단계적 활성화 기반 Tri-Engine 스마트 팩토리 위협 탐지 모델

이우곤*, 고성노*, 이재원*, 정은수**
*청주대학교 학부생, **청주대학교 교수
*ggungsu@naver.com **eunsujeong@cju.ac.kr

Phased Activation-based Tri-Engine Threat Detection Model for Smart Factories

Woo Gon Lee*, Sung No Go*, Jae Won Lee*, Eun Su Jeong**
*Cheongju Univ Undergraduate Student, **Cheongju Univ Professor.

요 약

스마트 팩토리 확산에 따른 IT/OT 융합은 생산 효율성을 증대시켰으나, 동시에 랜섬웨어 및 APT와 같은 지능형 사이버 위협의 공격 표면을 확장시켰다 기존의 OT 보안 연구는 주로 단일 모달리티 분석에 의존하였으며, 멀티 모달리티 연구의 경우 레이블 데이터가 부재한 초기 운영 환경에 적용하기 어렵다는 한계를 가진다. 본 논문에서는 이러한 OT 환경의 특성을 고려하여, 단계적 활성화 기반의 Tri-Engine 위협 탐지 모델을 제안한다. 제안 모델은 무결성, 제조 이상, 네트워크 트래픽 이상을 감지하는 3중 엔진으로 구성되며, 데이터 축적 수준에 따라 비지도 학습에서 지도 학습으로 탐지 알고리즘을 고도화한다. 이를 통해 초기 학습 데이터가 부족한 환경에서도 즉시 운용이 가능하며, 다계층 상관 분석을 통해 기존 단일 계층 모델을 우회하는 스텔스 공격 및 임계 값 내에서 조작되는 공격을 탐지할 수 있다.

Ⅰ. 서 론

스마트 팩토리의 확산으로 OT(Operational Technology) 공정 환경은 폐쇄망에서 벗어나 IIoT 센서, 클라우드 등과 연동되는 사이버-물리 시스템(Cyber Physical System)으로 확장되고 있다. 이러한 스마트 팩토리의 발전은 생산성 향상과 비용 절감에 기여하고 있지만, 동시에 랜섬웨어나 지능형 지속 위협(Advanced Persistent Threat, 이하 APT)과 같은 고도화된 공격에 목표가 되고 있다[1].

이를 위해 OT 공정 보안을 위한 연구가 선행되고 있으나, 기존의 OT 보안 연구는 주로 네트워크 트래픽이나 공정 센서 데이터 중 하나의 모달리티(Unimodal)에 의존하여 정상 패턴을 학습하는 데 집중했다[2]. 그러나 APT 공격은 내부 침투 후 권한 탈취, 제어 명령 조작, 센서 데이터 위•변조 등 다계층에 걸쳐 실행되므로, 단일 계층 방어 체계로는 한계를 가진다[3]. 또한, 실제 OT 환경은 초기 구축 단계(Cold-Start)에서 학습 가능한 공격 레이블(Label)이 부재하며, 기밀성보다 가용성을 우선하는 스마트 팩토리의 특이성으로 인해 실험적 공격 주입이 불가능하다[4].

본 논문에서는 스마트 팩토리의 초기 구축 단계(Cold-Start)을 극복하기 위한 단계적 활성화 기반 Tri-Engine 위협 탐지 모델을 제안한다. 제안 모델은 무결성 검증, 제조 이상 감지, 트래픽 보안을 위한 3중 AI 모델로 구성되며, 데이터 축적 수준에 따라 단계적으로 활성화 학습 전략을 실행한다. 이를 통한 본 논문의 기여는 다음과 같다. 초기 데이터가 부족한 환경에서도 적용하는 것이 가능한 모델을 제안한다. 또한, 특정 계층의 이상 탐지에는 유효하지만 계층 간 연관 관계를 분석이 불가하여 고도화된 APT 공격을 탐지하지 못하는 기존 보안 모델들의 한계를 극복한다.

Ⅱ. 관련연구

본 장에서는 기존 OT 환경에서 단일 모달리티 기반의 탐지 연구의 한계와 멀티 모달 연구의 제약을 분석한다.

2.1 단일 모달리티 기반 이상 탐지 연구

기존의 산업 제어 시스템 보안 연구는 주로 네트워크 트래픽이나 공정 센서 데이터 중 하나의 소스에 의존한다. Umer 등은 기계 학습 기반 IDS가 네트워크와 프로세스 레벨로 독립적으로 발전해왔음을 지적하며 통합 탐지의 필요성을 강조했다[2]. 또한, Burgetová 등은 네트워크 통계 기반 탐지를 제안하였으나, 프로토콜을 준수하는 스텔스 공격, 임계치 값 내에서 조작하는 공격에 대해서는 탐지 성능이 저하됨을 확인했다[5]. 이는 단일 소스만으로는 다계층에서 실행되는 APT 공격 식별에 한계를 가진다는 것을 시사한다.

