, https://www.espressif.com/

[3] Sensirion AG, https://www.sensirion.com/

[4] U-blox, https://www.u-blox.com/

[5] CodeZoo, https://www.codezoo.co.kr/

[6] ThingSpeak, https://thingspeak.com/

Research On The Development Of Suction Pump Packages for Installation and Dismantling of Co2 Subsea Injection System

Hyoun Kang
Maritime ICT·Mobility Research Department
Korea Institue of Ocean Science & Technology(KIOST)
Busan, Republic of Korea
hkang@kiost.ac.kr

Osoon Kwon
Maritime ICT·Mobility Research Department
Korea Institue of Ocean Science & Technology(KIOST)
Busan, Republic of Korea
oskwon@kiost.ac.kr

Abstract*— In this paper, we conducted research on a suction pump package, which serves as the foundational equipment for CO2 subsea injection systems aimed at reducing carbon emissions. In South Korea, where securing onshore CO2 storage sites is challenging, the development of deep-sea CO2 injection technology utilizing the East Sea and the West Sea is crucial. This technology is primarily employed to connect CO2 emission sources such as power plants and refineries located along the coast to storage sites, making CO2 subsea storage systems indispensable. While research on CO2 subsea storage systems is ongoing domestically as a vital component of marine plants, there is a lack of investigation into CO2 fluid subsea storage systems for deep-sea conditions. Moreover, existing CO2 subsea injection facilities face various challenges during installation and decommissioning processes. To address these challenges, we developed a CO2 subsea injection suction pump package and conducted research on design technologies optimized for South Korea's subsea soil conditions. Additionally, we derived the design and fabrication techniques for a suction and pile batch installation package tailored to domestic subsea soil conditions, which can be operated under deep-sea conditions

Keywords*— Suction Bucket Foundation, Suction Pump System, CO2, CO2 Subsea Injection System, Carbon Emissions Reduction

I. INTRODUCTION

In recent years, carbon emissions and the resulting climate change have become a serious concern for the global environment. As a result, technological measures to reduce carbon emissions and reduce atmospheric CO2 concentrations are gaining importance. Subsea injection of CO2 is proposed as a solution to these problems, especially in the deep oceanic regions due to the limitations of onshore CO2 storage.

In Korea, deep-sea CO2 injection technology utilizing the East and West Seas is of great significance as it is difficult to secure onshore CO2 storage. These technologies operate by directly connecting CO2 emission sources, such as power plants and refineries located on the coast, with storage, and an essential component is a CO2 subsea storage system.[1-2]

In Korea, subsea storage systems have been researched and developed as major equipment for offshore plants, but they have only been applied to offshore oil and gas field production, and CO2 fluid subsea storage systems for large depths have not been studied. In addition, the foundation for CO2 subsea injection facilities, which are rapidly developing along with offshore platforms, has only been installed in some subsea production facilities in the Donghae gas field, which is a subsea gas production facility with a similar structure. [3-4]

However, conventional CO2 subsea injection facility foundations suffer from a number of problems during installation and decommissioning. New technologies and facilities are needed to overcome these challenges and create an efficient system.[5]

Since it is difficult to pile as the water depth increases, a mat-type foundation is used to expand the seabed contact surface using steel plates, etc. in the case of rigid seabed, and in the case of soft seabed, a suction bucket foundation is used to drain the water inside the bucket by using a pump after self-drilling the seabed to drill using the pressure difference between the inside and outside of the bucket, and a large-scale foundation that does not require large equipment for drilling and can be installed quickly is widely used. [6-7]

Therefore, this study aims to develop a suction pump package using a suction bucket foundation as the basis of a subsea storage system for the installation and dismantling of CO2 subsea injection system. A suction pump package optimized for domestic seabed ground conditions was studied to derive design technology, and a suction pump package that can be operated in large water depth conditions was developed. In addition, the installation format and design technology suitable for domestic seabed ground conditions were studied. [8]

II. SUCTION PUMP PACKAGE

A. Basic Design_Suction Bucket Foundation

An increasingly popular technique in suction pump system design is to utilize a pump performance curve that varies with the size of the suction bucket foundation. This technology is used to optimize the performance of the suction pump, and by adjusting the operating characteristics of the pump according to the size of the suction foundation, efficiency and reliability can be maximized. Although the domestic CO2 subsea injection system is still in the pre-commercialization stage, research has increased over the past decade along with the increasing demand for the technology, and in preparation for the commercialization of CO2 subsea injection system technology in 2030, the domestic market size of CO2 subsea injection system technology utilizing the Donghae gas field

since the mid-2020s is expected to continue to expand, including the existing Gunsan basin in the western part of Korea. Considering these domestic subsea ground conditions, we designed a suction pump package with an integrated installation/withdrawal package optimized for domestic subsea ground conditions.



