Explainable Quantum-empowered Antispoofing Intelligence for Trustworthy Connected Autonomous Vehicles Communication

Simeon Okechukwu Ajakwe *MIEEE*, Dong-Seong Kim *SMIEEE*
Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
simeonajlove@gmail.com, dskim@kumoh.ac.kr

***Abstract*—Connected Autonomous Vehicles (CAVs) rely on secure and trustworthy communication for safe and efficient operation. Spoofing attacks, where malicious entities inject false information, pose a significant threat to CAV communication. This paper proposes a novel framework, Explainable Quantum-Empowered Antispoofing Intelligence (EQAI), to address this challenge. EQAI leverages the principles of quantum information theory to enhance spoofing detection capabilities and integrates explainable AI (XAI) to provide transparency and interpretability in the decision-making process. Results show that the EQAI exhibited better performance compared to classical approaches in accuracy, latency, robustness, and communication overhead.**

***Index Terms*—Connected Autonomous Vehicles (CAVs), Explainable AI (XAI), Spoofing Attacks, Quantum Information Theory, Trustworthy Communication, V2X.**

I. INTRODUCTION

Connected Autonomous Vehicles (CAVs) are poised to revolutionize transportation, offering enhanced safety, efficiency, and convenience. However, the reliance on Vehicle-to-Everything (V2X) communication makes them vulnerable to various cyberattacks, including spoofing [1]. In a spoofing attack, a malicious entity impersonates a legitimate entity to inject false information into the communication network. This can have catastrophic consequences in the context of CAVs, leading to incorrect driving decisions, accidents, and even loss of life. Traditional cybersecurity measures, such as encryption and authentication, can mitigate some aspects of spoofing attacks [2]. However, the dynamic and complex nature of V2X communication requires more sophisticated solutions. Classical machine learning-based methods have been explored for spoofing detection, but they often lack transparency and can be susceptible to adversarial attacks [3].

To address these limitations, this paper proposes a novel framework, Explainable Quantum-Empowered Antispoofing Intelligence (EQAI), which combines the principles of quantum information theory and explainable AI (XAI) [4]. Quantum information theory provides a theoretical foundation for secure communication and spoofing detection, leveraging the unique properties of quantum states [5]. XAI enhances the transparency and interpretability of the detection process, providing insights into why a particular message is classified as spoofed or legitimate for informed decision-making.

In this paper, Section II describes the system design and methodology, Section III presents the simulation results, and Section IV concludes the paper.

II. SYSTEM DESIGN AND METHODOLOGY

The proposed EQAI framework comprises three main components: (1) Quantum-Enhanced Spoofing Threat (QUEST) detection module, (2) Explainable Hybrid AI ($\widehat{x}$AI) module, and (3) Trust Assessment and Decision-making (TAD) module, as seen in Fig. 1.

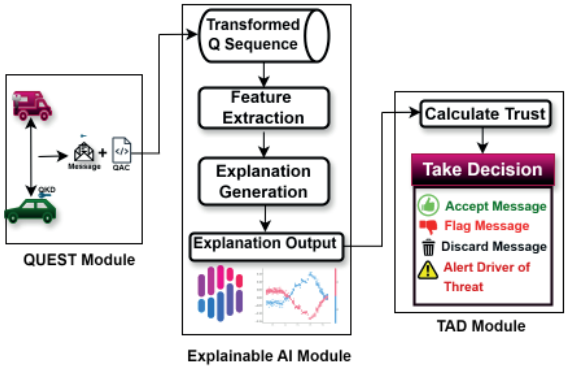

Fig. 1. EQAI Architecture highlighting the (1) QUEST module; (2) $\widehat{x}$AI module; and (3) TAD module for trustworthy V2V and V2X communication.

A. Quantum-Enhanced Spoofing Threat Detection (QUEST)

Unlike classical information, the QUEST component takes advantage of the superposition and entanglement properties of quantum information to detect subtle deviations caused by spoofing attacks. QUEST involves (4) processes to intercept spoofing attacks: (i) quantum key distribution (QKD) by establishing a shared secret key between communicating vehicles for authentication; (ii) generating a quantum authentication code (QAC) using the shared key and appending it to the message; (iii) Transmitting the message with the QAC as a sequence of quantum states; and (iv) verifying the QAC by measuring the quantum states using the shared secret key. If QAC is corrupted beyond a tolerable threshold, the message is flagged as potentially spoofed.

