https://doi.org/10.7840/kics.2025.50.12.1958

정적메모리 모니터링을 통한 딥러닝 기반 원자력 발전소내 사이버 공격 탐지

임 규 현◆, 신 수 용°

Detection of Cyber Attacks within Nuclear Power Plants Using Deep Learning-Based Monitoring of Static Memory

Gyu Hyun Lim◆, Soo Young Shin°

요 약

원전내의 사용되는 모듈을 포함하는 기기들은 U.S. NRC (Nuclear Regulatory Commission) 규제요건에 의해 동적 메모리가 아닌 정적 메모리를 사용하도록 규제하고 있다. 따라서, 본 연구에서는 정적 메모리의 사용도 변화에 따른 정상/비정상 상태를 확인하기 위한 딥러닝 기반의 공격 감지 시스템을 제안한다. 제안된 시스템은 메모리 사용량에 대한 시계열 데이터를 MTF(Markov Transition Field)를 이용하여 시각화한다. LSTM Auto-Encoder(AE)를 이용하여 공격을 감지하고, 이를 통해 재구성된 데이터를 시각화하여 공격 종류를 분류하는 모델을 제안하며 그 성능을 분석한다. 분석 결과 제안된 시스템이 정적 메모리 시스템에서의 공격에 대한 정확한 감지 및 분류를 수행함을 보여준다.

키워드 : LSTM 오토인코더, 사이버보안, 메모리 분석, 마르코프 전환, 합성곱 네트워크
Key Words : LSTM Auto-Encoder, Cyber Security, Memory Analysis, Markov Transition Field, CNN

ABSTRACT

The Device that include modules used within a nuclear power plant is regulated to employ static memory rather than dynamic memory, as per the U.S. NRC (Nuclear Regulatory Commission) regulatory requirements. Therefore, this research proposes a deep learning-based attack detection system to identify normal and abnormal states in accordance with changes in static memory usage. The proposed system visualizes time-series data of memory usage using Markov Transition Field (MTF) and employs LSTM Auto-Encoder (AE) for attack detection. It further proposes a model to classify attack types by visualizing the reconstructed data and analyzes its performance. The analysis results demonstrate that the proposed system effectively detects and classifies attacks in static memory systems.

◆ First Author : Kumho National Institute of Technology, Department of Digital Convergence Engineering, total8100@kepco-enc.com, 정회원
° Corresponding Author : School of Electronics Engineering, Kumoh National Institute of Technology, wdragon@kumoh.ac.kr, 종신회원
논문번호 : 202504-084-E-RN, Received April 9, 2025; Revised July 2, 2025; Accepted July 28, 2025

Ⅰ. 서 론

오늘날 점차적으로 연결되고 디지털화 되는 세상에서 원자력 발전소와 같은 중요한 기반 시설 시스템은 정교한 사이버 위협에 노출되고 있다[1]. 한국원자력통제기술원의 자료에 의하면 국내 원전에 대한 사이버 공격은 2018년부터 2022년 8월까지 918건 발생했다. 이들 중 대부분은 원전 홈페이지나 e메일 시스템 등 업무 지원 시스템이 주요 공격 대상이었다. 그러나 실제로 2019년 인도 쿠단 쿨람 원전이 악성코드에 감염된 사례도 발생했다. 해당 사례는 외부와 단절된 폐쇄 망이므로 안전하다고 생각한 사회적 통념에 허를 찌르는 사건이었고, 이에 대해 원전 사이버보안 규제 기관은 심층 방호체계를 정립하여 각 필수 기기들의 등급 부여를 통해 관리하도록 하였다.

위 사건과 같은 원전 시스템에 대한 사이버 공격은 운영 장애에서부터 안전 및 보안 저해까지 심각한 결과를 야기할 수 있다. 따라서 중요한 국가 시설을 보호하기 위한 견고한 사이버 공격 메커니즘이 필수적인 상황이다.

U.S. Reg. Guide 1.152 (Rev.3), “Criteria for Use of Computers in Safety Systems of Nuclear Power Plants”에서는 원전을 사이버 위협으로부터 보호하기 위해 Secure Develop ment and Operational Environment (SDOE)에 대한 규제요건을 정립하고 있으며, 중요 기능을 수행하는 핵심 장비 및 기기에는 동적 메모리가 아닌 정적 메모리를 활용하도록 규제하고 있다[2]. 이를 위해 기존의 연구들에서는 동적 스케줄링 분석을 통해 충분한 가용 메모리가 있는지를 분석하여 보이는 방법을 활용하거나 최소한 몇 개의 프로그램들을 일괄적으로 메모리에 상주시키는 메모리 폐쇄(memory locking)를 통해 안전 기능의 신뢰성을 유지하고자 하였다[3]. 그러나 이러한 방법들은 정적 메모리를 활용하는 방안을 제시하였으나 안전 기능의 신뢰도에 영향을 미칠 수 있는 사이버 위협에 대한 대응책은 제시하지 않았고, 현재 원전에 대해 관련한 선행 연구는 거의 전무하다 시피하다. 따라서 본 논문은 원자력 발전소 내에서 사이버 공격 탐지를 강화하기 위한 새로운 접근 방식으로 이상 탐지(Anomaly Detection) 기법을 활용하는 것을 제안한다.

