Journal of the Korea Academia-Industrial cooperation Society
Vol. 24, No. 10 pp. 829-835, 2023

https://doi.org/10.5762/KAIS.2023.24.10.829
ISSN 1975-4701 / eISSN 2288-4688

머신러닝 기술을 이용한 사이버위협 대응 방안에 관한 연구

이광형[1], 정용훈[2*]
[1]서일대학교 소프트웨어공학과, [2](주)투에이소프트

Research on cyber threat response methods using machine learning technology

Kwang Hyoung Lee[1], Young Hoon Jung[2*]
[1]Department of Software Engineering, Seoil University
[2]twoASoft

요 약 최근 IoT, ICT 환경에서 정교하게 진화하고 있는 사이버위협은 증가하고 있으며, 이에 보안 운영 환경이 감당하기 힘든 수준으로 복잡해지고, 이로 인해 분석할 데이터는 증가하고 있다. 또한 보안 전문 인력 부족 및 성숙도 부족에 따른 휴먼에러가 증가하고 있다. 본 논문에서는 머신러닝 비지도학습과 준지도학습 모델을 이용하여 정교하게 진화하는 사이버위협에 대응할 수 있도록 하였다. 본 논문에서 사용된 비지도학습 모델은 세션정보(L4), 프로토콜 헤더정보(L7), 파일 등의 정보를 수집하고, 이를 위협정보와 비교하여 유사한 위협을 매핑 및 라벨링하여 이상행위를 탐지하는데 사용하였다. 클러스터링 기술을 통해 모델링을 수행하고, 생성된 모델은 재학습을 통해 모델 업데이트하여 분석속도가 향상될 수 있으며, 생성된 모델은 준지도학습에 재학습하여 모델을 업데이트할 수 있도록 하였다. 준지도학습 모델은 비지도학습 모델의 시간별 클러스터링 기반 탐지기법은 평소와 다른 행위를 탐지하는데 유용하고, 공격유형별 지도학습 기반 탐지 기법은 네트워크 기반 행위가 특정 공격에 해당하는지 구별하여 탐지하는데 유용하다. 준지도학습은 지도학습 모델과 비지도학습 모델을 적절히 혼합하여 탐지 정확도와 노이즈(탐지)를 줄일 수 있는 장점이 있다.

Abstract Cyber threats are evolving more sophisticatedly in the IoT and ICT environments. As a result, the security operating environment is becoming unmanageably complex, and the data to be analyzed is increasing. In addition, human errors are increasing due to a lack of security professionals and maturity. This study used unsupervised and semi-supervised machine learning models to respond to sophisticated cyber threats. The unsupervised learning model used in this paper collects information, such as session information (L4), protocol header information (L7), and files, and compares this with threat information to map and label similar threats and detect abnormal behavior. Modeling is performed through clustering technology, and the analysis speed can be improved by updating the generated model through re-learning. The generated model can be re-trained through semi-supervised learning to update the model. The temporal clustering-based detection technique of the unsupervised learning model in the semi-supervised learning model is useful for detecting behavior that is different from usual, and the supervised learning-based detection technique for each attack type is useful for distinguishing and detecting whether network-based behavior corresponds to a specific attack. Semi-supervised learning can reduce the detection accuracy and noise (detection) by appropriately mixing supervised learning models and unsupervised learning models.

Keywords : Unsupervised Learning, Semi-Supervised Learning, Clustering, Classfication, Labeling

본 논문은 2023년도 서일대학교 학술연구비에 의해 연구 되었음.
*Corresponding Author : Young Hoon Jung(twoASoft)
email: jung7773@naver.com
Received September 1, 2023 Revised October 5, 2023
Accepted October 6, 2023 Published October 31, 2023

1. 서론

최근 IoT, ICT 환경의 발전과 관련 활용이 생활과 밀접하게 발전하면서 사이버위협 또한 증가하고 그 공격 기술 또한 진화하고 있다. 이러한 사이버위협은 현재의 보안관제 기술로 신・변종 사이버위협에 대응하는데 한계가 있음이 나타나고 있다. 국내외에서도 다양한 위협 사례가 발생하고 있으며, 공격 대상이 국가 또는 기업에서 개인까지 공격 대상이 확대되고 있다. 국내외 사이버 공격 사례는 제조, 의료, 교통 등 산업 전반에서 발생하고 있다.