B. Explainable Hybrid AI Module ($\widehat{x}$AI)

While quantum mechanics provides enhanced security, it can be opaque. The $\widehat{x}$AI module addresses this by providing explanations of the QUEST module decisions. This includes:

- **Extract Relevant Features** by defining a quantum circuit $U_\phi(x')$ that encodes the classical feature vector $x' \in \mathbb{R}^n$ into a q-qubit quantum state.

$$U_\phi(x') = \bigotimes_{i=1}^{q} R_z(x'_i) H_i \quad \text{if } q = n \tag{1}$$

- **Define Variational Quantum Circuit (VQC)** define a parameterized quantum circuit $V(\Theta)$ with L layers, where Θ represents the set of trainable parameters (rotation angles). Each layer consists of single-qubit rotations and entangling gates (e.g., CNOT).

$$V(\Theta) = V_L(\theta^{(L)}) \cdots V_1(\theta^{(1)}) \tag{2}$$

where $\theta^{(l)}$ are the parameters in the l-th layer.

- **Define Quantum Neural Network (QNN) Output** as the expectation value of a set of observables $\mathcal{O} = \{O_j\}_{j=1}^{M}$ on the final quantum state:

$$q_{out}(x';\Theta) = \langle 0|^q (U_\phi(x')^\dagger V(\Theta)^\dagger \mathcal{O} V(\Theta) U_\phi(x'))|0\rangle^q \tag{3}$$

where $|0\rangle^q$ is the initial state of the q qubits. For classification into C classes, we used Pauli Z ($M = q$) measurements on each qubit and then processed these measurements classically.

- **Define Classical Neural Network** $g(z; W)$ with weights W, which takes the output of the QNN (or a processed version of it) as input and produces the final classification.

$$z = \text{Process}(q_{out}(x';\Theta))\hat{y} \,||\, g(z;W) \tag{4}$$

where Process = desired output dimension with more complex layers; and g = final output layer with C neurons and C classes with a softmax activation function.

- **Define Hybrid Model** as $f(x';\Theta,W) = g(\text{Process}(q_{out}(x';\Theta));W)$ having a cross-entropy loss function defined as:

$$\mathcal{L}(\hat{y}, y) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{5}$$

- **Hybrid Model Training & Evaluation** using trainable parameters Θ and W and updating the parameters using the classical optimizer $\mathcal{O}$ with learning rate α:
$\Theta \leftarrow \Theta - \alpha \cdot \mathcal{O}(\nabla_\Theta L)$ and $W \leftarrow W - \alpha \cdot \mathcal{O}(\nabla_W L)$

C. Trust Assessment & Decision-Making (TAD) Module

The TAD module integrates the output of the QUEST and $\hat{x}$AI modules to assess the trustworthiness of the received messages and make informed decisions for drivers. The GPS spoofing detection for autonomous vehicles dataset from IEEE dataport [6] was used to train the proposed model. It has 158,170 samples, 13 features, 55% legitimate samples (0), and 45% spoof attacks; `Simplistic(1)`, `Intermediate(2)`, and `Sophisticated(3)`. Simulation was carried out in a Python environment using PennyLane and PyTorch frameworks.

III. Result and Performance Evaluation

TABLE I
EQAI PERFORMANCE EVALUATION USING CLASSICAL AI METRICS

Model	Acc (%)	Prc (%)	Rec (%)	F1 (%)	Loss	Time(s)
MLP	91.90	91.48	91.90	91.20	0.2746	1691.13
EQAI	80.36	77.19	80.04	75.82	0.8360	7506.04

From Table I, the EQAI performed relatively well in determining the legitimacy of a received vehicular message compared to classical ML, though with higher training time.