이상 탐지(Anomaly Detection)이란, 정상과 비정상으로 구분되는 상태를 탐지하는 것으로 크게 비지도 이상 탐지(Unsupervised Anomaly Detection)와 지도 이상 탐지(Supervised Anomaly Detection)로 구분된다. 이상 탐지는 시계열 데이터분석, 기기 상태 검증, 이상 거래 감지, 영상분석 등 다양한 분야에서 활용되고 있으며, 데이터 종류에 따라 다양한 접근법이 존재하고 있다.

본 연구에서는 실제 원전 운전 데이터의 경우 원자력안전법[4]에 의해 활용이 불가능하므로 원전에서의 이상 탐지를 위해 Kaggle에서 제공하는 ‘Malware Memory Analysis’ 라는 Malware별 메모리 사용량 분석 자료를 참조하여 필요한 열들을 수정하여 작성한 데이터 셋을 활용하였다[5].

해당 데이터는 CIC(Canadian Institute for Cyber security)에서 제공하는 자료이며, 산업계에서 일반적으로 사용하는 Malware의 메모리 점유율에 대한 데이터 셋이지만 원자력 발전소에서도 일반 산업계에 사용하는 상용품들을 안전 기능을 담당하는 기기에 활용하기 위한 방안으로 CGID(Com

mercial Grade Item Dedication)이라는 평가 제도를 국제 표준에 따라 EPRI TR-106439, “Guideline on Evaluation and Acceptance of Commercial-Grade Digital Equipment for Nuclear Safety Application[6]”를 통해 제시하고 있으므로 해당 데이터 셋을 분석한 자료가 원자력 발전소내의 안전 기기에도 적용이 가능할 것으로 판단된다.

규제 기관에서 특히 강조하고 있는 안전 기능을 수행하는 주요기기에 이와 같은 Malware가 침투되어 적시에 해당 기능이 동작하지 않는다면 천문학적인 재난 비용은 물론 인명 피해가 발생할 수도 있으므로 이를 위한 견고한 사이버 공격 감지가 필요하다.

본 연구의 주요 목표는 RNN(Recurrent Neural Network)의 일종인 LSTM(Long Short-Term Memory) 기반 Auto Encoder를 활용한 메모리 이상 탐지 시스템을 제안하여 원자력 발전소 내 주요 안전기기 또는 모듈의 사이버 위협 감지기능 및 신뢰성을 제고시키는 것이다.

Ⅱ. 시스템 구성

그림 1은 제안하고자 하는 사이버 공격 탐지 시스템의 개념도이다.

2.1 LSTM Auto-Encoder(LSTM AE)

LSTM AE[7]는 입력된 데이터를 재구성(복원)해내는 기능을 가진 알고리즘으로 입력 데이터인 정상 데이터에 대한 특징을 학습하고 학습된 모델에 데이터를 집어넣었을 때 재구성한 결과와 정상 특징과의 차이점을 비교해 이상 여부를 판단할 수 있다.

