논문 20-45-09-20 The Journal of Korean Institute of Communications and Information Sciences '20-09 Vol.45 No.09
https://doi.org/10.7840/kics.2020.45.9.1642

머신러닝을 이용한 웹 공격 탐지 방안에 관한 연구

류 호 균◆, 김 광 용°

A Study on Detection Method of Web Attack Using Machine Learning

Ho-Gun Rou◆, Gwang-Yong Kim°

요 약

우리 사회는 정보기술에 대한 의존도가 점차 높아지고 있으며, 기관과 기업은 사물인터넷, 클라우드, 빅데이터, 인공지능 등 혁신적인 기술을 이용하여 각종 맞춤형 서비스를 제공하고 있다. 수집되는 방대한 데이터는 컴퓨터 자원을 효율적으로 활용하는 분산기술을 통해 효과적으로 처리되고 있지만, 한편으로는 다양한 보안 취약점을 발생시키기도 한다. 보안 취약점은 사이버 공격의 통로가 되며, 이 경우 기존의 보안 체계를 우회하여 침투할 수 있는 확률이 높아진다. 시그니처 탐지 기반의 기존 정보보안시스템으로는 적절하게 대응하는데 한계가 있기 때문에, 대안으로 인공지능 기술을 활용하기 위한 다양한 노력이 이루어지고 있다. 본 연구에서 실제 운영 중인 홈페이지로 들어오는 공격을 인공지능 기술로 탐지 후 기존 정보보안시스템의 탐지 내역과 비교하는 방식으로 실험을 진행하였다. 실험 환경의 제약으로 인해 연구의 범위가 웹 공격으로 한정되었지만, 후속 연구에서는 인공지능 기술을 다양한 유형의 사이버 공격 대응에 적용하고 오 탐지에 대한 개선방안을 제시함으로써, 조직의 실질적인 정보보안 수준 강화에 기여할 것을 기대한다.

키워드 : 인공지능, 기계학습, 머신러닝, 알고리즘, 웹 공격, 탐지
Key Words : A.I., Artificial Intelligence, Machine Learning, Algorithm, Web Attack, Detection

ABSTRACT

Our society is becoming increasingly dependent on information technology, and institutions and companies are providing various customized services using innovative technologies such as the Internet of Things, the cloud, big data and artificial intelligence. The vast amount of data collected is effectively processed through distributed technology that efficiently utilizes computer resources, but on the other hand, it also creates a variety of security vulnerabilities. A security vulnerability is a pathway to cyberattacks, which increases the likelihood of infiltration by bypassing existing security systems. Since there is a limit to the proper response of the existing information security system based on signature detection, various efforts are being made to utilize artificial intelligence technology as an alternative. In this study, the experiment was conducted by detecting attacks coming into the actual operating homepage with artificial intelligence technology and comparing them with the detection details of the existing information security system. Although the scope of research has been limited to web attacks due to limitations in the experimental environment, follow-up research expects to contribute to strengthening the organization's actual level of information security by applying artificial intelligence technology to various types of cyber attack responses and presenting improvement measures for false detection.

◆ First Author : Soongsil University Graduate School of IT Policy & Management, hgr1203@soongsil.ac.kr, 학생(박사), 정회원
° Corresponding Author : Soongsil University Department of Management, gygim@ssu.ac.kr, 정교수, 정회원
논문번호 : 202007-156-0-SE, Received July 6, 2020; Revised July 20, 2020; Accepted July 20, 2020

Ⅰ. 서 론

오늘날 우리 사회는 컴퓨터와 인터넷 기술의 발달, 다양한 서비스의 보급으로 등으로 인해 공공·민간 전 분야에서 사이버 의존도가 점차 높아지고 있으며, 네트워크를 통한 정보의 양은 기하급수적으로 증가하고 있다. 특히 각 기관이나 기업에서 운영하고 있는 웹 서비스는 단순한 정보전달이나 홍보 목적 외에도 국민이나 고객의 알 권리, 마케팅, 개인정보전달 등 그 사용도가 높아지고 있으며[1], 인공지능(Artificial Intelligence), 사물인터넷(Internet of Things: IoT), 빅데이터(Big data) 등의 혁신적인 기술을 중심으로 고객 맞춤형 서비스를 제공하여 큰 변화를 이끌고 있다[2].

네트워크와 서비스가 상호 융·복합되고 있고 분산 기술을 이용하여 컴퓨터 자원을 효율적으로 공유할 수 있게 되어 방대한 데이터의 효과적인 분석이 가능하게 되었지만[3], 이는 공격자로 하여금 보다 다양한 공격방법과 기술, 표적 대상 등 수 많은 보안위협의 경로를 제공하고 있다[4]. 이러한 보안위협에 대응하기 위해 IT서비스를 운영하는 조직은 침입방지시스템(Intrusion Prevention System: IPS), 웹방화벽(Web Application Firewall: WAF), 보안 정보 및 이벤트 관리 시스템(Security Information & Event Management: SIEM) 등 다양한 정보보안시스템과 보안운영센터(Security Operations Center: SOC)를 운영하고 있으나, 시그니처 기반의 운영 특성 때문에, 최근의 신종(Zero-Day) 또는 우회 공격의 대응에 한계를 보이고 있다.

