https://doi.org/10.5392/JKCA.2021.21.12.057

인공지능기술의 IoT 통합보안관제를 위한 데이터모델링

Data Modeling for Cyber Security of IoT in Artificial Intelligence Technology

오영택, 조인준
배재대학교대학원 사이버보안학과

Young-Taek Oh(naturaloyt@gmail.com), In-June Jo(injune@pcu.ac.kr)

요약

산업 전 분야에 4차 산업혁명의 신기술인 IoT(Internet of Things), AI(Artificial Intelligence), Bigdata 등이 융합되어 새로운 가치를 창출하는 초연결 지능정보사회가 도래되고 있다. 모든 것이 네트워크에 연결되어 데이터가 폭발적으로 증가하고, 인공지능이 스스로 학습하여 지적 판단 기능까지도 가능하다. 특히 사물인터넷은 언제 어디서나 어느 것과도 연결될 수 있는 새로운 통신환경을 제공함에 따라 모든 것들이 연결되는 초 연결을 가능케 하고 있다. 인공지능 기술은 인간이 가진 지각, 학습, 추론, 자연어처리 등의 능력을 컴퓨터가 실행할 수 있도록 구현되고 있다. 인공지능은 기계학습, 딥러닝(Deep leearning), 자연어처리, 음성인식, 시각인식 등 첨단기술을 개발하는 방향으로 발전되고 있으며, 안전, 의료, 국방, 금융, 복지 등의 다양한 응용분야에 특화된 소프트웨어와 머신러닝(Machine learning), 클라우드(Cloud) 기술을 포함하고 있다. 이를 통해 인간의 편의와 새로운 가치를 제공하기 위해 산업 전반의 다양한 분야에 활용된다. 하지만, 이와는 반대로 지능적이고 정교해진 사이버 위협들이 증가하고 신기술의 기술적 안전성 확보와 같은 잠재적 역기능들을 동반함에 따라 이에 대한 대응이 필요한 시점이다. 본 논문에서는 이러한 역기능을 해결하기 위한 하나의 방안으로 인공지능기술을 활용하여 IoT 통합보안관제 가능하도록 새로운 데이터모델링(Data modelling) 방안을 제안하였다.

■ 중심어 : | 인공지능 | IoT | 통합보안관제 | 데이터모델링 | 머신러닝 |

Abstract

A hyper-connected intelligence information society is emerging that creates new value by converging IoT, AI, and Bigdata, which are new technologies of the fourth industrial revolution, in all industrial fields. Everything is connected to the network and data is exploding, and artificial intelligence can learn on its own and even intellectual judgment functions are possible. In particular, the Internet of Things provides a new communication environment that can be connected to anything, anytime, anywhere, enabling super-connections where everything is connected. Artificial intelligence technology is implemented so that computers can execute human perceptions, learning, reasoning, and natural language processing. Artificial intelligence is developing advanced technologies such as machine learning, deep learning, natural language processing, voice recognition, and visual recognition, and includes software, machine learning, and cloud technologies specialized in various applications such as safety, medical, defense, finance, and welfare. Through this, it is utilized in various fields throughout the industry to provide human convenience and new values. However, on the contrary, it is time to respond as intelligent and sophisticated cyber threats are increasing and accompanied by potential adverse functions such as securing the technical safety of new technologies. In this paper, we propose a new data modeling method to enable IoT integrated security control by utilizing artificial intelligence technology as a way to solve these adverse functions.

■ keyword : | Artificial Intelligence | IoT | Integrated Security Control | Data Modeling | Machine Learning |

접수일자 : 2021년 08월 02일
수정일자 : 2021년 08월 23일
심사완료일 : 2021년 08월 23일
교신저자 : 조인준, e-mail : injune@pcu.ac.kr

I. 서 론

1. 연구의 배경 및 목적

4차 산업혁명이 진행되는 과정에서 새로운 가치 창출은 IoT와 AI 등과 같은 신기술이 융합되어 초연결 지능정보사회로 진화를 통해서 이루어진다. 모든 사물이 네트워크에 연결(초연결) 됨에 따라 데이터가 폭발적으로 증가하고 있다. 이들 데이터를 인공지능이 스스로 학습하여 육체노동뿐만 아니라 지적 판단도 가능한 지능정보사회가 도래되고 있다. 즉, 네트워크(IoT, 5G), 데이터(Cloud, Big Data), 인공지능SW(기계학습, 알고리즘)등과 같은 지능화 기술이 각 분야의 기반 기술과 융합되어 새로운 산업영역으로 진화가 핵심이 되고 있다. 이로써, 4차 산업 혁명 시대에서 핵심 투입 요소인 데이터가 기존 생산요소(노동, 자본)를 압도하는 새로운 경쟁 원천으로 부각 되었다. 특히, 사회적 난제 해결에 지능화기술이 융합되는 현상이 보편화됨과 동시에 삶의 양식과 사회 인프라의 변화 등 사회 전반에 광범위한 파급력을 유발하고 있다.