2.2 지도 학습 기반 멀티 모달 탐지의 데이터 종속성 및 한계

단일 모델의 한계를 극복하기 위한 이종 데이터를 결합하는 연구가 시도되고 있다. Zhan 등은 네트워크와 물리 공정 데이터를 공동 임베딩 공간으로 투영하는 크로스 도메인 표현 학습을 제안했다[6]. 또한, Costanzino 등은 비전 데이터와 센서 데이터를 결합한 멀티 모달 모델을 통해 결함 검출 성능을 향상시켰다. 그러나 이러한 연구들은 공통적으로 충분한 양의 라벨링된 학습 데이터를 전제로 설계되었으며, 정상 데이터만 존재하는 초기 스마트 팩토리의 환경에서는 적용하기 어렵다. 따라서 본 논문에서는 초기 레이블 문제를 해결할 수 있는 단계적 활성화 기반의 Tri-Engine 모델을 제안한다.

Ⅱ. 단계적 활성화 기반의 Tri-Engine 시스템 설계

2.1 Tri-Engine 구조

제안 모델은 공정 환경의 이질성을 극복하기 위한 이중 수집 체계와, 초기 Label의 부재를 극복하기 위한 단계적 활성화 기반 Tri-Engine으로 구성된다. 그림 1은 제안 모델의 전체적인 구성도를 보여준다. 데이터 수집 계층은 현장마다 상이한 네트워크 정책과 설비 구성을 고려하여 Push와 Pull 방식을 채택한다. 실시간성이 요구되는 이벤트는 설비가 동적으로 전송하는 Push 방식으로, 외부 네트워크와의 연결이 제한적인 폐쇄망 환경에서는 시스템이 주기적으로 요청하는 Pull 방식으로 데이터를 수집

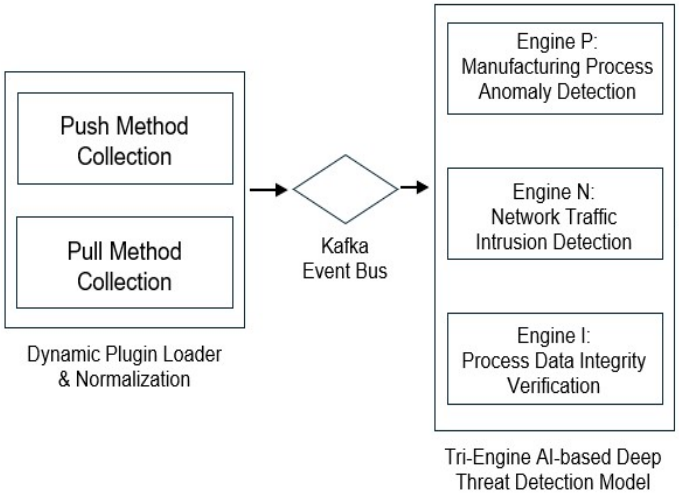

Fig 1. Overview of the Proposed Tri-Engine AI Architecture

한다. 수집된 이기종 데이터는 정규화 및 특징 추출 과정 후 Tri-Engine으로 분기된다. 표 1은 엔진의 단계적 활성화에 따른 사용 알고리즘과 학습 데이터를 보여준다. Engine-I는 HMAC-SHA256 기반의 해시 대조를 통해 전송 구간의 데이터 위·변조를 검증하고, 오염된 데이터를 분석 파이프라인에서 배제한다. Engine-P와 Engine-N은 초기 레이블이 부재한 OT 환경의 특성을 고려하여 비지도 학습의 경우 Isolation Forest, 데이터가 축적된 경우 Random Forest 알고리즘을 사용한다.

표 1. Specifications of Tri-Engine Models and Data

Engine	Algorithm	Input Data
Engine P (Phase 1)	Isolation Forest	MES Data
Engine P (Phase 2)	Random Forest	
Engine N (Phase 1)	Isolation Forest	Network Traffic
Engine N (Phase 2)	Random Forest	
Engine I	HMAC-SHA256	Data Hash

구체적으로 시스템 구축 초기인 Phase 0에서는 데이터 축적 및 기준점 수립에 집중하며 수동 모드로 작동한다. 데이터가 일정량 확보된 Phase 1에서는 Isolation Forest를 활성화하여 이상 징후를 식별하고 관리자에게서 피드백을 수집한다. 피드백을 통한 Label이 축적된 Phase 2에서는 Random Forest를 사용하여 위협을 정밀하게 분류 및 식별한다.

