https://doi.org/10.33778/kcsa.2024.24.3.067

순열 엔트로피 기반 사이버 물리 시스템의 조작된 운영 데이터 식별 방안 연구★

김 가 경*, 엄 익 채**

요 약

에너지 발전소 등을 포함한 주요 기반시설을 공격 대상으로 하는 공격자들은 지능적이고 정교화된 공격을 수행하는 동시에, 목표에 달성할 때까지 공격 흔적을 은닉한다. 특히 실제 물리적 환경과 연결되어 있는 사이버 물리 시스템의 운영 데이터를 조작하는 것은 사람의 안전에 직접적으로 영향을 줄 수 있다. 사이버 물리 시스템의 특성에 따라 일반적인 정보 기술 환경에서의 이상 식별 및 탐지 방법과는 차별적인 접근법이 필요하다. 이에 본 연구에서는 사이버 물리 시스템의 특성을 고려하기 위하여 재귀적 필터링을 수행하고, 악의적으로 조작된 운영 데이터를 식별하기 위한 엔트로피 기반의 접근법이 통합된 방법론을 제안한다. 공개된 산업제어시스템 보안 데이터셋을 기반으로 합성한 데이터에 제안하는 방법론을 적용한 결과, 조작된 운영 데이터를 효과적으로 식별할 수 있음을 검증하였다.

Research on Identifying Manipulated Operation Data of Cyber-Physical System Based on Permutation Entropy

Ka-Kyung Kim*, Ieck-Chae Euom**

ABSTRACT

Attackers targeting critical infrastructure, such as energy plants, conduct intelligent and sophisticated attacks that conceal their traces until their objectives are achieved. Manipulating measurement data of cyber-physical systems, which are connected to the physical environment, directly impacts human safety. Given the unique characteristics of cyber-physical systems, a differentiated approach is necessary, distinct from traditional IT environment anomaly detection and identification methods. This study proposes a methodology that integrates both recursive filtering and an entropy-based approach to identify maliciously manipulated measurement data, considering the characteristics of cyber-physical systems. By applying the proposed approach to synthesized data based on a publicly available industrial control system security dataset in our research environment, the results demonstrate its effectiveness in identifying manipulated operational data.

Key words : Cyber Physical System, Industrial Control System, Measurement Data, Anomaly Detect System, Recursive FIltering, Instrument and Control System, Permutation Entropy

접수일(2024년 08월 19일), 수정일(1차: 2024년 09월 06일), 게재확정일(2024년 09월 27일)

* 전남대학교/정보보안융합학과(주저자)
** 전남대학교/데이터사이언스대학원(교신저자)

★★ 본 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 지역전략융합보안핵심인재양성사업(IITP-RS-2022-II221203, 50%)과 원자력안전위원회의 재원으로 한국원자력안전재단의 지원을 받아 수행된 원자력안전연구사업(No. 2106061, 50%)으로 연구되었음

1. 서론

제어시스템의 디지털화로 에너지 발전소 등을 포함한 산업제어시스템 환경의 유연하고 효율적인 운영이 가능하게 되었다. 대표적인 예시로 자동화된 산업제어시스템 동작 구조로 일정한 범위 내로 제어 변수를 유지할 수 있으며, 기기 결함 발생 시 경보 발생 기능으로 운전자가 신속한 조치를 취할 수 있도록 도움으로써 유연하고 효율적으로 안전한 운영 환경을 도모할 수 있게 되었다[1].

이러한 시스템의 디지털화는 공급망 조달 글로벌화, 확장된 네트워크 및 데이터 통신, 원격 및 무선 연결, 보안 취약점 증가 등에 의한 사이버 위협 표면도 증가시킨다[1]. 기반시설의 망분리 구조 방식에도 불구하고, 2010년 이란의 나탄즈 우라늄 농축 시설에 '스턱스넷' 공격이 발생한 사례를 확인할 수 있다.

기반시설을 타깃으로 하는 사이버 공격자들은, 일반적인 IT 환경을 타깃으로 하는 공격보다 더 정교하고 지능화된 공격 기술을 사용한다. 또한 이들은 달성하고자 하는 목표를 이룰 때까지, 공격 흔적 들을 은닉하는 '데이터 조작, 삭제, 주입' 등의 전술도 함께 수행한다[2].

스턱스넷 공격 사례에서도 PLC(Programmable Logic Controller)의 코드를 의도적으로 조작하여 원심분리기에 물리적 스트레스를 계속적으로 가하는 한편, 운전자가 이를 인지하지 못하도록 HMI(Human-Machine Interface)에 표시되는 데이터를 정상 허용 범위인 것처럼 은닉하였다[3][4]. 정상적인 운영 상황에서 운전자는 모니터링 및 경보 시스템을 통하여 기기의 이상을 인지하고 수동적 조치를 수행함으로써 위험을 완화할 수 있으나[5], 공격자가 주입한 조작된 데이터로 인하여 운전자의 적절한 조치가 이루어지지 못하였다. 결과적으로 원심분리기의 파괴라는 피해가 초래되었다.

제어시스템 보안 강화를 위한 이상 탐지 기술이 발전하고 있으나, 시스템 로그, 통신 패킷 등을 대상으로 하는 네트워크 수준의 이상 탐지에 중점인 경우가 대다수이다[6]. 승인되지 않은 불법적 접근에 대한 초기 탐지 및 차단도 중요하나, 이미 이를 우회하여 내부에서 이미 진행되고 있는 공격을 식별 및 중단하는 것도 중요시되어야 한다.

기계학습 기반의 이상 탐지 기법은 사전 모델 훈련 여부에 따라 지도 학습, 반지도 학습, 비지도 학습으로 구분할 수 있다. 지도 학습의 경우 산업제어시스템 환경의 데이터 자체가 민감할뿐더러, 손쉽게 활용할 수 없다는 단점이 존재한다. 반지도 학습의 경우, 비지도 학습의 경우 훈련 데이터에 대한 보안 위협은 존재하지 않으나. 노이즈에 쉽게 영향을 받는 운영 데이터에의 곧바로 적용은 어려울 수 있다.

산업제어시스템의 운영 데이터는 네트워크 수준의 데이터와는 다르게 계절의 주기성, 시스템 재작동 등에도 영향을 크게 받으므로 이를 고려할 수 있는 이상 탐지 알고리즘을 활용하여야 함을 의미한다.

또한 현실과 맞닿아 있는 사이버 물리 시스템의 고유 특성인 비선형성과 불확실성을 고려하여야 한다. 비선형성이란 센서, 액추에이터, PLC, HMI 등이 제어 루프 내에서 상호작용을 거치며, 출력하고자 하는 제어 변수에 노이즈가 발생하는 것을 의미한다. 불확실성이란 측정을 위한 센서는 측정 대상 그 자체가 아니며, 환경적 요소에도 불구하고 강인하게 유지될 수 있는 가장 인접한 곳에 위치하여 물리적 특성을 측정하는 것에서 발생하는 노이즈를 의미한다.