이러한 순기능적인 변화추세는 반면에 해킹 등 사이버 위협 증대, 신기술에 의한 기술적 안전성 위협 증대 등과 같은 내재 된 역기능들을 동반한다. 신기술 중에서 사물인터넷(IoT)은 언제 어디서나 어느 것과도 연결될 수 있는 새로운 통신환경을 제공한다[1]. 인간과 인간, 인간과 사물, 사물과 사물을 연결하는 초연결사회를 가능하게 하는 핵심기술이다. 이 기술은 가트너(Gartner)가 선정하는 10대 전략기술에 2012년부터 매년 선정되었다. 2019년 발표된 전략기술 트렌드의 10개 항목 중, 스마트 스페이스(Smart space), 디지털 트윈(Digital twin), 증강분석(Renforcement analyse), 인공지능이 등이 모두 직간접적으로 IoT와 관련되어 있다. IBM사는 2021년까지 네트워크에 연결되는 사물들이 약 500억개 이상이 될 것으로 전망하였다. CISCO사는 네트워크에 연결된 사물의 수가 2014년 144억개에서 2020년 501억개로 3.5배 증가하였다고 발표하였다. 궁극적으로 사물인터넷은 인간의 편의와 새로운 가치를 제공하기 위해 산업 전반의 다양한 분야로 활용영역을 확대해 가고 있다.

본 논문에서는 IoT 서비스로 생성되는 수많은 보안로그를 대상으로 인공지능 기술을 적용하여 통합보안관제를 할 수 있도록 한다. 이를 통해서 기하급수적으로 증가하는 보안이벤트에 대한 신속하고 정합도가 높은 판단을 인공지능을 통해 할 수 있도록 한다. 통합보안관제는 침입 차단, 침입 방지, 악성코드 차단, 취약점 진단시스템이 종합적으로 침해 대응이 가능한 시스템을 말한다[9]. 이를 위해 데이터모델링을 하고, 데이터 모델링에 대한 검증을 통하여 제안기술의 효과성을 제시한다.

2. 연구의 내용 및 범위

국내·외 사이버공격은 해킹의 자동화로 무차별적인 공격과 사이버 전면전 위험 고조 등으로 나날이 지능화, 고도화되고 있다. 침해 이벤트 또한 기하급수적으로 증가 추세이다. 최근 해킹기법의 변화를 보면 제로데이 공격과 랜섬웨어와 같은 은닉기법을 비롯해 모바일 공격까지 확대 및 증가하고 있다. 또한 소셜 네트워크의 해킹으로 계정 탈취와 개인정보유출 등의 공격도 증가 추세이다[2]. 신규 서비스의 증가로 IoT 기기를 이용하여 정보를 수집하고 생성해내는 여러 장비가 기하급수적으로 늘어나고 있다. 이와 관련된 보안로그도 엄청나게 많이 쌓이고 있다. 그만큼 사이버공격의 유형도 다양해짐에 따라 새로운 공격이 생겨난다. 기존의 레거시 기반의 보안장비 로그뿐만 아니라, IoT 기기와 관련된 보안장비의 로그도 분석해야 한다. 이로써, 보안 전문인력들의 수준 높은 분석력이 더욱 필요하게 된다. 하지만, 전 세계적으로 사이버 전사와 같은 보안 전문인력을 보유하고 있으나, 한정된 인력과 시스템의 한계로 인해 적기 대응하기에는 미흡한 실정이다.

즉, 기하급수적으로 증가하는 추세에 있는 보안이벤트와 로그를 보안 전문인력이 모두 대응하기에는 한계에 부딪히고 있다. 따라서, 본 논문에서는 이러한 복합적인 문제를 해결하기 위한 하나의 해결책을 새롭게 제안하였다. 이는 인공지능 기반의 IoT 통합보안관제시스템을 효과적으로 구현하기 위한 제안이다. 개략적인 내용은 이를 위해 어떻게 IoT 기기의 로그를 수집하는가 이다. 또한, 기존의 레거시 기반의 보안로그와 어떻

게 함께 분석하고, 인공지능기술을 활용하는 방안을 제안하였다.