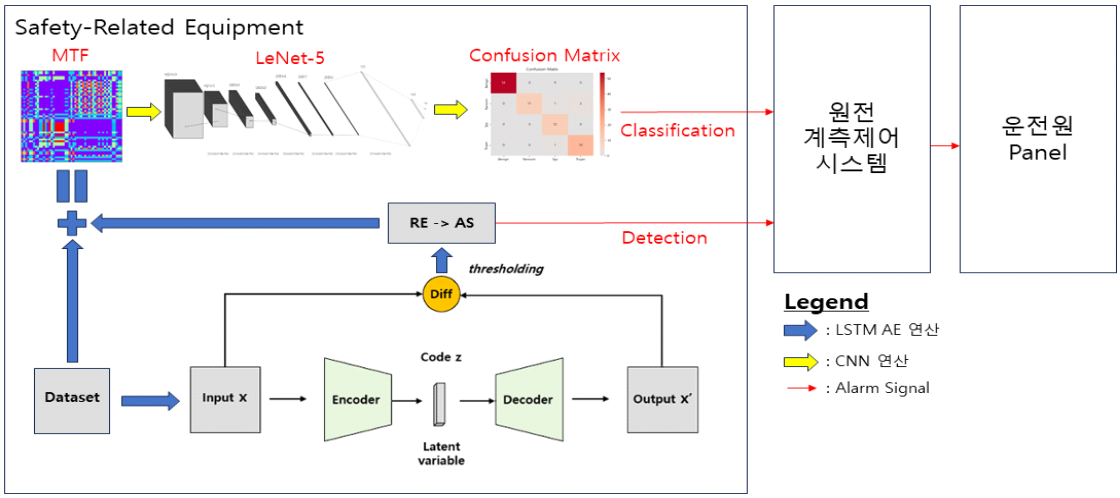


그림 1. 제안하는 사이버 공격 탐지 개념도
Fig. 1. System Configuration

2.2 Markov Transition Field(MTF)

MTF Algorithm[8]은 이산화한 시계열 데이터의 전이 확률을 나타내는 알고리즘이다. MTF를 구성하기 위해 주어진 시계열 데이터 집합 $\boldsymbol{X}$를 값에 따라 Q개의 구간으로 나눈 뒤, 시간 인덱스 t_i의 시계열 데이터 값 x_i에 맞는 구간 q_j(j∈[1,Q])에 배정한다. 시간 축을 따라 1차 마르코프 체인 방식으로 Q x Q 크기의 가중치 인접 행렬 W를 구성한다.

$w_{i,j}$는 q_i구간에서 q_j구간으로 전이하는 빈도를 나타낸다. $\boldsymbol{W}$의 각 열의 합을 1로 정규화 함으로서 마르코프 전이 매트릭스를 구성할 수 있다. 이 과정에서 $\boldsymbol{W}$는 각 $\boldsymbol{X}$의 분포와 시간 인덱스 t_i에 대한 시간 종속성이 제거된다. 이러한 $\boldsymbol{W}$의 정보 손실을 극복하기 위해, 시간 순서를 따라 각 확률을 정렬하여 MTF를 다음과 같이 정의한다.

MTF의 i행 j열값인 $\boldsymbol{M}_{i,j\,|\,|i-j|=k}$는 시간 인덱스 t_i일 때의 데이터 값이 속한 구간 q_i에서 시간 인덱스 t_j의 데이터 값이 속한 구간 q_j로 전이할 확률로 두 시간의 차이(|i - j|)가 k인 지점 간 전이확률을 나타낸다. 구간의 너비가 크면 대부분의 값을 평균에 가장 가까운 구간으로 집계하고, 구간의 너비가 작으면 극한 구간에서 집계되는 값이 적어진다.

$$W = \begin{bmatrix} w_{11} \mid P(x_t \in q_1 \mid x_{t-1} \in q_1) & \cdots & w_{1Q} \mid P(x_t \in q_1 \mid x_{t-1} \in q_Q) \\ w_{21} \mid P(x_t \in q_2 \mid x_{t-1} \in q_1) & \cdots & w_{2Q} \mid P(x_t \in q_2 \mid x_{t-1} \in q_Q) \\ \vdots & \ddots & \vdots \\ w_{Q1} \mid P(x_t \in q_Q \mid x_{t-1} \in q_1) & \cdots & w_{QQ} \mid P(x_t \in q_Q \mid x_{t-1} \in q_Q) \end{bmatrix}$$

그림 2. 가중치 인접 행렬 W의 구조
Fig. 2. Structure of Weight Adjacency Matrix W

$$M = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \ddots & \vdots & \\ M_{n1} & M_{n2} & \cdots & M_{nn} \end{bmatrix} = \begin{bmatrix} w_{ij} \mid x_1 \in q_i, x_1 \in q_j & \cdots & w_{ij} \mid & x_1 \in q_i, x_n \in q_j \\ w_{ij} \mid x_2 \in q_i, x_1 \in q_j & \cdots & w_{ij} & x_2 \in q_i, x_n \in q_j \\ \vdots & & \ddots & \vdots \\ w_{ij} \mid x_n \in q_i, x_1 \in q_j & \cdots & w_{ij} & x_n \in q_i, x_n \in q_j \end{bmatrix}$$

그림 3. MTF 행렬 M의 구조
Fig. 3. Structure of MTF Matrix M

위와 같이 MTF 연산을 통해 구해진 시계열 이미지는 pyts. image.MarkovTransi tionField API로 구현되며, pyts.image. MarkovTransitionField 클래스의 trans from() 메서드는 시계열 데이터 집합을 입력으로 받아 데이터 집합의 첫 번째 요소인 그라미안 각도 필드의 데이터로 구현된다.

