공학석사 학위논문

지도학습 알고리즘을 이용한 침입방지시스템의 페이로드 로그 보안관제 성능 개선

2021년 8월

부경대학교 대학원

정보보호학과

김 홍 경

공학석사 학위논문

지도학습 알고리즘을 이용한 침입방지시스템의 페이로드 로그 보안관제 성능 개선

지도교수 이 경 현

이 논문을 공학석사 학위논문으로 제출함.

2021년 8월

부경대학교 대학원

정보보호학과

김 홍 경

김홍경의 공학석사 학위논문을 인준함.

2021년 8월 27일

위 원 장 공학박사 김 창 수 (인)

위 원 이학박사 신 상 욱 (인)

위 원 이학박사 이 경 현 (인)

목 차

표 목 차

그 림 목 차

지도학습 알고리즘을 이용한 침입방지시스템의 페이로드 로그 보안관제 성능 개선

김 홍 경

부경대학교 일반대학원 정보보호학과

요 약

사이버 공격은 날로 지능화되고 대형화되는 가운데 현재의 비즈니스 환경은 정보통신과 정보기술시스템이 연결되어 있다. 네트워크 확장과 네트워크 속도 증가를 통한 생활의 편리함 속에는 사이버 공격이라는 위협이 항상 도사리고 있다. 더욱이 현재는 5G 네트워크의 상용화로 IoT 기기의 활성화는 개인정보 유출과 DDoS 공격 및 각종 다양한 사이버 공격에 노출되고 있고 그 규모 또한 확대되고 있다.

이러한 사이버 공격에 적절하게 대응하기 위하여 기업과 공공기관 등에서 많은 전담 사이버 공격을 방어하고 대응하는 담당자를 채용하고 적절한 업무를 수행하고 있지만. 현재의 보안관제의 능력은 대응에는 한계가 있다. 현재의 보안관제시스템은 로그, 이벤트, 네트워크 패킷 연동 기반에서 동작한다. 이러한 탐지 기준에 벗어나거나 예상치 못한 새로운 공격기법이 시도된다면 해당 이벤트를 탐지하지 못하거나 공격을 받은 것을 감지하지 못하는 경우도 많다.

현재 4차 산업혁명 시대에 인공지능, 5G 네트워크, 가상화 기술, 블록체인 등이 다양한 신기술들이 등장하고 있는데 사이버 공격도 이러한 신기술을 활용한 기법이 조금씩 발생하고 있다. 인공지능 기술을 활용한 악성코드 프로그램 제작 및 배포가 하나의 예일 수 있다. 만일 인공지능을 통해 악성코드 프로그램을 제작하는 기술이 발전한다면 앞으로 새로운 악성코드의 출현 · 빈도 · 공격 확산 속도 등은 크게 증가할 것으로 예상된다. 이러한 상황에서 머신러닝, 인공지능, 딥러닝[1] 기술들을 활용한 다양한 보안솔루션들도 등장하고 있다. 그러나 아직 걸어가야 할 길이

멀다. 머신러닝 기술을 이용한 보안솔루션을 도입하기 어려운 부분이 탐지 이벤트를 오탐 하거나 과하게 탐지하는 경우라 할 수 있다. 또한 이러한 보안 이벤트를 분석하고 대응할 수 있는 보안전문가도 부족한 것이 현실이다[2].

이를 극복하기 위하여 기존 보안관제시스템의 기술을 활용하되 머신러닝 기술을 활용한 보안관제시스템을 구축함으로써 기존의 보안관제시스템의 한계를 극복하고 진화하는 사이버 공격에 효과적인 대응이 가능한 보안관제체계를 만들 필요가 있다.

본 연구에서 기존 침입방지시스템에서 보안로그를 분석하되 더욱 상세한 원본 데이터를 체계적으로 분석할 수 있는 모델을 만들고 지도 학습(Supervised Learning) 알고리즘은 분류(Classification)작업을 위한 CNN 알고리즘과 회귀(Regression) 작업을 위한 랜덤 포레스트(Random Forest) 알고리즘을 활용하여 보안관제시스템의 탐지능력을 향상 시키는 결과를 도출하였다. 지도 학습에서는 보안 이벤트의 정확한 예측을 위하여 필드를 정리하고 각 필드마다 레이블(Label)을 설정하는 것이 중요하다. 보안관제 담당자는 보안 이벤트의 필드의 특성을 분석하고 피처(Feature)를 추출하여 머신러닝이 인식할 수 있도록 데이터를 변환해야 한다. 초기에 이 과정은 수고가 많이 따르지만 보안관제의 업무의 효율성과 머신러닝의 인공지능 기술의 예측률을 높이는데 중요한 부분이라 할 수 있다. 각 기업이나 공공기관 마다 운영하는 장비와 솔루션이 다르기 때문에 해당 조직에 적합한 보안관제 모델링을 결정하는 것은 보안담당자에게 필요한 연구라 할 있다.

본 연구를 통해 머신러닝 기술을 이용하여 보안장비가 탐지하지 못한 공격을 탐지하고 과탐·오탐을 최소화 하며 인공지능을 통해서 탐지된 예측률을 높임으로써 보안관제 담당자의 업무를 효율화 하고 보안관제시스템의 성능을 향상시키는데 기여할 수 있기를 기대한다.

---

1) 완전 새로운 기술이나 이론이 아닌 인공신경망의 한계였던 심층 구조 학습을 해결하기 위하여 심층 신경망을 이용한 이론(Yann LeCun · Yoshua Bengio · Geoffrey Hinton, 2015)[1]

Performance Enhancement for IPS payload log Monitoring and Control System using a supervised learning algorithm

Hong-Kyung Kim

Department of Information Security,
The Graduate School, Pukyong National University

Abstract

Cyber attacks are becoming more intelligent and larger, and information communication and information technology systems are connected in the current business environment. The threat of cyber attacks always lurks in the convenience of life through network expansion and network speed increase. Moreover, with the commercialization of 5G networks, the activation of IoT devices is now exposed to personal information leakage, DDoS attacks, and various cyber attacks, and the scale is also expanding.

In order to respond appropriately to such cyber attacks, companies and public institutions have hired a person in charge of defending and responding to many cyber attacks and perform appropriate tasks. Current security control capabilities have limitations in response. Current security control systems operate based on log, event and network packet inter-working. If these detection criteria are deviated or unexpected new attack techniques are attempted, the event may not be detected or the attack may not be detected in many cases.

In the era of the 4th industrial revolution, various new technologies such as artificial intelligence, 5G networks, virtualization technologies, and block

chains are emerging, and cyber attacks are also gradually occurring techniques using these new technologies. One example is the creation and distribution of malicious code programs using artificial intelligence technology. If the technology for creating malicious code programs through artificial intelligence develops, it is expected that the appearance, frequency, and attack rate of new malicious codes will increase significantly in the future. In this situation, various security solutions using machine learning, artificial intelligence, and deep learning technologies are also emerging. However, there is still a long way to walk. The most difficult part of introducing a security solution using machine learning technology is the case of false detection or excessive detection of detection events. In addition, there is a lack of security experts who can analyze and respond to these security events.

To overcome this, a security control system that overcomes the limitations of the existing security control system and proactively responds to the evolving cyber attack technique by building a security control system using machine learning technology while utilizing the technology of the existing security control system. I need to make it.

In this study, the security log is analyzed in the existing intrusion prevention system, but a model that can systematically analyze more detailed original data is created, and the supervised learning algorithm is a technique that utilizes CNN and Random Forest to detect attacks in the security control system. Results were derived to improve the ability and performance. In supervised learning, it is important to organize fields and label each field for accurate prediction of security events. The security control officer must analyze the characteristics of the field of the security event, extract the feature, and transform the data so that machine learning can recognize it. Initially, this process takes a lot of effort, but it can be said to be an important part in improving the efficiency of security control work and the predictability of artificial intelligence technology of machine

learning. Since the equipment and solutions operated by each company or public institution are different, it is a necessary study for the security officer to determine the appropriate security control modeling for the organization.

Through this study, the security control personnel's work is streamlined and the performance of the security control system is improved by detecting attacks that security equipment cannot detect using machine learning technology, minimizing over-detection and false detection, and increasing the detection rate through artificial intelligence. We look forward to contributing to improvement.

Ⅰ. 서 론

1. 연구의 목적

정보통신기술의 발전은 우리 삶의 편의성을 향상시키고 국가적 발전을 위한 다양한 순기능을 가져왔지만 이에 비례하여 사이버 공격, 정보 유출, 프라이버시 침해 등 정보화 역기능의 발생 가능성 또한 증가하였다[3].

또한 한국인터넷진흥원의 2020년 7대 사이버 공격 전망 발표 자료에 따르면 공공기관·기업으로 확대되는 렌섬웨어 공격, 문자·이메일 안으로 숨어드는 악성코드, 진화하는 지능형 표적공격, 융합 서비스 대상 보안 위협 등을 주요 사이버 공격이라고 전망했다[4].

이처럼 다양한 해킹 공격에 사용되는 지능형지속보안위협 APT(Advanced Persistent Threat)공격 기법이 기업 및 공공기관을 대상으로 확대되고 있기 때문에 최근 APT 기법의 사이버 공격에 적절한 대응방법의 중요성이 높아지고 있다.

그러나 기존의 단일 정보보호시스템인 침입차단시스템(Firewall), 침입방지시스템(Intrusion Prevention System), 웹방화벽(Web Application Firewall)으로는 APT 기법의 사이버 공격 대응에 미흡한 부분이 많다.

기존 보안관제시스템은 ESM(Enterprise Security Management)으로 이기종의 정보보호시스템의 로그를 수집하고 탐지 현황을 모니터링하고 정보보호시스템들의 실제 로그들을 통합하여 관리해주는 시스템으로 발전하였다.

현재 SIEM(Security Information & Event Management)은 보안장비의 로그와 이벤트의 데이터 필드들을 각각 정의하고 상관관계를 분석하기 때문에 빅 데이터를 활용한 시나리오 기반의 대응 프로세스를 수행할 수 있게 되어 ESM 보다 정교한 분석을 할 수 있도록 발전하였다[5].

그러나 인공지능 AI(Artificial Intelligence) 기술을 기반으로 고도화된 새로운 공격 기법이 빠르게 늘고 있다. 이러한 상황에 침해 대응 분석과 새로운 후속 조치도 매우 중요하지만 4차 산업 혁명 시대의 인공지능 기반으로 공격 데이터를 수집·가공 및 정제하여 공격 데이터의 특징을 실시간으로 추출하여 공격 여부를 판별할

수 있는 머신러닝 기술을 통하여 보안관제시스템의 성능개선을 기대한다.

2. 연구의 내용 및 범위

본 연구의 목적인 사이버 공격을 방어하는 보안관제시스템의 성능개선을 위해 기존 사이버 공격 탐지 기술의 한계를 확인하고, 보안관제 분야에서 활용이 가능한 머신러닝 기술을 분석하고, 머신러닝 기술에 적용 가능한 알고리즘들을 정리한다.

본 연구에서 지도학습(Supervised Learning) 방법을 사용한다. 기존 보안관제시스템 SIEM에서 침입방지시스템의 보안이벤트에서 페이로드(Payload[2)]) 로그 데이를 수집하고 데이터에서 각각의 필드들을 분류하고 각 특성에 따라 레이블 처리를 한다. 이렇게 만들어진 데이터는 머신러닝이 쉽게 인식할 수 있도록 데이터를 수치화 하는 등 전처리 과정을 거치게 된다. 데이터는 특성이나 데이터 간의 관계성을 고려하여 피처와 피처의 그룹을 생성한다. 이렇게 만들어진 데이터는 학습데이터가 되어 머신러닝에 업로드 한다. 기존 보안관제시스템의 경보(Alert)에 대한 과·오탐을 줄이기 위해 지도 학습(Supervised Learning) 방법을 사용하고, 피처의 특성을 머신러닝의 적절한 인공지능 알고리즘을 선정하여 교차학습을 시킨다.

연구의 검증은 실제 보안관제 업무에서 수집되는 빅 데이터 중심의 보안 이벤트를 기반으로 머신러닝에 알맞은 데이터로 변환하여, 특성을 추출하고, 모델링하여 이를 반복 학습하여 머신러닝 모델의 Accuracy(정확도)와 Precision(정밀도) 그리고 Recall(재현율)을 통해 예측률을 검증한다. 또한 예측률이 높은 탐지정책을 기존 보안관제시스템의 탐지조건에 추가 반영하여 성능 개선을 확인한다.

---

2) 헤더 정보를 제외한 목적지에 전달되는 데이터, 사이버보안에서는 악성코드, 공격패턴, 스팸 등 분석에 사용되는 실제 데이터를 의미한다.

3. 연구의 구성

본 연구의 구성은 다음과 같이 총 5장으로 구성된다.

Ⅰ. 서론에서는 연구의 목적, 내용 및 범위에 대하여 기술하였다.

Ⅱ. 이론적 배경에서는 선행연구를 바탕으로 사이버 공격에 대한 연구, 통합보안관제시스템에 대한 연구와 머신러닝 기술에 대한 연구를 기술하였다. 또한 머신러닝 기술 중 지도학습의 알고리즘에서 본 연구에 해당되는 알고리즘을 설명하고 관계에 대해 정리 기술하였다.

Ⅲ. 본 연구의 목적인 보안관제시스템의 성능을 개선하기 위한 보안관제시스템을 구성방법을 제시하고 각 구성에 대한 관계와 기술적인 사항을 정리하였다. 또한 각 구성에서 머신러닝의 탐지 및 예측모델을 구성하기 위한 절차와 방법들을 기술한다.

Ⅳ. 머신러닝 기술을 활용하여 사이버 공격 탐지에 대한 정확성과 오차를 추출하여 예측률을 실험을 통해 검증한다.

Ⅴ. 결론에서는 본 연구의 연구 결과를 요약하고 연구결과의 시사점을 기술하였다. 아울러 연구의 미흡한 점과 향후 연구방향을 제시하였다.

Ⅱ. 이론적 배경

1. 사이버 공격에 대한 연구

1) 사이버 공격의 정의

사이버 공격이라 함은 해킹 · 컴퓨터바이러스 · 논리폭탄 · 메일폭탄 · 서비스방해 등 전자적 수단에 의하여 국가정보통신망을 불법침입 · 교란 · 마비 · 파괴하거나 정보를 절취 · 훼손하는 일체의 공격행위를 말한다[6].

또는 사이버 공격은 크게 2가지 유형으로 분류될 수 있는데 하나는 대상 컴퓨터를 비활성화 시키거나 네트워크 통신을 오프라인으로 만드는 공격이고 다른 하나는 공격 대상 컴퓨터의 데이터에 비정상적인 방법으로 접근하여 관리자 권한을 취득하는 것이 목표인 공격 유형이다[7].

2) 사이버 공격의 유형

사이버 공격의 크게 3가지 유형으로 나눈다면 사이버 범죄, 사이버 테러, 사이버 전쟁으로 설명할 수 있다. 사이버 공격의 유형은 <표 Ⅱ-1>로 간단하게 정리를 하였다. 이에 의하면 주요 사이버 공격 유형과 최근 사이버위협 동향에 대해서 유형별로 확인할 수 있다. 사이버 공격의 종류와 특징을 유형별로 분류하여 머신러닝 기능을 활용한 보안관제시스템에 적용할 수 있는 탐지모델을 구성하기 위하여 유형별로 구분하고자 한다.