한국인터넷진흥원에서 발표한 최근 3년간 침해사고 신고 건수 통계를 보면 2021년 640건에서 2022년 1,142건으로 약 2배가 증가했으며, 2023년 상반기 침해사고 신고 건수는 664건으로 전년 대비 약 40%가 증가하였다고 한다.

시스코 발표 자료에 따르면 사이버 위협 경보 중 93%가 보안장비에서 발생한 경보이며, 나머지 7%는 보안장비가 경보를 발생하지 않은 것으로 조사되었다. 보안 경보가 발생한 93% 중 확인된 경보가 56%, 확인되지 않은 경보가 44%이다. 확인한 경보 56%중에서도 정상적인 경보는 34%에 그치며 정상적이지 않은 경보가 66%로 전체 위협 중 18%만이 정탐으로 판단된다고 한다. 또한 정상적인 경보 34% 중 해결된 경보는 51%, 해결되지 않은 경보는 49%로 전체 위협 중 약 9%만이 해결되고 있다고 한다.

이러한 사이버위협의 증가는 기존 보안관제에 활용되고 있는 침입탐지시스템, 방화벽, 딥러닝 등 시그니처기반 탐지 이벤트를 상관분석하여 탐지 이벤트를 줄이고 있다. 하지만 보안 위협은 더욱 정교해지고 진화하여 대응해야할 컴플라이언스는 늘어나고 있다. 또한 보안 운영 환경이 감당하기 어려운 수준으로 복잡해져 분석할 데이터는 증가하고 있다.

또한 기존 사이버위협 대응체계는 딥러닝 기반으로 충분한 양질의 데이터가 필요하며, 이러한 데이터들의 지도학습을 위해서는 라벨링 과정이 필수적이다. 하지만 현실에서 양질의 데이터를 충분히 수집하기란 매우 어려우며 많은 시간과 비용이 발생한다.

매년 기관 및 기업에서 운영하고 있는 전산장비와 보안장비는 증가하고 있으며, 이에 따라 대량의 로그가 발생되고 있다. 하지만 전문 인력 부족으로 대량의 로그를 처리하는데 한계가 있다.

본 논문에서는 머신러닝 기술 중 비지도학습과 준지도학습 이용한 사이버위협 대응체계를 제안하고자 한다. 2장은 관련연구로 머신러닝 기술과 보안관제의 진화에 대해 기술한다. 3장에서는 제안하는 시스템에 대한 역할, 기능 등을 기술하였다.

2. 관련연구

2.1 지도학습

지도학습(Supervised Learning)은 훈련 데이터(Training Data)로부터 하나의 함수를 유추해내기 위한 기계 학습(Machine Learning)의 한 방법이다. 훈련 데이터는 일반적으로 입력 객체에 대한 속성을 벡터 형태로 포함하고 있으며, 각각의 벡터에 대해 원하는 결과가 무엇인지 표시되어 있다. 이렇게 유추된 함수 중 연속적인 값을 출력하는 것을 회귀분석(Regression)이라 하고 주어진 입력 벡터가 어떤 종류의 값인지 표식하는 것을 분류(Classification)라 한다.

지도 학습기가 하는 작업은 훈련 데이터로부터 주어진 데이터에 대해 예측하고자 하는 값을 올바로 추측해내는 것이다.

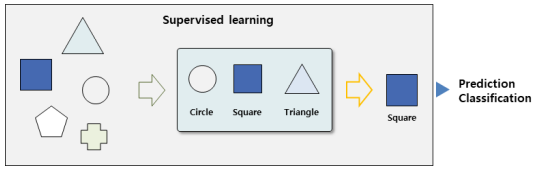

Fig. 1. Supervised Learning

이 목표를 달성하기 위해서는 학습기가 "알맞은" 방법을 통하여 기존의 훈련 데이터로부터 나타나지 않던 상황까지도 일반화하여 처리할 수 있어야 한다.

지도학습 방법은 주로 이메일의 스팸 여부 분류, 소셜미디어 공유 점수 및 성과 점수 예측, 이미지 인식 등에 활용된다.