특히, 인공지능의 범용적인 발전은 역설적으로 기존 정보보안 체계를 더욱 위험에 빠뜨리고 있는데, 공격자가 인공지능의 활용을 통해 기존 인간의 능력에 따른 제약을 받지 않음으로써 기존에는 가능하지 못했던 공격을 수행할 수 있게 되었다[5].

인공지능이 적용된 공격을 기존의 체계로 대응하는 것이 점차 불가해짐에 따라, 인공지능 기반 방어체계의 필요성이 커지고 있는데, 다양한 정보보안시스템에서 빅데이터 수준으로 데이터를 생성해서 머신러닝을 위한 학습데이터로 활용하고 사이버 공격 이전까지의 이런 데이터를 잘 축적해 놓은 다음 이상행위가 탐지되었을 경우 바로 방어를 한다면, 보안관제 분야에 인공지능이 효과적으로 사용될 수 있다[6].

최근 수년간 정보보안 분야에 인공지능을 적용하기 위한 노력이 있어왔는데, 인공지능을 이용하여 사이버 공격을 탐지하는 유형은 크게 세 가지로 요약할 수 있다. 첫째, 이미 알려진 공격에 대해 보안시스템에서 탐지한 이벤트가 정탐인지 오탐인지 확인하는 것이다. 둘째, 정상행위 기록에서 이상행위를 분석하고 새로운 공격 패턴을 찾아내는 것이다. 셋째, 국내외 보안연구소 및 기관으로부터 입수된 새로운 공격 패턴을 적용하여 유사한 행위나 흔적을 찾아내는 것이다[7].

본 연구는 위 세 가지 중 두 번째 유형인 정상행위 기록에서 이상행위를 분석하고 새로운 공격 패턴을 찾아내는 실험을 진행하고자 한다. 즉, 머신러닝 알고리즘으로 웹 로그를 분석함으로써, 기존 정보보안 시스템에서 탐지되지 않은 외부로부터의 웹 공격을 탐지해 낸 실험 결과를 제시한다.

본 논문은 다음과 같이 구성된다. 2장에서는 본 연구에서 다룰 정보보안시스템과 인공지능 기술의 배경을 살펴보고, 3장에서는 알고리즘을 이용하여 웹 공격을 탐지하기 위한 실험 환경을 제시한다. 4장에서는 실험 결과를, 마지막 5장에서는 본 논문의 결론과 향후 연구방향에 대해 기술한다.

Ⅱ. 관련 연구

2.1 정보보안시스템

정보보호시스템이라고도 하며, 정보통신망을 통해 수집·저장·검색 및 송수신되는 정보의 훼손·변조·유출 등을 방지하기 위한 기술이나 장치이다. 정보보안시스템의 분류는 [표 1]과 같다.

표 1. 정보보안시스템 분류(한국정보보호산업협회)
Table 1. Classification of Information Protection Systems

분류	항목
네트워크 보안	웹방화벽, 네트워크방화벽, 침입방지시스템, 디도스 차단 등
시스템 보안	백신, 안티 스파이웨어, 안티 피싱, 스팸차단, Secure OS 등
콘텐츠	DB 보안, DB 암호, PC 보안 등
암호/인증	OTP, SSO, HSM, EAM 등
보안관리	ESM/SIEM, PMS 등
기타	기타

2.1.1 침입방지시스템(IPS)

침입방지시스템은 네트워크 트래픽과 동일 선상인 in-line 형태로 존재하여 트래픽을 감시하고 악성으로 판단될 경우 해당 패킷을 차단하는 등 적극적으로 공격에 대처하는 능동형 시스템이다. 방화벽과는 달리 운영체제나 어플리케이션의 취약점을 능동적으로 사전에 정의하여 웜이나 바이러스, 비정상 트래픽을 차

단할 수 있어 한 단계 더 진보한 보안시스템이라 할 수 있다. 이러한 침입방지시스템은 웜, 바이러스, 악성코드 및 해킹으로부터 유발되는 유해트래픽을 사전 차단함으로써 인터넷 및 내부 네트워크 자원의 효율적 사용을 도모할 수 있다. 그러나 다양한 공격을 방어하기 위한 이벤트의 종류가 많고 공격유형에 따른 탐지정책의 복잡함으로 오 탐지가 발생할 가능성이 존재한다. 따라서 침입방지시스템을 도입한 전산환경에 따라 이벤트 발생조건에 대한 오차가 크므로, 최적화된 탐지규칙을 정의하기 까지 상당한 시간이 소요될 수 있다[8].