그러나 이상 탐지에 있어 성능이 검증되었다고 하더라도, 이러한 사이버 물리 시스템의 특성을 고려한 이상 탐지 연구는 매우 결핍한 실정에 있다.

사이버 물리 시스템의 특성을 고려한 이상 탐지를 위해서는 기계학습 알고리즘의 관점과는 차별적으로 접근하여야 한다. 본 연구에서는 사이버 물리 시스템의 비선형성과 불확실성을 고려하고, 악의적으로 조작된 데이터를 주입하는 공격자의 특성을 활용하여 허위 데이터를 식별하는 방법론을 제안하고자 한다.

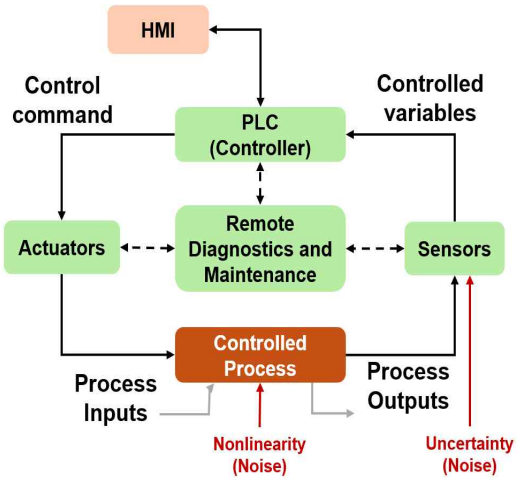
2장에서는 본 연구에서 제안하는 방법론의 이론적 배경과 산업제어시스템 운영 데이터 수준에서의 이상 탐지를 위한 관련 연구들을 분석한 내용을 나열하였다. 3장에서는 현실적인 사이버 물리 시스템의 특성을 고려한 노이즈 필터링 방법론과, 공격자의 악의적 의도로 조작된 데이터를 식별하는 방법론에 대하여 소개한다. 4장에서는 제안하는 방법론의 효과를 검증하기 위하여 수행한 실험 및 그 결과에 대하여 명시하였다. 5장에서는 본 연구의 포괄적 요약 및 향후 연구 방향에

대하여 제시하였다.

2. 연구 배경

2.1 연구 대상 환경

보편적인 산업제어시스템의 구성요소인 HMI, PLC, 액츄에이터, 센서 간의 상호작용은 Figure 1과 같이 이루어진다[7]. 사용자는 HMI를 통하여 PLC에서 전송된 공정 상황 데이터를 모니터링하고, 설정값을 조정한다. 사용자가 입력한 설정값에 도달 또는 유지하기 위하여 PLC는 액츄에이터에 제어 명령을 내리며, 액츄에이터는 이에 적합한 작업을 수행한다. 이러한 프로세스의 출력물로 센서 측정값이 존재하며, 센서의 측정값은 다시 PLC로 전달되는 반복적 제어 루프 구조를 가진다.



(Figure 1) Industrial Control Loop

이러한 제어 루프에서 시스템 간 상호작용 과정 중에 의도치 않은 노이즈가 유입되게 된다. 즉, 사용자가 입력한 설정값에 따른 PLC의 제어 명령에도 불구하고, 현실적 노이즈로 인하여 완벽히 정확한 값에 도달할 수 없음을 의미한다. 이를 사이버 물리 시스템의 현실적 특성으로 인해 발생하는 비선형성이라 정의한다.

또한 물리적 및 환경적 정보를 수집하여 PLC로 전달하는 센서에는 불확실성이 존재한다. 센서는 측정하고자 하는 그 시스템 대상 자체가 아니며, 계절적, 시간적, 상태적 조건에 따라 쉽게 영향을 받는다. 이는 측정 오차와 프로세스 오차로 정의되며, 현실 세계를 수학적으로 모델링하는 과정에서 발생하는 자연스러운 결과이다.

산업제어시스템의 운영 데이터 관점에서 이상 및 침입을 탐지하기 위한 기존의 방법론은 크게 2가지로 구분할 수 있다. 이는 [8], [9], [10], [11], [12], [13] 등과 같이 기계학습 기반으로 이상을 식별하는 방법론과, 실제 물리적 환경을 수학적으로 모델링하여 이상을 식별하는 [14], [15]과 같은 방법론으로 구분될 수 있다.

전자의 경우 고도화된 인공지능 모델만으로 이상을 탐지한다는 점에서 현실적인 한계가 존재하지만, 성능은 후자보다 뛰어나다는 장점이 존재한다. 반대로 후자의 경우는, 성능은 뒤처지더라도 더 현실적인 요건을 고려할 수 있다는 장점이 존재한다.

2.2 사이버 물리 시스템의 물리적 모델링

본 연구에서는 산업제어시스템 내 계측제어시스템 또한 사이버 물리 시스템에 해당하므로, 실제 물리적 환경을 수학적으로 모델링할 수 있는 방법론을 사용하였다. 물리적 환경을 수학적으로 모델링하는 방법론 중 하나인 '칼만 필터(Kalman Filter, KF)'를 적용하였다.

현실 잡음이 포함되어 있는 측정치를 바탕으로 선형 역학계의 상태를 추정하도록 '루돌프 칼만(Rudolf Kalman, 1930)'에 의해 '칼만 필터가 개발되었다. 현재까지도 로봇 공학, 제어 공학, 컴퓨터 비전, 레이다, 신호처리 등 다양한 분야에서 활발하게 활용되고 있다.

칼만 필터는 이산 시간 선형 동적 시스템을 기반으로 동작하며, 각 시간에서의 상태 벡터는 이전 시간의 상태 벡터를 통해서 결정된다는 마르코프 연쇄를 가정하고 있다. 과거의 측정값과 추정값을 바탕으로 현재 상태 변수의 결합분포를 예측하는 재귀적 방식으로 동작하며, 예측과 업데이트라는 두 단계를 반복적으로 수행한다.

그러나 칼만 필터는 선형 시스템만을 대상으로 개발되어, 비선형 구조를 가지는 시스템에는 적용하기에는 어려움이 존재한다. 이를 해결하기 위하여 '확장 칼만

필터(Extended Kalman Filter, EKF)'가 개발되었다. 확장 칼만 필터는 비선형 함수의 매 순간을 편미분하여 선형화된 값을 얻어내는 방식으로, 주요 변수 간의 관계를 나타낸 자코비안 행렬을 기반으로 한다. 다만, 상황에 따라 확장 칼만 필터를 적용할 수 없는 경우도 존재한다. 대표적인 예시로 주요 변수 간의 관계나 비선형 함수 자체를 알 수 없는 경우이다. 또한 자코비안 행렬을 기반으로 하므로, 이 행렬 자체가 잘못된 설정값이라면 칼만 필터의 목적을 달성할 수 없다.