<표 Ⅱ-1> 사이버 공격 유형[8]

분류	공격유형	설명
주요 사이버 공격	사회공학적 해킹	시스템이 아닌 사람의 취약점을 타겟으로 공략하여 신뢰를 바탕으로 대상자를 속이고 원하는 정보를 얻는 공격기법
	악성코드 유포	스팸메일, 유틸리티·동영상 파일 위장, 자동 업데이트, 특정 IP주소 대역 선별 등을 통하여 악성코드 유포
	랜섬웨어 유포	사회공학적 기법, 워터링홀, 인터넷 배너광고, 게임용 앱 등 랜섬웨어 유포 방식 다변화
	소셜미디어 해킹	피싱메일을 통해 이메일 계정·비밀번호를 탈취하여 SNS에 무단 접속한 후 악성코드 유포
최근 사이버위협 동향	IoT 대상 사이버위협 증가	ICT 발전에 따라 IoT 결합 CCTV, 스피커 등의 사용이 증가하면서 해킹 위협에 노출
	타깃형 랜섬웨어 확산	국가·공공기관, 기업을 타깃으로 하는 APT 공격으로 랜섬웨어 공격방식 변화
	모바일 공격 증가	모바일 메신저, 사진뷰어 등을 사칭한 악성 앱을 통하여 APT 공격
	코드서명인증서 악용 공격	보안이 취약한 소프트웨어 개발사의 코드서명 인증서를 탈취한 후 악성코드 유포 등 해킹에 악용
	공급망 공격 확산	모바일 소프트웨어, 모바일 기기 제조사 대상 코드서명 인증서 탈취, 스피어피싱 등과 결합한 APT 공격
	APT 공격 정교화	스피어피싱과 APT 공격이 결합되면서 고도화· 지능화
	폼재킹 증가	특정 프로그램으로 제작된 결제 웹페이지를 사용하는 쇼핑몰 웹페이지를 공격하여 카드정보 탈취

3) 사이버 공격의 피해 및 대응 사례(2017~2018)

(1) 웹호스팅 업체 “인터넷나야나” 랜섬웨어 감염사태

2016년 워너크라이 랜섬웨어 대란 이후 6월 중순 웹호스팅 업체 “인터넷나야나”에서 랜섬웨어 감염사태가 이슈가 되었는데, 당시 워너크라이 랜섬웨어가 한참

이슈화 되고 있을 시점에서 Erebus[3]라는 생소한 랜섬웨어에 감염되어 호스팅 중이던 웹사이트 데이터를 보관 중이던 DB서버의 데이터 대부분이 변조되어 정상적인 서비스를 할 수 없는 상태가 되어 회사 측에서 상황 파악 이후 사측 사이트를 이용하여 공지를 통해 피해상황을 알렸지만 이를 알게 된 고객들의 반발 및 불만 폭주로 인하여 업체 대표가 직접 해커와이 협상을 통해 개인 자산 및 회사 지분을 매각하여 사태를 해결하였는데, 사태 당시 해커가 요구한 금액은 한화 약 50억 상당의 비트코인을 요구하였지만 대표와 해커간의 꾸준한 협상을 통해 13억 원어치의 비트코인을 지불하여 감염된 서버들의 복호화 키를 받아 복호화 작업을 진행하였으며, 당시 미래창조과학부에서 해당 사태에 대한 기술적인 지원을 해주기로 하는 등 사회적으로 큰 이슈가 되었던 사건이다[10].

(2) 비트코인 거래소 해킹 사건

2016년 이후 비트코인을 이용한 재테크 및 금전적 이득을 위한 투자 사업이 활발해지면서 비트코인 거래소 또한 동시다발적으로 많이 생겨났으며 범죄의 수익을 비트코인으로 주고받는 경우 또한 지속적으로 발생하였다. 비트코인 계좌 개설 및 거래를 위한 거래소가 이슈화 되었는데 법적 · 제도적 규제가 마련되지 않은 상태에서 무분별한 거래가 이루어지다보니 많은 사건사고가 발생하였는데 거래소 자체를 공격대상으로 삼아 비트코인을 절취해가는 사례가 발생하기 시작했다.

아래 <그림 Ⅱ-1> 가상 통화 거래소 해킹 사고 관련 보도 기사 사진은 보는 바와 같이 최근 많은 경로를 통하여 거래소 공격이 실제로 이루어지고 피해 또한 상당히 큰 것을 볼 수 있는데 국내외를 가리지 않고 지속적으로 발생 중인 사실 또한 확인할 수 있었다.

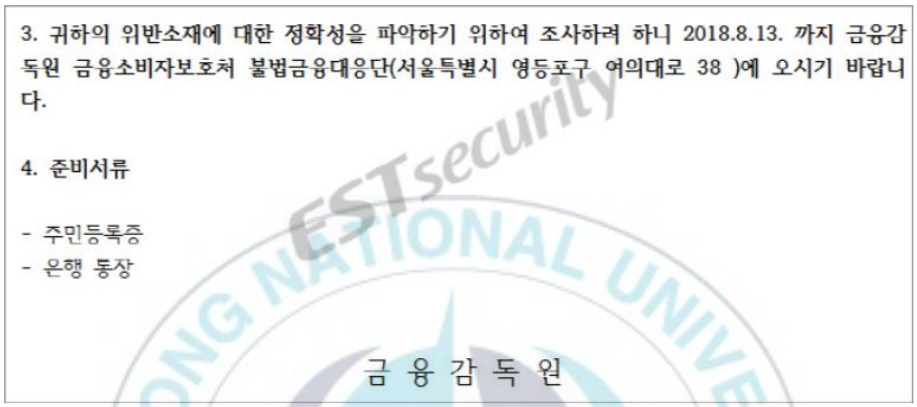
---

3) 2016년 9월에 발견된 랜섬웨어로 '사용자 계정제어(UAC) 보안 기능'을 우회하고, 스스로 '익명(Tor) 브라우저 클라이언트'를 다운받아 추적을 어렵게 하는 지능적인 랜섬위어임[9]

금감원 사칭해 암호화폐 거래소 해킹 시도.. 이메일 피싱 등장

이스트시큐리티 대응센터 발견..주의 당부

등록 2018-08-08 오전 11:19:57
수정 2018-08-08 오전 11:19:57


3. 귀하의 위반소재에 대한 정확성을 파악하기 위하여 조사하려 하니 2018.8.13. 까지 금융감독원 금융소비자보호처 불법금융대응단(서울특별시 영등포구 여의대로 38)에 오시기 바랍니다.

4. 준비서류

- 주민등록증
- 은행 통장

금 융 감 독 원

금융감독원 사칭 이메일 첨부파일 예시. 이스트시큐리티 제공

<그림 Ⅱ-1> 가상 통화 거래소 해킹 사고 관련 보도[11]

(3) 크립토재킹

가상화폐가 이슈화 되며 채굴 형 악성코드 또한 기하급수적으로 증가하여 피해 사례가 언론에서 자주 보도가 되었는데 이중에서 크립토재킹[4]이라는 신종 공격 방식이 이슈화가 되었는데 공격자가 사용자 몰래 컴퓨터나 서버시스템에 채굴용 악성코드를 설치하여 해당 장비의 자원을 무단으로 사용해서 수익을 창출했고 사용자들의 방문 빈도가 높으면서 상대적으로 보안이 허술한 홈페이지에 악성스크립트 등 다양한 경로에 해당 악성코드를 삽입하여 접근하는 시스템 및 그 이하 다른 시스템의 자원도 감염시키는 형태로 전파되었다.

4) 암호화폐(Crytocurrency)와 탈취(hijacking)의 합성어로 해커들이 사용자이 PC자원을 이용하여 가상통화를 채굴하는 공격 기법이다

아래 <그림 Ⅱ-2> 가상 통화 거래소 해킹 사고 관련 보도 기사는 웹사이트 이용자가 크립토재킹 악성 스크립트에 감염된 웹 페이지에 접속하게 되면, 이용자 컴퓨터는 가상화폐를 채굴하여 공격자의 지갑 주소로 가상화폐를 전송하는 공격을 보여주고 있다.

URL 주소	사이트 종류	유포 종류
http://www.sx10xx.kr	쇼핑몰	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://www.unaxxxx.net	쇼핑몰	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://www.nxxj.co.kr	회사 홈페이지	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://www.sixxxgxxx.kr	쇼핑몰	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://3xxhoxxx.com	쇼핑몰	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://yelxxloxxx.co.kr	회사 홈페이지	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://www.lxxapxx.co.kr	회사 홈페이지	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://www.kxxxib.com	쇼핑몰	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://www.kxxxxshinxxxx.co.kr	회사 홈페이지	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://mhosxxxl.co.kr	병원 홈페이지	채굴형 악성 스크립트, 사용자 정보 수집 코드
http://m.unaxxxx.net	모바일 쇼핑몰	채굴형 악성 스크립트, 사용자 정보 수집 코드

<그림 Ⅱ-2> 크립토재킹 악성 스크립트가 입력된 웹 페이지[12]

2. 통합보안관제시스템에 대한 연구

1) ESM(Enterprise Security Management) 개념

ESM은 정보보호시스템들의 로그를 통합 관리하고 유사한 보안 정책을 추출하고 적용하고 로그의 상관관계를 연관 분석하여 각 정보보호시스템들의 상호 운용성, 관리성 및 보안성을 최대화하여 위험요소를 최소화하는 관리 솔루션이라고 이해할 수 있다[13].

기업들에 구축되어 있는 대부분의 ESM시스템들은 이기종 네트워크 환경에서 발생하는 로그들의 오류를 최소화 하도록 패턴 매칭 하여 중앙집중식으로 관리하도록 개발되어 보안업무 담당자가 중요한 작업에 집중할 수 있도록 실시간으로 공격을 탐지하는 것에 목적을 두고 있다[14].

그러나 단조로운 패턴 매칭의 탐지 방식으로는 APT 공격과 같은 지능화 공격에 대한 탐지의 어려움이 나타남으로 고도화된 통합 보안관제 시스템의 필요성이 대두되고 있다[15].

2) SIEM(Security Information & Event Management) 개념

ESM의 한계를 개선하기 위해 나온 보안관제 시스템 개념이 SIEM(Security Information & Event Management)이다.

SIEM은 잠재적인 위협과 사이버 공격에 반응하고 보안 정책을 효과적으로 구성하기 위해 데이터를 보안 정보로 변화하는데 사용된다. SIEM은 자동 로그시스템, 내장 보고 및 방화벽 또는 안티바이러스 소프트웨어 등에서 생성된 경고와 같은 스트림 이벤트에서 데이터를 소싱한다. 이렇게 수집된 데이터는 필터링을 거쳐 머신러닝과 통계 방식을 사용하는 시스템에 제공되기도 한다. 이후 비정상 행위를 탐지하고 IT 전담 인력에게 우선순위에 따라 경보를 발생시킨다.

SIEM은 전체 IT 인프라에서 비롯된 모든 보안 데이터를 모으는 센터이다. 이렇게 수집된 데이터로 실시간으로 사건을 관리하며, 지난 문제를 자세하게 탐지하고 감사한다. 또한 데이터 컴플라이언스 요구 조건에 대한 문서를 작성한다.

네트워크의 소프트웨어와 애플리케이션이 제공하는 광대하고 고도로 세분화된 데이터들을 직접 분류하고 상호 연관시키는 것은 사실상 불가능하다. SIEM은 이러한 문제를 해결하기에 적합한 대안으로 간주된다[16].

일반적인 SIEM의 주요 기능은 통합 보안로그 수집 및 보관, 탐지 및 분석을 통한 대응체계 구축, 실시간 모니터링 및 관리로 구분할 수 있는데 <표 Ⅱ-2>로 간단하게 정리를 하였다.

<표 Ⅱ-2> SIEM 주요 기능

기 능	설 명
통합 보안로그 수집 및 보관	• 정보보호시스템 및 운영체계 등 대용량 보안 이벤트 관리 • 실시간 로그 수집 • 무결성 확보를 위한 웜스토리지 내 원본로그 보관
탐지 및 분석을 통한 대응체계 구축	• 동종 및 이 기종 간 주요 이벤트에 대한 연관 분석 • 실시간 보안이벤트 탐지 및 보안경보 • 보안로그의 위험도 분류 및 신규 위협 여부 분석
실시간 모니터링 및 관리	• 대시보드를 통한 실시간 모니터링 관리 • 로그 수집 · 보관 · 탐지 및 분석 등 현황 관리 및 조건 검색 • 위험도 분석을 통한 상관관계 특성 조건에 따른 알람 설정

3) 머신러닝 기반의 진화된 SIEM(Security Information & Event Management)의 개념

보안관제시스템은 보안 로그를 수집하거나 공격 탐지로그를 확인 할 수 있는 단방향성 정보만을 제공하기 때문에 새로운 보안 위험 감지나 분석에 소요되는 시간이 길고 적절한 대응처리에 오랜 시간이 걸린다. 각기 다른 형태의 로그를 분석하고 다양한 이벤트를 형식에 맞게 파싱하여 종합적으로 분석 할 수 있어야 하지만 현재 SIEM에서는 탐지 조건정의 분석에 의하기 때문에 새로운 의사 결정을 지원하는 기능이 부족하다. 로그 데이터는 매우 방대하지만 데이터 포맷은 다양하게 존재하고 기본단계에서의 분석이 필요하므로 전체적인 흐름을 파악하기에는 많은 어려움이 있다. 좀 더 다양한 분석기법을 반영한 머신러닝 기술을 이용하여 의사결정이 유연한 진화된 SIEM의 기술이 필요하다[15].

SIEM 보안관제와 인공지능 기반 보안관제시스템의 관제 업무를 <표 Ⅱ-3>에 간단하게 비교 정리 하였다.

<표 Ⅱ-3> SIEM과 인공지능 기반 SIEM 보안관제 비교[15]

기 능	SIEM	인공지능 기반 SIEM
탐지	시그니처 기반 이벤트 탐지	복합 위험도 산정에 의한 중요도별 이벤트 탐지
분석	보안 장비 로그 추적 분석	위험도 산정 분석, 자산취약점 분석, 데이터 모델링을 이용한 탐지
보안관제 영역	보안장비의 연계관계를 통한 공격과 공격간의 상관분석	인공지능의 산정, 위험도, IT 자산취약점 정보의 인과관계를 통한 상관분석
위험 학습주체	관제 요원이 분석하여 학습	머신러닝 기술을 활용한 지속적인 학습 및 모델링

3. 머신러닝에 대한 연구

1) 머신러닝의 개념

머신러닝은 "명시적으로 프로그램 되지 않고 컴퓨터 스스로 학습하는 능력"으로 정의되며 정보 보안 업계에 큰 의미를 내포하는 개념으로 악성코드, 로그 분석 뿐 아니라 조기 취약점 파악과 수정까지 보안에 큰 도움을 준다.

또한 컴퓨터의 보안을 개선하고 반복적인 업무를 자동화 시키고 정보 유출의 위험성을 낮출 수 있다. 이는 머신러닝 기술을 이용해 탑재한 보안 솔루션이 전통적인 방어 체계와 비교하여 지능형 공격을 더 빠르게 탐지할 것으로 예상하고 있으며 앞으로 머신러닝은 보안에 혁신적인 변화를 일으킬 것으로 간주된다. 그러나 문제는 해커도 이를 알고 있으며 해킹용 인공지능과 머신러닝 도구를 제작한다는 것이다. 그러므로 기존 방식의 보안관제의 방식으로는 머신러닝 기술을 활용한 신개념 사이버 공격 방어에는 한계가 있다[17].

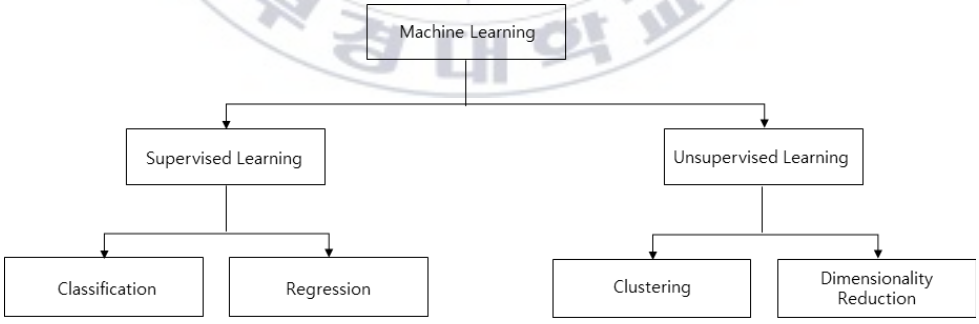
2) 머신러닝 알고리즘

머신러닝은 크게 지도학습(supervised learning)과 비지도학습(unsupervised learning)의 두 가지 주요 유형으로 나눌 수 있다.

지도학습은 데이터에서 추출한 특별한 특징(feature)과 데이터의 특징과 관련 있는 레이블 사이의 특수한 관계를 모델링하는 알고리즘이다. 따라서 모델이 결정되면 그 모델을 이용하여 탐지되지 않았던 새로운 데이터에 레이블 값을 적용할 수 있다. 그리고 이것은 분류(classification)와 회귀(regression)라는 작업으로 나뉘게 된다. 분류에서는 레이블은 이산적인 범주이나, 회귀에서 연속적인 수량이라 말할 수 있다.

비지도학습은 레이블을 참조하지 않는 대신 데이터의 특징 정보를 모델링하는 것이라 설명할 수 있다. 이 모델은 군집화(clustering)와 차원 축소(dimensionality reduction)같은 작업을 포함한다. 군집화 알고리즘은 데이터의 개별 그룹을 식별 하는 반면, 차원 축소 알고리즘은 데이터를 좀 더 간결하게 표현하는 방법을 찾는다[18].