2.3 LeNet-5 모델 구성

LSTM AE와 MTF를 통해 재구성한 시각화된 데이터를 활용하여 학습을 진행시키기 위한 모델로 가장 널리 알려진 CNN 모델 중 하나인 LeNet-5[9]의 모델 구성을 조정하여 활용하였다.

2.4 이상 탐지

본 연구에서는 지도 학습으로 CNN의 일종인 LeNet-5와 비지도 학습으로 LSTM AE를 사용하였다.

LSTM AE는 데이터의 라벨링이 필요하지 않으며, 입력 데이터와 출력 데이터가 유사한 방향으로 학습된다.

학습 시 인코더를 통해 입력 데이터의 특징을 추출하여 Hidden Layer를 거쳐 잠재공간(Latent space)으로 저장한 후 다른 한 부분인 디코더에서 원래의 데이터를 복원하는 과정을 거친다. 보통 복원하는 과정에서 손실이 발생하게 되고, 이 때문에 복원된 데이터와 원본 데이터 사이에 간극이 생기게 된다. 최종적으로 학습이 종료된 이후에는 새로 입력된 데이터와 원본 데이터 사이에는 주요하지 않는 특징들을 제외시키고, 기존 학습 원본 데이터와 유사한 방향으로 복원한다.

이 때 복원 시 계산되는 Reconstruction Error(RE)와 이를 통해 산출된 Anomaly Score(AS)값을 추가하여 재구성한 각 메모리 데이터를 MTF 과정을 통해 시각화하고, 시각화된 데이터를 CNN의 일종인 LeNet-5로 모델을 학습하여 비정상 상태를 판별할 수 있다.

2.5 원전 계측제어시스템(MMIS)[10]

원전 계측제어시스템(MMIS-Man Machine Interface System)은 원전의 상태를 계측, 제어하고 감시하며 보호하는 원전의 안전 운전에 있어 핵심적인 설비이다. 원자력 발전소 전체의 다양한 변수 또는 시스템의 연속적인 감시를 통해 기 설정된 운전범위 이내로 시스템을 유지하는 기능을 수행한다.

(Fig. 4)는 발전소보호계통의 단일 채널에 대한 단순 구성도이다. 발전소보호계통은 PLC (Programmable Logic Controller)를 통해 감시되는 변수의 입력값을 처리하고, 원자로 정지 등의 안전 기능을 수행한다. 이러한 발전소보호계통을 구성하는 PLC는 입출력 기기, 메모리, 통신모듈 등 다양한 하부기기를 감시 및 제어하고, 안전필수 기능을 수행하는 응용 프로그램을 실행시키기 위한 실시한 운영체제를 탑재하고 있다. 안전필수 계측제어시스템에 탑재되는 실시간 운영체제는 특히 결정론적 성능 특성을 유지하기 위하여 정적 메모리 할당 기법을 통해 메모리 및 자원에 대한 접근 시간 제한치를 가질 수 있도록 설계되고, 메모리 예측 가능성이 확보되도록 설계한다.

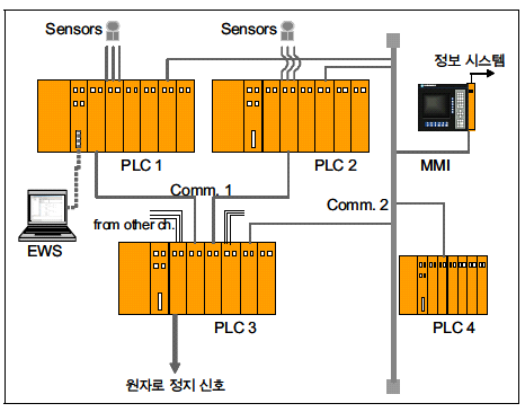


그림 4. 발전소 보호 시스템 단일 채널 구성도
Fig. 4. Power Plant Protection System Single Channel Configuration Diagram

Ⅲ. 학습 프로세스

3.1 데이터 수집

본 논문에서 제안하는 메모리 이상 탐지를 통한 사이버 공격 감지를 위해 학습 데이터는 케글(Kaggle)에 공개된 'Malware Memory Analysis / CIC-MalMem-2022' 자료를 수정하여 사용하였다. 해당 데이터셋에는 정상 상태(Benign)와 비정상 상태(Malware)의 프로세스별 메모리 사용 정도가 데이터화 되어 있고, 이에 시간 데이터(Timestamp)열을 추가하여 로그 파일과 같이 변형하여 활용하였다.