2.2 비지도학습

기계 학습의 일종으로 데이터가 어떻게 구성되었는지를 알아내는 문제의 범주에 속한다. 이 방법은 지도 학습 혹은 강화 학습(Reinforcement Learning)과는 달리 입력값에 대한 목표치가 주어지지 않는다. 자율 학습은 통계의 밀도 추정(Density Estimation)과 깊은 연관이 있다.

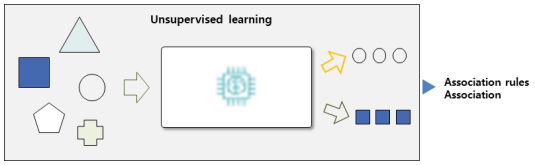

Fig. 2. Unsupervised learning

이러한 자율 학습은 데이터의 주요 특징을 요약하고 설명할 수 있다. 자율 학습의 예로는 클러스터링(Clustering)을 들 수 있다. 또 다른 하나의 예로는 독립 성분 분석(Independent Component Analysis)이 있다.

주로 활용되고 있는 분야는 구매 행동에 따른 고객 그룹화, 사진 목록에서 비슷한 얼굴로 그룹화, 고객 데이터에서 연관성 식별 등에 활용되고 있다[1].

2.3 준지도학습

준지도학습(Semi-supervised learning)이란 적은 레이블 데이터가 있으면서 추가로 활용할 수 있는 대용량의 레이블이 없는 데이터가 있다면 준지도학습을 고려할 수 있다[3,6,7]. 준지도학습은 소량의 레이블 데이터에는 지도학습(supervised learning)을 적용하고 대용량 레이블이 없는 데이터에는 비지도학습(unsupervised learning)을 적용해 추가적인 성능향상을 목표로 하는 방법론이다. 이런 방법론에 내재되는 믿음은 레이블을 맞추는 모델에서 벗어나 데이터 자체의 본질적인 특성이 모델링 된다면 소량의 레이블 데이터를 통한 약간의 가이드로 일반화 성능을 끌어올릴 수 있다는 것이다[2,4].

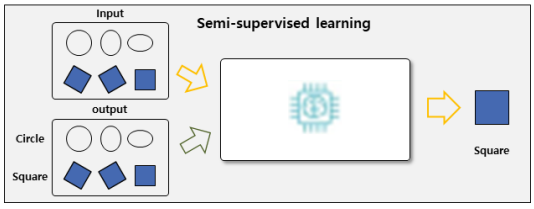

Fig. 3. Semi-supervised learning

2.4 보안관제의 진화

보안관제는 사이버공격과 내부정보 유출 사고를 체계적으로 대응하기 위해 구성된 인력과 프로세스 그리고 기술의 집합이다. 조직은 공격으로부터 생존하기 위해 잠재적 위협을 인식하고 사고를 조기에 탐지하여 신속하게 대응하기 위한 기술과 조직의 정보 도메인을 모니터링하여 공격을 예측하고 탐지 및 대응하는 보안 전문가가 운영하는 것으로 정의하고 있다.

보안관제센터 구성요소는 보안관제 및 관련 업무를 수행할 수 있는 전문조직과 업무의 흐름을 결정하는 프로세스 그리고 보안관제를 위해 필요한 시스템으로 구성된 기술로 구분된다.

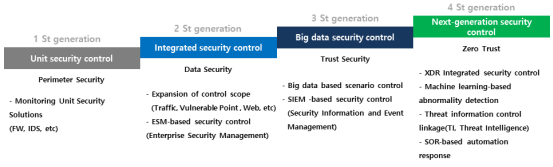

Fig. 4. Generational changes in security control

보안관제 기술은 Fig. 4와 같이 1세대 단위 보안 솔루션 모니터링으로 보안 장비별 발생하는 이벤트를 모니터링한다. 2세대는 관제 범위를 트래픽, 취약점 등으로 확대하고 보안장비에서 발생하는 이벤트 통합을 목적으로 ESM(Enterprise Security Management)기반 보안관제로 진화하였다. 3세대 빅데이터 보안관제는 로그와 이벤트에 대해 빅데이터 분석과 상관경보 분석을 수행하는 전문 분석가가 필요 하였으며, 실시간 처리가 가능한 빅데이터 기반 SIEM 보안관제를 수행하였다. 4세대는 머신러닝 기반의 이상행위 탐지와 위협정보를 관제와 연동 등 제로트러스트를 목적으로 진화하였다.