2.1.2 웹방화벽(WAF)

웹방화벽은 네트워크 트래픽과 동일 선상지금의 TCP/IP 기반의 인터넷 환경은 웹서비스의 발달과 밀접한 관계가 있다. TCP/IP의 허술한 보안체계에도 불구하고, 웹서비스의 기반기술인 HTTP 프로토콜이 바로 TCP/IP 환경에 최적화 되어 운영되기 때문이기도 하다. HTTP 프로토콜을 이용하여 서비스하는 것을 웹 어플리케이션이라고 한다. 주로 초기에는 정적인 HTML(HyperText Markup Language)언어로 제작되었다가 웹 화면 구현의 용이성, 각 객체 간 상호 연동, 게시판과 같은 실시간 동적 홈페이지 구현을 위해 ASP, JSP 등 다양한 프로그래밍 언어로 구현되고 있다. 인터넷 초기에 단순 정보공유 차원에서 운영되던 웹은 점차 기업의 주요 마케팅 도구로 발전하면서 아이디, 비밀번호 등 개인정보를 포함하여 각종 기업정보를 포함하게 되었다. 또한 웹서비스의 특성상 불특정 다수에게 조건 없이 접속을 허용해야 함에 따라 해커들은 각 기관이나 기업의 웹서버 취약점을 통해 내부 시스템으로의 접근을 시도하게 되었다. 특히, 웹 환경에서 이루어지는 전자상거래가 발달하면서 인터넷 쇼핑몰이 해커들의 주 공격대상이 되었고 고객의 단순 개인정보 뿐만 아니라 금융정보까지 획득하면서 금전거래와 같은 악의적 목적으로 발전되었다. 웹서비스는 외부 모든 사람에게 접속을 허용해야하기 때문에 방화벽은 웹서비스를 대상으로 하는 침입시도를 방어할 수 없다. 침입방지시스템 또한 홈페이지 공격을 방어하기에는 제한적일 수밖에 없는데 웹서비스를 이용하는 것 자체가 정상 행위이기 때문에 정상사용자와 해커의 분별이 어렵다.

이러한 웹서비스의 특징 때문에 웹에 특화되어 방어하는 보안시스템인 웹방화벽이 개발되었다.

대부분의 웹 공격은 웹 애플리케이션을 구축할 때 생기는 취약점을 이용하여 웹 서버를 공격하거나 데이터베이스 내용을 악용하는 형태이다[9]. 현재 웹 방화벽은 HTTP의 취약점을 기준으로 각 기관이나 기업의 웹 환경을 비교하여 정상과 공격을 정의하는 방식을 갖는다. 그러나 웹 방화벽은 모든 웹 트랜잭션에 대한 보안성을 확인하여야하기 때문에 시스템 성능이 중요한 요소가 된다. 뿐만 아니라 웹페이지 구현시 누가, 어떤 언어로, 어떻게 프로그래밍 하였는지에 따라 보안정책이 매우 다르게 적용되기 때문에 최적화 되기 까지 상당 시간이 소요된다. 웹 방화벽의 보안정책은 매우 복잡하다. 따라서 보안관제 관점에서도 오 탐지가 가장 많고, 공격 여부를 판단하는데 오랜 시간이 소요된다[8].

2.1.3 보안 정보 및 이벤트 관리 시스템(SIEM)

최근의 고도화/대규모화되는 네트워크 인프라를 통한 다양한 보안 위협을 조기에 탐지 및 대처하기 위해서는 체계적인 수단이 필수적이다. 이를 위하여 각종 보안 장비, 네트워크 인프라, 서버/스토리지 장비 및 서비스 응용들로부터 생성되는 로그, 패킷 등 대량의 이벤트 데이터를 수집하고, 이에 대하여 빅데이터 솔루션을 활용한 보안 분석을 수행하는 보안관제 체계가 필요한데, 보안 정보 관리(Security Information Management: SIM)와 보안 이벤트 관리(Security Event Management: SEM)가 결합된 보안 정보 및 이벤트 관리 시스템(Security Information and Event Management: SIEM)이 널리 활용되고 있다[10].

[그림 1]에 제시한 바와 같이 SIEM은 가상/실제 네트워크들, 서비스 응용들, 시스템 로그들과 이벤트 데이터를 수집한 후에 이를 분류하고 분석해서 빠른 리포팅을 제공하고, 추가 개입이나 변경된 대응이 필요한 경우는 경고를 수행한다.

그렇기 때문에, SIEM은 IT 조직의 보안 컨트롤 타워인 보안운영센터에서 핵심적인 역할을 담당한다 [11].

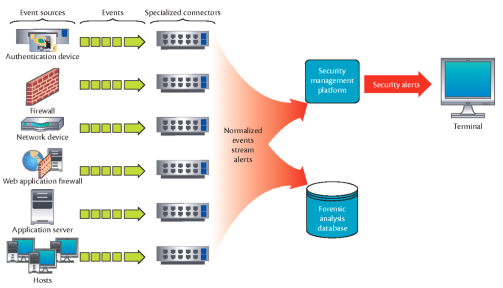

그림 1. 일반적인 SIEM 솔루션의 구조
Fig. 1. The structure of a common SIEM solution

2.2 인공지능

2016년 세계경제포럼(World Economic Forum)에서 인공지능 기술이 주요 기반 기술로 부상될 것으로 예측되었으며, 국내외 주요국에서 인공지능 기술 발전을 위한 다양한 국가 정책이 발표되었다[12].