머신러닝의 구성은 <그림 Ⅱ-3>와 같다.



<그림 Ⅱ-3> 머신러닝의 구성

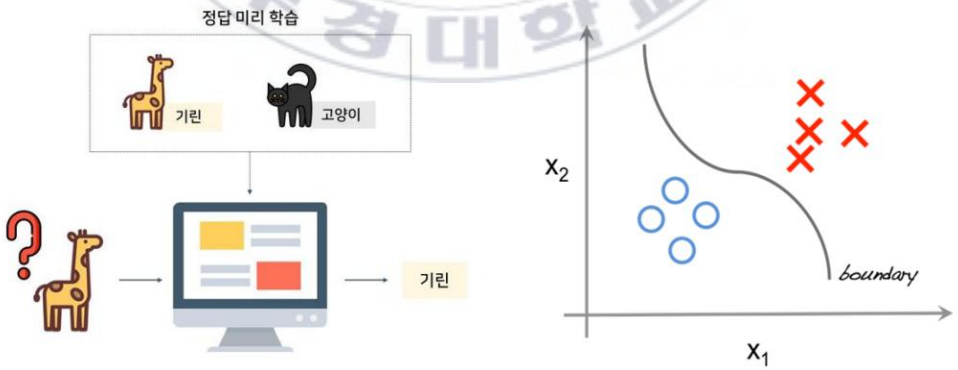
위 <그림 Ⅱ-3> 머신러닝 구성에 대한 자세한 설명을 위해서 머신러닝 학습 방법의 기본유형에 대한 정의는 아래 <표 Ⅱ-4>와 같다.

<표 Ⅱ-4> 머신러닝 학습방법의 기본유형

구분	종류	정의
지도학습	분류	두개 이상의 이산적인 범주로 레이블을 예측 하는 모델
	회귀	연속적인 레이블을 예측 하는 모델
비지도학습	군집화	데이터의 개별 그룹을 탐지하고 식별하는 모델
	차원축소	고차원 데이터의 저차원 구조를 탐지하고 식별하는 모델

(1) 지도학습(Supervised Learning)

지도학습 알고리즘은 알려진 입력 데이터 세트 및 해당 데이터에 대해서는 알려진 출력을 사용하고, 새 데이터에 대한 응답을 위해 합리적인 예측이 생성되도록 모델을 학습한다[19].



<그림 Ⅱ-4> 지도학습 개념도[20]

위의 <그림 Ⅱ-4>은 지도학습의 개념도이다. 각 객체(instance)의 결과(answer)를 설계자가 모두 알고 있으며, 해당 내용을 참조하여 머신러닝을 수행한다. 회귀분석(Regression Analysis)으로 데이터의 함수 관계를 예측하고, 의사결정나무(Decision Tree)로 데이터 속성에 따라서 의사결정을 하는 학습 모델을 만들고 계속해서 반복하여 최종 결정을 도출한다.

(2) 비지도학습(Unsupervised Learning)

비지도학습은 데이터에서 숨겨진 패턴이나 고유 구조체를 찾는다. 이러한 패턴이나 구조체는 분류되지 않은 입력 데이터로 구성된 데이터 세트의 추론에 사용된다. 클러스터링은 가장 일반적인 비지도학습 기법이다. 이 기법은 탐색적 데이터 분석을 통해 데이터에서 숨겨진 패턴이나 그룹을 찾는 데 사용 된다[19].

아래 <그림 Ⅱ-5>은 비지도학습의 구성이다.

<그림 Ⅱ-5> 비지도학습 개념도[20]

각 객체(instance)의 결과(answer)를 설계자가 모두 알고 있지 않아 기계가 직접 내용을 추론한다. 군집화(Clustering)는 비슷한 관측치 끼리 군집을 하고, 차원 축소(Dimensionality Reduction)는 데이터간의 연관 규칙을 찾는다.

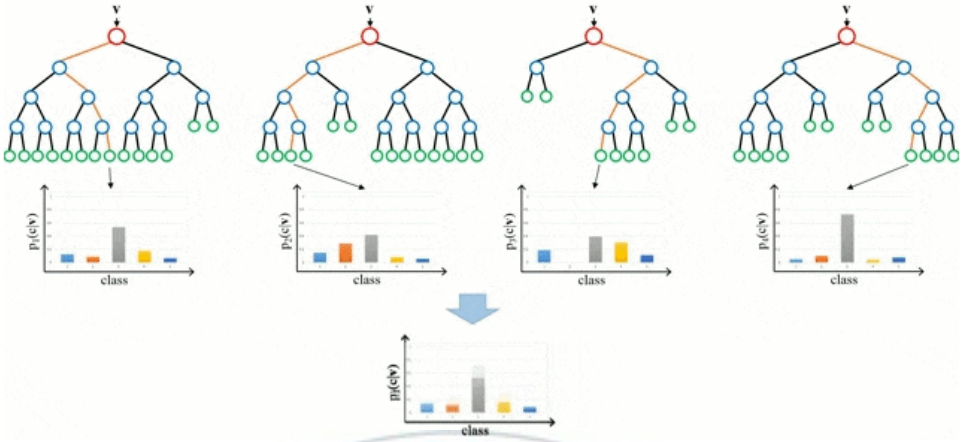
4. 지도학습(Supervised Learning)의 알고리즘 연구

1) 의사결정나무(Decision Tree)

의사결정 문제를 나무에 비교하여 나무의 가지를 가지고 목표와 상황과의 상호 관련성을 나타내어 최종적인 의사결정을 하는 방법으로, 의사결정 규칙을 마치 나무와 같이 구성하여 자료를 여러 개의 소집단으로 분류(classification) 혹은 예측(prediction)하는 분석기술이다. 의사결정나무 알고리즘은 특정 기준(질문)에 따라 데이터를 구분하고 분석하여 데이터들 사이에서 발견할 수 있는 각각의 특성을 조합한 후에 나무모양의 분류 모형을 다시 만들고 이 분류모형을 토대로 새로운 Record를 다시 분류하여 해당 속성 값을 예측한다. 의사결정나무의 나무모양에서 나무를 거꾸로 뒤집었을 때 맨 위쪽을 뿌리 노드(Root Node), Attribute를 분리하는 기준인 내부 노드(Internal Nodes), 노드 사이를 연결해 주는 Link 노드, 맨 마지막 Class를 의미하는 잎(Leaf) 혹은 Terminal 노드들로 구성된다. 의사결정나무 알고리즘을 활용하는 이유는 5 가지로 요약할 수 있다.

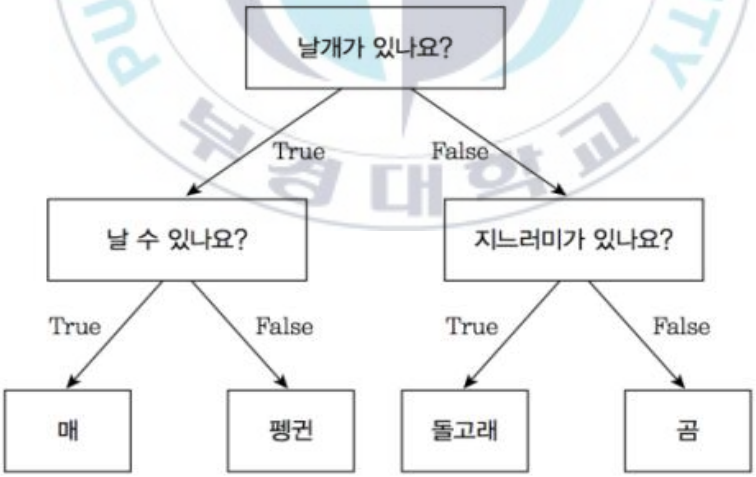
첫째, 분류나 예측 근거를 제공하므로 이해하기 쉽다. 둘째, 불필요하게 데이터를 구성하는 속성들이 많아도 분류에 별다른 영향이 없는 속성들을 자동적으로 제외시키기 때문에 데이터 선정에 용이하다. 셋째, 데이터의 변환단계의 속도를 단축시킨다. 넷째, 속성을 분류할 때 영향을 쉽게 파악할 수 있다. 다섯째, 모형 구축 시간이 짧다[19].

아래 <그림 Ⅱ-6>은 의사결정나무 알고리즘 모식도이다.



<그림 Ⅱ-6> 의사결정나무 알고리즘 모식도[19]

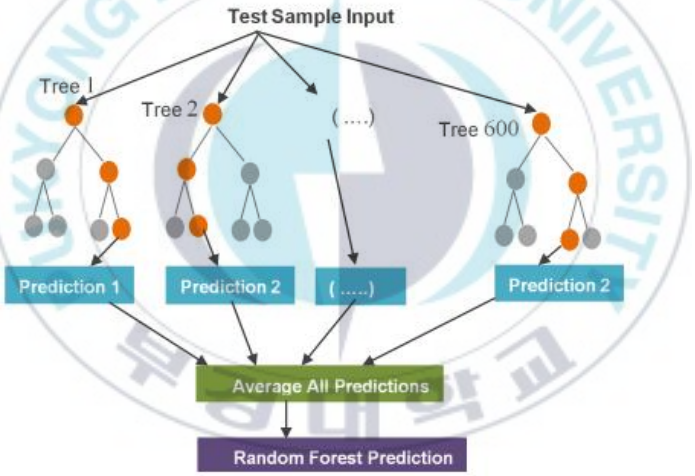
의사결정나무 알고리즘은 특정 기준이나 질문에 따라서 데이터를 계속 구분하고 그 결정에 따라 의사결정나무 모델로 표현되는데 아래 <그림 Ⅱ-7>은 의사결정나무 알고리즘을 설명하는 예이다.



<그림 Ⅱ-7> 의사결정나무 알고리즘 설명

2) 랜덤 포레스트(Random Forest)

랜덤 포레스트(Random Forest) 알고리즘은 의사결정나무의 분류보다 더 정확도를 높이기 위해, <그림 Ⅱ-8>과 같이 복수 개의 나무를 나열하고 각각의 나무에서 나온 예측을 모두 조합하여 합리적인 결론을 내리는 구조이다. 의사결정나무가 특정 데이터에만 잘 작동할 가능성이 크다는 단점을 극복하기 위해 만들어진 알고리즘으로, 같은 데이터에 대하여 의사결정나무를 여러 개 만들어, 그 결과를 종합해 내는 방식이다. 이와 같은 기법을 앙상블[5] 이라고 하는데, 이를 통해 정확도와 안정성을 높일 수 있다[21].



<그림 Ⅱ-8> Random forest decision rules[21]

랜덤 포레스트 알고리즘의 장점으로는, 의사결정나무의 쉽고 직관적인 장점을 그대로 가지고 있고, 앙상블 알고리즘 중에서 속도가 빠른 특성이 있으며 다양한 분야에서 좋은 성능을 나타내는 것이다[22].

그러나 하이퍼 파라미터가 많아 튜닝을 위한 시간이 많이 소요되는 단점도 있다.

5) 머신러닝에서 여러 개의 모델들을 학습시켜, 그 모델들의 예측 결과를 통해 하나의 모델에서 나온 값보다 더 나은 값을 예측하는 방법을 말함

3) 앙상블 학습(Ensemble Learning)

앙상블 학습은 조화를 통해서 독립적인 모델을 생성하는 것을 말한다.

어떤 데이터의 결과를 예측할 때, 주로 한가지의 모델을 활용한다. 앙상블 학습은 개별로 학습한 여러 개의 모델을 조화롭게 학습시켜서 여러 개의 모델들의 예측 결과들을 이용하여 더 정확한 예측 값을 도출해 내는 장점이 있다.

또한 앙상블 학습은 여러 개의 의사결정나무를 결합하여 하나의 의사결정나무보다 더 좋은 성능을 얻어내는 머신러닝 기법이다.

앙상블 학습법에는 배깅(Bagging)과 부스팅(Boosting)이라는 두 가지 방법이 있다[23].

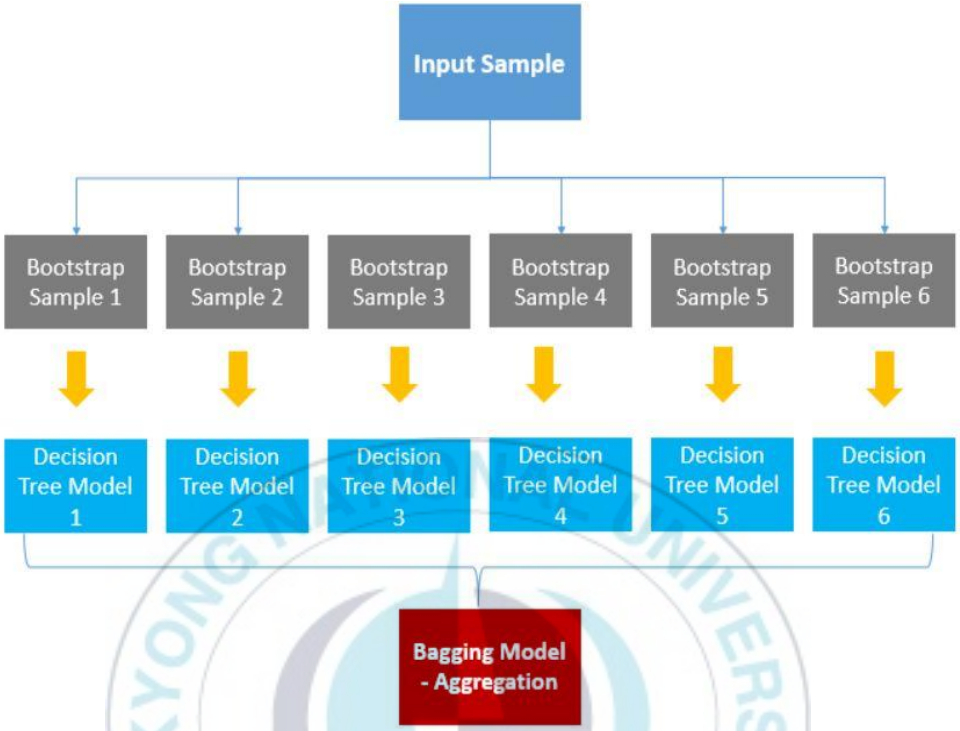
(1) 배깅(Bagging)

배깅은 bootstrap aggregating의 줄임말로 통계적 분류와 회귀 분석에서 사용되는 머신러닝 알고리즘의 안정성과 정확도를 제고하기 위해 고안된 앙상블 학습의 메타 알고리즘이다. 또한 배깅은 분산을 줄이고 과적합(overfitting)을 피하도록 해준다. 의사결정나무 학습법이나 랜덤 포레스트에만 적용되는 것이 일반적이기는 하나, 그 외의 다른 방법들과 함께 사용할 수 있다[24].

아래의 <그림 Ⅱ-9> Bagging Learning Method는 입력된 데이터로부터 Bootstrap[6]을 하고 의사결정나무를 통해 부트스트랩 처리를 한 데이터 모델을 학습시켜 학습된 모델의 결과를 집계하여 예측한 값 중에 가장 많은 값을 최종 예측 값으로 선정하는 절차를 나타내고 있다.

---

6) 부트스트랩(Bootstrap)이란, 일반적으로 한 번 시작되면 알아서 진행되는 일련의 과정을 뜻한다.



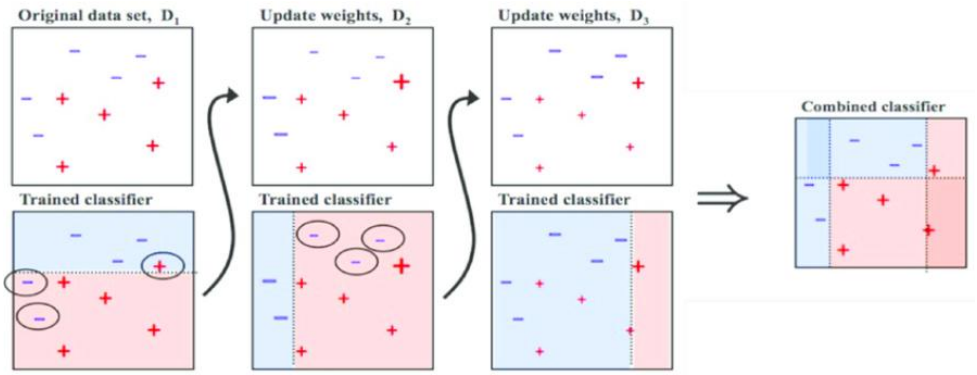
<그림 Ⅱ-9> Bagging Learning Method[25]

(2) 부스팅(Boosting)

부스팅은 모델 간의 연결고리를 의미한다. 이 방식은 가중치가 다음 모델에 영향을 주는 방식이다.