2.5 보안관제 문제점

기존의 보안관제는 침입탐지시스템, 웹 어플리케이션 방화벽, 딥러닝 등 시그니처 기반 탐지 이벤트를 상관분석하여 탐지 이벤트를 줄이고 있다. 그러나 보안 위협은 더욱 정교해지고 진화하여 대응해야할 컴플라이언스는 늘어나고 있으며, 보안 운영 환경이 감당하기 힘든 수준으로 복잡해져 분석할 데이터는 증가하고 있다. 또한 이를 대응하기 위한 보안관제 인력은 한계에 다다르고 있다[8].

관리적 측면의 문제점 세 가지는 다음과 같다. 첫 번째 IT 인프라 복잡성 증가에 따른 공격 요인 증가, 두 번째 대응해야 하는 컴플라이언스의 증가, 세 번째 보안 예산 부족으로 인한 보안 담당 인력 부족으로 꼽을 수 있다.

기술적 측면의 문제점 세 가지는 다음과 같다. 첫 번째 신규 보안위협 증가로 인한 보안솔루션 복잡도 증가, 두 번째 보안솔루션 경보 급증으로 보안관제 업무 증가, 세 번째 보안 전문 인력 부족 및 성숙도 부족에 따른 휴먼에러 증가를 들 수 있다[3,5].

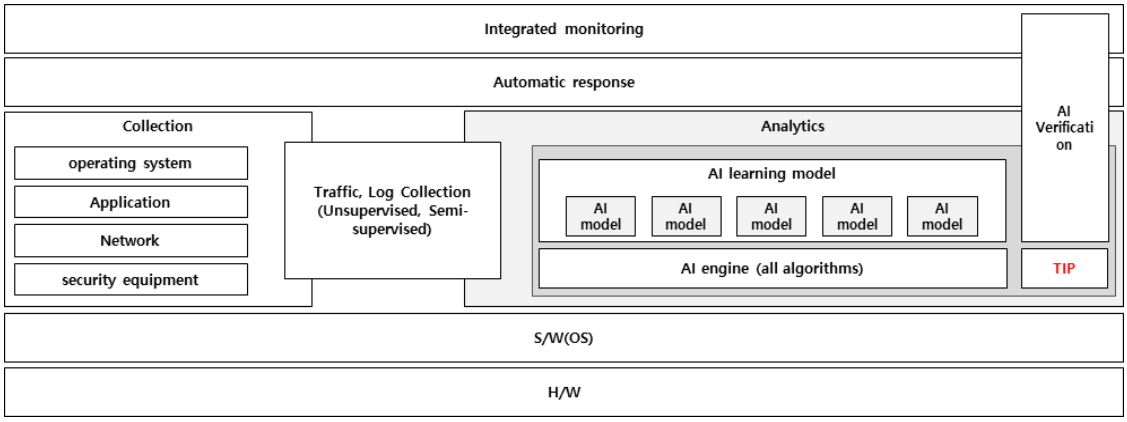

Fig. 5. Proposed system configuration diagram

3. 머신러닝 기반 사이버위협 대응체계

본 논문에서는 머신러닝 기술 중 비지도학습을 기반으로 대량의 로그 및 트래픽 처리가 가능하며, 빠른 분석을 통해 신·변종 사이버위협으로부터 선제적 대응과 침해사고 최소화가 가능한 시스템을 제안하고자 한다.

제안하는 시스템은 통합 모니터링, 자동대응, 수집, 분석 모듈로 구성된다. 향후 AI 기술 발달로 새로운 AI 학습 알고리즘이 나오는 경우 보다 쉽게 추가할 수 있는 모듈 형태로 설계하였다.

3.1 수집 서버

사이버 공격이 점차 고도화되고 정교해짐에 따라 기존 보안솔루션들의 공격 대응에 점차 한계점이 드러고 있다. 이를 개선하기 위해서는 다양한 보안 시스템과 트래픽에서 이벤트 데이터를 수집하고 이상 이벤트 데이터에 대한 상관관계 분석이 필요하다. 보안 이벤트 로그 수집은 보안 사고 탐지와 분석을 위한 필수 요소로 트래픽, 로그, 파일 등을 수집할 수 있어야 한다.