제4차 산업혁명의 물결 속에서 모든 분야에서 혁신을 추동하는 대표 기술로 자리매김한 인공지능을 학계에서는 [표 2]와 같이 다양하게 정의하고 있다. 인공지능의 아버지로 불리는 존 메카시(John McCarthy)는 1956년 재직 중이던 다트머스대학교에서 개최한 회의에서 인공지능이라는 용어를 사용하였고, 2007년 11월 스탠퍼드대학교 홈페이지에 공개한 질의/답변 형태의 문서에서 "지능적인 기계를 만드는 과학 및 엔지니어링"이라고 정의하였다. 이후 많은 연구자들은 인간과 같은 지적 사고와 활동을 하는 시스템 및 이를 구현하기 위한 연구라는 큰 틀에서 인공지능을 정의하였고, 우리나라는 관계부처 합동(2016.12.27.)으로 수립한 「제4차 산업혁명에 대응한 지능정보사회 중장기 종합대책」에서 인공지능을 "인간의 인지능력(언어·음성·시각·감성 등)과 학습, 추론 등 지능을 구현하는 기술로 인공지능 SW/HW, 기초기술(뇌과학·산업수학 등)을 포괄"한다고 정의하였다[13].

표 2. 인공지능에 대한 다양한 정의[13]
Table 2. Various definitions of artificial intelligence

정의	연구자/기관
지능적인 기계를 만드는 과학 및 엔지니어링	McCarthy (1956)
인간의 사고와 의사결정, 문제해결, 학습과 같은 활동의 자동화	Bellman, R.(1978)
컴퓨터가 사고하도록 만드는 흥분되는 새로운 시도 ... 마음을 가진 기계	Haugeland, J.(1985)
계산 모델 활용을 통한 정신적 역량 연구	Charniak, E. and D. McDermott(1985)
사람처럼 지능이 요구되는 기능들을 수행할 수 있는 기계를 만드는 기술	Kurzweil, R.(1990)
지능형 행위를 계산 프로세스 측면에서 설명하고 에뮬레이션하는 연구 분야	Robert J. S.(1990)
사람이 더 잘하는 무엇인가를 어느 순간 컴퓨터가 할 수 있도록 하는 연구	Rich. E. and K. Knight(1991)
지각, 추론 및 행동을 가능하게 하는 계산에 관한 연구	Winston, P.E.(1992)
지능적 행위의 자동화과 관련된 컴퓨터 과학 분야	Luger, G. F & W.A. Stubblefield(1993)
인공적으로 만들어진 지능을 가지는 실체, 또는 그것을 만들고자 함으로써 지능 자체를 연구하는 분야	마쓰오 유타카 (2015)
인공적으로 만든 지적인 활동을 하는 물건(시스템)	
인간의 인지능력(언어·음성·시각·감성 등)과 학습, 추론 등 지능을 구현하는 기술로 인공지능 SW/HW, 기초기술(뇌과학·산업수학 등)을 포괄	관계부처합동 (2016.12.27.)

인공지능 기술을 정보보안 분야에 적용하려는 다양한 노력이 있어왔는데, 주로 다뤄지고 있는 기술은 [표 3]과 같이 머신러닝, 딥러닝, 알고리즘과 피쳐 등이 있다.

표 3. 인공지능을 정보보안 분야에 적용한 연구들
Table 3. Studies on the Application of AI to Information Security

연구자	연구 내용
정의섭 외 2인(2019)	익명 네트워크 이상징후 탐지
주영지 외 4인(2019)	(AI 기반) 보안관제시스템 정탐률 최대화
문재웅 외 3인(2019)	인공지능과 빅데이터 기반의 정보보안시스템 개발
이춘근(2019)	차세대 보안관제 체계 개발
오영택(2018)	통합보안관제 서비스모델 개발
정진영(2018)	통합보안관제 자동화
이국진(2018)	네트워크 이상징후 탐지
이상혁(2017)	봇넷 탐지
이현욱(2011)	비정상활동 패킷 분석을 통한 해킹 탐지

2.2.1 머신러닝(Machine Learning, 기계학습)

머신러닝은 명시적으로 작성된 프로그램에 따라 동작하는 것이 아닌 데이터로부터 학습을 통해 작업을 수행할 수 있도록 가르치는 것이다. 이러한 머신러닝은 인공지능의 한 분야로 컴퓨터가 여러 데이터를 이용하여 학습한 내용을 바탕으로 예측이나 결정을 도출하기 위해 특정한 모델을 구축한다. 머신러닝은 학습하는 환경 혹은 데이터의 속성 형태에 따라 지도학습, 비지도학습, 강화학습으로 나뉜다. 지도학습 방식은 정답이 정해진 문제에 대해 정해진 특성 정보를 통해 학습을 하고 모델을 생성한다. 그리고 생성된 모델을 통해 주어진 문제의 정답을 분류(Classification)한다. 비지도학습 방식은 지도학습의 정답의 분류와는

다르게 정답이 없는 데이터의 패턴을 분석 및 학습하여 유사한 데이터로 군집화(Clustering)하거나 유사한 데이터의 군집과 동떨어진 이상 데이터의 탐지에 활용된다[14]. 강화학습은 지도학습과 관련된 하위 분류이지만, 지도학습 기법이 사전에 '정답'을 학습에 이용하는 것과 달리, 강화학습은 시간에 따라 변화하는 외부 환경의 '반응'을 학습에 이용한다는 측면에서 지도학습과 구별된다[13].