부스팅이 배깅과 동일하게 랜덤하게 샘플링을 하지만, 가중치를 부여한다는 차이점이 있다. 또한 배깅은 병렬로 학습시키는 반면, 부스팅은 순차적으로 학습시키고, 학습이 끝나고 나온 결과에 따라 가중치를 재분배하게 된다.

가중치 부여는 오답에 대해 높은 가중치를 부여하고, 정답에 대해 낮은 가중치를 부여하기 때문에 오답에 더욱 집중하게 된다. 부스팅 기법의 경우, 정확도가 높은 반면 데이터 이상치(data outlier)에 취약할 수 있다는 단점도 있다[26].



<그림 Ⅱ-10> Boosting Learning Method[27]

위의 <그림 Ⅱ-10> Boosting Learning Method 는 +와 -로 구성된 데이터 셋을 분류하는 그림이다.

D1은 2/5 영역의 구분선에서 데이터를 나누었다. 위쪽 영역의 + 하나는 분류가 잘못되었고, 아래쪽 영역의 - 두개도 분류가 잘못되었다. 잘못 분류된 데이터의 가중치는 높여주고, 분류가 잘된 데이터는 가중치는 낮춘다.

D2를 보면 D1에서 분류가 잘된 데이터는 가중치가 낮아졌고 분류가 잘못된 데이터의 가중치는 커졌다. 분류가 잘못 된 데이터에 더 높은 가중치를 부여하여 다음 모델에서 더 집중하여 분류시킨다. D2에서는 오른쪽 세 개의 -가 분류가 잘못되었다.

따라서 D3에서는 세 개의 -의 가중치가 커졌다. D1 모델에서 가중치를 부여했던 각각의 +와 -는 D2에서는 분류가 잘 되었으므로 D3에서는 다시 가중치가 작아진 것이다.

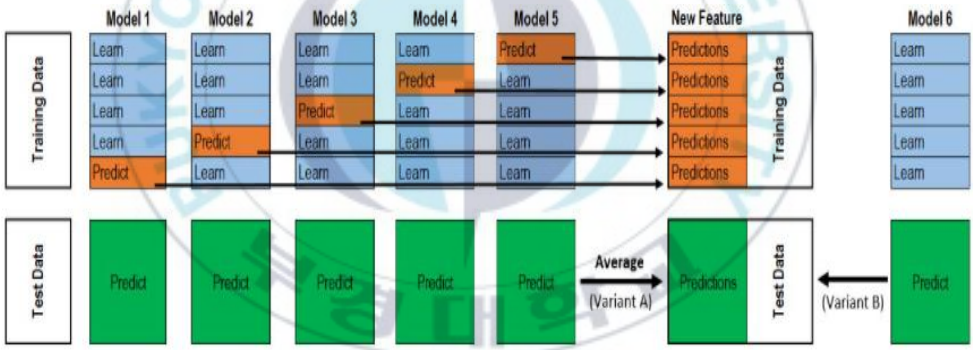
D1, D2 그리고 D3의 Classifier를 합쳐서 최종의 Classifier를 구하게 된다. 최종의 Classifier는 +와 -를 정확하게 구분해 주는 것을 볼 수 있다.

(3) 스태킹(Stacking)

스태킹 기법은 앙상블 학습과 같은 원리에 속하면서 약간의 차이가 있다. 스태킹 기법은 학습 데이터를 통해 예측 데이터들을 생성하고 그 예측 데이터들을 다시 학습 데이터로 입력받아 최종 예측 모델을 만드는 방식이다.

스태킹 기법은 성능이 향상된다는 장점은 있는 반면 예측 값이 다시 입력 값으로 들어가기 때문에 과적합(overfitting[7])이 잘된다는 단점도 있다. 이러한 과적합의 단점을 완화시키기 위하여 스태킹 기법과 교차기법을 동시에 적용하기도 한다. 교차검증과 함께 적용한 경우, 메타 모델이 학습 데이터로 사용되고 예측 값은 테스트 데이터로 사용되는 모델이 된다.

아래 <그림 Ⅱ-11> 은 Stacking Learning Method 모형이다.



<그림 Ⅱ-11> Stacking Learning Method[28]

랜덤 포레스트와 CNN 알고리즘의 결과 값을 합치는 일반적인 앙상블 학습 보다 효과를 높이기 위하여 Text-CNN 알고리즘의 결과 값을 랜덤 포레스트의 학습 값과 예측 값에 다시 사용 하는 스태킹 방식을 사용한다.

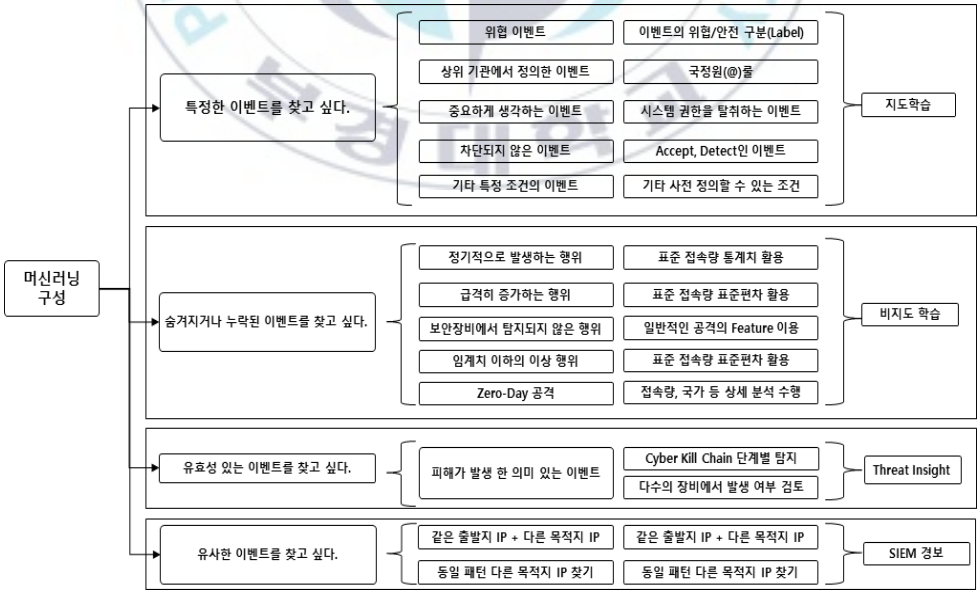
7) 머신러닝이 학습 데이터를 과하게 학습하여 실제 데이터에 대한 오차가 증가하는 현상

Ⅲ. 머신러닝 기술을 이용한 보안관제시스템 구성

본 연구는 머신러닝 기술을 이용한 보안관제시스템을 구성하기 위하여 기존에 운영하는 보안관제시스템인 SIEM의 원본 데이터(Raw Data)의 활용이 필요하다. SIEM 데이터를 머신러닝 기술에 적용하기 위하여 수집영역, 전처리 영역, 학습 및 탐지 영역, 저장 영역 단계별로 데이터를 처리하여 보안관제 요원의 분석결과를 학습하고 자동 분류 및 예측된 값을 활용하여 보안관제 업무를 효율화하고 보안관제의 분석시간을 단축하는 것이 목적이다.

1. 머신러닝 보안관제의 목표

보안담당자는 보안관제 업무를 효율화하고 보안관제의 분석시간을 단축하기 위하여 시스템을 구성하기에 앞서 조직에 적합한 보안관제의 목표를 두고 구성을 계획할 수 있다. 아래 <그림 Ⅲ-1>은 보안관제 목표설정을 위한 계획단계이다.



<그림 Ⅲ-1> 보안관제 목표설정

2. 머신러닝 플랫폼 구성

머신러닝 플랫폼 구성은 아래 <그림 Ⅲ-2>과 같다.

초기 설정 및 시스템 관리를 함에 있어서 구성도의 어떤 부분에 해당하는지 알기 위함이다. 머신러닝 플랫폼의 데이터 흐름은 크게 학습용 데이터 흐름과 탐지용 데이터 흐름이 있으며 각 흐름별 수집, 전처리, 학습 및 탐지 순으로 흘러감을 알고 해당 단계별 설정 관리를 설명함에 있어서 전체 흐름을 보기 위함이다.

<그림 Ⅲ-2> 머신러닝 플랫폼 데이터 흐름도

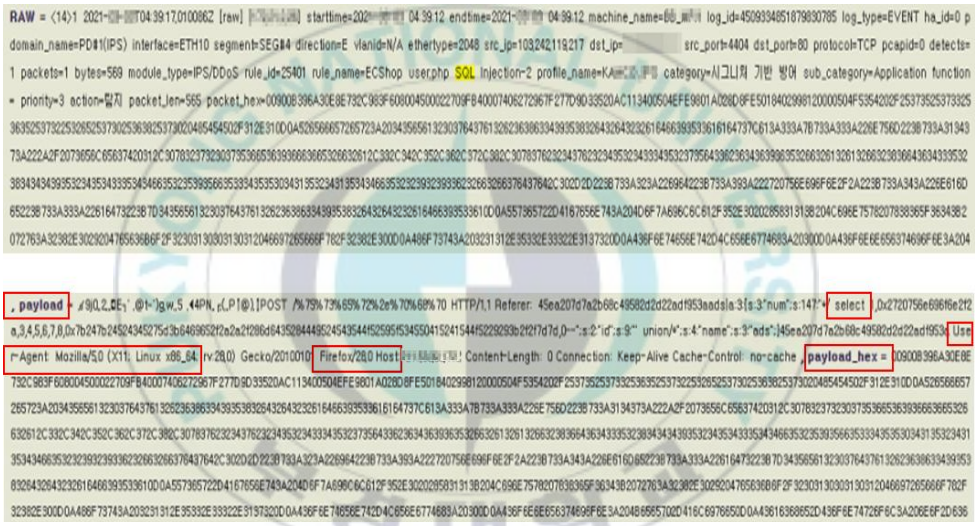
3. 페이로드(Payload) 학습 데이터 수집

머신러닝의 학습 데이터 수집은 SIEM, SmartGuard, CTI등과 같은 로그 발생기로부터 어떤 항목을 어떻게 수집할 것인가를 결정하는 단계이다.

본 연구에서는 데이터 흐름도 상에서 보안관제시스템인 SIEM 으로부터 데이터를 수집하는 단계에 해당한다.

1) 페이로드 데이터 수집 설정

머신러닝의 학습 데이터 수집 설정은 수집하려는 필드 설정, 해당 필드가 속한 로그 유형 설정을 완료하고 추가된 유형과 필드를 이용하여 데이터 수집 시스템 설정 단계로 이어지게 된다. 본 연구에서는 침입방지시스템의 원본 hex형태의 페이로드 데이터를 수집한다. 추가로 데이터를 수집하기 위해서는 침입방지 시스템의 페이로드 로그를 전송하도록 먼저 설정을 해야 한다.



<그림 Ⅲ-3> 페이로드 데이터 로그

위의 <그림 Ⅲ-3>는 침입방지시스템에서 SQL Injection[8] 공격을 차단한 RAW데이터와 페이로드 로그 데이터를 비교한 실제 예시 데이터이다. 위쪽의 데이터는 침입방지시스템의 일반 원본데이터로써 전통적인 시그니처를 통해 공격을 탐지한 그림이고, 아래쪽의 페이로드 데이터를 보면 SQL Injection 공격에 대하여 머신러닝이 분석하기 위한 추가적으로 정보를 제공한다.

발생한 보안 이벤트에 대해서 페이로드 등을 보고 관제 요원이 수동으로 직접

8) 공격자가 보안 취약점을 이용하여, 의도적으로 SQL 문을 주입하고 실행되도록 하여 데이터베이스에서 중요 정보를 노출시키거나 비정상적으로 조작하는 행위

정탐 · 오탐을 판단하게 된다. 많은 보안 이벤트에 대해서 중요 탐지 패턴 명 및 다양한 Field 값을 가지고 경보를 설정하는 데이터 샘플링 데이터로 활용하는 것이다.

<표 Ⅲ-1>은 <그림 Ⅲ-3>의 페이로드에서 제공하는 추가정보를 간략하게 정리한 것이다.

<표 Ⅲ-1> 페이로드 데이터 정보

구 분	정 의
Payload	머신러닝이 Payload_hex 분석하고 파싱하여 정렬된 데이터로 제공한 정보, 공격 Method 관련 정보
select	SQL Injection 공격의 select 쿼리문
user-agent	공격자의 운영체제 정보
Firefox	공격자의 브라우저 정보
Payload_hex	머신러닝이 분석하는 hex 원본데이터 영역

2) 페이로드 데이터 수집 필드 설정

머신러닝이 지도학습 알고리즘을 통해 분석할 침입방지시스템의 원본 데이터와 페이로드 데이터를 수집하기 위해서 보안관제시스템의 SIEM으로부터 지도학습용 침입방지시스템의 보안로그 데이터의 필드를 사전에 정의하여야 한다.

아래 <표 Ⅲ-2>는 SIEM에서 침입방지시스템의 데이터 필드를 정의한 값이다.

<표 Ⅲ-2> 데이터 필드 정보

NO	머신러닝 필드명	필드 설명	IPS 필드명
1	event_time	이벤트 발생 시간	event_time
2	origin	탐지 장비	origin
3	origin_id	탐지 장비의 ID	origin_id
4	origin_name	탐지 장비 이름	origin_name
5	logtype	로그 구분	logtype
6	product	탐지 장비 제품명	product
7	s_ip	출발지 IP	s_ip
8	s_port	출발지 PORT	s_port
9	d_ip	목적지 IP	d_ip
10	d_port	목적지 PORT	d_port
11	action	대응방법(detect/block)	sublog
12	method	탐지 유형	method
13	s_country	출발지 IP 국가	s_country
14	d_country	목적지 IP의 국가	d_country
15	direction	통신방향(10가지구분) 1:내부>내부, 2:내부>외부, 3:내부->DMZ, 4:외부>내부, 5:외부>외부, 6:외부->DMZ, 7:DMZ->내부, 8:DMZ->외부, 9:DMZ->DMZ 0:알 수 없음.	direction
16	risk	제품에서 정의하는 위험도(4가지구분) 0:정보,1:높음,2:중간,3:낮음	-
17	pkt_size	요청 패킷 사이즈 (null 값은 0으로 치환)	pkt_bytes
18	payload	Payload	payload
19	duration	접속 유지 시간	duration
20	count	시도 이벤트 개수	count

위 <표 Ⅲ-2>에 머신러닝 분석을 위해서 페이로드 정보가 추가된 것을 볼 수 있다.

3) 페이로드 데이터 레이블(Data Label)

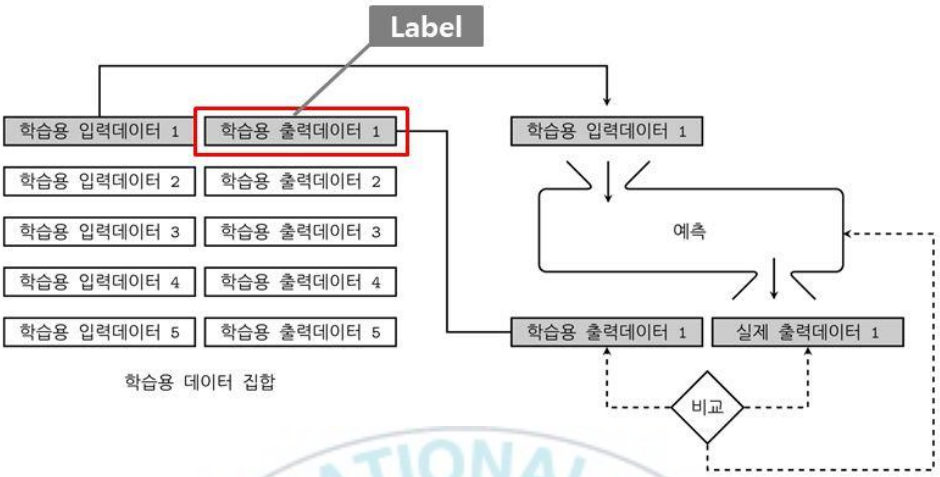
침입방지시스템의 보안로그의 페이로드 정보를 학습데이터로 활용하여 레이블 작업을 한다. 레이블 작업은 탐지이벤트의 목표 값을 정의하여 정탐(공격 : O), 오탐(공격 : X), 중요이벤트 O, X를 활용하여 주요 공격유형별 100개 이상의 데이터를 레이블 처리하여 최적의 학습결과를 도출하는 단계이다.

