논문 2019-2-15 http://dx.doi.org/10.29056/jsav.2019.12.15

머신 러닝을 활용한 IDS 구축 방안 연구

강현선*†

A Study on the Establishment of the IDS Using Machine Learning

Hyun-Sun Kang*†

요 약

컴퓨팅 시스템들은 사이버공격에 대한 다양한 취약점을 가지고 있다. 특히 정보화 사회에서 지능화된 다양한 사이버공격은 사회적으로 심각한 문제와 경제적 손실을 초래한다. 전통적인 침입탐지시스템은 오용침입탐지(misuse)기반의 기술로 사이버공격을 정확하게 탐지하기 위해서는 지속적인 새로운 공격 패턴 갱신과 수많은 보안 장비에서 생성되는 방대한 양의 데이터에 대한 실시간 분석을 해야만 한다. 하지만 전통적인 보안 시스템은 실시간으로 탐지 및 분석을 통한 대응을 할 수 없기 때문에 침해 사고의 인지시점이 지체되어 많은 피해를 야기할 수도 있다. 따라서 머신 러닝과 빅데이터 분석 모델 기반으로 끊임없이 증가하는 사이버 보안 위협을 신속하게 탐지, 분석을 통한 대응과 예측할 수 있는 새로운 보안 시스템이 필요하다. 본 논문에서는 머신 러닝과 빅데이터 기술을 활용한 IDS 구축 방안을 제시한다.

Abstract

Computing systems have various vulnerabilities to cyber attacks. In particular, various cyber attacks that are intelligent in the information society have caused serious social problems and economic losses. Traditional security systems are based on misuse-based technology, which requires the continuous updating of new attack patterns and the real-time analysis of vast amounts of data generated by numerous security devices in order to accurately detect. However, traditional security systems are unable to respond through detection and analysis in real time, which can delay the recognition of intrusions and cause a lot of damage. Therefore, there is a need for a new security system that can quickly detect, analyze, and predict the ever-increasing cyber security threats based on machine learning and big data analysis models. In this paper, we present a IDS model that combines machine learning and big data technology.

한글키워드 : 침입탐지시스템, 기계학습, 사이버공격, 빅데이터, 보안

keywords : IDS, machine learning, cyber attack, big data, security

1. 서 론

점점 고도화되어가는 정보화 사회에서 사이버 공격은 정부 기관 및 기업이 직면한 가장 큰 위

* 남서울대학교 교양대학 교수
† 교신저자: 강현선(email: sshskang@nsu.ac.kr)
접수일자: 2019.11.30. 심사완료: 2019.12.16.
게재확정: 2019.12.20.

협으로 급부상 하고 있다. 사이버공격은 컴퓨팅 시스템의 보안 취약점을 악용하여 다수의 경로와 단계로 중요한 자산 유출 및 정보 시스템을 저해한다. 현재 사이버공격은 점점 지능화되고 새로운 공격 기법을 접목하고 있으며, 공격 대상과 목표도 다양해지고 있다. 이와 같은 사이버 공격으로부터 손실을 최소화하기 위해 많은 비용과 시간, 자원을 집중 투자하여 위협을 차단하고 있다. 일반적으로 사이버공격은 사전 계획에 의하여 다단계로 천천히 수행하기 때문에 침입이 발견되기 까지 많은 시간이 소요된다. 따라서 사이버공격 위협에 대한 신속한 가시성과 분석을 통한 예측으로 정확한 공격 위협을 발견하고 대응할 수 있는 보안 시스템이 필요하다. 침입탐지시스템(Intrusion Detection System, IDS)과 침입방지시스템(Intrusion Prevention System, IPS)은 방화벽 정책에 따라 악의적인 해킹, 악성코드, 스팸메일 등을 여과 없이 통과시키는 한계점을 보완하기 위한 보안 시스템이다. 본 논문에서는 머신 러닝을 접목한 IDS 구축 방안을 제안한다. 먼저 2장에서는 사이버 보안 위협에 따른 국내외 대응 동향과 기본적인 머신 러닝 기술을 설명한다. 3장에서는 전통적인 컴퓨팅 보안 시스템과 문제점을 살펴보고, 머신 러닝 기술을 활용한 IDS 구축 방안을 제시한다. 4장에서는 기존 보안 시스템과 제안 보안 시스템의 특징 및 성능을 비교하고, 마지막으로 5장에서 결론을 맺는다.