수집 서버는 다양한 방법으로 데이터를 수집하여 사이버위협으로부터 대응할 수 있어야 한다.

- 공개된 출처에서 URL, IP, 도메인 이름, 파일 등의 정보를 수집하고 인공지능 기술을 통해 새로운 TI(Threat Intelligence)를 도출 한다.
- 네트워크 구간의 트래픽 중 세션정보, 주요 서비스의 헤더정보에서 IP, URL, File 등을 분석하여 신·변종 공격을 탐지하고, 내부 단말기의 이상행위를 탐지 한다.
- 기존 보안장비에서 발생하는 이벤트 탐지로그를 수집하여 인공지능 기법을 통해 분석 한다.

3.2 분석 서버

분석 서버는 머신러닝 기반의 다양한 학습 모델을 지원할 수 있도록 모듈화 하여 쉽게 추가할 수 있도록 설계하였다. 분석 서버는 기본적으로 비지도학습과 준지도학습을 기본한다.

3.2.1 비지도학습

비지도학습 모델은 대량의 로그 및 트래픽 처리가 가능하며, 빠른 분석을 통해 신·변종 사이버위협으로부터 선제적 대응과 침해사고 최소화가 가능하다.

비지도학습은 웹 트래픽 내 공격자로 의심되는 이상행위 탐지 및 라벨링을 위해 사용한다.

비지도학습은 세션정보(L4), 프로토콜 헤더정보(L7), 파일 등의 정보를 수집하고, 이를 위협정보와 비교하여 유사한 위협을 매핑 및 라벨링하여 이상행위를 탐지하는데 사용된다.

라벨링 전 클러스터 정보는 인공지능 학습 모델을 업데이트하는데 사용된다.

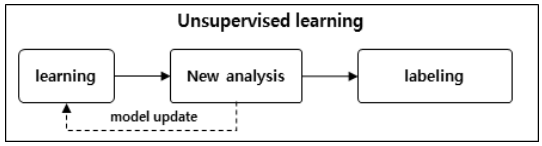

Fig. 6. Unsupervised learning process

비지도학습 특성상 학습데이터는 운영 데이터가 되어야 하며, 업데이트 시 신규 운영 데이터를 지속적으로 학습한 결과가 모델에 반영될 수 있어야 한다. 비지도학습 모델 생성 과정은 다음과 같다.

- 적절한 클러스터링 알고리즘 선택 후 파라미터와 하이퍼 파라미터를 설정한다.
- 일정 기간 동안 데이터를 취합한 후 적절한 피처를 선택하고 각 피처의 중요도를 결정한다.
- 비지도학습 머신러닝 엔진으로 데이터를 학습시켜 모델을 생성한다.
- 생성된 모델이 부적합할 경우 수정 사항 반영 후 결정한다.

클러스터링 기술은 이상행위를 탐지하기 위해 매우 적합한 기술 중 하나이다. 일정 기간 동안 트래픽 데이터를 학습(클러스터링)하여 모델링을 수행하며, 초기 모델링을 거쳐 생성된 모델은 신규 데이터가 어떤 클러스터링에 속하는지를 재학습하여 모델을 업데이트한다.

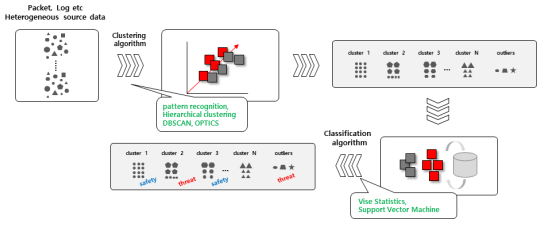

Fig. 7. Clustering process

이 과정을 지속적으로 반복하여 모델은 정교해지고 트래픽의 특성이 모델에 잘 반영되어 분석 속도를 향상 시킬 수 있다. 다음 Fig. 7은 클러스터링 과정을 나타내고 있다.