아래 <그림 Ⅲ-4> 은 데이터 레이블의 예를 설명하고 있다. 정해진 학습 데이터를 미리 정해서 레이블 처리를 하고 머신러닝에 반복 학습시키면 입력 값의 성별이 없어도 학습 데이터의 레이블을 통해 결과 값을 도출해 낸다.

<그림 Ⅲ-4> 데이터 레이블 설명

이 과정을 반복하기 위하여 학습용 입력데이터와 학습용 출력데이터를 사전에 만들어 준비한다. 학습용 입력데이터를 넣어 예측한 실제 출력데이터에 레이블 값을 입힌 학습용 출력 데이터를 대입시켜 비교하고 다시 예측 값을 도출시킨다.

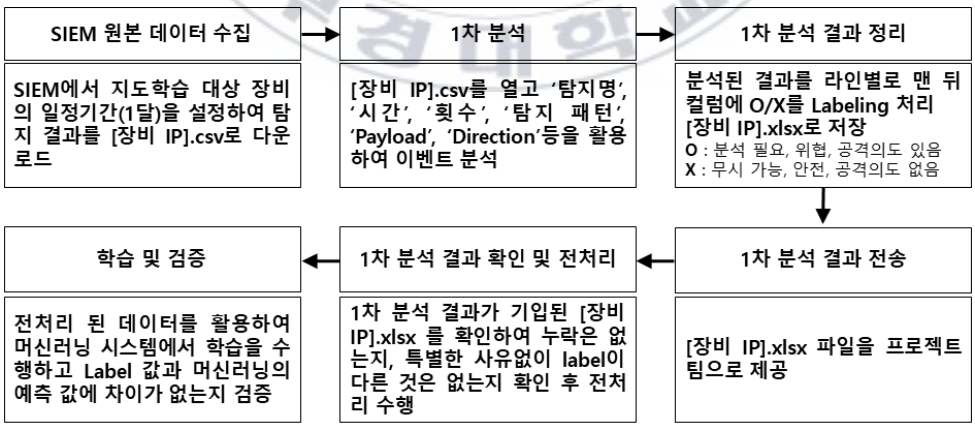
아래 <그림 Ⅲ-5> 와 같이 예측된 값을 다시 학습용 데이터로 만들고 레이블 하는 과정을 반복하여 지도학습의 정확성을 높이게 된다.



<그림 Ⅲ-5> 데이터 레이블 학습 과정

(1) 데이터 레이블 절차

SIEM 수집 데이터를 이용해 1차 분석결과를 기입하고 학습 및 예측 후의 결과 검증하는 절차를 <그림 Ⅲ-6> 로 설명할 수 있다.



<그림 Ⅲ-6> 데이터 레이블 절차

(2) 데이터 레이블 예시

레이블 처리는 SIEM에서 다운로드 받은 지도학습용 데이터로 탐지명, 페이로드, 관련 IP, 시간, 통신방향 등 다양한 정보를 종합 판단하여 아래 <표 Ⅲ-3>와 같이 상세 분석 필요 여부를 레이블로 O/X를 기입한다.

<표 Ⅲ-3> SIEM 학습 데이터 레이블 예시 #1

TIME (시간)	S_IP (출발지 IP)	S_PORT (출발지 포트)	D_IP (목적지 IP)	D_PORT (목적지 포트)	ATTACK (탐지명)	Payload (판단근거)	.....	Label
2019-07-19	1.1.1.1	44413	2.2.2.2	80	SQL Injection	GET /search.asp?keyward=aa'OR 1=1 HTTP1.1		O
2019-07-16	1.1.1.1	11324	2.2.2.2	80	SQL Injeciton	GET /sql injection.pdf HTTP 1.1		X

아래 <표 Ⅲ-4>는 보안담당자가 보안로그에서 로그의 각 항목과 페이로드로 부터 특이사항에 확인하고 레이블을 붙이는 과정이다. 레이블을 붙이는 과정은 기존 보안관제시스템을 운영하면서 탐지결과에 대한 누적된 데이터와 경험을 토대로 진행된다고 볼 수 있다.

<표 Ⅲ-4> SIEM 학습 데이터 레이블 예시 #2

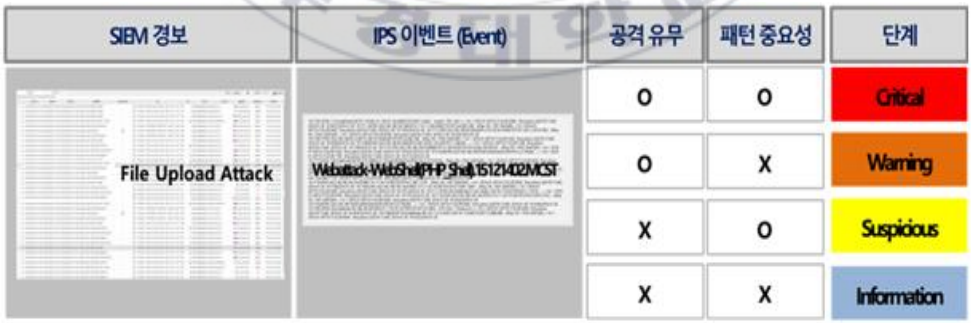
구분	특징	Label
탐지명	탐지명이 mdaemon webconfig overflow이고 jquery ui accordion x.xx.x 문자열이 포함된 경우	X
	탐지명이 WebAttack-AdminPage(PHPInfo) 이고 payload에 php 버전이 포함된 경우	O
	탐지명이 WebAttack-AdminPage(PHPInfo) 이고 모든 문자열이 붙어 있는 경우	X
	탐지명이 icmp ping X-scan scan인 경우	O
	탐지명이 Webattack-Struts2(FileOutputStream) 이고 모두 소스코드 오류이거나 전송에서 발생됨	X
	탐지명에 tcp syn flooding, ssh login burte force, udp port scan 등이 포함된 경우	O
user agent	User Agent에 ZmEu가 포함되어 스캔 공격으로 판단되는 경우	O
	User Agent에 Mindspark MIP 가 포함되어 애드웨어가 설치된 이후 발생되는 패킷으로 판단된 경우	O
	User Agent에 Medunja Solodunnja 6.0.0가 포함되어 내부에서	O

	외부로 통신하는 C&C 통신임으로 판단된 경우	
	User Agent에 Mozilla/5.0 (compatible; Baiduspider/2.0; 가 포함된 코드가 크롤링도구 라고 판단된 경우	O
페이로드(payload)	Payload에 ls%20-al 이 포함되어 디렉터리 목록을 조회하는 것으로 판단되는 경우	O
	Payload에 wget+http이 포함되어 외부에서 파일을 다운로드하는 시도로 판단된 경우	O
	Payload에 cmd.exe가 있고 외부에서 내부 시스템 명령어 사용을 할 수 있다고 판단된 경우	O
	Payload에 /Manager/css/admin_main.css 포함된 경우	X
	Payload에 fckeditor가 포함된 경우	O
	Payload에 manager-gui 가 있고 tomcat 관리자 페이지 로 접근 시도로 판단될 경우	O
	Payload에 BitTorrent protocol이 포함된 경우	O
	Payload에 %22%20and%20%22x%22%3D%22는 " and "x"="y 를 의미 하고 sql injection으로 판단될 경우	O
	Payload에 PROPFIND / HTTP/1.1 등 get/post 이외의 메소드가있는 경우	O
	Payload에 uploadDir=../../upload 이 포함되어 있으나 JPG를 다운로드 하는 경우	X
	Payload에 onload=confirm(/OPENBUGBOUNTY/ 이 포함된 경우	O
	Payload에 muieblackcat이 포함된 경우	O
	Payload에 SELECT가 포함되어 SQL 문을 의미하는 경우	O
	Payload에 XP_CMDSHELL이 있는 경우	O
	Payload에 ECHO가 있어 추가 명령을 실행하는 것으로 판단된 경우	O
	Payload에 %20AnD%20 가 있어 sql injection 공격 테스트로 판단된 경우	O
	Payload에 %20or%20 가 있는 경우 O, 하지만 뒤에 값을 비교하는 구문이 없는 경우	X
	Payload에 ..%2F..%2F가 포함되어 있더라도 실제 파일을 다운로드 시도가 아닌 경우	X
	Payload에 gf_page=upload가 포함된 경우	X
	Payload에 referer = naver.com 가 포함된 경우	X
	Payload에 <methodCall>이 포함된 경우는 외부에서 xml기능으로 계정정보를 수집하는경우	O
HTTP 헤더	HTTP 헤더에 Cache-Control: no-store, must-revalidate 이 포함된 경우는 ddos cc attack 으로 판단된 경우	O

	HTTP 헤더 쿠키나 세션에 eval 문자열이 포함될 경우	X
기타	,s=new java.io.BufferedReader 문자열이 포함되어 자바 오류 페이지로 보이는 경우	X
	LOCALCOMPUTER 라는 문자열이 들어가 악성코드 감염 후 발생시키는 악성코드 감염 신호로 판단된 경우	O
	200 ok 응답에 .baidu.com 등 악성코드 유포에 이용될 수 있는 링크가 포함된 경우	O
	/../../가 포함된 경우 경로 조작 취약점을 활용할 수 있는 경우	O
	eval 문자열이 포함되어 비정상적인 문자열 실행으로 판단된 경우	O
	URL 등에 admin이 포함된 경우	O
	로그를 주고받는 패킷이 탐지되는 경우	X
	xml에 포함된 eval이 있는 경우	X

(3) 데이터 레이블을 통한 보안위협 단계 구분

탐지된 보안이벤트에 대해 정·오탐 여부를 레이블 처리 하고, 공격 의도, 유효성, 사고 이관 등 각 관제 환경을 고려하여 레이블 한 결과를 <그림 Ⅲ-7> 과 같이 보안위협의 단계별로 학습 데이터를 구분하여 정리하거나



<그림 Ⅲ-7> 보안위협 단계 구분

또는 <그림 Ⅲ-8> 과 같이 두 가지 조건을 조합하여 보안위협 단계를 구분할 수 도 있다.

<그림 Ⅲ-8> 레이블 Matrix 보안위협 단계 구분

4) 페이로드 데이터 연계 설정

머신러닝이 보안관제시스템 SIEM 으로부터 보안로그 데이터를 수집하기 위하여 연계설정을 하여야 한다. <표 Ⅲ-5> 는 데이터 수집 시스템 설정은 수집필드, 로그 유형별 필드 추가 설정이 완료된 상태에서 데이터 수집 시스템을 어떻게 구성할지 선택하기 위하여 항목을 정리한 것이다.

<표 Ⅲ-5> 보안관제시스템 연계 데이터 필드 정보

분류	항목	내용
기본정보	연계타입	SIEM 연계 선택 (Syslog, FTP, RDB 도 선택 가능)
고급설정 1	Output streaming	실시간 처리를 원하는 IPS 로그타입 선택 (IPS, WAF, TMS 중 선택)
	Datetime field	Default mgr_time 필요시 변경

	Datetime format	Default yyyyMMddHHmmssSSS
	Logtype field	SIEM의 필드 “logtype”, “log” 중 수집할 log type 필드 선택 (기본값: logtype)
고급설정 2	Index fields	Raw 데이터 저장 시 사용할 path, 각 로그타입 별 origin별 hdfs에 데이터 저장 (기본 값:origin)
	Filters	수집 시 각 로그타입 별로 origin으로 필터 설정 가능
	Parser config	csv 포맷 이외에 json 등의 파일을 파싱할 때 사용 (기본값은 빈칸으로 수집 생성)
	Default values	기본 필드를 추가 할 때 사용 (기본값은 빈칸으로 수집 생성)
	Generated Values	해당 필드에 대한 내부 값을 입력 (기본값은 빈칸으로 수집 생성)

4. 데이터 전처리

데이터 전처리는 데이터 수집 설정에서 추가한 구성을 기반으로 수집 데이터를 머신러닝의 엔진과 모델에서 활용할 수 있는 형태로 그룹화 또는 변형하는 단계이다. 데이터 전처리는 곧 수집된 데이터에서 필드별 가중치 점수를 주고 머신러닝이 분석할 수 있도록 로그 유형별로 가공된 피처를 설정하는 과정이라 말할 수 있다.

아래 <표 Ⅲ-6>는 데이터 전처리 유형을 보여준다.

<표 Ⅲ-6> 데이터 전처리 유형[29]

유 형	설 명
처리	• 수집 시스템의 모델별 데이터 연동 처리 • 학습 시 데이터 샘플링(언더/오버/복합)을 통한 데이터 비대칭 해결
정제	• 특징적인 필드 정제 추출 • 오류발견, 보정, 삭제 및 중복 제거 • 구분자 정의, 정규식 정의 등 기본 정제 • 데이터 분할, 병합, 코드변환 등에 의한 데이터 타입으로 변경
변환	• 결측치 보정, 이상치 식별 및 제거 • 정책 기반 하여 데이터 태깅, 기본 및 정밀 수치화 등 데이터 변환 • 정규화, 군집화, 요약, 계측생성 등의 방법 활용 • 알고리즘이 데이터 분석이 용이한 형태로 변환
전송	• 연계가 필요한 저장소로 통합 • 수치 변환된 데이터를 학습 및 탐지부에 로드벨런싱 하여 분산처리 • 학습 완료 된 전처리 데이터의 폐기

1) 피처(Feature)의 개념

피처는 '특징' 이라는 뜻이다. 만약 공격의 의도를 가진 IP주소를 찾아내고자 한다면 공격 의도를 가진 IP주소의 특징들을 우선 파악해야 할 필요가 있다. 만약 지속적으로 공격을 시도하는 위험도가 높은 IP를 찾아내는 것이 목표이면 공격

위험도, 공격 횟수, 공격을 시도한 기간 등을 포함하여 IP Black Listing을 만든다면 이러한 것을 IP 데이터의 피처라고 말할 수 있다. 피처가 중요한 이유는 분석 이후 분류, 회귀, 군집과 같이 데이터에서 특성을 찾는 작업이 각 피처의 값이 되고 그 결과는 어떤 피처를 사용하였는지에 따라 달라진다.

그러므로 보안 이벤트를 활용하여 사이버 공격의 의도를 찾아내거나 비정상적인 행위를 사전에 탐지하기 위하여 올바른 피처를 선별해 낼 수 있어야 한다[30].

(1) 피처 선정 방법

2012년 미 국방성 산하 SANDIA 연구소에서 실시한 사이버 위협과 관련된 데이터 분석에 필요한 측정 항목 및 측정 방법에 대한 연구인 Cyber Security Metrics를 살펴보면, 5가지의 사이버 보안 분야 데이터의 피처 선정 기준을 말하고 있다[31].

첫째, 의미가 명확하고 모호하지 않아야 한다. 둘째, 목적에 맞고 올바른 의사결정을 지원하는 척도를 이용해야 한다. 셋째, 데이터 수집이 용이해야 한다. 넷째, 정성적 척도보다 정량적인 척도가 좋다. (숫자로 나타낼 수 있는 요소가 더 좋다.) 다섯째, 데이터 전체를 표현할 수 있는 척도가 중요하고 하나의 척도로 표현하기 불가능하다면 여러 개의 척도로 균형을 바로 잡아야 한다.

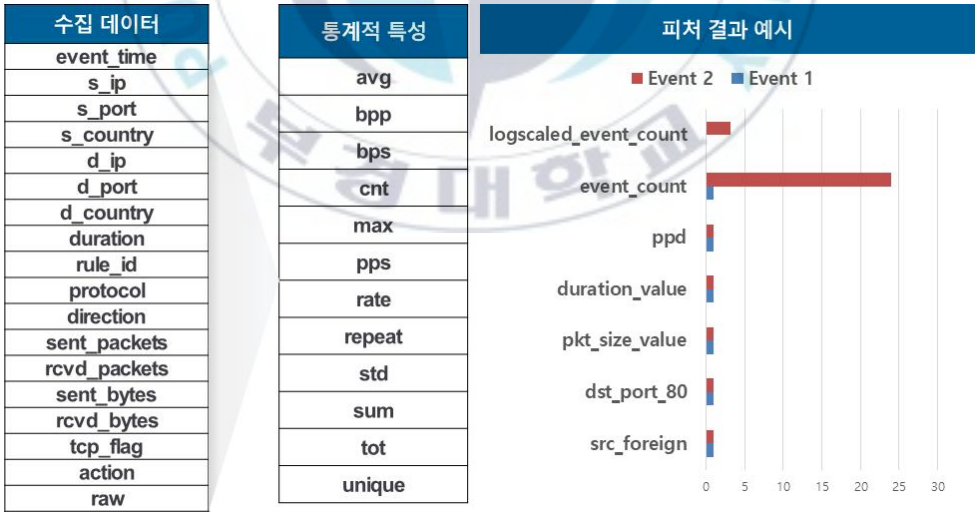
(2) 피처의 카테고리 목록

아래 <표 Ⅲ-7>는 보안로그에서 피처를 생성하기 위해서 카테고리를 선정하고 거기에서 해당되는 하위 피처를 뽑아서 보안담당자가 보안 로그를 보며 세부적인 사항들을 완성하며 생성하는 과정의 예를 보여준다.