Ⅲ. 시나리오 기반 위협 탐지 메커니즘

APT 공격은 단일 계층에서 관찰되지 않는다. 네트워크 침투부터 시작하여 권한 상승, 공정 데이터 조작 등 여러 계층에 걸쳐 복합적으로 공격이 실행된다.이를 구체화하기 위해, 발열 반응을 제어하는 화학 공정의 냉각 시스템을 대상으로한 공격 시나리오를 가정한다. 공격자는 냉각 펌프를 강제로 정지시켜 반응기의 과열을 유도하면서도, 관제 시스템의 감시를 회피하기 위해 펌프가 정상 가동 중이며, 유량이 흐르는 것처럼 조작된 허위 데이터를 전송한다. 기존의 네트워크 기반 IDS는 패킷의 구조와 프로토콜 필드의 값이 정책을 준수하므로 해당 트래픽을 정상으로 인식한다. 또한, 공정 기반 이상 탐지 모델은 센서 값이 설정된 임계치 범위 내에 존재하기 때문에 이를 이상으로 식별하지 못한다. 그러나 제안하는 Tri-Engine 모델은 다음과 같은 방식으로 작동하며 공격을 식별한다.

첫째, Engine-I는 공격자가 관제 시스템으로 전송되는 패킷을 가로채 값을 정상으로 변조하는 과정에서 무결성이 훼손된다. 엔진은 수신된 데이터에 대해 Hash 값을 파악하고 이를 원본 해시값과 대조한다. 이 과정에서 해시 불일치가 발생한다면 해당 패킷을 차단한다.

둘째, 만약 해시를 우회할 경우, Engine-P는 제어 명령으로 냉각 펌프가 정지했음에도 유량이 정상보고 되었거나, 정상 유량임에도 반응기 온도가 상승하는 물리적 상관관계의 이상을 감지한다.

셋째, Engine-N은 공정 도구가 하위 데이터를 주입하기 위해 생성하는 트래픽의 주기나 세션 지속 시간 등의 이상치를 탐지한다.

제안 모델은 개별 엔진의 위험 점수를 종합하여, 이를 고위험 위협으로 판단하여 해당 공격을 차단한다.

Ⅳ. 결론

스마트 팩토리 환경을 폐쇄성에서 벗어난 사이버-물리 시스템으로 진화하고 있다. 그러나 데이터 가용성의 제약과 초기 학습 레이블의 부재라는 한계로 인해 고도화된 위협에 대응하지 못하고 있다. 기존의 단일 계층 보안 모델은 특정 모달리티에 국한되어 다계층 기반의 APT 공격을 식별하지 못한다. 이에 본 논문은 OT 환경의 제약을 극복하는 단계적 활성화 기반 Tri-Engine 위협 탐지 모델을 제안한다. 제안하는 모델은 공정, 네트워크, 무결성 검증의 3중 AI 모델로 구성된다. 각 엔진은 초기 데이터가 없는 환경에서는 비지도 학습으로 미지의 위협을 식별하고, 데이터가 축적된 환경에서는 지도 학습을 통해 이상을 탐지한다.

향후 연구에서는 SwaT 등 공개된 데이터셋을 활용하여 제안 모델의 탐지 정확도와 재현율을 정량적으로 검증할 계획이다. 또한 Tri-Engine의 탐지 결과를 SOAR 플레이북과 연동하여 탐지부터 대응까지의 평균시간을 단축하는 자동화된 보안 인프라로 발전시킬 계획이다.

참 고 문 헌

[1] Nilufer Tuptuk, Stephen Hailes. "Security of smart manufacturing systems". Journal of Manufacturing Systems. Volume 47. 2018.

[2] Muhammad Azmi Umer, Khurum Nazir Junejo, Muhammad Taha Jilani, Aditya P. Mathur. "Machine learning for intrusion detection in industrial control systems: Applications, challenges, and recommendations". International Journal of Critical Infrastructure Protection. Volume 38. 2022.

[3] Duraid Thamer Salim, Manmeet Mahinderjit Singh, Pantea Keikhosrokiani. "A systematic literature review for APT detection and Effective Cyber Situational Awareness (ECSA) conceptual model". Heliyon. Volume 9, Issue 7. 2023.

[4] Deval Bhamare, Maede Zolanvari, Aiman Erbad, Raj Jain, Khaled Khan, Nader Meskin. "Cybersecurity for industrial control systems: A survey". Computers & Security. Volume 89. 2020.

[5]I. Burgetová, P. Matoušek and O. Ryšavý. "Anomaly Detection of ICS Communication Using Statistical Models" 2021 17th International Conference on Network and Service Management (CNSM). 2021.

[6] D. Zhan, W. Zhang, L. Ye, X. Yu, H. Zhang and Z. He. "Anomaly Detection in Industrial Control Systems Based on Cross-Domain Representation Learning". in IEEE Transactions on Dependable and Secure Computing. vol. 22, no. 3. 2025.

[7] A. Costanzino, P. Z. Ramirez, G. Lisanti and L. Di Stefano. "Multimodal Industrial Anomaly Detection by Crossmodal Feature Mapping" 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024.