클러스터링 알고리즘을 통해 여러 클러스터를 분류하고 분류 알고리즘을 통해 안전한 클러스터와 위협 클러스터로 분류한다. 생성된 클러스터 내의 값들이 특정 위협의 속성 정보와 유사한 경우 클러스터 데이터의 라벨을 위협의 이름으로 제시한다. 이렇게 위협으로 라벨링된 클러스터와 비교하여 유사도 측정 속도각 실시간 탐지 및 대응이 가능할 것으로 판단된다.

3.2.2 준지도학습

준지도학습은 정찰, 준비, 감염, C2 서버 접속, 내부 통신, 정보유출 등 일련의 단계에서 이상행위를 탐지한다. 준지도학습은 비지도학습의 문제점을 최소화하고 성능 향상을 위해 사용하였으며, 준지도학습은 제한된 수의 레이블이 있는 데이터와 많은 수의 레이블이 없는 데이터를 훈련하는데 적합한 학습 모델이다. 또한 준지도학습은 지도학습 모델과 비지도학습 모델을 적절히 혼합하여 탐지 정확도와 노이즈(탐지)를 줄일 수 있는 장점이 있다.

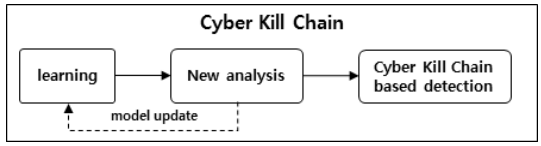

Fig. 8. Cyber Kill Chain process

학습데이터는 트래픽 세션정보와 주요 프로토콜 헤더 정보의 세션 단위 테스트, 모델링은 특정 네트워크 행위 기준으로 시계열 분석을 위한 주기적 모니터링, 분석 및 탐지는 패킷 또는 세션 단위로 사이버 킬 체인 기반 룰셋을 기반으로 탐지한다.

준지도학습은 N개의 비지도학습 모델과 M개의 지도학습 모델을 이용하여 클러스터링할 수 있다. 특정 시간대의 통신 형태를 과거와 비교하여 비정상 행위를 탐지 할 수 있으며, 하루 특정 시간대의 통신 형태를 요일에 따라 비교하여 비정상 행위 탐지가 가능하다. 또한 특정 애플리케이션 또는 접속 외부 사이트 별로 사용 추이가 과거와 비교하여 비정상 행위를 탐지할 수 있다.

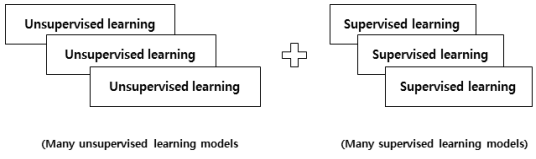

Fig. 9. Clustering using many abnormal behavior detection models and many supervised learning models

M개의 지도학습은 위협 유형에 따라 롤을 등록하여 사이버 킬체인 단계에 따라 처리하게 된다.

준지도학습 모델 업데이트는 실제 환경에 적용 시 모델이 유용하게 작동하는지 평가하여 유용하지 않다고 판단될 경우, 해당 유형의 위협 사례가 상당 수량 확보되었을 때 이를 모델에 반영할 필요가 있다고 판단될 경우 업데이트 되어야 한다. 준지도학습은 모델 특성상 개발사

에서 새로운 데이터 분석 결과를 반영한 신규 모델을 자동 또는 수동으로 분석 서버에 업데이트하는 것이 효과적이다.

4. 결론

기존 사이버위협 대응체계는 침입탐지시스템, 방화벽, 딥러닝 등으로 충분한 양질의 데이터가 필요하며, 이러한 데이터들의 지도학습을 위해서는 라벨링 과정이 필수적이다. 하지만 현실에서 데이터를 충분히 수집하기란 매우 어려우며, 많은 시간과 비용이 필요하다. 또한 수집된 데이터에 라벨링을 하는 과정에서도 시간과 노력이 필요하다.

본 논문에서는 점차 정교하게 진화하는 신규 사이버위협에 대응하기 위해 비지도학습과 준지도학습을 이용한 사이버위협 대응체계를 제시하였다. 최근 비지도학습 기반의 라벨링 기법들은 진화하여 많은 시간과 비용이 소요되는 수동 라벨 작업 방식보다 효과적이며 확장성을 가지고 있다는 평가를 받고 있다.