<표 Ⅲ-7> 피처 카테고리 및 목록

구 분	통계적 특성
IP 식별	IPR Score, IP User Code, 국내/해외 여부
타이밍	탐지 일수, 비업무시간 탐지 횟수, 탐지 간격
공격 Vector	목적지 IP 개수, 공격 네트워크 개수
공격 강도	이벤트 개수, 기간별 dip/이벤트 개수, 탐지/차단 비율
공격 다양성	발생 이벤트 종류 개수, 장비별 탐지/차단 건수, 탐지 Agent 개수

위의 카테고리 특성을 고려하여 SIEM에서 수집된 데이터 필드들을 군집화 하여 통계적 특성을 적용하면 아래 <그림 Ⅲ-9> 과 같이 피처 결과 예시와 같이 피처를 추출할 수 있다.



<그림 Ⅲ-9> 피처 추출 결과 예시

2) 데이터 전처리 피처 설정

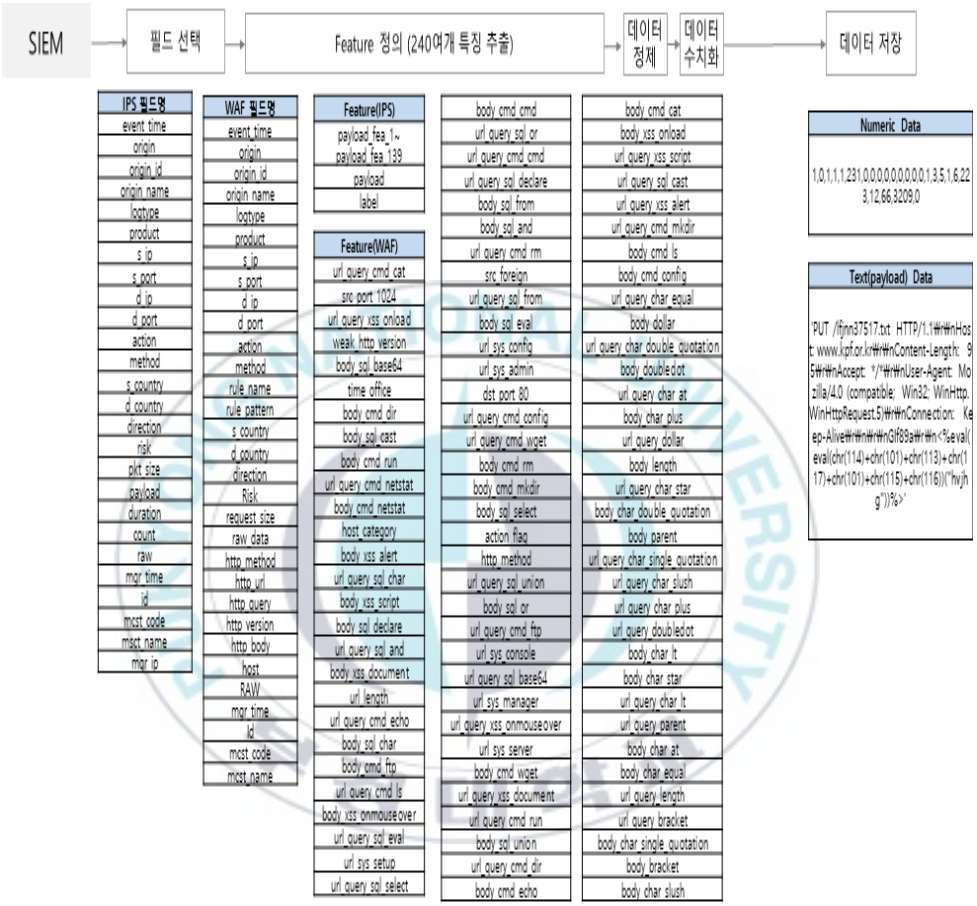
데이터 전처리 피처 설정은 데이터 수집 설정에서 추가한 구성을 기반으로 수집 데이터를 머신러닝의 엔진과 모델에서 활용할 수 있는 형태로 그룹화 또는 변형하기 위함이다. 아래 <표 Ⅲ-8>은 데이터 전처리를 위한 피처 설정 방법이다. 데이터 수집 로그 유형별 설정에 따라 전처리 단계에서 수집된 데이터를 어떻게 변형할지 구성하는 단계이다. 위 <표 Ⅲ-2> 지도학습에 필요한 데이터 필드 정보를 바탕으로 로그 유형별 필드를 선정하여 피처를 설정한다.

<표 Ⅲ-8> 데이터 전처리 피처 설정

구분	주요 필드	피처 이름	설명	전처리함수	Label
attack	attack	tomcat_attack_flag	공격명별 심각도 정의	case when sum(case when ${field1} = 8009 and ${field2} = 0 and ${field3} > 150 and !(${field4} in ('0', '64')) then 1 else 0 end) > 0 then 1 else 0 end	
s_port	src_port_1024	src_port_1024	출발지 1024이하 포트 구분	case when(${field1} > 1 and ${field1} < 1024) then 1 else 0 end	
d_port	dst_port_80	dst_port_80	비일반적인 목적지 포트	case when(${field1} != 80 and ${field1} != 443) then 1 else 0 end	
http_url	http_url	url_length	'http_url'의 문자열 길이	CHAR_LENGTH(CASE WHEN ISNULL(${field1}) THEN '' WHEN ISNAN(${field1}) THEN '' WHEN LOWER(TRIM('\r',TRIM('\n',TRIM(${field1})))) IN ('', 'null', 'nan') THEN '' ELSE TRIM('\r',TRIM('\n', ${field1})) END)	
action	action	action_cnt	대응방법	avg(case when(${field1} == '102') then 1 else 0 end)	
direction	direction	direction	통신 방향	${field1}	
method	method	http_method	시그니처 명	sum(case when(${field1} == 'GET') then 1 else 0 end)	
	method	http_method	시그니처 명	sum(case when(${field1} == 'POST') then 1 else 0 end)	
pkt	pkt_size	pkt_size	패킷 사이즈	${field1}	

pkt_iize	logscaled_pkt_size	logscaled_pkt_size	로그 스케일 패킷 사이즈	ln(${field1})	
duration	sduration	duration_avg	평균 듀레이션	avg(${field1})	
	mduration	duration_sum	전체 듀레이션	sum(${field1})	
ppd	ppd	duration, pkt_size	1분당 패킷 사이즈, Packet per minute	${field2} / ${field1}	
ppd	logscaled_ppd	logscaled_ppd	로그 스케일된 1분당 패킷 사이즈	ln(${field2}) / ${field1}	
event	event_count	event_count	이벤트 개수	count	
payload	payload payload	url_query_length url_query_char_lt	'http_query'의 문자열 길이 http_query'에 특수문자 '<'의 개수/MAX/AVG/Repeat	CHAR_LENGTH(CASE WHEN ISNULL(${field1}) THEN '' WHEN ISNAN(${field1}) THEN '' WHEN LOWER(TRIM('\r',TRIM('\n',TRIM(${field1})))) IN ('', 'null', 'nan') THEN '' ELSE TRIM('\r',TRIM('\n',${field1})) END) UDF_COUNT_CHARACTERS(${field1}, '<')	
	payload	url_query_dollar	http_query'에 특수문자 '$'의 개수/MAX/AVG/Repeat	UDF_COUNT_CHARACTERS(${field1}, '$')	
	payload	url_query_sql_and	'http_query'에 단어 ' and ' or '%20and%20'의 or '+and+' 포함 여부(양옆에 스페이스 포함)	CASE WHEN ISNULL(${field1}) THEN 0 WHEN ISNAN(${field1}) THEN 0 WHEN LOWER(TRIM('\r',TRIM('\n',TRIM(${field1})))) IN ('', 'null', 'nan') THEN 0 WHEN LOWER(${field1}) LIKE '% and %' == true THEN 1 WHEN LOWER(${field1}) LIKE '%20and%20' == true THEN 1 WHEN LOWER(${field1}) LIKE '%+and+%' == true THEN 1 ELSE 0 END	
	payload	weak_http_version	'http_version 1.1인지 여부	CASE WHEN ISNULL(${field1}) THEN 1 WHEN ISNAN(${field1}) THEN 1 WHEN LOWER(TRIM(${field1})) IN ('', 'null', 'nan') THEN 1 WHEN LOWER(${field1}) IN ('http/1.1') THEN 0 ELSE 1 END	

아래 <그림 Ⅲ-10> 은 데이터 전처리를 통해 피처를 생성하여 Numeric Data를 추출하는 과정을 보여주는 구성도이다.



<그림 Ⅲ-10> Numeric Data

3) 데이터 정제 및 수치화

데이터 수집 로그 유형별 설정에 따라 피처를 추출하고 각 피처의 특성을 재조합 혹은 분리하여 전처리 함수 값에 대입시켜 수치화한다. 전처리 함수 값에 의해 간략하게 산출된 0 or 1 값은 Numeric 데이터로 저장하고 문자로 처리되는 값은 출력 값을 간소화하여 저장한다.

<그림 Ⅲ-11> One-hot encoder는 문자 데이터를 간소화하기 위해 쓰이는 인공지능의 알고리즘인데 벡터의 하나의 축에서 자기 자신은 1로 표현되고 나머지는 모두 0 으로 처리하여 머신러닝이 이해할 수 있는 방식으로 부호로 처리하여 학습의 성능과 정확도를 높이는 방식이다.

(1) One-hot encoder

<그림 Ⅲ-11> One-hot encoder[32]

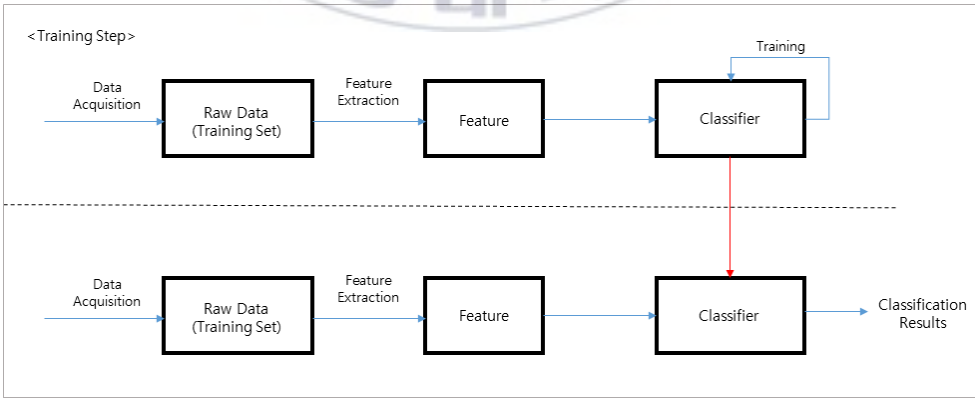
h,i,e,l,o 5개의 글자는 각자 중복 없이 독립적이다. 여기서 input 에서는 맨 마지막 데이터인 'o' 가 포함되어 있지 않고, output에서는 가장 첫 번째 데이터인 'h' 가 포함되어 있지 않다. 그리고 확률 상 input이 'h' 일 때 output이 'i' 가 될 수도 있고 'e' 가 될 수도 있다.

One-hot encoder로 나온 벡터의 값을 통해서 CNN 알고리즘을 이용하여 정확도를 높일 수 있다.

(2) CNN 알고리즘

CNN(Convolutional Neural Network) 알고리즘은 1989년에 제안된 이후로 현재까지도 많이 쓰이고 있는 Deep Network 모델로 인간의 시신경 vision 처리를 모방하여 Computer vision에 특화된 모델이다. 데이터 전처리는 잘되었다는 가정하에 데이터를 사용하는 알고리즘 개발이 이전의 머신러닝 framework의 연구라면, CNN은 전처리(pretraining)가 성능에 미치는 영향을 고려하여 가장 성능이 좋은 특징 지도를 선택하는 합성곱 계층(convolution Layer)필터를 학습하는 모델이다 [33].

<그림 Ⅲ-12> 는 CNN의 학습 및 테스트 구성도이다.

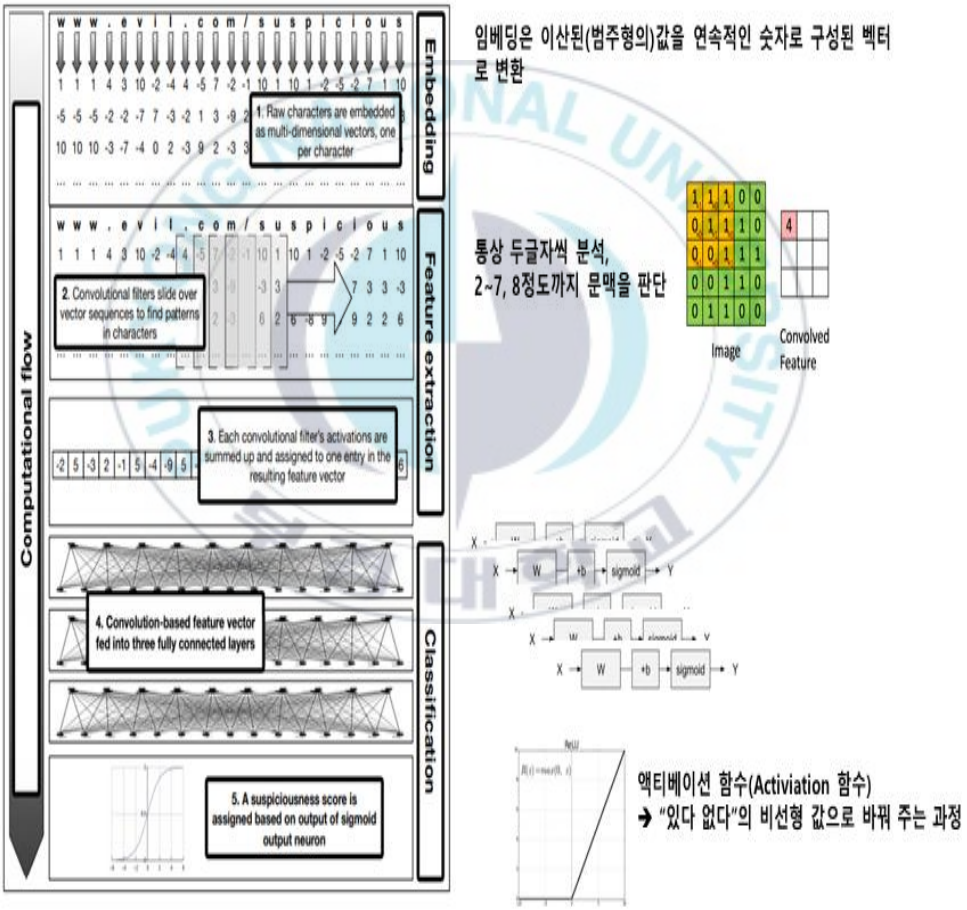


<그림 Ⅲ-12> CNN의 학습 및 테스트 구성[33]

합성곱 계층(convolution Layer)은 데이터 입력 값에 필터링을 적용하여 입력 값과 필터링의 가중치를 곱하고 그 값을 모두 더한다. 이 값이 활성화 함수(activation function)를 거쳐 하나의 합성 갑을 생성한다[34].

또한, convolution 계층에서 특징이 추출 되었다면 그 추출된 특징 값을 인공 신경 지능망에 넣어서 분류한다.

<그림 Ⅲ-13> 은 CNN 모델 아키텍처이다.



<그림 Ⅲ-13> CNN 모델 아키텍처[35]

5. 학습 및 탐지

머신러닝의 학습 및 탐지 모델 설정은 데이터 수집과 전처리 설정을 기반으로 머신러닝 학습, 머신러닝 탐지 모델 설정을 하고 학습 및 탐지 시스템 설정에서 모델을 이용한 구성하기 위함이다.