이러한 한계점을 보완할 수 있는 또 하나의 칼만 필터로 '무향 칼만 필터(Unscented Kalman Filter, UKF)'가 존재한다. 무향 칼만 필터는 비선형 함수를 편미분하여 선형화하는 확장 칼만 필터와 달리, 비선형 함수 그 자체를 알아내고자 하는 목적을 가진다. 무향 칼만 필터는 비선형 함수의 가우시안 분포를 전제로 하는 소수의 시그마 포인트 분석을 통하여 평균 및 분산값을 예측한다.

< Table 1 > Comparison of characteristics among Kalman Filters

	KF	EKF	UKF
Operation Functions	선형	비선형	비선형
Mathematical Features	선형대수	야코비안 행렬	무향 변환, 시그마 포인트
Usability	낮은 연산량	구조화된 공식 사용	선형화 오류 낮음
Computation (relative)	비교적 낮음	중간	비교적 높음

예측한 시점이 'k'라고 하였을 때, 그 다음 시점 'k+1'에서는 'k' 시점의 예측값과 'k+1' 시점의 실제 관측값 간의 차이를 계산한다. 이 차이의 정도에 따라 '칼만 이득'이라 정의되는 가중치를 조정하여 최적의 추정값을 도출하는 방향으로 동작하며, 이 추정값은 다시 다음 'k+2' 시점의 예측값이 되는 동작 구조를 가진다. 칼만 이득은 예측과 측정의 불확실성을 비교하여 시스템의 상태 추정을 최적화하기 위한 목적의 가중치로, 어떠한 값에 더 높은 신뢰도를 부여할 지를 결정한다.

칼만 필터(KF), 확장 칼만 필터(EKF), 무향 칼만 필터(UKF)는 각각 상이한 장점과 단점을 가지고 있으며, 적용하고자 하는 상황에 따라 적합한 필터를 사용하여야 한다. 이 3가지 칼만 필터의 주요 특징을 Table 1에 나타내었다.

2.3 악의적으로 조작된 운영 데이터

산업제어시스템 환경의 물리적 손상을 목표로 하는 공격자는 PLC와 액츄에이터 등을 악의적으로 조작하는 한편, HMI에는 정상 범위인 것처럼 조작된 운영 데이터를 주입하여 모니터링 감독자 및 시스템을 기만하는 경우가 존재한다[16][17][18]. 공격자는 목표 달성을 위한 공격을 계속적으로 수행하기 위하여, 모니터링 감독자 및 시스템이 수행할 수 있는 적시에 적합한 안전 또는 보안조치가 이루어질 수 없도록 한다.

정상 운영 상황에서의 센서로 측정한 데이터는 사이버 물리 시스템의 특성으로 인하여 무작위적인 패턴을 가진다. 이에 비하여 공격자는 모니터링 감독자를 기만하기 위하여 데이터 변동을 최소화하고, 이전 측정값을 유지하거나 반복하기 때문에 낮은 무작위성을 띄게 된다. 이러한 전제를 검증하기 위하여 자동화된 조작 데이터 주입 상황과 수동적인 조작 데이터 주입 상황을 분류하여 실험을 진행하였다. 실험은 HIL(Hardware In The Loop) 시뮬레이터 기반의 산업제어시스템 테스트베드에서 수집된 보안 데이터셋 'HAI(HIL-based Augmented Industrial Control System, HAI)[19]'를 활용하였다.

'HAI'는 GE사의 터빈 테스트베드, EMERSON사의 보일러 테스트베드, FESTO사의 모듈형 생산 수처리 시스템 테스트베드를 통합적으로 구현한 시뮬레이터에서 수집된 데이터이다.

2010년 HAI 1.0을 시작으로, 가장 최신의 HAI 23.05가 공개적으로 배포되어 있다. HAI 23.05에는 249시간의 정상 작동 상태에서 수집한 데이터와 52개의 의도적 공격이 수행된 79시간의 비정상 상태에서 수집한 데이터가 포함되어 있다. 그 중 보일러 공정의 회수 물탱크 수위 제어 루프 내의 센서값 'P1_LIT01'을 조작한 공격 시나리오 5가지를 식별하였다.

< Table 2 > Description of 'HAI' Attack Scenario targeting 'P1_LIT01' data manipulation

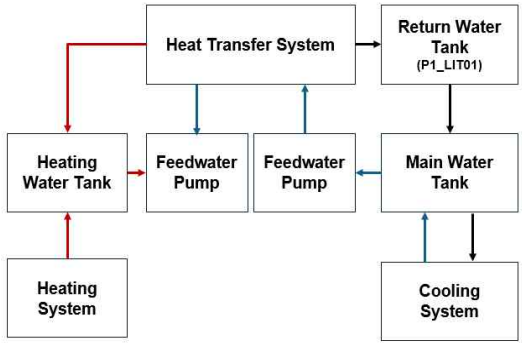
Attack Scenario ID	Attack Target Points	Description
AP 10	P1_B3005	유량 제어 루프의 설정값을 조작하고, HMI 화면의 설정값 변경 사항 은닉
	P1_LIT01	이전 센서 값 유지
AP 15	P1_B3004	수위 제어 루프의 설정값을 조작하고, HMI 화면의 설정값 변경 사항 은닉
	P1_LIT01	이전 센서 값 반복
AP 17	P1_LCV01D	수위 제어 루프의 통제 변수 값 조작 후 정상으로 복구
	P1_LIT01	이전 센서 값 반복
AP 42	P1_LCV01D	수위 제어 루프의 통제 변수 값 조작 후 정상으로 복구
	P1_LIT01	이전 센서 값 반복
	P1_FT03	이전 센서 값 유지
AP 44	P1_LCV01D	수위 제어 루프의 통제 변수 값 조작 후 정상으로 복구
	P1_LIT01	이전 센서 값 반복

'P1_LIT01'가 비정상적으로 높거나 낮다는 것은 보일러 공정 내 결함이 발생하였다는 것을 의미한다. 'P1_LIT01'은 0~720mm의 범위 내에서 수위 측정이 가능한 운영 데이터이며, 보편적으로 회수 물탱크의 정상 범위는 전체 용량의 50%~75% 사이에 위치한다. 'HAI' 데이터의 공격 시나리오 중, 'P1_LIT01'를 대상으로 데이터 조작 공격에 대한 설명은 Table 2에 나타내었으며, 공격 시나리오들이 수행된 타임스탬프는 Table 3에 나타내었다.