하지만 비지도학습은 라벨링 기법의 신뢰도 개선을 위한 지속적인 개선이 필요하다.

제안하는 논문에서는 비지도학습을 기반으로 대량의 로그 및 트래픽 처리가 가능하도록 하여 신·변종 사이버위협으로부터 선제적 대응과 침해사고 최소화가 가능한 시스템을 제안하였다. 또한 비지도학습의 문제점을 최소화하고 성능 향상을 위해 준지도학습을 혼합하여 사용하였다. 준지도학습은 제한된 수의 레이블이 있는 데이터와 많은 수의 레이블이 없는 데이터를 훈련하는데 적합한 기법이다.

제안하는 사이버위협 대응체계는 진화하는 사이버위협으로 보안솔루션 경보 급증으로 보안관제 업무 급증과 보안 전문인력 부족 및 성숙도 부족에 따른 휴먼에러를 최소화할 수 있는 방법을 제안하였다.

향후 자기지도학습 및 강화학습을 적용하여 스스로 학습데이터에 없는 정의와 규칙을 찾아 분류, 의미를 부여할 수 있는 사이버위협 대응체계를 만들어 진화하는 사이버위협으로부터 대응할 수 있는 방법이 필요하다.

References

[1] H. G. Shon, *Automatic Analysis of Intrusion-Detection Big-Data Using Unsupervised Learning and Deep Learning*, Master's thesis, Kookmin university, 2022.

[2] J. H. Park, "*The Analysis of Semi-supervised Learning Technique of Deep earning-based Classification Mode"l*, The Korean Institute of Broadcast and Media Engineers, Volume 26, No 1, 81~83, 2021.
DOI: http://dx.doi.org/10.5909/JBE.2021.26.1.79

[3] J. I. ok, *"Cyber security control system problems and machine learning application technology status"*, Korea Institute of information security, Volume 31, No3, 2021.

[4] J. S. Lee, *"A Study on Defense and Attack Model for Cyber Command Control System based Cyber Kill Chain"*, Korea, society for internet information, Volume 22, No1, 2021.
DOI: http://dx.doi.org/10.7472/jksii.2021.22.1.41

[5] S. B. Park, "A Study on Defense and Attack Model for Cyber Command Control System based Cyber Kill Chain", Weekly technology trends, Special series, Volume 3. No3. www.iitp.kr

[6] K. W. Nam, *"A Study on Security operation Using Cyber KillChain (focusing on Heinrich's law)"*, Master's thesis, Dongguk university, 2020.

[7] Y. T. Oh, *"Data Modeling for Cyber Security of IoT in Artificial Intelligence Technology"*, Master's thesis, Pai chai university, 2021.

[8] J. H. Hong, *"Artificial Intelligence-based Security Control Construction and Countermeasures"*, The Korea Contents Association, Volume 21, No1, 2021. pp.531-540.
DOI: http://dx.doi.org/10.5392/JKCA.2021.21.01.531

이 광 형(Kwang-Hyoung Lee) [종신회원]

- 1998년 2월 : 광주대학교 컴퓨터공학과 졸업 (공학사)
- 2002년 2월 : 숭실대학교 컴퓨터공학과 (공학석사)
- 2005년 2원 : 숭실대학교 컴퓨터공학과 (공학박사)
- 2005년 3월 ~ 현재 : 서일대학 소프트웨어공학과 정교수

〈관심분야〉
영상인식, 네트워크보안, AI, ICT

정 용 훈(Yong-Hoon Jung) [종신회원]

- 2006년 8월 : 숭실대학교 일반대학원 컴퓨터학과 (공학석사)
- 2010년 2월 : 숭실대학교 일반대학원 컴퓨터학과 (공학박사)
- 2011년 3월 ~ 2014년 2월 : 서일대학교 조교수
- 2018년 8월 ~ 2021년 3월 : 바스랩 연구소장
- 2021년 4월 ~ 현재 : 유니허브랩 연구소장
- 2022년 6월 ~ 현재 : 투에이소프트 수석연구원

〈관심분야〉
블록체인, DID, 사용자 인증, 네트워크 보안, 융합 보안, AI