1) 학습데이터 업로드

학습데이터를 생성하였다면 머신러닝이 학습할 수 있도록 <표 Ⅲ-9>처럼 머신러닝 서버에 학습데이터를 업로드 한다.

<표 Ⅲ-9> 학습데이터 업로드 방법

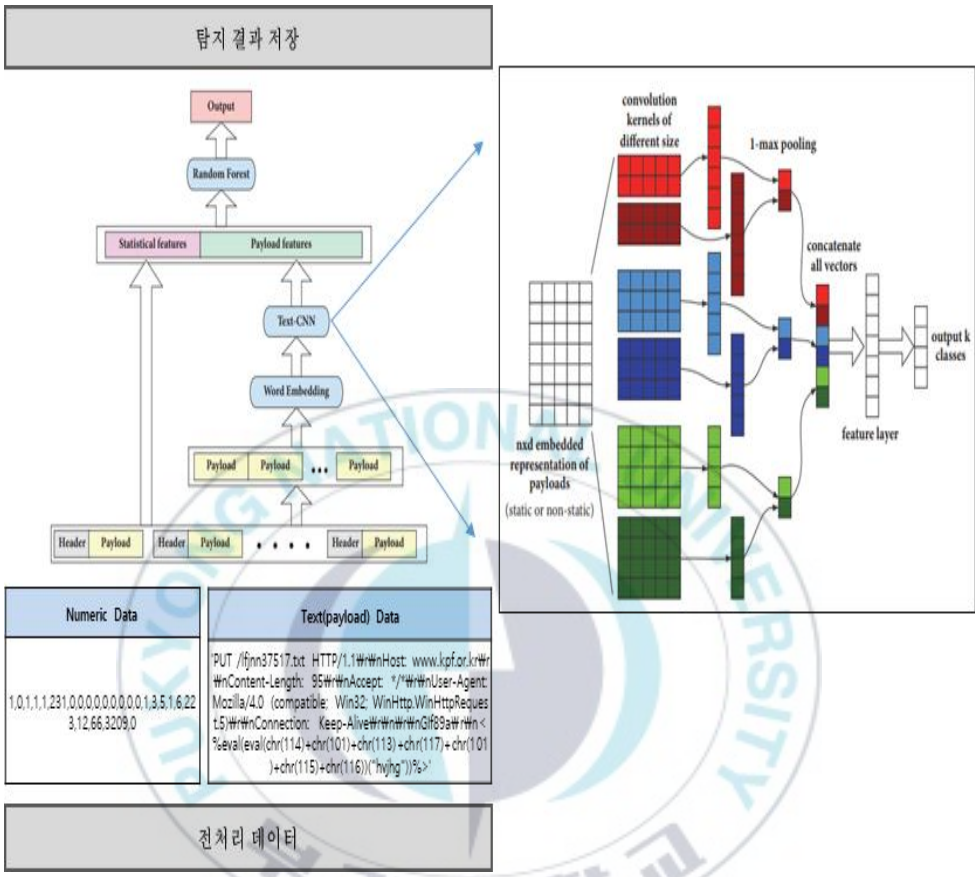
서버	업로드 방법

머신러닝 서버	# 압축 해제하고 해당디렉터리로 이동합니다.
	$ tar -zxvf ~/training_dataset.tar.gz
	$ cd ~/training_dataset
	
	# HDFS 파일 시스템 내에 /data/uploaded 생성
	$ hdfs dfs -mkdir -p /data/uploaded
	
	# training_dataset 디렉터리 안에 있는 파일 및 디렉터리들을 hdfs 파일 시스템으로 업로드
	$ hdfs dfs -put * /data/uploaded

2) 학습 및 예측 알고리즘

학습모델을 설정하기 위해 피처와 모델의 유형과 그룹을 계획하고 선정하여야 한다. 또한, 전처리된 수치형 데이터와 텍스트형 데이터는 랜덤 포레스트, CNN 조합의 RF-CNN 알고리즘을 사용하였고, 데이터에서 통계적인 피처들을 추출하고, 페이로드 등의 텍스트는 인공지능 스스로 피처를 추출하여 결과를 저장토록 했다.

<그림 Ⅲ-14> 은 학습 및 예측 알고리즘 구성을 한눈에 보여준다.



<그림 Ⅲ-14> 학습 및 예측 알고리즘 구성[36]

3) 평가 및 예측

데이터 학습을 통한 평가 및 예측치를 검증하기 위하여 Confusion Matrix를 활용하여 검증한다. 이 방법은 Accuracy(정확도), Precision(정밀도), Recall(재현율) 등 알고리즘 및 머신러닝 모델의 성능을 평가하는 지표로 많이 사용되는 방법이다. Confusion Matrix는 학습을 통해서 예측 정확도를 측정하기 위하여 예측 값과 실제 값을 비교하기 위한 표를 말한다.

<표 Ⅲ-10>은 Confusion matrix 측정지표이다.

		예측	
		공격(Positive)	정상(Negative)
실제	공격(True)	TP	FN
	정상(False)	FP	TN

T는 TRUE, F는 FALSE, P은 POSITIVE, N은 NEGATIVE

TP와 TN은 실제 값을 맞게 예측한 부분, FP와 FN은 실제 값과 다르게 예측한 부분

$$\{\text{Accuracy (정확도)}\} = \frac{TP + TN}{TP + FP + TN + FN}$$ 전체 샘플 중 맞게 예측한 샘플 수

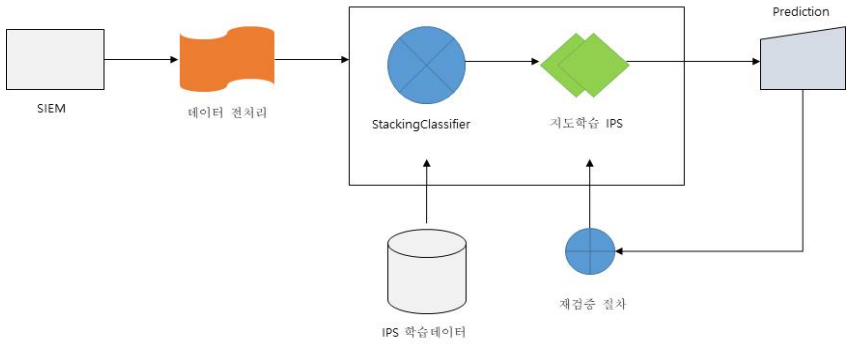
$$\{\text{Precision (정밀도)}\} = \frac{TP}{TP + FN}$$ 실제 공격 중 공격이라고 예측한 수

$$\{\text{Recall (재현도)}\} = \frac{TP}{TP + FP}$$ 공격이라고 예측한 것 중 실제 공격인 것

<표 Ⅲ-10> Confusion matrix 측정지표[37]

4) 지도학습 Model Flow chart

아래 <그림 Ⅲ-15>은 머신러닝의 지도학습 Model Flow chart이며, 아래 <그림 Ⅲ-16>은 머신러닝 Model에 Training 진행의 소스코드 일부이다.



<그림 Ⅲ-15> 지도학습 탐지 결과

```
"training_data": {
   "path": "/data/uploaded/ips/tr-01",
   "type": "hdfs"

},
"overwirte": false,
"algorithm": {
                "parameters": {
                                "category_encoding": "one-hot",
                                "depth": 6,
                                "leaf_reg": 3,
                                "model_setting": {
                                                "line_end": "₩n",
                                                "drop_duplicates": false,
                                }
                                "learning_rate": 0.001,
                                "grow policy": "DecisionTree",
                                "random_strength": 1
                },
                "key": "CategoricalClassfier"
},
"model_id": "model00000",
"features": {
                "src_port_1024": "categorical",
                "dst_port_80": "categorical"
                "url_length": "categorical",
.....
```

<그림 Ⅲ-16> 지도학습 모델 Training 소스코드

출처 : 이글루시큐리티 AI 교육 매뉴얼

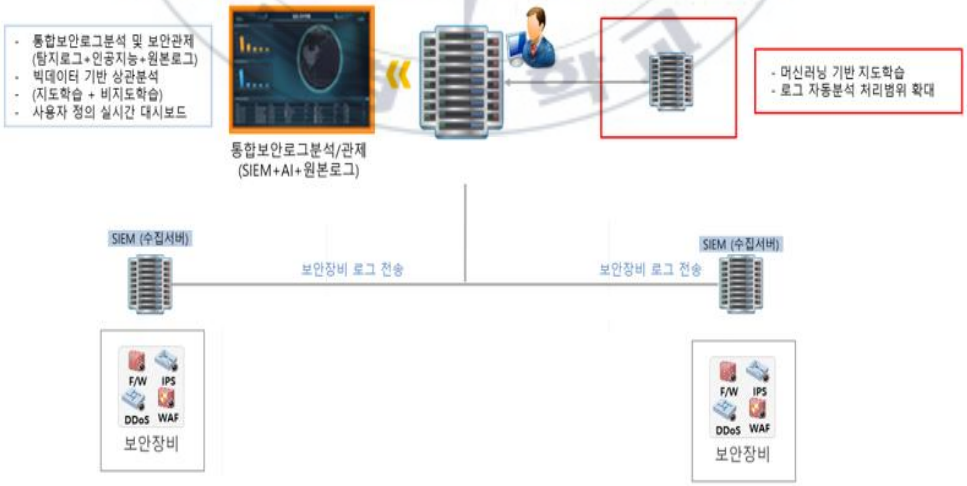
Ⅳ. 실험 및 결과

1. 실험 및 환경

실제 운영 중인 침입방지시스템, 보안관제시스템 SIEM의 장비로부터 학습로그를 추출하고, 머신러닝 장비를 추가로 설치하여 보안관제시스템의 성능 향상을 검증한다.

1) 실험 환경 구성

연구 환경은 네트워크 보안장비의 로그 수집 대상은 <그림 Ⅳ-1>과 같이 외부망 웹서버를 보호하기 위해 관제를 하는 구성이며 이중 수집되는 로그 정보는 침입방지시스템, 기존 보안관제시스템의 보안로그에서 제공받아 분석한다. 침입방지시스템에서 탐지된 경보 이벤트는 머신러닝에 정·오탐 예측 모델로 구현하였다.



<그림 Ⅳ-1> 시스템 환경 구성

연구에 사용되는 보안장비는 보안상 제원 정보만 공개 한다.

<표 Ⅳ-1> 연구 장비 구성

항목	제원 정보	수량
머신러닝	• CPU : Intel(R)Silver 4214R @2.40GHZ, 24core • RAM : 256GB (32G * 8) • HDD : 600GB * 2EA (Raid 1) 4TB * 3EA (Raid 0) • OS : Ubuntu 16.0.4.6	1
수집서버 (SIEM)	• CPU : Intel(R)Silver 4214R @2.40GHZ, 24core • RAM : 256GB (32G * 8) • HDD : 600GB * 2EA (Raid 1) 4TB * 3EA (Raid 0) • OS : CentOS 7.5.1810.2	2
콘솔장비	• CPU : Intel i7 4core • RAM : 16GB (32G * 8) • HDD : SSD 500GB • OS : Windows 10 Pro	1

2) 연구 데이터 준비

연구를 위해 수집된 데이터는 아래 <표 Ⅳ-2> 와 같다

<표 Ⅳ-2> 머신러닝 학습 데이터

항목	침입방지시스템 머신러닝 학습 데이터 수집 건수
수집 기간	2021.2.1. ~ 2021.2.28
총 데이터	1,388,891 건
학습 데이터(60%)	833,355 건
테스트 데이터(40%)	555,556 건

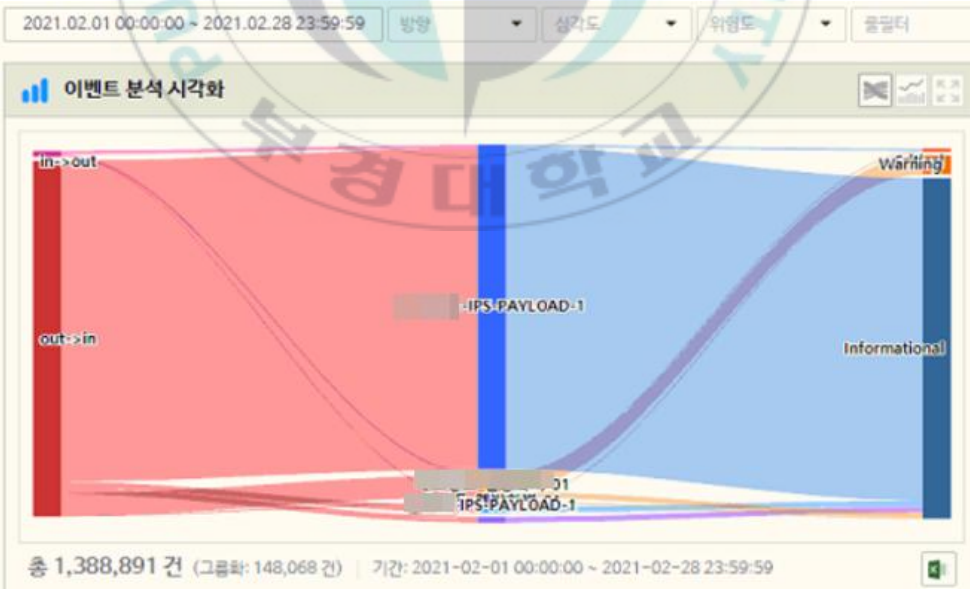
2. 정탐 · 오탐 예측 결과

지도학습의 정탐 · 오탐 예측은 보안관제 담당자가 신뢰할 수 있는 실제 예측치와 비교해서 정확도 수치를 높이고 기존 보안관제시스템에 탐지정책에 반영하여 업무 효율을 높일 수 있는지 확인하는 연구이다.

아래 <그림 Ⅳ-2>은 침입방지시스템의 탐지로그와 페이로드 특성을 추출하여 레이블 작업을 통해서 피처를 추출한 값인 학습용 데이터와 Text-CNN 알고리즘으로 추출된 예측 테스트 데이터를 다시 입력하여 랜덤 포레스트를 통해 최종 예측한 결과이다.

1) 지도학습 탐지 결과

아래 <그림 Ⅳ-2>은 침입방지시스템에서 보안로그를 머신러닝의 지도학습에서 탐지한 결과이다.



<그림 Ⅳ-2> 지도학습 탐지 결과

출처 : 이글루시큐리티 SIEM

(1) 위험 단계별 탐지결과

아래 <표 Ⅳ-3>은 머신러닝이 분석한 데이터에 대해서 위험 단계별로 분류한 것이다. critical 단계에서 9,367 건이 분석이 되었는데 이 결과는 보안장비의 시그니처에서 탐지된 정보와 중복될 수도 있다.

<표 Ⅳ-3> 머신러닝 위험 단계 분석

수집 기간 : 2021.2.1.~2.28.