본 논문의 내용을 살펴보면 다음과 같다. 제 II 장에서는 인공지능 기술의 IoT 통합보안관제의 필요성과 그 활용방안을 알아본다. 이를 위해 통합보안관제의 체계와 능동형 IoT 통합보안관제 프로세스를 정리하였다. 최종 목표 플랫폼 방안을 설계하고 실무에서 최종적인 활용을 위해 요구되는 문제에 대해 극복방안들에 대해서도 정리하였다. 제 III 장에서는 인공지능 기술 활용에 있어서 핵심적인 데이터 모델링에 대한 알고리즘의 동향과 데이터모델링 원칙을 선정하게 된 내용을 다루었다. 또한, 데이터셋 구성기준에 따라 데이터의 수집부터 전처리과정까지 알아보았다. 제 IV장에서는 인공지능 기술이 반영된 IoT 통합보안관제 데이터모델링 검증을 위해 알고리즘을 선정하였다. 마지막으로 효과검증방안을 수립하여 데이터모델링에 대한 정합성을 검증하였다. 제 V장은 본 연구의 데이터모델링에 대한 결론으로 연구의 결과를 정리하고 앞으로 인공지능 기술의 IoT 통합보안관제 분야에 추가적 연구가 필요한 부분을 기술하였다.

II. 인공지능기술의 IoT 통합보안관제의 필요성 및 활용방안

1. 인공지능 기술의 IoT 통합보안관제 개요

IoT와 인공지능 기술을 기반으로 다양한 서비스를 활용하려는 사업이 [표 1]과 같이 국내뿐 아니라 전 세계에서 진행 중이다. 정부는 2019년 경제부총리 주관으로 데이터의 수집 유통 활용 전 단계를 활성화를 발표했다[8]. 세계적 수준의 인공지능 생태계 조성, 산업 전 분야와 인공지능 간 융·복합을 촉진하기 위한 “데이터·AI 경제 활성화 계획” 발표 하였다. 국내에서는 스마트시티 국가 시범도시라는 사업으로 부산 에코델타시티, 세종시 등의 사업을 국토교통부에서 추진 중이다[3]. 이는 백지상태 부지의 장점을 살려 세계적 수준의 국가 시범도시 조성하여 4차 산업혁명 관련 융복합 新기술 테스트베드(Test bed), 도시문제를 해결하여 삶의 질을 제고 하기 위해서다. 이로써 혁신 산업생태계 조성을 균형 있게 추진 등의 주요 내용이 있다.

해외에서는 도시문제 해결을 위해 일정지역을 실증구역으로 테스트 베드를 조성하는 사업이 네델란드 암스테르담, 캐나다 토론토, 스페인 산탄데르에서 추진 중이다. 미국 시카고 ‘AoT(Array of Things) 프로젝트’, 싱가포르 ‘버츄얼 싱가폴에서는 도시 플랫폼을 구현했다. 이는 도시문제에 사전 대응하기 위해 센서를 설치하고 데이터를 분석하여 예방 서비스들을 창출하고 있다.

표 1. IoT 서비스 분야별 구현기능

분야	구현기능
헬스케어	• 활동상태 관리, 보호자 모니터링
에너지	• 시스템/배터리 소비전력 모니터링
제조	• 반복업무자동화 및 실시간 모니터링
스마트홈	• 원격 홈관리 및 보안향상
금융	• 결제 간소화, 생체 인증보안
교육	• 자동 출결 시스템, 전자도서관
국방	• 무인 이동체, 감시, 정찰
농림/축산	• 산업환경 데이터 수집, 스마트팜 등
교통	• 자율주행 자동차, 주차장 자동관리
관광	• 맞춤형 관광 상품 패키지, 위치기반
유통	• 물류창고 관리시스템, 운송 및 무인택배
건설	• 건물 네트워크 보안, 에너리관리

위와 같이 IoT와 인공지능 기술이 다양한 분야에 활용됨에 따라 동반되는 역기능적인 부작용은 이들의 보안취약점을 노리는 각종 사이버공격 들을 들 수 있다. 즉, 적대적 인공지능의 출현으로 공격자는 침투한 네트워크를 더욱 효과적으로 탐색할 수 있게 된다. 이에 대응하기 위해서는 더욱 자동화되고 지능적인 방어방법이 필요하다. 주요 위협을 살펴보면, IoT 서비스를 위해 개방된 장소에 설치된 센서노드에 대한 비인가자의 물리적 접근 및 파손을 들 수 있다. 사용자가 소유한 스마트 폰, 스마트 기기 등 단말의 분실 및 도난에 의한 통신기능 상실로 인해 IoT 서비스가 중단될 수도 있다[4]. 단말을 분실 할 경우엔 정보 유출 사고 등을 들 수 있다. 이런 위협에 대응하기 위한 하나의 방안으로 인공지능기술의 IoT 통합보안관제시스템을 들 수 있다.

2. 인공지능 기술의 IoT 통합보안관제 체계

앞장에서 설명한 신기술에 의해 발생하는 다양하고 복잡한 서비스별 위협을 선제적으로 대응이 필요하여 보안관제기관의 사이버 위기를 능동적이고 효과적으로 관리할 수 있는 능동형 통합보안관제를 제시하였다.