3.2 데이터 구성

구현한 LSTM AE모델의 학습 및 시험을 위해 본 연구에서는 5만 8천여 개의 55개 프로세스 데이터 세트를 활용하여 시계열 데이터를 생성하였다. 데이터는 크게 정상 데이터(Normal)과 비정상 데이터(Abnormal)로 구성하였고, 정상구간을 4개(Sn, Vn_1, Vn_2, Tn), 비정상구간을 2개(V_A, T_A)로 나누어 학습에 활용하였다.

학습용 데이터(S_N)는 모델을 학습하는데 사용하는 데이터이다. 이 데이터를 활용하여 정상구간의 정보를 압축할 수 있도록 모델을 학습한다. 또한, 학습된 모델을 활용하면 특정 데이터 구간이 입력으로 들어왔을 때 추론 과정을 통해 구간별 RE를 구할 수도 있다.

파라미터 추정용 데이터(V_{N1})는 RE분포의 파라미터를 추정하는데 활용한다. RE가 정규분포를 따른다고 가정하고 정규분포의 파라미터 N(μ, ∑)를 Maximum Likelihood Estima tion(MSE)을 활용하여 구한다. 이후 아래와 같은 식을 활용하여 각 구간의 비정상 점수(AS)를 계산할 수 있다.

[Anomaly Score 계산식]

$$a^{(i)} = (e^{(i)} - \mu)^T \sum^{-1} (e^{(i)} - \mu) \qquad (1)$$

$a^{(i)}$: i지점의 비정상 점수

이 비정상 점수($a^{(i)}$)가 사용자가 지정한 Threshold (τ)를 상회하여 $a^{(i)}>\tau$ 인 이 지점을 비정상이라고 정의한다. 이와 같은 방법으로 추론 단계에서 각 지점 및 구간의 비정상 여부를 판단할 수 있다. 학습용 데이터를 이용하여 모델을 학습하고, 파라미터 추정 데이터를 활용하여 비정상을 정의한 뒤 검증용 데이터(V_{N2}, V_A)를 활용하여 정상과 비정상 구간을 잘 분류하는지 확인한다. 검증 데이터를 활용하여 최종 학습모델과 최종 파라미터를 도출한 뒤 테스트 데이터(T_N, T_A)에서 모델의 최종 성능을 도출한다.

3.3 모델훈련 과정

인코더는 입력으로 n개의 연속한 벡터 $x^{(1)}$, $x^{(2)}$, … , $x^{(n)}$를 사용한다. 입력 $x^{(k)}$는 다변량 데이터이므로 변수의 개수 m개로 구성된 벡터($x^{(k)} \in R^m$)이다. 인코더는 매 단계의 입력으로 $x^{(t)}$와 이전 단계에서 인코더로부터 받은 은닉 벡터인 $h_E^{(t-1)}$을 활용하여 정보를 압축하고 다음 단계의 은닉 벡터인 $h_E^{(t)}$를 생성한다. 이와 같은 방식으로 인코더의 마지막에 단계에서 생성된 $h_E^{(t)}$는 특징 벡터로 부르며 디코더의 초기 은닉 벡터로서 활용된다.

디코더는 입력으로 인코더에서 생성한 특징 벡터를 받아 원본 데이터를 역순으로 재구성한다. 즉, 디코더는 $x^{(\widehat{n})}$, $x^{(\widehat{n-1})}$, .., $x^{(\widehat{1})}$를 차례로 생성한다. 이때 디코더의 매 단계 입력으로 원본 데이터의 역순인 $x^{(t+1)}$과 이전 단계에서 디코더로부터 받은 은닉 벡터인 $h_D^{(t-1)}$을 활용하여 정보를 압축하고 다음 단계의 은닉 벡터인 $h_D^{(t)}$를 생성한다. 다음 단계의 디코더에 은닉 벡터 $h_D^{(t)}$를 전달하기 전에 완전연결 선형계층를 통과시켜 복원 데이터인 $x^{(n-\widehat{t+1})}$를 생성한다.

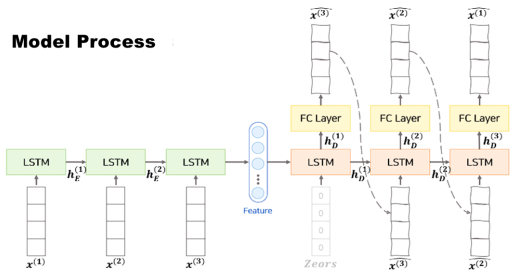


그림 5. 모델 프로세스
Fig. 5. Model Process

AE는 이때 입력인 $x^{(1)}$, $x^{(2)}$, … , $x^{(n)}$와 출력인 $x^{(\widehat{1})}$, $x^{(\widehat{2})}$, …, $x^{(\widehat{n})}$의 차이인 Mean Squared Error(MSE)를 최소화하는 방향으로 학습한다.