위험 단계	빅데이터 SIEM	머신러닝 분석 건수	머신러닝 추가탐지 (누적포함)
전체	4,865,146 건	1,388,891 건	29% 추가탐지
critical	25,146 건	9,637 건	37% 추가탐지
Warning	186,547 건	71,660 건	39% 추가탐지
Suspicious	995,442 건	235,756 건	24% 추가탐지
Information	3,667,301건	1,072,131 건	29% 추가탐지

(2) 지도학습 예측 정확도 평가

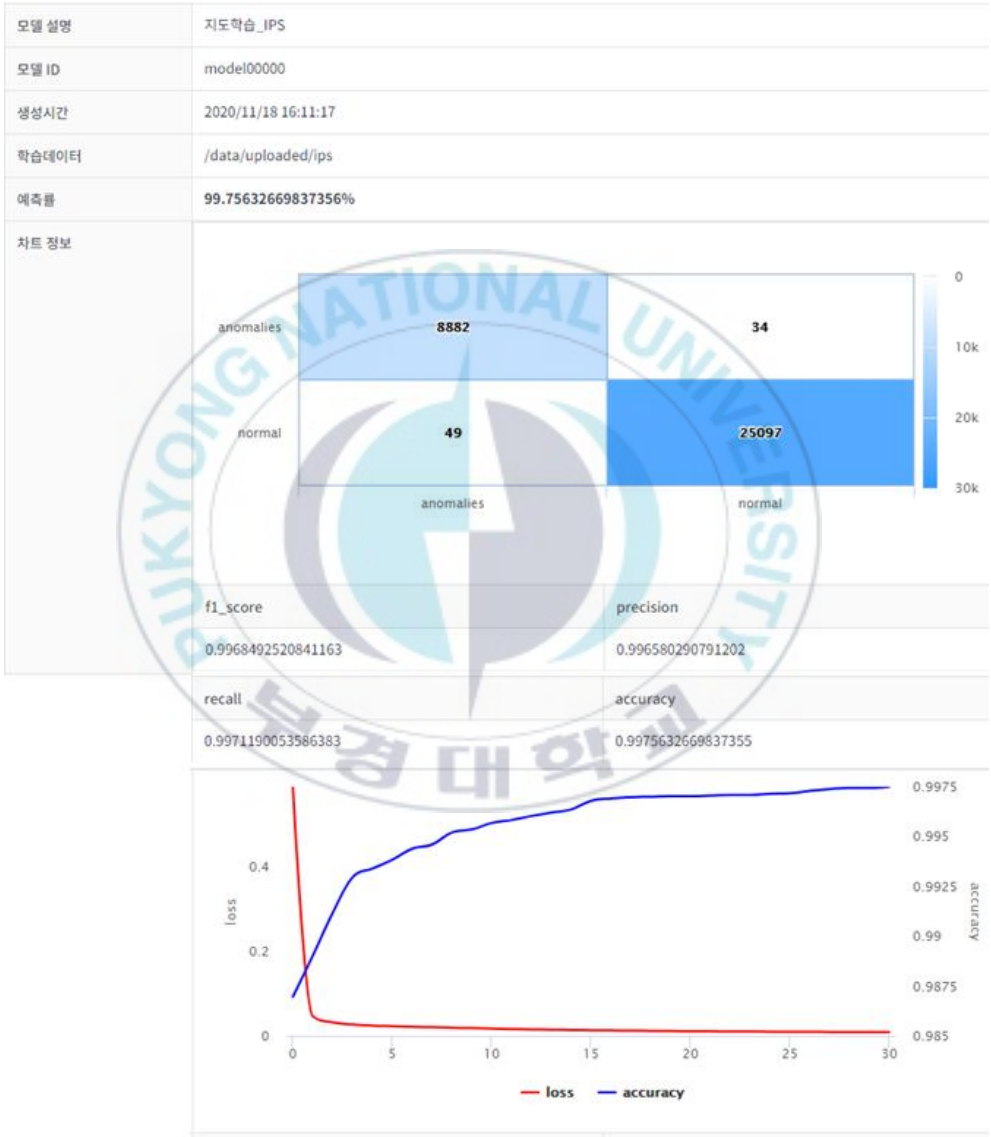
<표 Ⅳ-4>는 머신러닝의 침입방지시스템의 보안로그와 페이로드 추출 데이터에서 예측한 critical 결과를 Confusion Matrix를 통해 평가한 결과이다.

<표 Ⅳ-4> Confusion Matrix 예측 결과

		머신러닝 예측		
		정탐(Normal)	오탐(Abnormal)	합계
실제 이벤트	정탐(Normal)	25,097	49	25,146
	오탐(Abnormal)	34	8,882	8,916
	합계	25,131	8,931	34,062

<그림 Ⅳ-3>는 지도학습의 예측 결과이다.

정확도(accuracy)는 99.75%이고, 정밀도(precision)는 99.65%이고, 재현도(recall)는 99.71%를 보이고 있다.



<그림 Ⅳ-3> 지도학습 예측 결과

출처 : 이글루시큐리티 AI

2) SIEM 지도학습 결과 반영

보안관제 담당자는 보안장비가 탐지하지 못하는 관제 사각지대의 영역에 머신러닝 지도학습 결과의 예측률을 관제업무에 반영할 필요가 있다. 예측률을 계속적으로 높여야 하는 과제는 있으나 예측 결과를 보안관제시스템의 탐지 룰에 적용하여 기존 SIEM 장비의 룰 과의 상관관계를 재분석하여 그 예측률을 더욱 높일 수 있을 것으로 기대할 수 있다.

아래 <표 Ⅳ-5>는 머신러닝의 critical 예측 결과를 SIEM 탐지 룰에 13개의 조건정의를 통해서 추가로 반영한 정보이다. 보안관제 담당자는 SIEM 탐지정보에 대한 관제를 우선적으로 수행하여 업무의 효율성과 성능을 높일 수 있을 것이다.

<표 Ⅳ-5> SIEM 머신러닝 예측 결과 추가

번호	위험도	발생주기	조건	조건 정의
1	critical	보안상 숨김 처리	s_ip	[IPS미탐지] SQL Injection
2	critical		s_ip	[IPS미탐지] URL 인코딩 이상행위
3	critical		s_ip	[IPS미탐지] 개인정보 노출(계정정보)
4	critical		s_ip	[IPS미탐지] 개인정보 노출(사용자정보)
5	critical		s_ip	[IPS미탐지] 게시판 에디터
6	critical		s_ip	[IPS미탐지] 관리자 페이지 접속 시도
7	critical		s_ip	[IPS미탐지] 백업파일 접근시도
8	critical		s_ip	[IPS미탐지] 비정상적인 접근
9	critical		s_ip	[IPS미탐지] 원격 명령어 실행
10	critical		s_ip	[IPS미탐지] 이상 Parameter 감지
11	critical		s_ip	[IPS미탐지] 이상 URL 접근시도
12	critical		s_ip	[IPS미탐지] 제한된 WEB Method 탐지
13	critical		s_ip	[IPS미탐지] 크로스사이트스크립팅

Ⅴ. 결론

본 연구는 기존 보안관제시스템을 통해 진화하는 사이버 공격 방어의 한계를 확인하고 인공지능 기술을 이용한 사이버 공격에 적극적인 대응을 위하여 머신러닝 기반의 보안관제시스템을 구성하기 위함이다. 머신러닝 기반 보안관제시스템의 성능개선을 위해서 크게 침입방지시스템과 웹 방화벽의 보안로그를 분석하여 탐지 알고리즘을 연구하는 지도학습 방법과 방화벽과 웹서버 등을 이용하여 인공지능 알고리즘을 통해 보안로그를 분석하는 비지도학습으로 나눌 수 있다.

비지도학습 알고리즘은 학습데이터를 분석하면서 도출되는 오탐 · 과탐의 비율을 줄이는데 목적이 있다. 그러나 보안관제 담당자가 비지도학습의 머신러닝 기반 보안관제시스템을 구축하다 보면 오탐 · 과탐의 양이 예상보다 훨씬 많이 도출되면서 보안관제 담당자의 고민이 시작된다. 대량의 오탐 · 과탐 결과를 사람이 일일이 확인하는 수작업으로 정탐여부를 가려내는 절차가 반드시 필요하다.

따라서 비지도학습을 통해서 인공지능이 탐지한 공격 로그를 바로 차단할 것인가의 결정은 더욱 부담스러울 수 있다. 그러므로 머신러닝이 탐지한 공격 로그를 신뢰하고 차단하기 보다는 탐지용으로 활용되는 경우도 많다. 물론 이 과도기를 거치며 머신러닝이 탐지기술의 예측률을 높일 수 있다면 높은 수준의 머신러닝 보안관제가 가능할 것이다. 그러나 이 과도기를 넘기기 까지는 보안관제 담당자의 업무의 양은 오히려 증가할 수도 있고 예측하지 못한 사고도 발생할 수 있다.

본 연구에서 지도학습을 통해서 기존 보안장비의 로그 수집 영역을 넓히고 기존에 탐지했던 탐지 룰을 재활용하여 머신러닝 기술을 접목하는 방법을 채택하였다.침입방지시스템의 원본로그 RAW 데이터에서 페이로드 영역을 추가로 수집하여 더욱 심도 있는 보안로그 분석을 시도하였고 기존 탐지 룰과 비교하여 의미 있는 정보를 인식할 수 있도록 레이블 처리를 하여 피처와 피처의 그룹을 만들어 지도학습을 통해 머신러닝을 학습하는 데이터를 생산하였다. 물론 이

정보를 머신러닝이 쉽게 인식할 수 있도록 데이터를 전처리 하고 보안관제 담당자가 보안로그를 확인하는 것은 시간과 노력이 드는 과정은 마찬가지다. 그래서 이 문제를 해결하기 위하여 원본로그 데이터 분석을 철저히 하여 오탐·과탐의 양을 줄이기로 하였다.

지도학습 기반 머신러닝 보안관제시스템을 구축하는 과정으로 첫째, 탐지로그가 많고 비교적 signature의 DB정보가 풍성한 침입방지시스템의 로그를 선택하여 보안로그와 페이로드 영역을 추가로 추출하여 재분석하였다. 둘째, 머신러닝이 인식할 수 있는 정보로 데이터전처리 과정을 거치기 위하여 보안데이터의 필드와 특성을 정리하고 각 필드마다 의미 있는 정보에 레이블 하여 데이터의 특성인 피처를 추출하였다. 셋째, 머신러닝의 데이터 분석의 효율성과 속도향상을 위하여 데이터를 수치화하고 간단하게 Numeric 데이터로 변환하여 학습 데이터를 머신러닝에 전달한다. 넷째, 머신러닝의 분석 알고리즘인 랜덤포레스트와 CNN알고리즘을 함께 활용하여 알고리즘의 교차검증을 하여 정확률을 높이고 Confusion Matrix를 통해 예측률을 높인다. 다섯째, 머신러닝 알고리즘을 통하여 예측률을 기존 보안관제시스템인 SIEM장비의 조건정의에 대입시켜 탐지 룰을 추가 생성하여 머신러닝 기술을 활용한 보안관제시스템을 구축하였다.

본 연구가 보안관제 담당자의 업무의 효율성과 정확성을 제고하는 부분에 도움을 줄 수 있을 것이라 생각한다. 현재 인공지능 기술을 이용한 악성코드 제작이 가능해 지고 공격이 시도되기도 한다. 신기술의 사이버 공격에 적절하게 대응하기 위하여 보안관제에 활용할 수 있는 인공지능 기술과 딥러닝 기술 발전은 더더욱 필요하다. 향후 인공지능 기반 보안관제 기술이 더욱 진화한다면 보안장비 뿐 아니라 IoT 장비를 비롯한 보안관제 영역 확대와 지도학습, 비지도학습 기반 보안관제도 더욱 확대될 것이라 기대된다.

본 연구에서 다양한 보안장비를 모두 테스트 해 보지 못한 한계가 있다. 방화벽, 웹 방화벽, 내부정보유출차단시스템 등 다양한 보안장비의 보안로그를 분석하고, 웹서버, 웹 애플리케이션 서버, 데이터베이스 서버 등 운영 장비의 로그의 상관관계를 분석하여 보안관제 담당자의 업무 효율성을 높일 수 있는 체계적인 연구가 필요하다.

참 고 문 헌

[1] Yann LeCun · Yoshua Bengio · Geoffrey Hinton, Deep learning, nature Vol.251 436-444, 2015.

[2] Apruzzese · Giovanni · Michele Colajanni · Luca Ferretti · Alessandro Guido and Mirco Marchetti, On the Effectiveness of Machine and Deep Learning for Cyber Security, 10th International Conference on Cyber Conflict(NATO CCDCOE),371-389, 2018.

[3] 국가정보원 · 과학기술정보통신부 · 행정안전부 · 방송통신위원회 · 금융위원회, 국가정보보호백서, 2019.

[4] 한국인터넷진흥원, https://www.kisa.or.kr/notice/press_View.jsp?mode=view&p_No=8&b_No=8&d_No=1851, 2019.

[5] 엄진국 · 권헌영, "SIEM을 이용한 침해사고 탐지방법 모델 제안," JIIBC, 2016-6-6, 43-54, 2016.

[6] 국가사이버안전관리규정 제2조, 대통령훈령 제316호, 2013. 9. 2.

[7] itworld, https://www.itworld.co.kr/news/145522. 2020.3.3.

[8] 한국정책학회, 사이버 공격을 통한 첨단사업비밀 유출 실태 및 대응방안 보고서, 2019

[9] 보안뉴스, https://www.boannews.com/media/view.asp?idx=55215, 2017.6.10.

[10] 한국과학기술정보연구원, 최신 사이버위협 동향 및 대응 방안 분석, 2018.

[11] 이데일리, https://www.edaily.co.kr/news/read?newsId=0279456619305352&mediaCodeNo=257, 2021.4.9.

[12] 디지털투데이, https://www.digitaltoday.co.kr/news/articleView.html?idxno=202359, 2018.8.27.

[13] 차영환 · 양해술, 기업보안관리(ESM) 제품의 보안성 평가모델 및 시험방법론 개발. 한국콘텐츠학회논문지 10(6), 2010.6, 156-165, 2010.

[14] Seoksoo Kim · Wooyoung Soh, Design of Security Management System. nternational JOURNAL OF CONTENTS 1(2), 2005.10, 22-25, 2005.

[15] 홍석원, A Study on the Efficient Security Control Method through the Evolutionary SIEM System. 성균관대학교 석사학위논문, 2019.

[16] Mokalled, Hassan; Catelli, Rosario; Casola, Valentina; Debertol, Daniele; Meda, Ermete; Zunino, Rodolfo, The Applicability of a SIEM Solution: Requirements and Evaluation. IEEE, 12-14, June 2019.

[17] 한복동 · 우성희, 머신러닝을 활용한 사이버 공격과 보안 방안. 국정보통신학회 종합학술대회 논문집 24(1), 551-553, 2020.
[18] 유주경, 영상관제에 있어 데이터 전처리기를 통한 이미지 분류의 성능 비교. 숭실대학교 석사학위논문, p8, 2017.
[19] 안광민, 빅 데이터 분석 아키텍처를 사용한 DDoS 공격 탐지, 대전대학교 박사학위논문, 42-46, 2018.
[20] 한국지역정보개발원, 인공지능기반 지방자치단체 보안관제시스템 구축 인공지능교육, 2021.
[21] Guabassi, Inssaf El; Bousalem, Zakaria; Marah, Rim; Qazdar, Aimad, International Journal of Online & Biomedical Engineering; 2021, Vol. 17 Issue 2, p90-105, 16p, 2021.
[22] Liu · Fei Tony · Kai Ming Ting and Zhi-Hua Zhou, Isolation Forest, ICDM: The 8th IEEE International Conference on Data Mining, 2008.
[23] Zwane · Tarwireyi · Matthew, Ensemble Learning Approach for Flow-based Intrusion Detection System. IEEE. 1-8, 2019.
[24] 위키백과, https://ko.wikipedia.org/wiki/%EB%B0%B0%EA%B9%85, 2019.4.3.
[25] Bagging Learning Method, https://wjddyd66.github.io/handson/Ch7.Ensemble_Learning_and_Random_Forest/, 2021.
[26] github, https://swalloow.github.io/bagging-boosting/, 2017.7.13.
[27] Boosting and Bagging explained with examples, https://www.analyticsvidhya.com/blog/2015/11/quick-introduction-boosting-algorithms-machine-learning/, 2015.
[28] Stacking Learning Method, https://www.kaggle.com/getting-started/18153, 2016.
[29] 조창섭, 사이버 공격 탐지 성능 개선을 위한 머신러닝 기반 보안관제시스템. 숭실대학교 박사학위논문, 2019.
[30] 이글루시큐리티, 보안로그 분석 및 블랙리스팅을 위한 피처 엔지니어링(Feature Engineering) 보고서, 2019.
[31] Mark Mateski · Cassandra M, Cyber Threat Metrics, SANDIA REPORT SAND2012-2427, 2012.
[32] 금융보안원, AI를 이용한 금융데이터 분석 교육, 2021.
[33] 이희준, 합성곱신경망(CNN)의 문자인식을 이용한 수기 작성된 주민등록번호 탐지 및 추출 기법. 숭실대학교 석사학위논문, 2016.
[34] 서정은 · 문종섭, 비관계형 데이터베이스 환경에서 CNN과 RNN을 활용한 NoSQL 삽입 공격 탐지 모델, 정보보호학회논문지 30(3), 455-464, 2020.

[35] Joshua Saxe · Konstantin Berlin, A Character-Level Convolutional Neural Network with Embeddings For Detecting Malicious URLs, File Paths and Registry Keys, aiXiv:1702.08568v1, 2017.

[36] Erxue Min · Jun Long · Qiang Liu · Jianjing Cui and Wei Chen, Anomaly-Based Intrusion Detection through Text-Convolutional Neural Network and Random Forest , Hindawi Security and Communication Networks Volunm 2018, Article ID 4943509, p9, 2018.

[37] muhammad hasnain · muhammad fermi pahsa, M.FEvaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking, IEEE Access Access, IEEE. 8:90847-90861, 2020.