Fig. 1. Subsea Structures Based on Suction Bucket Foundation

The suction bucket foundation is a foundation that uses the pressure difference between the inside and outside of the bucket to drain the water inside the bucket by using a pump after self-injection of the bucket foundation on the seabed, and it is a foundation that is highly utilized to support subsea equipment and structures on the seabed because it does not require large-scale equipment for injection, can install large-scale foundations at a fast speed, and has excellent load-bearing performance due to ground penetration.

B. Designing the Suction Pump Package

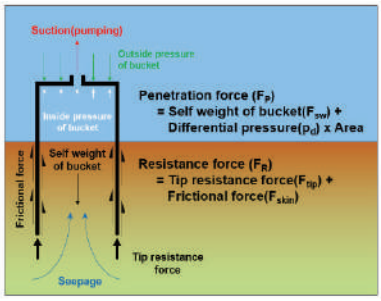


Fig. 2. Principle Of Suction Bucket Foundation

The design of the suction pump package optimized for domestic subsea ground conditions is based on the domestic ground environment. The magnetic field conditions are characterized by widespread distribution of thick sedimentary layers and soft ground, and multi-layered structure due to complex sedimentary environment. Therefore, it is important to derive foundation formats and design techniques that are suitable for this environment.

In addition, there is a need for a method to install suction pump systems at great depths without the use of ROVs. To this end, systems have been designed that can be installed remotely, eliminating the need for human (diver) and ROV operations. The system consists of a suction penetration pump, a suction withdrawal pump, and a control system, and is designed to operate reliably in deep water conditions.

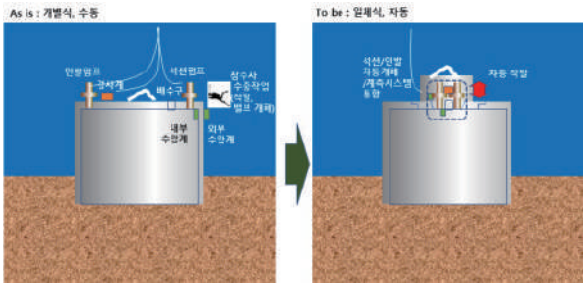


Fig. 3. Suction Pump Integration Package

The K1 model is used in shallow waters and is suction-only in the way it is used domestically, with the diver performing the suction installation and removal.

The K2 model is used in deep water and is used as a conventional method overseas, with suction only at home and ROVs used for removal overseas.

The K3 model is a conventional method used in deep water, which can only be suctioned domestically and fixed overseas.

The K4 model is the latest development in deep water technology, which can only be suctioned domestically and automatically separated overseas.

The K5 model is used as a deep water plant foundation and can only be suctioned domestically, but can also be suctioned and withdrawn and anchored offshore.

The K6 model is used as a plant foundation and anchor, and can only be suctioned domestically, but can also be suctioned and lifted offshore, and can be removed using an ROV.

The K7 model is used as a plant foundation and anchor, and can only be suctioned domestically, but can be suctioned and pulled out overseas, and can be automatically separated for domestic and overseas use.

In this study, the conventional K3 model used in deep water was designed as a four-pod system and a hydraulic drive system was developed to enable efficient operation of the suction pump even under deep water conditions. To this end, the stability and reliability of the suction pump was secured by designing and implementing an appropriate hydraulic drive system in consideration of the characteristics of the subsea ground. In addition, the internal structure and materials of the suction pump package were optimized through simulation to ensure pressure resistance performance. This ensured stable pressure-resistant performance that can cope with various environmental factors that may occur in deep water conditions.

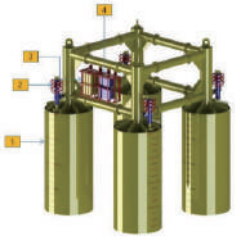


Fig. 4. 4-Pod Pump System : K3

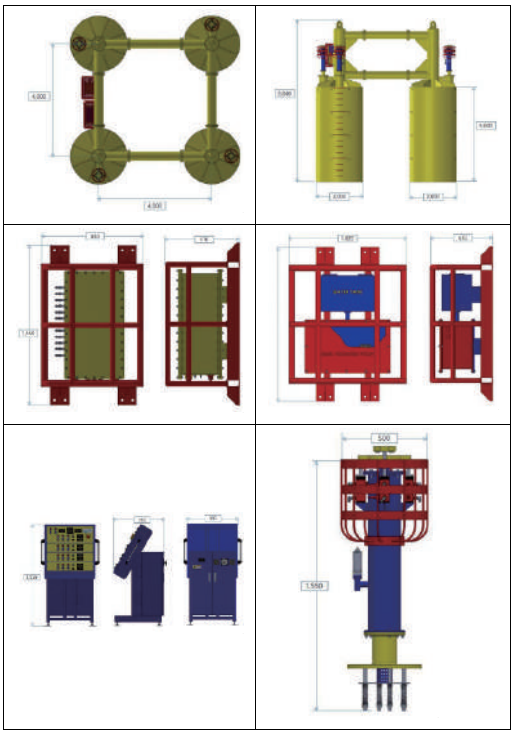


Fig. 5. Details of 4-Pod Pump System

i

C. Suction Pump Package System

i

Before entering the water, the suction pump closure and air vents are opened from the control panel. As you descend underwater, all the air in the suction base is expelled. Close the air vent when all air has been expelled and the suction base has settled to the bottom. After the suction operation is complete, close the suction pump closure and the operation is complete.



a. Venting Air b. Suction Operations c. Close Suction Pump

Fig. 6. Suction Pump System

The suction pump is responsible for pumping the water inside the suction pile by engaging the main body with the suction pile.