'HAI' 데이터가 수집된 HIL(Hardware In The Loop) 기반 테스트 환경 내, 보일러 공정은 Figure 2와 같이 구성되어 있다. 가열 장치(Heating System)에서 물을 가열하여 난방수 탱크(Heating Water Tank)에 저장한다. 난방수 탱크(Heating Water Tank)에 저장된 가열된 물은 급수 펌프(Feedwater Pump) 등을 사용하여 열 전달 시스템(Heat Transfer System)으로 보내진다. 주 물탱크(Main Water Tank)의 물은 필요로 하는 시스템들로 공급되며, 사용된 물은 냉각 장치(Cooling System)로 반환된다. 냉각 장치(Cooling System)에서 냉각된 물은 회수 물탱크(Return Water Tank)로 돌아오는 구조를 가진다

< Table 3 > Timestamp of 'HAI' attack scenario targeting 'P1_LIT01' data manipulation

Timestamp	Attack Scenario ID	Attack Start Time	Attack Duration Seconds
2022.08.12	-	-	-
2022.08.13	AP 10	4:43	133
2022.08.17	AP 15	3:37	131
	AP 17	5:46	122
	AP 42	10:36	133
2022.08.18	-	-	-
2022.08.19	AP 44	6:46	2051



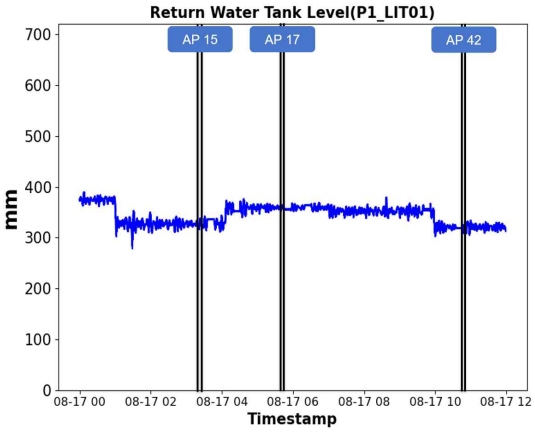
(Figure 2) Boiler Process of 'HAI' Data Testbed

2.3.1 자동화된 조작 데이터 주입

네트워크 계층의 방화벽, 침입탐지시스템 등의 경계보호시스템을 우회하여 침투에 성공한 공격자가 HMI에 'P1_LIT01'에 대한 조작된 데이터를 주입하고 있는 상황을 가정한다. 'HAI'의 2022.08.13.에 수행된 공격 AP 10이 포함된 데이터 파일은 자정부터 약 7시간 가량의 데이터만 시간적 연속성을 만족하고 있다. 2022.08.17.에 수행된 공격 AP 15, 17, 42이 포함된 데이터 파일은 24시간의 시간적 연속성을 만족하였으며, 2022.08.19.에 수행된 AP 44가 포함된 데이터 파일은 16시간의 시간적 연속성을 만족하고 있었다.

운영 데이터는 시계열 데이터이므로, 본 연구에서 제안하는 방법론 검증을 위하여 시간적 연속성을 만족하는 2022.08.17.의 자정부터 약 12시간 가량 수집된 데이터 '43,199'건을 분석 대상으로 하였다. 자동적으로 조작된 데이터를 주입하는 시스템 요소는 GAN(Generative Adversarial Network) 모델로 갈음하였다.

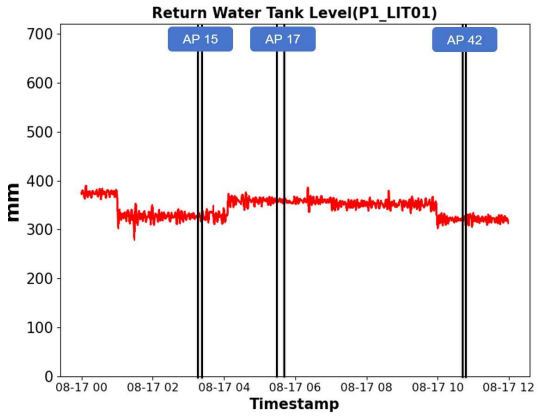
AP 15, 17, 42 각각의 공격이 수행되기 전 60초의 데이터를 모델에 동일하게 훈련시키되, 산업제어시스템 운영 구조상의 지연 시간을 고려하여 공격 지속 시간의 2배의 데이터를 생성하도록 하였다. HMI 디스플레이에 나타나는 정보와 같이 GAN 모델에 의해 합성된 'P1_LIT01' 데이터를 시각화하여 나타낸 결과는 Figure 3과 같다. 운전자가 감시하는 모니터링 화면상으로는 이상이 식별되지 않을 정도로 정교하게 조작되었음을 확인할 수 있다.



(Figure 3) 'Return water tank level' data with manipulated data of GAN model

2.3.2 수동적 조작 데이터 주입

동일한 환경과 대상에서, 공격자가 수동적으로 'P1_LIT01' 데이터를 조작하고 주입하는 상황을 가정한다. 자동적으로 조작된 데이터를 주입하는 상황과의 동일한 조건을 충족하기 위하여, AP 15, 17, 42 각각의 공격이 수행되기 전 60초 동안의 데이터를 그대로 쉬프트하여 공격 타임스탬프 기간에 주입하였다. 수동적으로 조작된 데이터를 주입한 결과는 Figure 4와 같다. 자동적으로 조작된 데이터를 주입하는 경우와 동일하게, 운전자가 감시하는 모니터링 화면상으로는 이상이 식별되지 않음을 확인할 수 있다.



(Figure 4) 'Return water tank level' data by Manual manipulated data injection

2.3.3 악의적으로 조작된 운영 데이터 특징 분석

공격자가 공격 흔적을 은닉하기 위한 목적으로 운영 데이터를 조작할 시, 변동을 최소화하고 이전 패턴을 유지 또는 모방할 시 Figure 3 및 Figure 4와 같은 결과가 나타나게 된다. 정상적인 운영 상황에서의 데이터 패턴과 비교하고자, 'HAI' 구성 파일 내 정상 운영에서 수집된 데이터를 랜덤적으로 '43,199'건을 추출하였다.

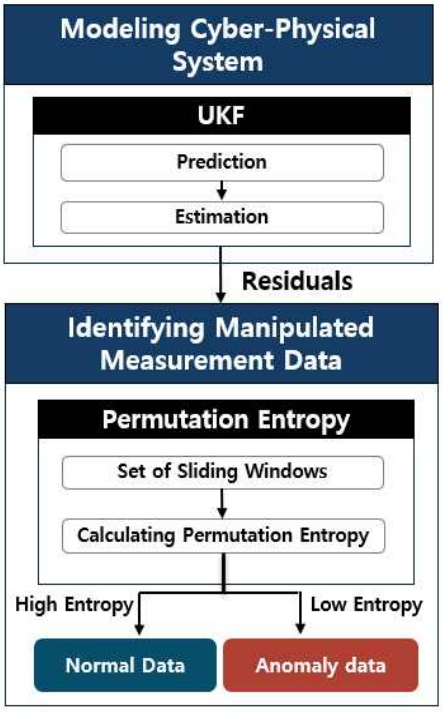
< Table 4 > Comparison of Data Variation

구분		데이터 변동량 (60초당 평균, mm)	
정상 운영 상황		10.93	
비정상 운영 상황	자동화된 조작 데이터 주입	AP 10	2.17
		AP 15	2.9
		AP 17	3.84
	수동적인 조작 데이터 주입	AP 10	0.4
		AP 15	0.81
		AP 17	0.44

공격 시나리오 수행 시간과의 균형화를 위하여 'k'초의 리턴 물탱크 수위와 'k-1'초의 리턴 물탱크 수위 간의 차이를 다 합하여, 60초 단위로 평균을 낸 결과는 Table 4와 같다. 60초는 슬라이딩 윈도우로, 사용

자가 지정 가능한 하이퍼 파라미터값이다. 본 연구에서는 GAN 모델 및 수동적으로 조작한 데이터를 주입하였을 때 활용한 데이터 크기에 기반하여 60초로 설정하였다. 이와 같은 결과는 공격자가 공격 흔적 은닉을 위하여 조작한 운영 데이터는 데이터 변동량이 정상 운영 상황보다 낮고 패턴적임을 확인할 수 있도록 한다.