AE는 학습 과정에서 비정상 데이터가 없는 정상 데이터만을 사용하고, 학습과정에서 디코더의 입력으로 원본 입력 데이터($x^{(1)}$, $x^{(2)}$, … , $x^{(n)}$)를 활용하는 교사 강요 기법을 적용한다.

[모델 학습 계산식]

$$Minimize \sum_{X \in S_N} \sum_{i=1}^{N} ||x^{(i)} - x^{(\widehat{i})}||^2 \quad (2)$$

S_N : 정상 데이터
$x^{(i)}$: 원본 데이터(input)
$x^{(\widehat{i})}$: i 지점의 복원된 데이터(output)
N : 입력 데이터 길이

학습 과정과 비교하여 동일한 점은 인코더에서 특징 벡터를 생성하고 디코더의 초기 은닉 벡터로 인코더의 특징 벡터를 사용한다는 점이다. 차이점은 디코더의 입력으로 원본 데이터($x^{(n-t+1)}$)가 아닌 이전 단계의 디코더에서 생성한 복원된 출력($x^{(n-\widehat{t+1})}$)을 사용한다는 점이다.

[복원 오차 계산식]

$$e^{(i)} = ||x^{(i)} - x^{(\widehat{i})}|| \quad (3)$$

$e^{(i)}$: i지점의 복원 오차

MSE Loss를 이용하여 학습을 진행하였지만 추론 시에는 Mean Absolute Error (MAE)를 활용하였다.

3.4 데이터 재구성

학습 과정과 추론 과정에 의해 생성된 RE값과 AS값을 새로운 열로 추가하여 새로운 데이터 셋을 구성하였고, 이를 각 클래스별로 구분한 시계열 데이터에 대해 MTF의 API를 활용하여 이미지화하는 작업을 수행하였다.

3.5 이미지 기반 이상탐지

재구성된 데이터를 활용하여 각 레이블 클래스별 이

미지 분류를 위해 다음과 같은 일련의 과정을 수행하였다.

이미지 데이터 셋을 수집하고 레이블 클래스(Benign, Trojan, Spy, Ransom)으로 분류하여 지정하였다. 이에 대한 이미지 데이터는 5분 단위의 시계열 데이터를 묶어서 이미지화 하는 과정을 수행하였고, 이미지 크기를 표준화하고 픽셀 값을 정규화 하였다.

데이터를 학습 세트와 검증 세트로 나누고, Data Generator를 활용하여 생성한 학습 세트를 사용하여 모델을 훈련하고 검증 세트를 사용하여 모델의 성능을 평가하였다.

기본적인 모델 구성은 가장 잘 알려진 CNN 구조 중 하나인 LeNet-5의 구성에서 약간의 튜닝 과정을 거치는 방식으로 아키텍처를 설계하였다. 아키텍처 설계 단계에서 구성한 합성곱과 풀링 연산을 통해 이미지의 특징을 추출하고, 완전 연결 계층에서 이러한 특징을 기반으로 클래스를 분류하였고, 정의한 손실 함수를 토대로 최적화 알고리즘(Adam)을 사용하여 모델의 가중치를 조정하면서 손실을 최소화하도록 하였다. 모델의 테스트를 위해 100개의 테스트 세트를 사용하여 분류된 클래스의 정확도를 평가하였다.

Ⅳ. 성능 평가

4.1 이상상태 검출 결과

[Fig. 6]은 LSTM AE 모델의 이상 점수 분포값을 나타낸다. 입력값이 이상 상태인지 식별을 위한 이상 점수(AS) 계산에는 데이터의 평균값(Mean Value), 표준 편차값(Standard Deviation Value)을 활용한 마할라노비스 거리 계산법을 사용하였다.

[마할라 노비스 거리 계산식]

$$a^i = (e^i - \mu)^T \sum^{-1} (e^i - \mu) \qquad (4)$$

a^i : i지점의 AS
e^i : i지점의 복원 오차
μ : 데이터의 평균
$\sum$: 데이터의 표준편차

또한, 테스트 데이터를 통해 재구성 손실을 계산하고 해당 임계값을 초과할 경우 이상 상태로 판별하였으며, 이때 임계값은 정상 구간에서의 재구성 손실값 중 최대값을 기준으로 설정하였다. 이는 정상 상태에서