2. 관련 연구

2.1 사이버공격에 대한 대응 동향 분석

2019년 한국인터넷진흥원(KISA)의 사이버 보안위협 분석에 따르면 홈페이지 변조는 자기 과시 및 핵티비즘(hacktivism) 목적을 위해 해킹 공격이 지속될 것이며, 침해 사고는 해킹 공격 방법이 지능화 및 다양화됨에 따라 증가할 것으로 예상하고 있다. 또한 악성코드는 지능화되고 새로운 공격 기법을 접목하여 공격 목표와 대상이 다양화되고, 위협은 갈수록 커질 것으로 전망하고 있다. 또한 해당 공격 기법은 보안 취약 및 민감한 사회적 이슈 대상에 대하여 인공지능을 활용하여 침입탐지를 교묘하게 우회하는 지능화된 위협을 주요 보안 이슈로 전망하고 있다[1].

이와 같은 사이버 보안 위협을 다각적인 협력과 효과적인 대응을 위하여 대부분 국가들은 관련 법제도 정비와 관련기관을 설치하여 국가전반의 사이버 보안 수준을 강화하고 있다. 국내에서는 2015년 사이버 위협을 효과적으로 대처할 수 있도록 공공기관과 민간이 함께 사이버 위협을 조기 탐지하여 전파할 수 있는 체계 구축하기 위하여 “사이버 위협 정보 공유에 관한 법률안”을 근거로 인터넷진흥원과 사이버 안전 센터를 운영하고 있다. 인터넷 진흥원은 사이버 위협 정보 분석 공유 시스템(C-TAS, Cyber Threat Analysis & Sharing)을 구축하여 정보공유 참여 회원사들이 수집한 악성코드와 각종 사이버 위협 정보를 공유하고 있으며, 공공기관에서는 각 단위 기관들이 조직적으로 범국가적인 공동 대응체계를 구축하기 위하여 국가정보원 산하기관인 사이버 안전 센터(National Cyber Security Center)를 운영하고 있다. 미국은 2015년부터 국외 사이버 위협을 사고관련 분석 및 체계적으로 대응하고 정부 대책을 조율할 목적으로 국가정보국 내에 사이버 위협 정보 통합센터(Cyber Threat Intelligence Integration Center, CTIIC)을 설립하였으며, EU는 2004년에 유럽 네트워크 정보 보안청(European Network and Information Security Agency, ENISA)을 설립하여 네트워크 보안 기능 조정 역할 수행과 사이버 범죄 침해 사고를 효과적으로 지원하고 있다. 일본은 2014

년부터 사이버 보안 기본법 제정으로 사이버 보안 정책 수립과 전략을 관장하는 사이버 보안 센터(National center of Incident readiness and Strategy for Cybersecurity, NISC)을 신설하여 민관 협력을 강화하고 있다. 하지만 현재 운영 중인 정보보호 시스템만으로는 지능화되고 첨단화된 공격수법에 대한 실시간 분석, 대응 및 대책 마련에는 한계가 있다.

2.2 인공지능과 머신 러닝

인공지능(AI, Artificial Intelligence)은 인간처럼 사고하고 감지하고 행동하도록 설계된 일련의 알고리즘 체계라 할 수 있다. 인공지능은 1950년 초기에는 인간의 문제 해결을 컴퓨터 언어로 구현하고자 하는 시도가 주를 이루어졌으며, 5세대 컴퓨터의 등장과 다양한 분야의 데이터 축적으로 특정 분야 전문 지식을 학습시키기 위한 전문가 시스템을 활발하게 연구가 되었다. 컴퓨팅 기술 발달과 빅데이터가 등장하면서 인공지능은 머신 러닝(Machine Learning: ML) 학습을 통해 패턴을 찾아가는 방식으로 진화했다[2,3]. 머신 러닝을 활용하여 컴퓨팅 시스템은 다양하고 많은 데이터의 분석을 통해 최신 트렌드와 시장 동향 등과 같은 유용한 정보를 얻으려고 노력하고 있으며, 학습을 통한 데이터 기반 통찰력과 지식 습득을 가속화하여 비즈니스 성공과 경쟁 우위를 결정할 수 있다. 프로그래밍은 하나 이상의 알고리즘과 프로그래밍 언어를 이용하여 프로그램을 구현하여 어떠한 새로운 가치를 전달하기 위해서는 각각의 단계마다 사람이 프로그램 수정 및 추가 구현을 해야 하지만, 머신 러닝은 컴퓨팅 장치에 의해 다양하게 수집된 많은 양의 데이터를 가지고 사람이 직접 프로그램 및 분석할 필요 없이 자율학습 능력으로 데이터를 신속하고 효율적인 분석을 수행할 수 있기 때문에 잠재적인 통찰력을 제공한다. 머신 러닝은 인력 관리, 고객 관리, 예방관리, 예측시스템 등 다양한 분야에서 활용하고 있다. 컴퓨터 보안 분야에서는 패턴 인식과 비정상 탐지 속성 집합을 정의한 뒤 스팸(spam)을 탐지하기도 한다. 본 논문에서는 사이버공격에 대해 실시간 침입을 탐지 및 분석함으로써 대응과 예측이 가능할 수 있도록 보안 시스템에 머신 러닝을 활용한 IDS 구축 방안을 제안한다.