인공지능 IoT 통합보안관제시스템을 구축하기 위해서는 다음과 같은 4가지 사항을 고려해야 한다.

첫 번째, IoT센서망 취약점 진단, 침해사고 조사 등을 담당하는 업무가 필요하다. 센서 기술, 무선 기술 등과 같은 새로운 기술에 의해 발생할 수 있는 신규 사이버 위협을 전담으로 대응할 업무조직이 필요하다.

두 번째는 알려지지 않은 위협탐지에 머신/딥러닝 기술 적용 장비 도입이 필요하다[5]. 방대한 위협데이터를 한정된 보안관제요원이 처리할 수 있도록 머신/딥러닝 모델을 개발해야 한다.

세 번째, 보안관제요원이 일 평균 수십만 건의 이벤트를 실시간 대응하는 것은 현실적으로 불가능하다. 정확한 위협 식별과 관제 효율성을 높이기 위해 [표 2]처럼 공격유형을 정의하고 실제 상황에 맞게 정의가 필요하다.

표 2. 공격유형의 정의

공격유형	내용
웹해킹	• 홈페이지 위·변조, 개인정보 유출 등 웹 애플리케이션의 취약점을 공격
악성코드	• 백도어, 스파이웨어, 랜섬웨어, 트로이목마, 웜, 바이러스 등 유포 및 감염
비인가접근	• 시스템 정보수집을 위한 스캐닝, 불법적인 네트워크 및 시스템 침입시도
서비스거부	• 네트워크 대역폭 및 시스템 자원을 고갈시키고, 어플리케이션/서버 프로그램을 공격하여 정상적인 서비스를 방해하는 공격 • 시스템 동작 중단, 어플리케이션 동작 중단 등
오남용	• 보안정책 위반 비보안 서비스 사용 등, • 시스템 접근권한의 오용 및 남용

현재 보안 분야에 AI를 적용하여 구현된 사례는 [표 3]과 같으며, 본 논문에서는 취약점 현황과 공격대상의 특성을 반영하고, IoT 로그까지 포함하도록 차별성을 두었다.

표 3. AI 보안관제 시스템 구현사례

제조사	Dark Trace	Vectra Networks	IBM	이글루 시큐리티	시큐레이어	티쓰리큐	RSA
강점	감사에 특화	N/W 트래픽 중심데이터 분석	A.I. Watson 사용	머신러닝 기반국내 사례보유	보안데이터 전문 기계학습	빅데이터, AI 통합 플랫폼	암호화 데이터 패킷분석
학습 모델	머신 러닝, 딥러닝	머신러닝, 딥러닝	머신 러닝, 딥러닝	머신 러닝	머신 러닝	머신 러닝, 딥러닝	머신 러닝
국내/외	국외	국외	국외	국내	국내	국내	국외
정·오탐 식별	X	X	X	Ids/IPS/ WAF	IDS/IPS	X	X
T.I. 연동	자체	자체	자체	자체	별도	별도	자체
외부 위협	O	O	O	O	O	O	O
내부 위협	X	X	O	O	O	O	X
이상 탐지	N/W	N/W	N/W	웹로그 방화벽	웹로그	X	X

네 번째, 위험도는 '상'(침해사고가 발생한 경우)만 실시간 대응하고, '중'(위협이 탐지 및 차단된 경우) 및 '하'(단순 정보성인 경우)는 시간을 가지고 중요시스템에 대한 공격 간의 연관성을 분석(상관분석)하여 대응하고 있으며 3등급 위험도는 [표 4]과 같다.

표 4. 3등급 위험도

위험도	상태	표현
상 (High)	• 침해사고가 발생한 경우로 실시간 대응 ex) 정보유출, 서비스마비 등	가청/가시 (빨강색)
중 (Medium)	• 위협이 탐지 및 차단된 경우로 시간을 가지고 대응 ex) 침입방지시스템에서 위협탐지 및 차단 등	가시 (노란색)
하 (Low)	• 단순 정보성인 경우로 시간을 가지고 대응 ex) 유지보수가 방화벽에 접속 등	가시 (초록색)

IoT 기기의 원초적인 보안환경을 확보하기 위해서 [표 5]와 같이 보안원칙이 지켜져서 설계, 배포, 운영된다.