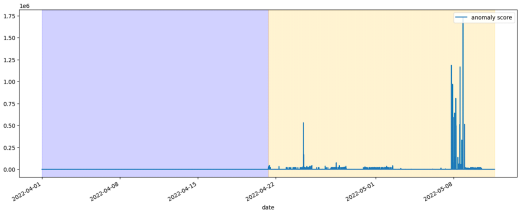


그림 6. 이상 점수 분포
Fig. 6. Anomaly Score Distribution

발생 가능한 손실률의 최대치를 넘는 경우를 비정상으로 간주하기 위한 기준이며, 이상 탐지 민감도를 조정할 수 있는 근거가 된다. [Fig. 6]의 이상 점수 분포에서 푸른색 배경은 원본 데이터상의 정상 구간을, 주황색 배경은 비정상 구간을 나타낸다. 데이터셋의 2022년 4월 21일 시점부터 사이버 공격이 시작되었음을 메모리 분석을 통해 확인하였고, 이는 원본 데이터셋의 내용과도 일치하였다. 그러나, 원본 데이터상으로는 비정상 상태임에도 이상 점수가 정상 구간과 큰 차이를 보이지 않아 시각적으로 구분이 어려운 부분이 존재했다. 이러한 경우에도, 설정된 임계값을 초과하는 데이터만을 추출해 분석한 결과, 차이는 미미했지만 이상 상태임을 명확히 판단할 수 있었다. 이러한 문제는 임계값을 보다 적절한 최적값으로 조정함으로써 해결 가능하며, 이를 통해 탐지 성능을 개선할 수 있다. 또한, [Fig. 6]의 결과를 통해 Malware의 종류에 따라 메모리 분포 양상이 상이함을 확인하였고, 메모리 분포 데이터만을 활용하여 Malware의 종류를 분류하였다.

4.2 이상점수에 따른 Malware 분류 결과

학습 데이터는 MTFA를 활용하여 '3.4 데이터 재구성'과 같이 이미지화하는 작업을 수행하였다. 테스트는 동일한 모델에 대해서 LSTM AE를 거쳐 생성된 재구성 정보(RE와 AS) 데이터를 포함하지 않고, 메모리 정보만 가지고 MTFA를 한 결과와 재구성 정보를 모두 포함한 결과를 비교하여 학습 모델이 개선되는 정도를 평가한 결과는 (Table 1)와 같다.

[Table 1]의 학습 모델 결과와 같이 재구성 정보를 포함한 결과가 정확도 측면에서 약 5%정도 개선되는 것을 확인할 수 있었고, 일반적으로 딥러닝 모델의 주평가지표로 활용되는 precision, recall, f1-score만 비교하여 보았을 때에도 성능이 개선되었다.

또한, 재구성 정보를 포함하는 것이 Vali dation Loss의 수렴도 측면에서도 (Table 2)와 같이 개선됨을 확인할 수 있었다. 또한, [Table 2]의 히트맵 그래프는 100개의 테스트용 이미지 데이터를 생성하여 모델 성능을 테

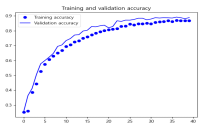
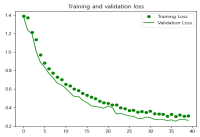
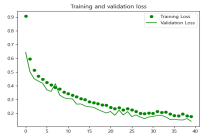
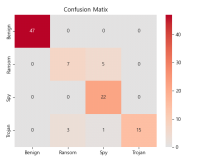
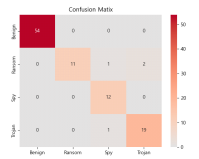
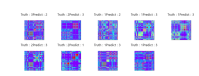
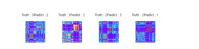
스트한 결과로 Error Image에서 보이는 차이와 같이 분류 오류율이 확연이 줄어듦을 알 수 있었다.

표 1. 주요 평가 지표
Table 1. Main Evaluation Metric

평가 지표 \ 구 분		MTF	LSTM-AE MTF
precision	0	1.00	1.00
	1	0.92	1.00
	2	0.89	0.86
	3	0.68	0.90
recall	0	1.00	1.00
	1	0.69	0.79
	2	0.84	1.00
	3	0.93	0.95
f1-score	0	1.00	1.00
	1	0.79	0.88
	2	0.86	0.92
	3	0.79	0.93
accuracy		0.91	0.96

※ Benign(0), Ransom(1), Spy(2), Trojan(3)

표 2. 훈련 및 검증 결과
Table 2. Training and Validation Result

MTF	LSTM-AE+MTF
	
	
	
[Error Image]	[Error Image]
	

V. 결 론

원자력 발전소의 경우 안전 기능을 담당하는 Category A 기능을 수행하는 기기에 대해서는 정적 주소 및 정적 메모리를 사용하는 것을 권고하고 있다.