3. 본론

3.1 전통적인 컴퓨팅 보안 시스템의 문제점

컴퓨팅 보안은 컴퓨팅 시스템의 다양한 취약점을 공격자에게 이용되는 것을 방지하는 기술로서 기밀성(confidentiality), 무결성(integrity), 가용성(availability)의 세 가지 목표를 충족해야 한다. 일반적으로 컴퓨팅 시스템은 하드웨어, 소프트웨어, 데이터의 상호 작용으로 인한 다양한 보안 취약점이 존재한다. 악의적인 공격으로 정보시스템의 가용성 제한, 소프트웨어의 대체 및 삭제될 수 있으며, 특히 데이터 공격은 사회적으로 심각한 문제를 야기할 수 있다. 다음 그림 1은 시그니처 기반(signature-based) 전통적인 컴퓨팅 시스템 보안 구조를 나타낸다. 방화벽(firewall)은 특정 IP나 Port의 트래픽을 차단하여 사이버공격을 차단하고 사이버공격의 주요 침투 대상이기 때문에 내부 및 외부 네트워크 환경을 올바르게 구성되어야 한다. 웹 방화벽(web application firewall)은 웹 애플리케이션 보안에 특화되어 개발된 솔루션으로 SQL Injection, Cross-Site Scripting 등 웹 공격을 탐지하고 차단한다. 침입탐지시스템(IDS)과 침입방지시스템(IPS)은 방화벽 정책에 따라 악의적인 해킹, 악성

코드, 스팸메일 등을 여과 없이 통과시키는 한계점을 보완하기 위한 보안 시스템이다.

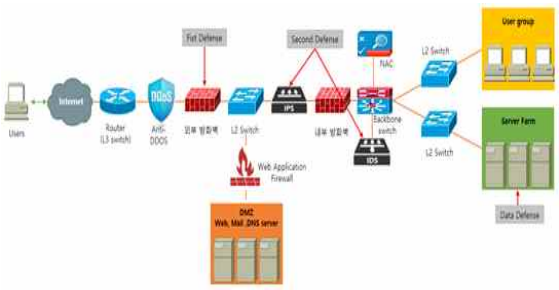

그림 1. 컴퓨팅 시스템 보안 계층 구조
Fig. 1. Computing System Security Hierarchy

3.1.1 침입탐지시스템(IDS)