표 5. IoT 공통 보안원칙

단계	IoT 공통 보안 원칙
설계 개발	• 정보보호와 프라이버시를 고려한 IoT 제품·서비스 설계
배포 설치 구성	• 안전한 초기 보안 설정 방안 제공 • 안전한 설치를 위한 보안 프로토콜 준수 및 안전한 파라메터 설정
운영 관리 폐기	• IoT 제품·서비스 취약점 패치 및 업데이트 지속 이행 • IoT 제품·서비스 취약점 패치 및 업데이트 지속 이행 • IoT 침해사고 대응체계 및 책임 추적성 확보 방안 마련

3. 인공지능 기술의 IoT 통합보안관제 프로세스

앞장에서 설명된, 체계와 함께 구현되어야 하는 부분이 바로 프로세스이다. 현재의 수동형 보안관제 단계는 [그림 1]과 같다. 이의 구조적 문제점은 단계 간 업무 연계성이 미흡하고, 단계별 수행업무가 불명확하다. 이는 보안관제와 운영 간에 역할 및 책임이 모호한 구조이다. 즉, 예방단계에서는 웹 서비스 위주로 취약점을 점검하여 네트워크 및 서버와 인프라에 대한 취약점 점검이 필요하다. 탐지단계에서는 알려진 공격을 차단하는 패턴 탐지기반의 보안시스템에서 최신공격기법이 적용되어 보안정책을 위회하는 지능형 공격이 탐지되지 않는다. 대응 단계에서는 급증하고 있는 사이버 위협을 한정된 인원 및 역량이 서로 다른 보안관제 요원이 대응하기에는 역부족이다. 사후 관리단계에서는 재발 방지대책 위주의 프로세스로 알려지지 않은 패턴을 이용한 지능형 공격에 대처할 수 없다.

그림 1. 수동형 통합보안관제 프로세스

이를 개선하기 위해서는 [그림 2]와 같은 능동형 보안관제 프로세스가 적용되어야 한다. 이로써, 방대한 위협데이터를 자동 분석하여 머신/딥러닝 시켜 새로운 위협을 예측/제거할 수 있도록 한다. 즉, 기존 사후관리 프로세스를 예측 프로세스로 개선이 필요하다.

그림 2. 능동형 통합보안관제 프로세스

앞서 설명한 능동형 통합보안관제 프로세스는 알려지지 않은 보안 위협탐지와 방대한 위협데이터를 머신러닝, 딥러닝 기술이 적용 된 보안시스템도 추가되어야 한다. 추가 적으로 네트워크 및 시스템에 은닉된 위협을 적극적으로 찾아 제거할 수 있는 보안시스템도 필요하다.

4. 목표 플랫폼 설계

앞에서 살펴본 사항들을 반영하여 [그림 3]과 같은 인공지능 IoT 통합보안관제 목표플랫폼을 새롭게 제안하였다. [그림 3]에서 보듯이 기존처럼 보안장비에서 도출되는 각종 보안 로그를 포함한다. 또한, IoT 관련 센서 및 각종 중요시설에 대한 상태정보 그리고 IoT 기기들의 상태까지도 포함한다. 보안관제를 하면서 생성된 취약점 및 공격에 대한 통계정보와 최신 보안 동향도 필요하다. 주요 시스템으로는 개인정보 처리시스템, 주요 정보통신 기반시설, 주요 홈페이지 서버들의 현황을 포함하여 보안관제 시스템인 SIEM(security information and event management)으로 모두 수집한다. 이렇게 수집된 데이터를 기반으로 사이버공격에 대한 탐지 및 분석이 이루어진다[6]. 이때, 데이터에 대한 필터링, 정규화, 분류 하게 된다. 이로써, 공격유형을 정의하여 분류하며, 위험도 기준에 따라서 이벤트 발생 시 공격과 오탐을 판단하게 된다.

인공지능 IoT 통합보안관제 플랫폼으로 수집된 데이터를 주고 받을 때는 JSON(Java Script Object Notation)을 사용하여 데이터의 외곡을 최소화한다. 이로써, 새로은 프로세스인 능동형 통합보안관제 프로세스가 적용되게 된다.