2.2.2 딥러닝

딥러닝은 사람의 학습방법을 기계에게 가르침으로써, 사물이나 데이터를 군집화하여 분류하는 데 사용하는 기술이다. 종종 인공신경망(Neural Network)이라는 용어가 사용되기도 하는데, 이것은 인간의 뇌를 모방하는 것을 뜻한다. 인공신경망은 수학적 논리학이 아닌 인간의 두뇌를 모방하여 수많은 간단한 처리기들의 네트워크로 구성된 신경망 구조를 상정하는 것이다. 인공신경망을 이용하면 분류나 군집화를 원하는 데이터 위에 여러 층을 얹어서 많은 데이터를 서로 비교해 유사도를 구해주거나, 라벨링 되어있는 데이터를 기반으로 분류 학습하여 자동으로 데이터를 분류할 수 있다. 이러한 기술은 컴퓨터비전, 음성인식 등에서 큰 성과를 보이고 있다. 이 기술이 탑재된 인공지능 바둑 알파고는 4주 만에 인간이 1천년동안 학습해야 하는 양을 학습하여 인간 챔피언을 이길 수 있었으며. 딥뉴럴 네트워크는 많은 데이터를 통해 데이터의 핵심내용을 요약하고 학습하는 알고리즘 모델을 의미하는 것으로 인공지능 기술에서 중요한 부분을 차지한다[15].

2.2.3 알고리즘(algorithm)과 피쳐(feature)

알고리즘은 "어떤 값이나 값의 세트를 입력(input)으로 취해서 또 다른 어떤 값이나 값의 세트를 출력(output)로 만들어 내는 명확히 정의된 계산 절차" 즉, 컴퓨터를 사용한 계산 절차이다. 알고리즘은 입력을 출력으로, 다시 말해 투입 값을 산출 값으로 변형하는 연속적인 컴퓨터의 계산 단계인 것이다. 다른 측면에서 알고리즘은 정확히 명시된 컴퓨터 계산 문제를 해결하기 위한 수단으로 볼 수도 있다. 문제의 진술은 원하는 투입과 산출의 관계를 일상 언어로 명시하고, 알고리즘은 그러한 "입력(투입)과 출력(산출)의 관계를 획득할 수 있는 컴퓨터의 특수한 계산 절차를 서술하는 것"이다. 또한 알고리즘은 기본적으로 '곱(and)', '합(or)', '부정(not)'이라는 세 가지 기본 연산 기호에 의해 설계될 수 있는데, 이때 사용되는 연산 규칙은 입력된 기호를 출력된 기호로 변형하는 역할을 한다[16].

피쳐는 알고리즘을 구성하는 요소로 머신러닝 모델이 데이터를 분류하기 위해 사용하는 데이터의 정보 혹은 속성으로, 사용하는 피쳐에 따라 머신러닝 알고리즘의 분류 성능이 크게 달라진다[17].

Ⅲ. 실험 방법

실험은 머신러닝을 이용하여 XYZ社(가명)의 2개 홈페이지[표 4]를 대상으로 하였으며, 웹서비스(웹 로그)에 대한 이상 사용자의 접근이나 행위의 식별을 목표로 [표 5]의 계획과 같이 진행하였다.

머신러닝이 적용된 플랫폼(이하 'AI 시스템')이 설치된 서버를 기 구축된 보안인프라에 연동하였다. 2차례의 개념 검증(Proof of Concept: PoC)을 진행한 후에 탐지 결과를 알고리즘에 반영 후, 1개의 홈페이지(B)에 모의 공격을 2차례 수행하였다.

웹 공격을 탐지할 수 있는 IPS와 WAF에서 탐지되지 않는 비정상 행위를 식별하기 위하여 행위기반 공격을 수행하였다.

표 4. 실험 대상
Table 4. subject of experiment

구분	로그량(일)	수집방법
홈페이지(A)	4Gbyte	SYSLOG
홈페이지(B)	300Mbyte	SYSLOG

표 5. 실험 순서
Table 5. experimental sequence

순서	단계	내용
1	현황 분석	웹 서비스 대상 선정 로그 형식 및 네트워크 구성
2	환경 구축	AI PoC용 서버 입고 SIEM 로그 수집, AI 플랫폼 설치 및 연동
3	데이터 분석	SIEM 로그 분석 특징(피쳐) 추출, 탐지 모델 설계
4	전처리/모델링	학습데이터 적재 및 모델링
5	탐지모델	탐지 모델 예측
6	1차 결과 분석	탐지 결과 분석 및 보고서 작성 (A 홈페이지)
7	2차 결과 분석	탐지 결과 분석 및 보고서 작성 (B 홈페이지)
8	3차 결과 분석	모의 공격 및 예측 결과 분석 (B 홈페이지)