IDS는 자료 수집, 자료 필터링 및 축약, 침입탐지, 책임 추적 및 대응의 기능으로 설치 위치와 목적에 따라 네트워크 기반 IDS(Network-based IDS: NIDS)와 호스트 기반 IDS(Host-based IDS: HIDS)로 구분한다. 1세대 NIDS는 순수 시그니처(signature) 기반 모델로 침입탐지를 위하여 센서를 사용하였고, 네트워크 속도의 진화로 센서가 패킷을 조사할 수 없는 심각한 성능적인 이슈가 발생하였다. 2세대 NIDS는 규칙적인 시그니처 집합을 비교하는 룰(rule) 기반으로 정확한 트래픽 처리와 탐지를 위하여 IP-defragmentation, TCP 재결합, 정규화(normalization)등 다양한 기술을 적용하였다. 하지만 패킷 시그니처 탐지를 위하여 플로우 안의 모든 패킷 바이트 정보를 조사하기 때문에 성능의 심각한 저하와 부정확한 탐지 가능성이 높아졌다. 제3세대 NIDS는 이상 침입탐지(anomaly)를 사용하여 공격 가능성 있는 네트워크 프로토콜들의 대상만 조사함으로써 성능 개선과 탐지 정확도를 개선하였다. 최근 NIDS는 순수한 통계적 분석을 사용하여 통상적인 통신 패턴을 학습하고 비정상에 대하여 공격을 탐지한다. 하이브리드 방식은 시그니처 기반, 규칙 기반, 프로토콜 이상 침입탐지 등과 같은 방법을 혼용하여 부정확한 탐지를 줄이고 있다. HIDS는 호스트 상에서 발생하는 이벤트에 따라 공격을 탐지한다. 초창기 HIDS는 공격을 탐지하기 위하여 “파일 감시” 개념으로 서버의 중요한 시스템 파일 변경을 감시하는 방식이었다. 현재의 HIDS는 보안 관리자로 하여금 시스템과 자원 사용에 대하여 엄격한 정책을 설정하여 HIDS 에이전트의 권한이 없는 사용자를 운영체제 시스템 레지스트리(registry)와 이벤트 로그 변경을 룰 셋(rule set)으로 비교 감시한다. 그러나 이 방법은 주기적으로 보안 관리자가 중요 서버의 HIDS 에이전트에 룰 셋을 구성해야 하기 때문에 항상 침입이 발생 후 사후 조치를 해야 한다. 따라서 현재 운영 중인 IDS는 지능화되고 다양해진 사이버공격에 대한 실시간 탐지 및 분석 그리고 대응 및 예측에 어려움이 있다.

3.1.2 침입방지시스템(IPS)

침입방지시스템(IPS)도 IDS와 마찬가지로 호스트 기반과 네트워크 기반 시스템으로 구분한다. 호스트 기반 IPS(Host-based IPS: HIPS)는 우선 소프트웨어 제품으로 취약한 응용 프로그램을 보호하는 시그니처 행위 기반 분석 알고리즘을 이용하며, 특정 규칙에 위배되는 이벤트 필터링(filtering)방식과 접근제어 기능의 트러스트(trust) 운영체제 방식이 있다. NIDS는 네트워크 트래픽을 탐지 기능만 제공하지만 NIPS는 공격 탐지를 위하여 원하지 않는 트래픽을 차단할 수 있는 인라인 장치이다. IDS와 IPS를 구분할 수 있는 핵심 요소는 자동 차단(automatic blocking)과 인라인 위치(inline position)라 할 수 있다. 따라서 NIPS는 인라인 솔루션으로 성능, 네트워크 재설계, 가용성에 대한 네트워크 사항을 충분히 고려해야 한다. 만약 NIPS가 정상적으로 동작하

지 않을 경우 네트워크 트래픽은 계속 통과하지만 보안은 상실된다. 따라서 IPS는 반드시 데이터 트래픽 경로 상에 인라인 운영해야 하기 때문에 장비에 대한 가용성 및 신뢰성이 필요하다. IPS도 IDS와 마찬가지로 주기적으로 보안 관리자가 침입을 방지하기 위하여 미리 정해진 규칙 목록을 시스템에 구성해야 하며, 알려지지 않은 침입에 대해서는 발생 후 사후 조치를 해야 하기 때문에 사이버공격에 대한 실시간 탐지 및 분석 그리고 대응 및 예측에 어려움이 있다.

3.2 머신 러닝을 활용한 IDS 구축 방안

전통적인 침입탐지 기술은 일반적으로 다음 그림 2와 같이 동작한다. 먼저 운영 중인 네트워크 및 서버에서 수집한 데이터(raw data)를 침입 판정이 가능하도록 정보의 가공 및 축약을 실시한다. 침입탐지 시스템의 핵심 단계인 분석 및 침입탐지에서는 수집된 데이터를 잘 가공하여 의미 있는 정보를 분석하여 침입 여부를 판정하고, 최종적으로 관리자에게 보고하여 최종 조치를 취한다. 이와 같은 시그니처 기반(signature based)과 이상 기반(anomaly based) 침입탐지 기술은 특정 공격의 유형에 대한 침입탐지 비교를 위해 데이터베이스와 설정 기준이 주기적으로 관리 및 유지되어야 하기 때문에 실시간 탐지 및 분석이 어렵다.