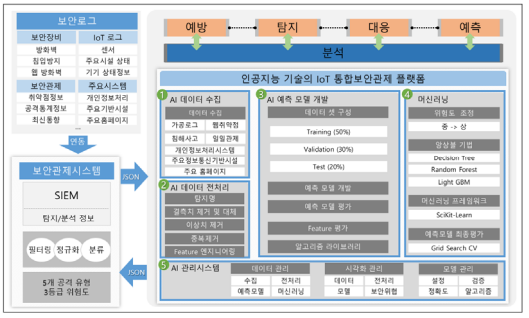

그림 3. 인공지능기술의 IoT 통합보안관제 목표플랫폼

이들 프로세스를 요약정리하면, ① 데이터 수집을 하면서 가공로그뿐 아니라, 원본로그까지 포함하여 수집하게 된다. ② 데이터 전처리에서는 정의된 탐지명을 기준으로 처리되며, 결측치와 이상치, 중복제거까지 이루어진다. 특히 중요한 데이터에서 중요한 특성값을 선정하게 된다. ③ 예측모델을 개발하기 위해서는 전체 데이터셋에서 트레이닝 데이터셋을 50% 사용한다. 또, 정합성 확인을 위한 벨리데이션 데이터셋으로 30% 한다. 나머지 테스트 데이터셋으로 20%를 사용하여 예측모델 및 피쳐값(Feature value)에 대한 평가를 수행한다. 마지막으로 머신러닝을 위한 알고리즘을 선정한다. ④ 머신러닝 시에는 위험도를 예측모델에 맞게 조정할 수 있다. 이때 머신러닝을 위해 알상블기법을 활용하여 최고의 학습방법을 도출하여 사용된다. 머신러닝 프레임워크는 'SciKit-Lean'을 사용하고, 최종 예측모델 평가방식은 'Grid Search CV'로 하게 된다. ⑤ AI 관리시스템에서는 데이터 관리에서 시각화와 모델관리까지 일련의 과정을 관리하게 된다.

5. 활용을 위한 극복방안

인공지능기술을 활용하여 IoT 서비스환경의 보안이 모두 해결된 것은 아니다, [그림 4]와 같이 그 외의 관리적, 기술적인 보안 요구사항도 함께 보강되어 사이버공격에 대해 더욱 철저한 보안환경이 구축될 수 있다[10]. 특히 주요 정보시스템에 대한 보안을 법적 준거성을 확보하기 위해 '행정기관 및 공공기관 정보시스템 구축·운영 지침'을 기반으로 수립된 [그림 4]의 보안대책을 이행한다. 이로써, 더 완벽한 보안환경을 구축할 수 있다.

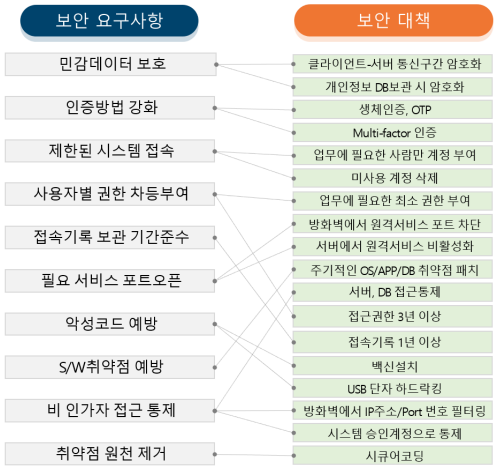

그림 4. 법적 준거성 확보를 위한 보안대책

실무환경에서 중요한 건 인공지능 기술과 IoT 관련 보안 위협에 대한 실 업무에 적용할 수 있도록 적극적인 협조가 필요하다. 이런 신기술은 사람만이 할수 있는 창의적이거나 복합적인 일을 생성한다. 결국 더욱 많은 일을 효율적으로 할 수 있도록 한다.

Ⅲ. 인공지능 기술의 IoT 통합보안관제 데이터모델링

1. 데이터모델링 개요

데이터모델링을 할 경우 그 목표를 선정하고 목표에 맞는 방법으로 하는 것이 무엇보다 중요하다. 본 논문에서는 인공지능 기술을 활용하여 보안이벤트에 대한 정합도를 인공지능을 통해 높일 수 있도록 머신러닝에 적합한 모델링을 제안하였다. 그 안에는 IoT 관련 데이터가 포함되도록 하는 것이 핵심이라고 할 수 있다.

2. 데이터 수집

본 논문의 실습을 위해 사용된 데이터는 국내 공공기관의 광역시급의 6개월간의 보안로그 데이터량으로 데이터 수집은 실제 업무에서 사용되는 보안관제시스템의 가공로그(6개월), 취약점 정보(6개월), 침해사고 조사 결과보고서(6개월), 일일 관제일지(6개월)를 자료를 활용하였다.

2. 데이터 전처리

데이터 전처리는 [그림 5]와 같이 6개월간의 많은 데이터 중 실효성 있는 필요 데이터셋을 구성하기 위해서 결측값, 이상치값을 제거하였다. 결과값은 'one-hot encoding' 하여 처리하였다.

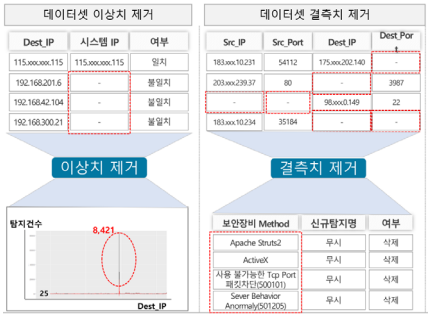

그림 5. 데이터 전처리 과정

앞의 데이터 전처리 과정을 거쳐 데이터는 정규화와 표준화 되었다. 수집 데이터 건수는 15,711,286 건에서 전처리 후 침해 공격명에 의해 재분류된 데이터 건수는 18,818,9978건로 데이터 건수가 변경되었다.