Furthermore, Fig. 2(a) demonstrates that the EQAI predicted legitimate messages (`Class 0`) with the highest probability of 0.52 as against 0.24, 0.16, and 0.05 for simplistic (`Class 1`), intermediate (`Class 2`), and sophisticated (`Class 3`) spoofing attacks. The features PD (pseudo-range), CP (carrier phase cycles), and CNO (carrier-to-noise ratio) provided the strongest support for determining a legitimate message (class 0), outweighing the relatively minor opposing influence of TOW, RX, and TCD. Hence, class 0 received the highest confidence score of 0.54 and was selected as the final prediction by EQAI.

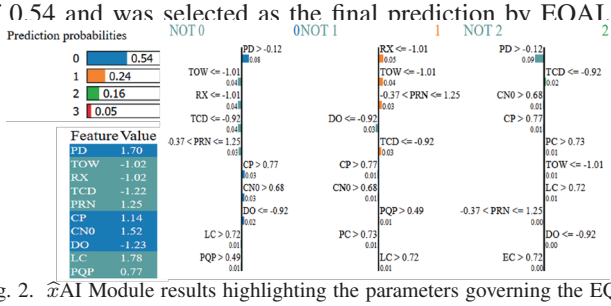

Fig. 2. $\hat{x}$AI Module results highlighting the parameters governing the EQAI-VQC decision to determine the authenticity of a received message by vehicles.

Finally, the EQAI had a total communication overhead of 0.0030MB and trainable parameters of 0.0004MB, better than MLP with 3108MB, meaning an increase in convergence speed. The estimated computational complexity of EQAI is $\mathtt{O(13log13)} \approx \mathtt{O}(4.7 \times 10^5)$ due to its hybrid classical-quantum structure, unlike the polynomial complexity of MLP $[\mathtt{O(N * 6016)} \approx \mathtt{O}(7.61 \times 10^8)]$, and it has a moderate robustness to sophisticated attacks.

IV. Conclusion

This paper presented a novel Explainable Quantum-Empowered Antispoofing Intelligence (EQAI) framework for trustworthy connected autonomous vehicles communication. By leveraging the explainability capability of XAI, the EQAI framework is more robust to sophisticated spoofing attacks that might evade classical detection methods by mimicking legitimate signals. Future work will improve the framework and adapt it to other attacks on CAVs for informed decision.

Acknowledgment

This research was supported by the Priority Research Centers Program through the NRF funded by the MEST (2018R1A6A1A03024003) (50%) and by MSIT under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2024-2020-0-01612) (50%) supervised by the IITP

References

[1] S. O. Ajakwe and D.-S. Kim, "Facets of security and safety problems and paradigms for smart aerial mobility and intelligent logistics," *IET Intelligent Transport Systems*, vol. 18, pp. 2827–2855, 2024.

[2] M. M. Khan, M. Kamal, M. Shabbir, and S. Alahmari, "Enhancing autonomous vehicle security: Federated learning for detecting gps spoofing attack," *Transactions on Emerging Telecommunications Technologies*, vol. 36, no. 4, p. e70138, 2025.

[3] M. Aledhari, R. Razzak, M. Rahouti, A. Yazdinejad, R. M. Parizi, B. Qolomany, M. Guizani, J. Qadir, and A. Al-Fuqaha, "Safeguarding connected autonomous vehicle communication: Protocols, intra-and inter-vehicular attacks and defenses," *Computers & Security*, p. 104352, 2025.

[4] V. U. Ihekoronye, S. O. Ajakwe, J. M. Lee, and D.-S. Kim, "Droneguard: An explainable and efficient machine learning framework for intrusion detection in drone networks," *IEEE Internet of Things Journal*, 2024.

[5] U. Ahmad, M. Han, and S. Mahmood, "Enhancing security in connected and autonomous vehicles: a pairing approach and machine learning integration," *Applied Sciences*, vol. 14, no. 13, p. 5648, 2024.

[6] G. Aissou, S. Benouadah, H. E. ALAMI, and N. Kaabouch, "A dataset for gps spoofing detection on autonomous vehicles," 2022. [Online]. Available: https://dx.doi.org/10.21227/8x3h-2817