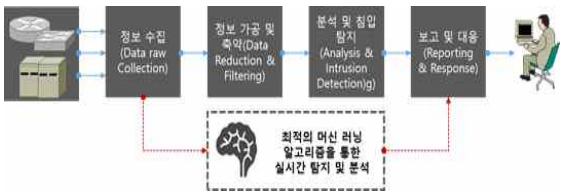

그림 2. 전통적인 침입탐지시스템 기술적 구성 요소
Fig. 2. Technical components using traditional intrusion detection system

본 논문의 머신 러닝 알고리즘 활용한 시스템에서는 정보수집 단계부터 분석된 침입탐지에 대한 보고 및 대응 단계를 최적의 머신 러닝 알고리즘을 통해 학습하고 실시간으로 탐지하고 분석한다. 다음의 그림 3은 머신 러닝 알고리즘을 활용한 실시간 탐지 및 분석 그리고 예측을 위한 IDS 구성도를 나타낸다. 해당 시스템에서는 과거의 공격 및 지속적으로 증가하는 사이버공격에 대한 대용량 정보인 빅데이터에 대해 분석과 학습을 함으로써, 비정상적인 트래픽을 탐지 또는 차단하여 사이버 보안 위협을 감소시킬 수 있다. 머신 러닝을 성공적으로 적용하기 위해서는 빅데이터와 관련해 대용량 데이터 보관 및 분석 기술과 머신 러닝의 정교한 알고리즘 실행을 위한 고성능 및 가용성의 컴퓨터가 필요하다. 다음 절에서는 머신 러닝을 활용한 IDS 적용 시 필요한 최적의 알고리즘 모델 선정 방법과 최적의 머신 러닝 알고리즘 모델에 대한 평가와 선정 방안에 활용할 수 있는 방법을 소개한다.

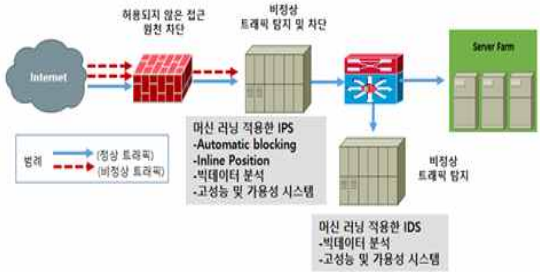

그림 3. 머신 러닝 활용한 IDS
Fig. 3. IDS using Machine Learning

머신 러닝은 다음의 그림 4와 같이 데이터 수집(data acquisition), 처리(data processing), 모델링(model engineering), 머신 러닝 루틴 실행(execution), 결과 배포(deployment) 단계로 진행된다. 데이터 수집은 다양한 컴퓨터 장치에서 많

은 양의 데이터가 수집되기 때문에 머신 러닝 데이터 처리 플랫폼의 신뢰성이 매우 중요한 부분이다. 수집한 데이터 필드 소스에서 가능한 한 많은 데이터 요소들을 추출하는 것이 필요하다. IDS 시스템을 위한 데이터 수집은 주로 네트워크 및 네트워크 시스템에서 수집된다. 데이터 처리는 머신 러닝 실행의 데이터 준비 과정으로 사전 통합 및 처리를 위한 데이터 변환 및 정규화, 인코딩(encoding)을 수행하는 모듈을 포함하며, 알고리즘 훈련을 위한 데이터 세트 표본을 선택한다. 또한 수집된 데이터는 중복되거나 부적절한 기능이 포함될 수 있기 때문에 기능 분석(feature analysis)을 실시하여 학습 시간을 줄이고 모델을 단순화 한다.