본 논문에서는 이러한 안전 기능을 수행하는 기기들의 모듈 혹은 프로세스의 메모리를 분석하여 사이버 위협의 예방과 조속한 조치를 위하여 딥 러닝 기술의 일환인 LSTM AE 및 CNN 알고리즘 그리고 MTF 시각화 알고리즘을 활용한 분석 모델을 제안하였다.

LSTM AE를 이용하여 이상 감지 시스템을 구현하였고 사이버 위협의 종류 분류를 위해 MTF 알고리즘과 CNN 모델을 통해 시각화하는 방식을 제안하였다. 실험 결과 메모리 분포에 MTF와 CNN 모델을 적용한 결과는 약 91%의 정확도를 도출하였으나 검출 연산 시 생성한 AS와 RE데이터를 활용하여 분류한 결과에서는 약 96%로 5% 정도의 모델 성능 개선을 확인할 수 있었다. 따라서 본 논문에서 제안하는 방식을 통해 대상 기기의 정적 메모리 상태를 모니터링 함으로서 원자력 발전소의 사이버 위협 검출 및 분류에 효과적인 방법이 될 수 있음을 실험을 통해 증명하였다. 향후 원자력 발전소에서 운영하는 시뮬레이터의 자료를 활용할 수 있다면 실적용성에 대한 연구를 진행할 예정이다.

References

[1] H. Kim, Legal Research Team in Korea Internet & Security Agency, "*Internet · Information Security Legislation Trends*," vol. 175, Retrieved Apr. 2022, from https://www.kisa.or.kr/skin/doc.html?fn=20230112_111002_376.pdf&rs=/result/2023-01/

[2] U.S. NRC, Reg. Guide 1.152 (Rev.3), "*Criteria for Use of Computers in Safety Systems of Nuclear Power Plants*," Retrieved Jun. 2010, from https://www.nrc.gov/docs/ML1028/ML102870022.pdf

[3] K. I. Jeong, J. K. Lee, J. C. Lee, Y. S. Choi, J. W. Choi, S. B. Hong, J. E. Jung, and I. S. Koo, "High reliable and real-time data communication network technology for nuclear power plant," in *Proc. Korea Atomic Energy Res. Insit.*, Daejeon Korea, Mar. 2008, from https://inis.iaea.org/search/41067566

[4] NUCLEAR SAFETY ACT, Enforcement Date 31, Oct. 2023, from https://www.law.go.kr/법령/원자력안전법

[5] L. Godoy, "*Malware memory analysis / CIC-MalMem-2022*," Retrieved Kaggle, 2022, from https://www.unb.ca/cic/datasets/malmem-2022.html

[6] EPRI TR-106439, "*Guideline on Evaluation and Acceptance of Commercial-Grade Digital Equipment for Nuclear Safety Applications*," Retrieved Nov. 1996, from https://www.epri.com/research/products/TR-106439

[7] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder-decoder for multi-sensor anomaly detection," in Proc. *ICML 2016 Anomaly Detection Wkshp.*, New York, NY, USA, 2016. (https://doi.org/10.48550/arXiv.1607.00148)

[8] J.-Y. Park, D.-H. Seo, and H.-W. Nam, "Deep-learning-based automatic modulation classification using imaging algorithm," *J. KIEIES,* 2021. (https://doi.org/10.5515/KJKIEES.2021.32.4.328)

[9] Y. LeCun, L. Bottou, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998. (https://doi.org/10.1109/5.726791)

[10] Y. Kang and K. Chong, "Safety evaluation on real time operating systems for safety-critical systems," in *Proc. KAIS*, vol. 11, no. 10, pp. 3885-3892, 2010. (https://doi.org/10.5762/KAIS.2010.11.10.3885)

임 규 현 (Gyu Hyun Lim)

2012년 8월 : 순천향대학교 컴퓨터 소프트웨어공학과 졸업
2024년 2월 : 금오공과대학교 디지털융합공학과 석사
2016년 6월~현재 : ㈜한국전력기술 엔지니어

<관심분야> 사이버보안, 소프트웨어V&V, 통신공학, 기계학습

[ORCID:0009-0005-8480-2290]

신 수 용 (Soo Young Shin)

1999년 2월 : 서울대학교 전기공 학부 졸업
2001년 2월 : 서울대학교 전기공 학부 석사
2006년 2월 : 서울대학교 전기공 학부 박사
2010년~현재 : 국립금오공과 대학교 전자공학부 교수

<관심분야> 차세대 무선통신 기술, 무인 이동체, 딥러닝, 영상처리, 블록 체인, 혼합현실

[ORCID:0000-0002-2526-2395]