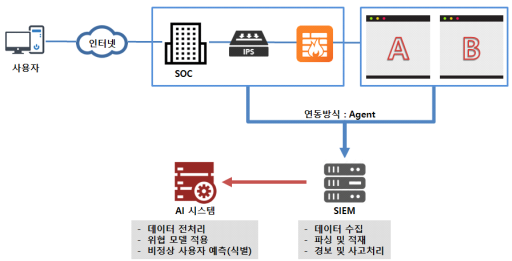

그림 2. AI 시스템 PoC 구성도
Fig. 2. AI System PoC Configuration Diagram

탐지모델 설계 시, 3가지 알고리즘을 사용하였고, 수집되는 로그를 분석 후 다수의 피쳐를 개발하여 적용하였다.[표 6]

표 6. 탐지 모델, 알고리즘, 주요 feature
Table 6. detection model, algorithm, key features

유형	탐지모델	사용 알고리즘	주요 feature
외부 사용자 이상 접근	비정상적인 접근 시도, 웹서버 과다접속 탐지, 이상 사용자 접근시도, BOT의 접근 시도, 취약한 XP사용자의 접근 시도 등	Isolation Forest	weak_refer, ip_count, user_pc_rate, user_etc_by_src_out_cnt, user_os_win_risk_cnt, user_bot_cnt 등
알려진 이상행위 탐지	SQL Injection, 크로스사이트 스크립팅, 원격시스템명령 실행, 관리자페이지 접근 시도, 자동화된 툴을 이용한 스캐닝 접근시도 등	Character-Level CNN, K-means	char_single_quotation_cnt, sql_kwd_and_ratio char_lt_cnt, xss_kwd_alert_cusum, cmd_echo_rate, sys_manager_cnt 등
의심스러운 데이터전송 탐지	대량의 데이터 유출, 지속적인 파일다운로드 등	Character-Level CNN, K-means	pkt_bytes_std, repeat_uri_cnt_max
잘못된 설정	잘못된 설정	Isolation Forest	repeat_uri_cnt_max

3.1 Character-Level CNN

Raw data를 글자단위로 학습하는 CNN 모델로써 보안 데이터에 특화된 모델이다. 새롭게 발생하는 text 형의 데이터에도 준수한 예측 성능을 보장하며, 엄청난 숫자의 파라미터를 가지는 모델을 생성하므로 메모리 사용량이 기타 모델보다 높은 편이지만 CNN은 GPU에서 가동하기 때문에, 시스템 메모리와 분산처리를 통해 성능을 보장할 수 있다.

3.2 Isolation Forest

의사결정나무를 앙상블 하는 방식의 알고리즘으로, 데이터셋 크기에 상관없이 작은 서브샘플을 사용하여 학습 시간이 빠르고 다차원 데이터에서도 쉽게 적용이 가능하다. 샘플링 수가 적어도 잘 동작하고 큰 정상 군집을 끝까지 분리할 필요가 없음으로 큰 집합에서도 빠른 성능을 내며, 메모리 요구양이 적다.

3.3 K-means

지정된 군집화의 개수만큼 군집화를 행하여, 정확도를 높이기 위해 군집된 분포간의 거리까지 확인하는 모델이다. I-Forest를 보완하여 보다 정밀한 예측이 가능하다.

[표 7]과 같이 시나리오생성, 공격 수행, 검증, 분석 절차로 모의 공격을 수행하였다.

표 7. 모의 공격 수행 절차
Table 7. Procedure for Performing Simulated Hacking

순서	단계	내용
1	시나리오 작성	• 이상행위 탐지모델 별 모의공격 시나리오 작성
2	모의 공격 수행	• 작성된 시나리오 기반의 모의공격 수행
3	이벤트 발생 여부 검증	• AI 시스템에서 해당 이상행위 예측/탐지 여부 검증
4	결과 분석	• 모의 공격 결과 비교 분석 - WAF, IPS, SIEM 탐지 결과 - AI 시스템 탐지 결과

Ⅳ. 실험 결과

홈페이지(A)를 대상으로 한 1차 PoC와 홈페이지(B)를 대상으로 한 2차 PoC 결과는 [표 8], [표 9]와 같다.

1, 2차 PoC 결과, 기존 보안시스템에서 탐지하지 못했지만 AI 시스템에서 탐지한 결과를 기존 보안시스템에 정책으로 반영한 후, 2차례 모의 공격을 진행

표 8. 1차 PoC 결과
Table 8. Primary PoC Result

탐지 모델	설명	AI 탐지	기존 탐지
관리자페이지 접근 시도	알려진 Black List IP에서 관리자 페이지 접근	O	X
대량의 데이터 유출	html 파일로 대량 접속, 대용량 데이터 전송 확인	O	X
지속적인 파일 전송	알려진 Black List IP에서 비정상 접속 시도	O	X
비정상적 접근 시도	외부 IP에서 특정 URL에 접근 시도	O	X
비정상 User-Agent 사용 탐지	알려진 Black List IP에서 비정상 User-agent, Http Version의 요청 시도	O	X