3. 특성선정 및 데이터셋 구성

데이터 전처리를 통해 재분류된 데이터를 기반으로 [그림 6]과 같이 1차 탐색에서 기술통계/빈도분석과 상관분석을 통해 대표 값을 선정했다. 2차 탐색은 등급화 산정과 이상 값 존재 탐색을 해서 최종 특성을 선정했다.

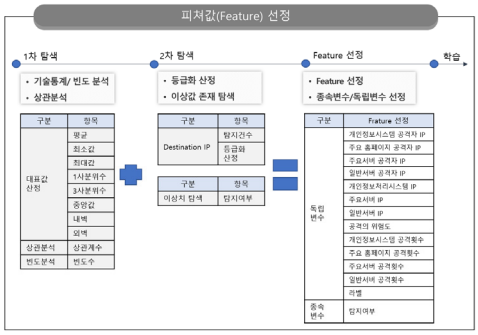

그림 6. 특성선정 절차

데이터셋 구성 시 [표 6]와 같이 컬럼은 특성 명칭을 영문 소문자로 구성하였으며, 2개 이상의 단어를 합성하여 구성하였다.

표 6. 데이터셋 구성내용

컬럼명	설명	
pis_src_ip	개인정보시스템 공격자 IP 주소	one-hot en
ws_src_ip	주요 홈페이지 공격자 IP 주소	one-hot en
ics_ws_src_ip	주요서버 공격자 IP 주소	one-hot en
gs_src_ip	일반서버 공격자 IP 주소	one-hot en
pis_dst_ip	개인정보처리시스템 IP 주소	one-hot en
ws_dst_ip	주요 홈페이지 IP 주소	one-hot en
ics_ws_dst_ip	주요서버 IP 주소	one-hot en
gs_dst_ip	일반서버 IP 주소	one-hot en
attack_risk	공격의 위험도	3등급
pis_ac_day_cnt	개인정보시스템 공격 횟수	평균값
ws_ ac_day_cnt	주요 홈페이지 공격 횟수	평균값
ics_w_ac_day_cnt	주요서버 공격 횟수	평균값
gs_ ac_day_count	일반서버공격 횟수	평균값
breach_proba	Label	0/1

Ⅳ. 인공지능 기술의 IoT 통합보안관제 데이터모델링 검증

1. 알고리즘의 선정

앞장에서의 데이터모델링을 검증하기 위해 사용하는 결정트리 기반의 앙상블 알고리즘은 현업에서 자주 사용하는 'LightGBM'을 선정하였다. 빠른 학습속도 및 높은 효율성을 보장하는 히스토리그램 기반 알고리즘으로 메모리 사양량이 적기 때문이다[7]. 높은 정확도를 만들고, 과적합 될 가능성이 있지만 매개변수로 조정할 수 있다. 범주형으로서 현재 데이터셋을 머신러닝하기에 최적이다.

2. 데이터모델링 정합성 검증

머신러닝에서 모델에 대한 성능평가 지표에는 정확도, 정밀도, 재현율, '오차행렬', 'GridSearchCV' 의 지표를 사용하여 수행한다. 'GridSearchCV' 는 교차 검증을 기반으로 하이퍼 파라미터의 최적값을 찾는다. 데이터 세트를 교차 타당성 검증을 위해 학습/테스트 세

트로 자동으로 분할한 뒤에 하이퍼 파라미터 그리드에 기술된 모든 파라미터를 순차적으로 적용하였다. 이로써, 최적의 파라미터를 찾을 수 있다. 'GridSearchCV'는 사용자가 조율하고자 하는 여러 종류의 하이퍼 파라미터를 다양하게 테스트하면서 최적의 파라미터를 편리하게 찾게 해준다. 동시에 순차적으로 파라미터를 테스트하므로 수행시간이 상대적으로 오래 걸리는 단점은 있다. 정확도는 [표 7]처럼 산정하고, 직관적으로 모델 예측 성능을 나타내는 평가지표이다. 이진 분류의 경우 데이터의 구성에 따라 머신러닝 모델의 성능을 왜곡할 수 있기 때문에 정확도 수치 하나만 가지고 성능을 평가하지 않는다.

표 7. 정합도 평가지표 공식

정확도 (Accuracy)	=	예측 결과가 동일한 데이터 건수 / 전체 예측 데이터 건수

3. 모델학습 및 최종평가

앞에서 선정된 알고리즘을 통한 모델학습을 위해서 [그림 7]과 같이 파이선에서 사이킷런(scikit-learn)을 사용하였다. 파이선 기반의 머신러닝을 위한 가장 쉽고 효율적인 개발 라이브러리를 제공한다.