3. 사이버 물리 시스템 특성을 고려한 악의적으로 조작된 데이터 식별 방안



(Figure 5) Process for identify manipulated measurement data

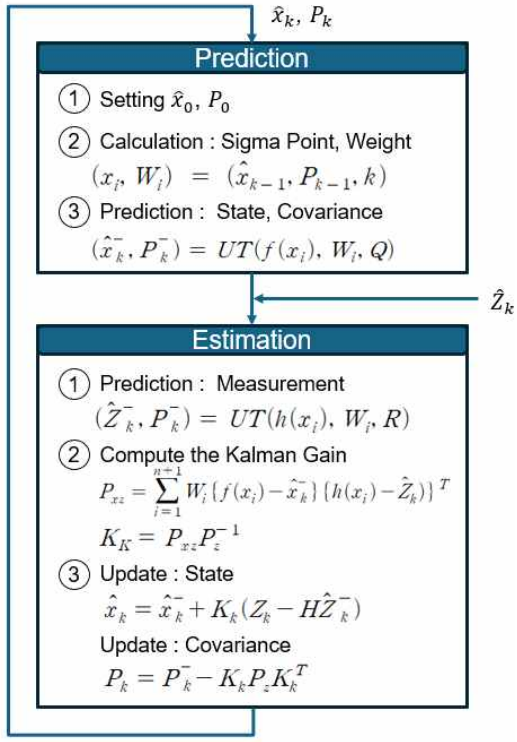
본 연구에서 제안하는 방법론은 Figure 5와 같은 프로세스를 가진다. 2장에서 소개한 3가지의 칼만 필터 유형 중, 비선형 동적 시스템에 적용 가능하고 변수 간의 관계를 파악할 수 없을 때 효율적으로 사용이 가능한 무향 칼만 필터를 선택하였다.

3.1 사이버 물리 시스템 모델링

사이버 물리 시스템을 수식 (1)과 (2)를 활용하여 상태 공간에 대한 모델링을 수행한다. 수식(1)은 상태 방정식이며, 수식 (2)는 관측 방정식을 나타낸다.

$$X_{k+1} = f(x_k) + w_k \quad (1)$$

$$Z_k = h(x_k) + v_k \quad (2)$$



(Figure 6) UKF Operation Process

사이버 물리 시스템 모델링 방정식을 기반으로, UKF는 Figure 6과 같이 작동하는 구조를 가진다. 예측(Prediction) 단계의 첫 번째 프로세스는 시스템의 출력 행렬과 오차 공분산을 초기값으로 설정하는 것이다. 이후 소수의 시그마 포인트를 기반으로 비선형 동적 시스템의 상태와 가중치를 계산하여 상태와 공분산에 대한 예측을 생성한다. 이러한 예측값은 추정 단계의 입력으로 사용된다.

추정(Estimation) 단계에서는 무향 변환(Unscented

Transformer, UT)을 수행하여 시간 'k'에서 관측된 값과 공분산을 예측한다. 이후 실제 관측값을 기반으로 '칼만 이득(Kalman Gain)'이라 불리는 가중치를 조정하여 최적의 예측값으로 나아간다. 이러한 과정을 거쳐 상태 및 공분산 값이 업데이트되며, 이는 다시 예측 단계로 피드백되는 반복적인 동작 구조를 가진다. 비선형 시스템 모델링 및 UKF 수식에 사용된 기호에 대한 설명은 Table 5에 나타내었다.

< Table 5 > Symbol Descriptions of Equation

Symbol	Description
x	state
z	observed value
W	weight
P	covariance
UT	Unscented Transform
w	system noise
v	measurement noise
Q	Covariance matrix of system noise
R	Covariance matrix of measurement noise
f(x)	nonlinear system state function
h(x)	Nonlinear system observation function
K	Kalman Gain
k	time point
k-1	time point
T	transpose of a matrix
—	predicted value
^	estimate
i	Sigma Point Index

이러한 UKF 알고리즘을 데이터셋 또는 실시간 모니터링 시스템에 적용하였을 때, 'Residuals'라는 잔차를 추출할 수 있다. UKF 잔차는 UKF의 예측값과 실제 관측값 사이의 차이를 의미한다. 산업제어시스템이 정상 운영 상태일 때, 비선형 시스템 모델링과 프로세스 및 측정 오차가 고려된 잔차는 낮은 값일수록 안정적인 상황을 의미한다. 시스템 재부팅이나 온도, 습도, 압력 등에 의한 영향을 받는다면, 높은 잔차를 가질 가능성이 높다. 언급한 경우와 같이 단순 기기 결함이나 환경적 영향에 의한 이상값은 UKF와 같은 필터링 수행 결과로 식별될 수 있다.

그러나 분석 대상으로 하는 운영 데이터가 악의적으로 조작된 경우라면, 이상을 식별할 수 없게 된다. 따라서 보완적인 방법론을 활용하여야 하며, 본 연구에서는 사이버 물리 시스템의 특성에 기인하여 순열 엔트로피 적용을 제안하고자 한다.