표 9. 2차 PoC 결과
Table 9. Secondary PoC Result

탐지 모델	설명	AI 탐지	기존 탐지
관리자페이지 접근 시도	관리자 페이지로 접속 시도 (Scanning)	O	O
이상 사용자 접근 시도	특정 페이지 접속 시도 (Scanning)	O	O
	특정 경로의 비정상 페이지 접속 시도	O	X
지속적인 파일 다운로드	매일 일정시간 동안 특정 페이지에 정기적 접속 시도 (Web Crawling)	O	O
비정상적 접근 시도	특정OS에서 사용하는 명령어를 활용하여 특정 페이지 접속 시도 (Web Crawling)	O	X
자동화된 툴을 이용한 접근 시도	동일한 페이지에 접속 시도, 오류 다수 반복	O	X

하였다. 모의 공격 진행 중 서비스가 중단될 수 있는 위험성을 고려하여, 상대적으로 영향이 작은 홈페이지(B)를 대상으로 하였다.

1차 모의 공격은 임계치, 행위(Behavior), 시그니처 기반의 5개 시나리오를 가지고 실시하였다. 그 결과 [표 10]에서와 같이 임계치, 시그니처 기반 공격은 기존 정보보안시스템에서 탐지 되었으나, 행위 기반 공격은 탐지 되지 않는 것으로 확인되었다. 다만, '봇넷 악용' 시나리오의 경우, 공격이 IPS에서 차단됨에 따라, AI 시스템으로 탐지가 가능한지 확인되지 않았다.

표 10. 1차 모의 공격 결과
Table 10. First Simulated Hacking Result

시나리오 명	공격 결과	IPS	WAF	SIEM	AI
회원 가입 계정 목록 추출	1만개의 계정 입력 결과 1,000개의 계정 사용 확인	O	O	O	O
로그인 페이지 사전 대입	계정 별 5회 이상 패스워드 입력 불가	O	-	O	O
웹쉘 접근 시도	동일 파일에 지속적 접근 시도에 대한 탐지, 차단 안됨	-	-	-	O
데이터 유출 공격 시도	공격 패턴 탐지로 IP 차단(시스템 명령 실행)	O	-	O	O
봇넷 악용*	BOTNET User-agent 입력 차단됨	O	O	O	-

표 11. 2차 모의 공격 결과
Table 11. Secondary Simulated Hacking Result

유형	공격명	공격 결과	IPS/WAF	SIEM	AI
행위+임계치	웹 파라미터 변조를 통한 권한 우회	• 1,000회의 패스워드 무작위 대입, 타인의 패스워드 변경 성공	X	X	O
행위+임계치	사용자 정보 수집	• 1,000명의 사용자에 대한 정보 수집 완료	X	X	O
시그니처	XSS(Cross Site Script)	• XSS 스크립트 입력/실행 성공	X	X	O

1차 모의 공격 결과에 따라 임계치 이하, 새로운 공격 패턴, 행위 기반 공격에 대한 추가 확인을 위해 [표 11]과 같이 3개의 추가 시나리오로 2차 모의 공격을 수행하였고, 그 결과, 모든 공격에 대해 IPS와 WAF, SIEM에서 탐지하거나 차단하지 못하였으나 AI 시스템에서는 이상행위를 식별하는데 성공한 것으로 확인되었다.

이 실험 결과는 공격자가 보안 취약점을 악용하여 임계치 이하의 공격, 행위기반 공격, 우회 패턴을 이용한 공격을 교묘하게 수행하였을 시, 정보보안시스템을 우회하여 웹 어플리케이션에 피해를 입힐 수 있으며, 심지어 침해 사실을 인지하지 못할 수 있다는 것을 의미한다. AI 시스템으로 보안 수준을 강화하는 것이 하나의 방법일 수는 있겠지만, 신규 시스템을 도입하고 최적화하는 것은 그리 간단하지 않다. 따라서, 침해사고 가능성을 줄이기 위해서는 신규 IT 자산 도입 시 보안성 검토를 실시하고, 운영 중인 주요 IT 자산에 대해서는 보안 점검을 주기적으로 수행하는 등 보안 관리를 통해 공격의 통로가 되는 보안 취약점을 최소화시키는 노력이 필요하다.

V. 결 론

본 연구는 갈수록 지능화되는 사이버 위협과 공격에 보다 능동적으로 대응하기 위해 AI 기술을 적용한 탐지 사례를 제시하였다.

국내 XYZ社의 홈페이지를 대상으로 PoC와 모의

공격을 수행하였고, AI 시스템을 연동하여 웹로그를 분석한 결과, 기존 정보보안시스템에서 탐지되지 않았던 몇 가지 공격 유형 AI 시스템에서는 이상행위로 탐지하는 것을 확인할 수 있었다. 이를 통해 시그니처 기반의 탐지정책 위주로 운영되는 보안관제 체계에서는 인공지능 기술을 실제로 자주 발생하고 있는 신규 또는 우회 공격에 대한 미탐지 가능성을 낮추고 예측을 통한 탐지율을 높이는데 활용할 수 있을 것이다.