최종 성능평가지표를 위해서 Training Data(70%)를 통해서 학습하고, Test Data(30%)를 기준으로 학습 성능을 최종 평가 한 결과는 아래와 같이 매우 높은 정합도를 보였다. 하지만, [표 8]과 같이 너무 높은 정합도는 오탐을 야기할 수 있고, 이를 개선하기 위해서 더욱 다양한 로그를 통한 연구가 지속되어 개선할 필요가 있다.

표 8. 탐지건별 학습성능 평가지표

오차행렬 [[287165]] 정확도 :1.0000, 정밀도 : 0.0000, 재현율 : 0.0000, F1 : 0.0000 ROC AUC 값 0.0000

Ⅴ. 결론

본 논문에서 새롭게 개발된 예측 모델은 IoT에 대한 통합보안관제를 인공지능기술을 적용할 수 있도록 하였다. 하지만, 인공지능기술을 통해 사람처럼 판단하려면 더욱 많은 로그를 통해 수많은 라벨링을 하여 정합도를 높여야 한다. 한정적인 로그로 학습을 한 결과는 정합도는 매우 높지만, 오버피팅(Overfitting)으로 인한 오탐을 유발하는 한계점이 있다. 추후 더 다양한 로그와 실제 IoT 로그까지 활용하여 학습할 수 있도록 지속적인 연구를 해야 한다. 이를 통해 실무에서 보안관제 요원 수십 명이 오랜 시간 분석해야 할 정·오탐에 대해 단시간에 인공지능이 정합도가 높은 판단을 하여 단순분석 인력은 현저하게 줄일 수 있을 것이다.나아가 더 다양한 환경에서 적용할 수 있는 연구의 기초자료가 되어 연구 분야를 확대할 수 있을 것이다.

구현 된 결과는 소프트웨어 모듈 형태로 탑재하여 사용할 수 있도록하고, 추가로 위협데이터를 기반으로 더 많이 학습하여 적용할 수 있는 예측모델이 지속적으로 개발되어야 한다.

참 고 문 헌

[1] 김진성, *안전한 AIoT 서비스를 위한 자율 사물의 보안성 강화 방안,* 아주대학교 정보통신, 석사학위논문, 2021.

[2] 정지만, *4차 산업혁명을 대비한 딥러닝 기술의 금융보안 적용 연구,* 국민대학교 컴퓨터공학, 석사학위논문, 2018.

[3] 김인원, *지능형 영상보안시스템의 수용의도 영향 요인에 관한 연구,* 숭실대학교대학원 IT정책경영학과, 박사학위논문, 2017.

[4] 홍준혁, *인공지능기반 보안관제 구축 및 대응 방안,* 배재대학교 사이버보안학과, 석사학위논문, 2021.

[5] 이보라, *금융 보안에서 휴먼팩터를 고려한 인간과 인공지능의 역할 및 협업 모델,* 고려대학교 정보보호대학원, 연구논문, 2018.

[6] 정진영, *인공지능을 활용한 금융권 통합보안관제 자동화 방안,* 건국대학교 정보통신대학원, 석사학위논문, 2018.

[7] 윤상필, *인공지능 시대의 보안 패러다임과 책임구조의 변화_규범의 역할과 보안정책의 원칙,* 고려대학교 정보보호대학원, 석사학위논문, 2018.

[8] 김기환, *인공지능 신경망을 활용한 AI보안 모델게반 실시간 상관전력분석 공격*, 동서대학교 유비쿼터스IT학과, 박사학위논문, 2021.
[9] 이세호, *사이버보안 프레임워크 기반의 보안 오케스트레이션 서비스 모델 제안*, 배재대학교 사이버보안학과, 박사학위논문, 2021.
[10] 박성준, *인공지능기반 엔드포인트 보안관제 구축 방안*, 동국대학교 국제정보보호대학원 정보보호, 연구논문, 2021.

저 자 소 개

오 영 택(Young-Taek Oh) 정회원

- 2019년 2월 : 배재대학교 사이버보안과(공학석사)
- 2014년 7월 ~ 현재 : 이글루시큐리티 (보안컨설턴트)
- 2019년 3월 ~ 현재 : 배재대학교 사이버보안과(공학박사)

〈관심분야〉 : 정보보호, 인공지능, 보안컨설팅, IoT

조 인 준(In-June Jo) 정회원

- 1982년 2월 : 전남대학교 계산통계학과 졸업
- 1985년 2월 : 전남대학교 전자계산학과 석사
- 1999년 2월 : 아주대학교 컴퓨터공학과 박사
- 1983년 ~ 1993년 : 한국전자통신연구원 선임연구원
- 1994년 ~ 현재 : 배재대학교 사이버보안학과 교수

〈관심분야〉 : 정보보호, 컴퓨터네트워크보안, 전산조직응용