3.2 순열 엔트로피를 활용한 조작된 운영 데이터 식별

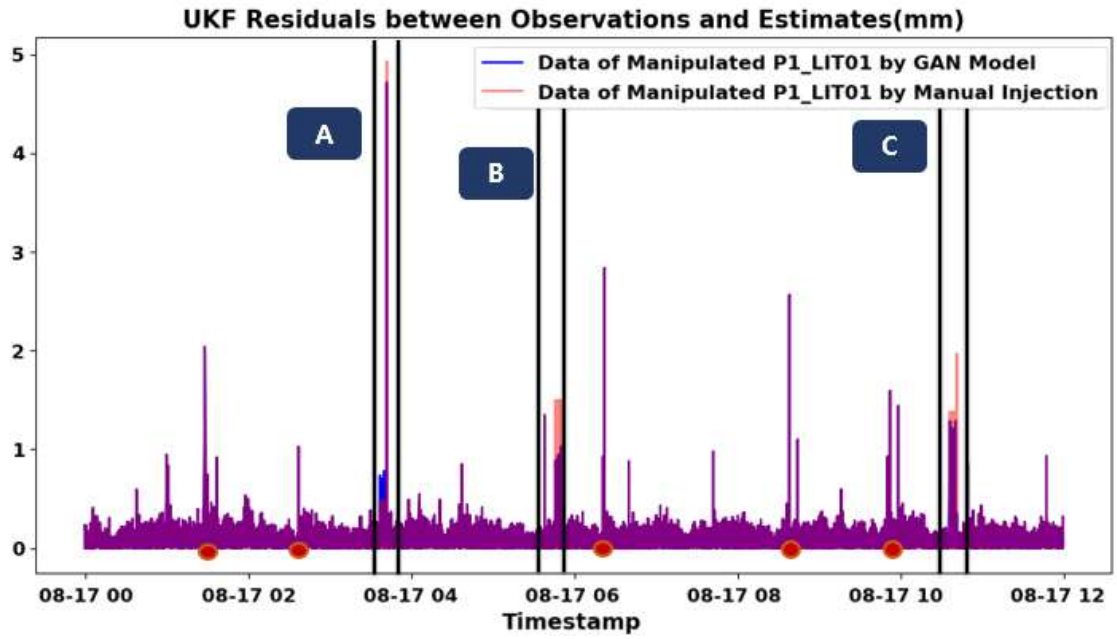
정상 운영 환경에서의 운영 데이터는 사이버 물리 시스템의 특성상 무작위성을 가진다. 이와 반대로 공격자에 의해 악의적으로 조작된 데이터는 변동량이 낮고, 이전 계측값을 반복하거나 유지한다. 이러한 특징을 활용하여 조작된 운영 데이터를 식별하기 위한 보완적인 방법론을 제안한다.

무작위 변수의 예측 불가능성을 나타내는 클로드 새넌(Claude Shannon)의 정보 엔트로피는 수식 (3)과 같이 정의된다.

$$H(P) = -\Sigma P(x_i)\log_2 P(x_i) \quad (3)$$

순열 엔트로피(Permutation Entropy) 이론은 시계열 데이터의 복잡성을 측정할 수 있다. 순열 패턴의 빈도 분포를 기반으로, 데이터의 불규칙성이나 무작위성을 정량화할 수 있다. 본 연구에서 활용하는 순열 엔트로피는 정보 엔트로피의 정의에서 파생하여, 사이버 물리 시스템 특성에 따른 계측 패턴의 무작위성으로 정의한다. 시계열 데이터의 순열 엔트로피를 산출하기 위하여, 수식 (4)와 같이 지정된 크기의 슬라이딩 윈도우로 시간을 연속적으로 분할한다.

순열 엔트로피를 산출하기 위한 사전 단계로, 비선형 함수에서 길이가 n인 슬라이딩 윈도우 집합을 생성한다. '$_j$ '는 첫 슬라이딩 윈도우 지점(s)을 나타낸다. 'T '는 임의 데이터 속성의 전체 튜플 수를 의미하며, '$_j$ '는 '1'에서 '$T-n+1$'까지 범위를 갖는다. 'X_i'는 슬라이딩 윈도우의 전체 집합을 의미하며, 'x'는 특정 시점의 슬라이딩 윈도우 집합을 의미



(Figure 7) UKF Residuals between Observations and Estimates

한다.

$$X_i = [x(j), x(j+1), \cdots, x(j+n-1)] \qquad (4)$$

이후 슬라이딩 윈도우 집합의 각 임베딩 벡터의 값을 오름차순으로 정렬하여 벡터의 순서를 결정하고, 이를 기반으로 가능한 순열 패턴의 모든 빈도를 확률로 산출한다. 슬라이딩 윈도우를 시간적 연속성에 맞추어 1초씩 이동시키어, 시점별 순열 엔트로피를 산출하도록 한다. x 집합 내의 가능한 모든 순열 패턴의 빈도 확률을 $p(\pi_i)$로 나타낸다.

4. 방법론 효과성 검증

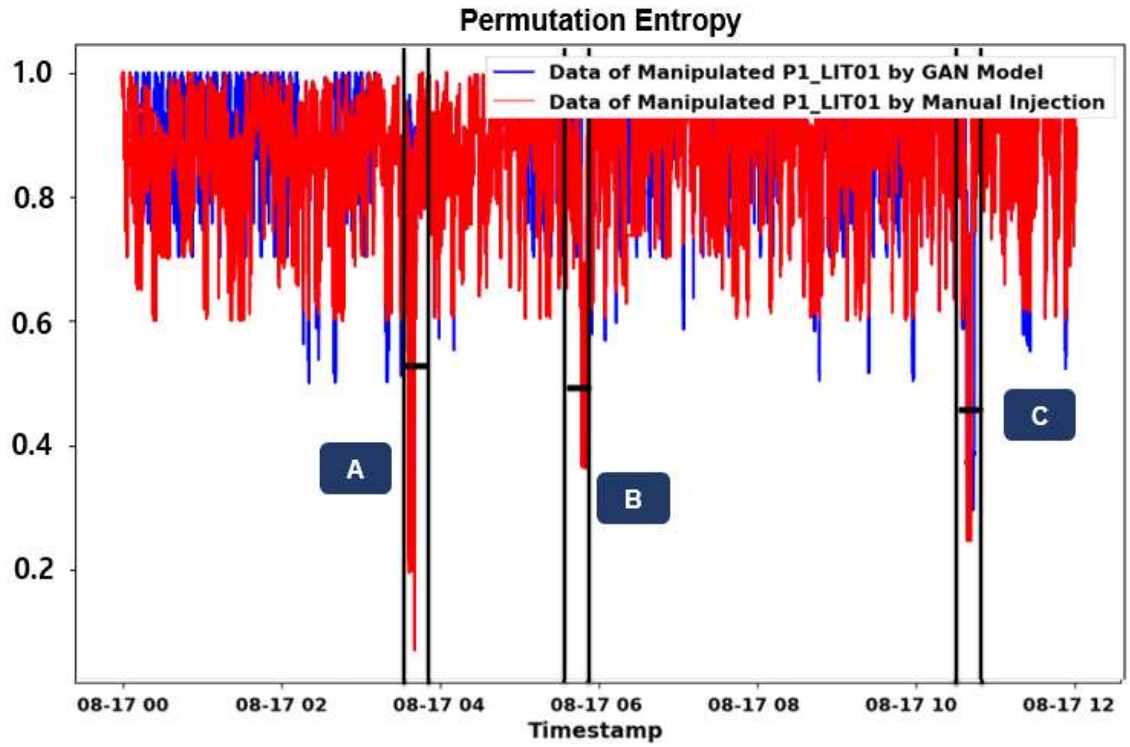
본 연구에서는 무향 칼만 필터와 순열 엔트로피의 결합 모델을 활용한 조작된 운영 데이터의 식별 방법론을 제안한다. 방법론이 달성하고자 하는 목적은 사이버 물리 시스템 특성을 고려하고, 산업제어시스템 환경에서 악의적으로 조작된 운영 데이터를 식별하는 것이다.