또한, AI 솔루션에서 생성하는 이벤트의 출발지 IP에 대하여 방화벽, 웹로그 등에서 추가 분석을 통해 잠재적 위험이 있거나 탐지하지 못한 공격을 선제적으로 식별 및 대응하거나, AI에서 식별된 공격을 토대로 SIEM의 경보를 정교화하거나, IPS 탐지 정책을 수정(정책 최적화)하는 용도로도 활용하는 등 다양하고 지능화된 사이버 공격을 AI 시스템을 활용하여 탐지함으로써, IT 조직과 인프라에 대한 보안 수준이 한층 강화될 것으로 기대한다.

이번 실험은 제한된 환경에서 웹 공격의 탐지에 대해서만 진행되었기 때문에 다양한 방법으로 다른 유형의 사이버 공격을 탐지하는 방안과 더불어 정확성(미 탐지와 오 탐지)에 대한 현황 및 개선방안에 대한 후속 연구가 필요할 것으로 보인다.

References

[1] H. Jang, “A design and implementation of deep learning-based intrusion detection system for web applications,” Hoseo Univ., Feb. 2019.

[2] C. Baek, S. Lim, and J. Choe, “A study on major characteristic analysis and quality evaluation attributes of artificial intelligence service,” *J. Korean Soc. Qual. Manag.*, vol. 47, no. 4, pp. 837-846, Dec. 2019.

[3] Y. Cho, “Understanding big data and its main issues,” *J. Korean Assoc. Regional Info. Soc.*, vol. 16, no. 3, pp. 43-65, Sep. 2013.

[4] D. Choi and Y. Kim, “Big data and enterprise security 2.0,” *Commun. Korean Inst. Info. Sci. and Eng.*, vol. 30, no. 6, pp. 65-72, Jun. 2012.

[5] G. Allen and T. Chan, *Artificial Intelligence and National Security*, Harvard Kennedy School Belfer Center for Science and International Affairs, Jul. 2017.

[6] J. Kim, “Relation between artificial intelligence and information security,” *Commun. Korean Inst. Info. Sci. and Eng.*, vol. 36, no. 2, pp. 14-17, Feb. 2018.

[7] C. Lee, “Next-generation security control based on artificial intelligence in the 4th industry,” *Mag. IEEE*, vol. 46, no. 3, pp. 25-30, Mar. 2019.

[8] H. Jang, “A study on integrated security management method of heterogeneous security systems,” Konkuk Univ., Feb. 2016.

[9] B. Kang, “Research about quality analysis of web fire wall system,” Hoseo Univ., Aug. 2010.

[10] S. Bhatt, P. K. Manadhata, and L. Zomlot, “The operational role of security information and event management systems,” *IEEE Secur. & Privacy*, vol. 12, no. 5, pp. 35-41, Jan. 2014.

[11] B. Cha, M. Choi, E. Kang, S. Park, and J. Kim, “Trends of SOC & SIEM technology for cybersecurity,” *Smart Media J.*, vol. 6, no. 4, pp. 41-49, 2017.

[12] Global Agenda Council on the Future of Software & Society, *Deep Shift: Technology Tipping Points and Societal Impact*, World Economic Forum, Sep. 2015.

[13] H. Yang, B. Choi, J. Lee, H. Jang, S. Baek, and D. Kim, *A Prospective Analysis of Artificial Intelligence(AI) Technology and Innovation Policies : Focused on Improving Korea’s National AI R&D Policy*, Science and Technology Policy Institute, Dec. 2019.

[14] S. Oh, W. Go, M. Kim, J. Lee, H. Kim, and S. Park, “Research on IoT threat detection technology through artificial neural network algorithm,” *Rev. KIISC*, vol. 29, no. 6, pp. 59-65, Dec. 2019.

[15] Y. Choi and K. Kim, “Artificial intelligence overview and application examples,” *Industrial Eng. Mag.,* vol. 23, no. 2, pp. 23-29, Jun. 2016.

[16] J. Nam, “Machine learning algorithms and discrimination,” Korea Univ., Feb. 2019.

[17] I. Shin, J. Song, J. Choi, and T. Kwon, “A practical feature extraction for improving

accuracy and speed of IDS alerts classification models based on machine learning," *Rev. KIISC*, vol. 28, no. 2, pp. 385-395, Apr. 2018.

류 호 균 (Ho-Gun Rou)

2005년 : 고려대학교 경영학사
2008년 : 고려대학교 경영학 석사
2009년~2011년 : 시큐베이스㈜
2011년~현재 : ㈜이글루시큐리티
2011년~현재 : 숭실대학교 IT정책경영학과 박사과정

<관심분야> 인공지능, 빅데이터, 정보보안, IT정책
[ORCID:0000-0002-8911-4615]

김 광 용 (Gwang-Yong Kim)

1984년 : 고려대학교 공학사
1991년 : 미국 조지아 주립대학 보험수리학 석사
1995년 : 미국 조지아 주립대학 박사
1999년~현재 : 숭실대학교 경영학부 교수

<관심분야> 데이터사이언스, 디지털트랜스포메이션, 인공지능, 빅데이터, 블록체인, 클라우드, 핀테크, 전자정부, IOT, 비즈니스 모델링(디자인싱킹, TRIZ, 캔버스모델 등) 등
[ORCID:0000-0002-6921-1071]