The air vent cylinder opens the air vent to release the air inside the suction pile and closes the seal hider by lowering it when all the air is released.

The Suction Pump Closure cylinder is used to close the suction pump to prevent water from escaping once the suctioning process is complete.

The valve pack consists of a block of 16 high-pressure solenoid valves with 4 high-pressure solenoid valves per cylinder (2 air vent cylinders and 2 suction pump cylinders). It receives signals from Air Vent ON/OFF and Closure ON/OFF and opens and closes the valves according to the respective commands.

The power pack consists of a water tank for use underwater as well as in the atmosphere, and a high-pressure pump that generates the water pressure to operate each cylinder.

The pressure sensor (inner) measures the pressure inside the suction pile, and the pressure sensor (outer) measures the pressure outside the suction pile.

The tilt sensor is connected to the main unit to measure the tilt of the suction pile.



i

Fig. 7. Details of Suction Pump Systemi

III. Conclusion

In this paper, a suction pump package technology utilizing a suction bucket base for CO2 subsea injection facility to respond to environmental problems caused by carbon emissions and climate change is studied.

In Korea, due to the limitations of onshore CO2 storage, the basic technology of CO2 subsea injection facility storage system in large depth areas of the ocean is becoming important. Therefore, we developed a basic system for CO2 subsea injection facility storage system and studied various technologies to design a suction pump system suitable for multi-layer structure utilizing the subsea ground in the East Sea and West Sea.

As a result, the installation format and design technology of the suction pump package of the CO2 subsea injection system were derived considering the subsea ground conditions in Korea. In particular, by developing a suction pump package that can be operated in large depth conditions, it was possible to install and dismantle the CO2 subsea injection system suitable for the large depth environment in the domestic region, and the hydraulic drive system was developed to secure the stability and pressure resistance of the suction pump and eco-friendly technology.

ACKNOWLEDGMENT

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry, and Energy (MOTIE) of the Republic of Korea (No. 20226A10100020)

REFERENCES

[1] Cao, Y., Yang, C., Yan, Z., Xu, J., Liu, H., & Xiong, Y. (2019). Design and implementation of intelligent subsea pumping system for CO2 sequestration. Journal of Petroleum Science and Engineering, 181, 106317i

[2] [2] Gao, C., Yang, C., Xu, J., Yan, Z., Liu, H., & Xiong, Y. (2020). Experimental study of an underwater pumping system for CO2 injection. Applied Energy, 268, 115016i

[3] [3] Cai, H., Yang, C., Guo, L., Wei, H., & Wang, Y. (2018). Development of a subsea pumping system for CO2 injection in deep water. Energy Procedia, 152, 1062-1067i

[4] [4] Yang, Z., Dong, J., Xu, H., Jiang, G., Li, X., Wang, K., & Zhang, X. (2019). Design and Experiment of a Downhole Extraction System for Carbon Dioxide Sequestration. Energies, 12(21), 4095.i

[5] [5] Kim, S., Cho, H., Jeon, D., & Park, H. (2017). Design of a subsea pumping system for CO2 injection. International Journal of Precision Engineering and Manufacturing-Green Technology, 4(3), 311-317i

[6] [6] Li, J., Zhang, Y., Li, W., Huang, H., & Chen, Y. (2019). Research on the key technologies of subsea pumping system for CO2 injection. Energy Procedia, 158, 6051-6056i

[7] [7] Ji Seong-Hyeon, Lee Gyu-Yeol, Yun Jun-Ung, Kim Dong-Jun, and Choe Jae-Hyeong. "Trends in Submarine Marine Plant and Development of Submarine Foundation Technology." Geotechnical Engineering Journal 30.11 (2014): 12-17.i

[8] [8] Zhang, G., Zhang, X., Zhang, Y., Zhang, B., Wang, W., & Xue, Z. (2016). Preliminary Study on Monitoring and Evaluation of CO2 Geological Storage in Jilin Oilfield. Energy Procedia, 104, 256-261.i

ICMIC 2024

The 3rd International Conference on
Mobile, Military, and Maritime IT Convergence

Proceedings of International Conference on Mobile, Military, and Maritime IT Convergence (eISSN: 2983-3132)

http://icmic-conf.org/