이를 검증하기 위하여 2장에서 다루었던 HAI 데이터 내 회수 물탱크 수위를 나타내는 'P1_LIT01'을 대상으로 제안하는 방법론을 적용하였다.

무향 칼만 필터의 시그마 포인트 분석을 위하여 추가적으로 'P1_LCV01Z' 데이터를 상태 변수로 활용하였다. 'P1_LCV01Z'는 물탱크 수위를 조절하는 'LCV01' 밸브의 현재 위치에 대한 데이터로, 물탱크 수위를 유지 및 조절하기 위한 제어 명령의 출력값을 나타낸다.

HAI 데이터의 비정상 운영 상황에서 수집된 2022.08.17.의 원본 데이터에 GAN 모델(GAN Model)로 조작된 데이터 주입, 수동적으로 조작된 데이터 주입을 수행하여 생성한 합성데이터(Manual Injection)를 대상으로 한다. 조작된 데이터의 주입 지점은 Table 3에 나타난 AP 15, 17, 42와 동일하다. 이는 편의상 각각의 지점을 'A', 'B', 'C'로 나타내었다.

데이터에 무향 칼만 필터를 적용하여 식별한 잔차는 Figure 7과 같다. Figure 7의 그래프 선은 각각 2.3.1의 자동적 조작 데이터 주입 과정(Data of Manipulated P1_LIT01 by GAN Model), 2.3.2의 수동적 조작 데이터 주입 과정에 따라 생성한 데이터(Data od Manipulated P1_LIT01 by Manual Injection)의 추정치와의 잔차를 나타낸다. 잔차의 임계값을 1mm라고 가



(Figure 8) Permutation Entropy

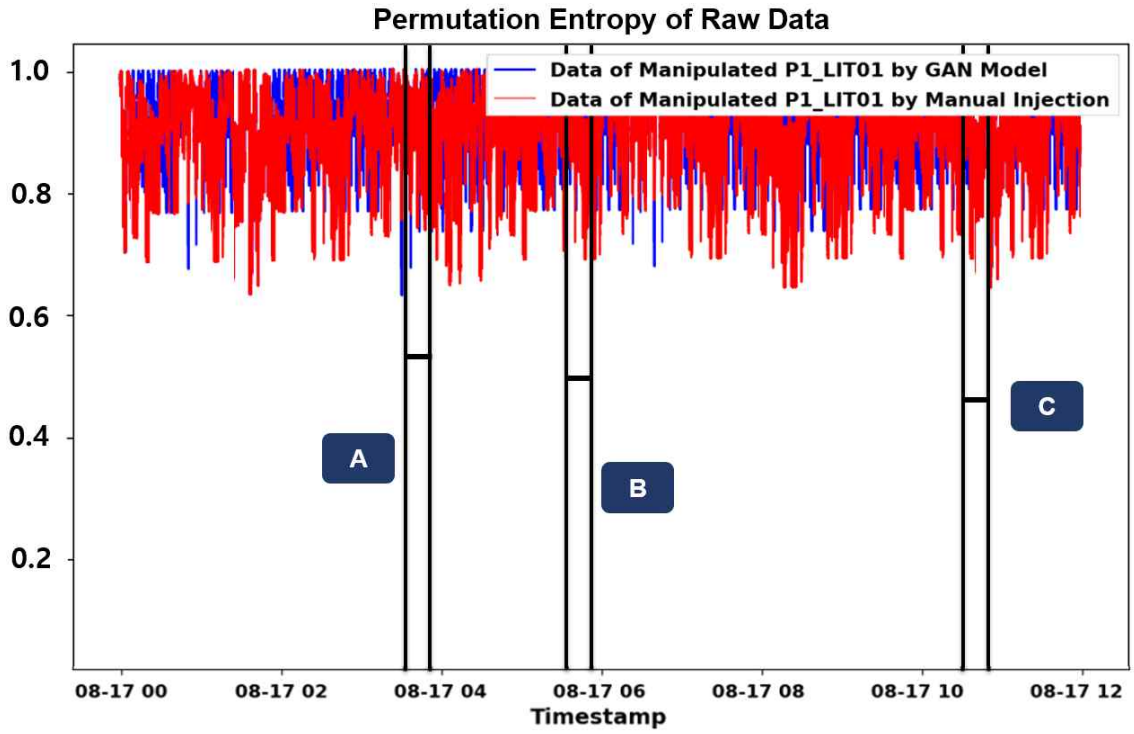
정하였을 때, 총 8가지 지점이 식별됨을 확인할 수 있다. 이 8가지 지점을 무향 칼만 필터에 의해 식별된 이상 징후라고 판단할 수 있으며, 이는 물리적 환경에 의한 영향, 기기 결함 및 사이버 공격이 분류되지 않고 모두 포함되어 있는 상태이다.

8가지 지점 중, 조작한 데이터를 주입하는 사이버 공격을 식별하기 위하여, Figure 7의 잔차에 제안하는 순열 엔트로피를 적용한 결과는 Figure 8과 같다. 잔차는 'abs' 함수를 적용하여 절대값으로 변환하였으며, 순열 엔트로피는 0~1로 정규화 작업을 수행하였다. Figure 8의 그래프 선은 각각 2.3.1의 자동적 조작 데이터 주입 과정(Data of Manipulated P1_LIT01 by GAN Model), 2.3.2의 수동적 조작 데이터 주입 과정에 따라 생성한 데이터(Data od Manipulated P1_LIT01 by Manual Injection)의 잔차에 적용한 순열 엔트로피 값을 의미한다.

Figure 8에서 다른 지점과는 달리 확연하게 식별 가능한 순열 엔트로피가 낮은 지점은, 조작된 운영 데이터를 주입한 사이버 공격이 수행된 지점임을 확인할 수 있다. 또한 Figure 7 내 임계값 '1mm'을 넘어서는 지점도, 사이버-물리 시스템 특성상 Figure 8 내 높은 순열 엔트로피를 가짐을 확인할 수 있다. 기존 무향 칼만 필터로 8가지 이상 징후를 식별하였다면, 순열 엔트로피를 적용함으로써 조작된 운영 데이터를 주입하는 사이버 공격이 수행된 3가지 지점을 확인할 수 있다.

즉, 무향 칼만 필터만 적용하였을 때는, 잔차의 크기에 따라 물리적 환경에 의한 영향, 기기 결함 및 사이버 공격 등이 포함된 전체적 관점에서의 이상 징후 식별 수준에만 그침을 의미한다. 임계값을 초과한 이상 데이터들이 기기 결함 및 환경적 영향에 의한 것인지, 사이버 보안 공격에 의한 것인지 분류할 수 없다.

또한 무향 칼만 필터로 노이즈를 고려하지 않은 원본 상태의 합성데이터에, 순열 엔트로피만 적용하였을 때는 Figure 9와 같은 결과가 나타난다. 즉, 무향 칼만 필터와 순열 엔트로피를 결합적으로 활용하지 않고 독립적으로 적용한다면, 사이버-물리 시스템 환경



(Figure 9) Permutation Entropy of Raw Data

에서 조작된 운영 데이터는 물론 이상도 식별할 수 없음을 확인할 수 있다. 이러한 점에 따라 본 연구에서 제안하는 무향 칼만 필터와 순열 엔트로피를 결합하여 조작된 운영 데이터를 식별하는 방법론의 효과성을 입증할 수 있다.

5. 결론

본 연구의 목적은 사이버 물리 시스템 환경에서 공격자에 의해 악의적으로 조작된 운영 데이터를 식별하는 것이다. 제안하는 방법론은 운영 데이터에 무향 칼만 필터를 적용하여 사이버 물리 시스템의 특성을 고려한 필터링 단계와, 윈도우 순열 엔트로피를 산출하는 단계로 이루어진다. 이러한 프로세스는 이상 식별 목적 또는 허위 데이터 식별을 위한 이전 연구들의 한계점을 보완하기 위함에 있다.

본 연구에서 보완하고자 하는 이전 연구들의 한계점은 크게 두 가지이다. 첫째, 칼만 필터 등의 재귀적 알고리즘으로 이상을 식별할 시, 그 이상이 기기 결함인지 사이버 보안 공격인지 원인을 알 수 없다는 점이다. 둘째, 보편적인 인공지능 알고리즘을 적용한 이상 탐지 방안들은 사이버 물리 시스템의 불확실성과 비선형성을 고려하지 않았다는 점이다.

제안하는 방법론은 공개된 산업제어시스템 보안 데이터셋 기반의 합성데이터에 적용하여 효과성을 검증하였다. 운영 데이터 모니터링 관점에서 이상 식별 및 원인 분류함으로써, 특히 시스템의 즉각적 상태 변경이 어렵고 운영자와의 실시간 상호작용이 필요한 산업제어시스템 환경에서의 높은 활용성을 기대할 수 있다.

참고문헌

[1] International Electrotechnical Commision(IEC), “Security for industrial automation and control systems”, 2021.

[2] W. Wang, Y. Xie, L. Ren, X. Zhu, R. Chang and Q. Yin, "Detection of data injection attack in industrial control system using long short

term memory recurrent neural network," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, China, 2018.

[3] S. Karnouskos, "Stuxnet worm impact on industrial cyber-physical system security," IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society, Melbourne, VIC, Australia, pp. 4490-4494, 2011.

[4] J. Tian, R. Tan, X. Guan, Z. Xu and T. Liu, "Moving Target Defense Approach to Detecting Stuxnet-Like Attacks," in IEEE Transactions on Smart Grid, vol. 11, no. 1, pp. 291-300, 2020.

[5] Institute of Electrical and Electronics Engineers Standard Association(IEEE SA). IEEE Standard Criteria for Safety Systems for Nuclear Power Generating Stations, 2018.

[6] Mokhtari S, Abbaspour A, Yen KK, Sargolzaei A. A Machine Learning Approach for Anomaly Detection in Industrial Control Systems Based on Measurement Data. Electronics. 10(4):407, 2021.

[7] National Institute of Standards and Technology, Guide to Operational Technology Security, 2023.

[8] Tanmoy Kanti Das, Sridhar Adepu, Jianying Zhou, “Anomaly detection in Industrial Control Systems using Logical Analysis of Data”, Computers & Security, Volume 96, 2020.

[9] M. Abdelaty, R. Doriguzzi-Corin and D. Siracusa, "DAICS: A Deep Learning Solution for Anomaly Detection in Industrial Control Systems," in IEEE Transactions on Emerging Topics in Computing, vol. 10, no. 2, pp. 1117-1129, 2022.

[10] C. Feng, T. Li and D. Chana, "Multi-level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Networks," 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Denver, CO, USA, pp. 261-272, 2017.

[11] M. R. G. Raman and A. P. Mathur, "A Hybrid Physics-Based Data-Driven Framework for Anomaly Detection in Industrial Control Systems," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 9, pp. 6003-6014, 2022.

[12] M. M. N. Aboelwafa, K. G. Seddik, M. H. Eldefrawy, Y. Gadallah and M. Gidlund, "A Machine-Learning-Based Technique for False Data Injection Attacks Detection in Industrial IoT," in IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8462-8471, 2020.

[13] S. Potluri, C. Diedrich and G. K. R. Sangala, "Identifying false data injection attacks in industrial control systems using artificial neural network", 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus, pp. 1-8, 2017.

[14] K. Manandhar, X. Cao, F. Hu and Y. Liu, "Detection of Faults and Attacks Including False Data Injection Attack in Smart Grid Using Kalman Filter," in IEEE Transactions on Control of Network Systems, vol. 1, no. 4, pp. 370-379, 2014.

[15] Y. Mo, S. Weerakkody and B. Sinopoli, "Physical Authentication of Control Systems: Designing Watermarked Control Inputs to Detect Counterfeit Sensor Outputs," in IEEE Control Systems Magazine, vol. 35, no. 1, pp. 93-109, 2015.

[16] S. Amin, X. Litrico, S. Sastry and A. M. Baye n, "Cyber Security of Water SCADA Systems -Part I: Analysis and Experimentation of Ste althy Deception Attacks," in IEEE Transaction s on Control Systems Technology, vol. 21, no. 5, pp. 1963-1970, 2013.

[17] R. Langner, "Stuxnet: Dissecting a Cyberwarfa re Weapon," in IEEE Security & Privacy, vol.

9, no. 3, pp. 49-51, 2011.

[18] G. Liang, J. Zhao, F. Luo, S. R. Weller and Z. Y. Dong, "A Review of False Data Injection Attacks Against Modern Power Systems," in IEEE Transactions on Smart Grid, vol. 8, no. 4, pp. 1630-1638, 2017.

[19] National Security Research Institute, "HAI Security Dataset Technical Details," Technical Report, Version 4.0, 2023.

〔저자소개〕

김 가 경 (Ka-Kyung Kim)

2022년 8월 충북대학교 정치외교학과 학사
2023년 3월 ~ 전남대학교 정보보안융합학과 석사
email : kakyung98@gmail.com

< 관심분야 > 산업제어시스템 보안, 데이터사이언스, 인공지능, 취약점 분석

엄 익 채 (Ieck-Chae Euom)

2003년 8월 전남대학교 컴퓨터정보학부 학사
2015년 2월 한국과학기술원 소프트웨어대학원 석사
2019년 2월 전남대학교 정보보안협동과정 박사
2019년 10월 ~ 전남대학교 데이터사이언스대학원 교수

email : iceuom@jnu.ac.kr

< 관심분야 > 산업제어시스템 보안, 취약점 분석, IoT 보안, 차세대인프라