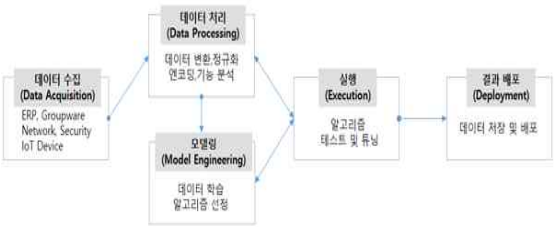

그림 4. 머신 러닝 알고리즘
Fig. 4. Machine Learning Algorithm

데이터 모델링은 실행 단계에서 발생할 수 있는 문제를 해결하기 위해 데이터 학습 알고리즘을 선택하고 조정하는 부분이다. 머신 러닝 학습 유형은 지도 학습(supervised learning), 비지도 학습(unsupervised learning), 강화 학습(reinforcement learning)의 세 가지 형태로 구분하며, 상황에 따라 특정 과제를 해결하기 위해 다른 범주의 기법을 조합해 원하는 결과를 얻기도 한다. 지도 학습은 목표 변수(target variable)와 변수 관계를 학습하도록 가르치는 것이며, 비지도 학습은 지도나 목표 변수를 설정과정 없이 알고리즘이 스스로 학습을 하는 것이다. 강화 학습은 머신 러닝 에이전트가 환경의 피드백으로 학습하는 것으로 보편적 머신 러닝 알고리즘보다 인공지능 기법에 더 가깝다. 변수의 개수가 너무 많은 경우에는 차원을 줄이기 위해 비지도 학습을 먼저 수행하다가 지도학습으로 바꾸기도 하며, 일부 인공 지능 애플리케이션에서는 지도 학습과 강화 학습을 조합하여 문제를 해결하기도 한다[2,3,4]. 머신 러닝을 활용한 IDS 보안 시스템에는 보안 및 침입탐지를 위해 비지도 학습 형태와 다양한 입력 또는 변수 간의 관계를 판별하고 예측하는데 사용되는 연관 규칙 학습 알고리즘의 적용이 적합하다. 실행 단계는 준비된 학습 데이터와 알고리즘 모델링이 완료되면 머신 러닝 루틴이 실행된다. 또한 의사 결정을 위해 알고리즘 테스트 및 튜닝을 통해 알고리즘 성능을 최적화한다. 이 과정에서 생성된 머신 러닝 알고리즘 학습 모델은 다양한 데이터에서 제대로 동작하는지에 대한 일반적인 평가 기준이 필요하다. 머신 러닝 알고리즘 모델 평가와 최종 모델 선정에 활용할 수 있는 방법이 제안되었다[5,6,7,8,9,10]. 다양한 형태의 장비에서 수집된 데이터 세트는 계층적 샘플링을 통해 모델 지도 학습을 위한 훈련 세트(training dataset)와 모델 선택을 위한 검증 세트(validation dataset), 최종 선택된 모델을 평가하기 위한 테스트 세트(testing dataset)로 세분화하여 머신 러닝 알고리즘 최적 평가 및 모델 선정을 위하여 사용한다. 훈련 데이터는 학습 알고리즘을 사용하여 알고리즘의 하이퍼파라미터(hyper parameter) 튜닝(tuning)을 통해 모델을 학습한다. 검증 데이터 세트는 최적의 하이퍼파라미터 설정하여 상호 모델 성능 비교를 통해 가장 성능이 좋은 모델을 선택하여 테스트 데이터로 최종 모델을 선정한다. 훈련 데이터로 사용한 적이 없는 테스트 데이터 세트는 일반화 정확도를 추정할 때 어떠한 편향을 만들지 않도록 한번만 사용해야한다. 모델 평가와 모델 선택을 위해

가장 널리 사용하는 K겹 교차 검증(k-Fold Cross-Validation)기법은 학습 알고리즘의 훈련과 테스트를 위해 수집된 모든 데이터 샘플을 훈련과 검증 단계를 연속적으로 교차함으로써 비관적 편향을 감소시키고, 반복적인 하이퍼파라미터 설정으로 알고리즘을 학습을 통한 모델을 평가한다[5,6,7,8]. 마지막으로 위와 같은 과정을 거친 머신 러닝 실행 결과는 IDS 시스템에 적용되어 실시간 탐지 및 대응에 활용하게 된다.

4. 성능 비교

이번 장에서는 전통적인 보안 시스템 방식과 본 논문에서 제시한 머신 러닝 알고리즘을 활용한 방식을 간단히 비교 분석한다. 다음의 표 1은 현재 운영 중인 전통적인 보안시스템 방식과 머신 러닝 알고리즘 활용한 보안 시스템을 비교한 것이다. 기존 방식은 시그니처 기반의 오용탐지 기술이 핵심 기술이고, 제안 방식은 머신 러닝 학습 알고리즘과 빅데이터 분석이 핵심기술이다.

기존 방식은 사이버 공격에 대해 지속적인 탐지와 알려지지 않은 신규 또는 변종 위협에 대한 탐지, 실시간 분석이 취약한 반면 제안 방식은 빅데이터 및 머신 러닝 알고리즘을 통한 실시간 탐지 및 실시간 분석이 우수하다. 즉, 제안 방식은 빅데이터 분석 및 머신 러닝 알고리즘을 통한 실시간 탐지 및 분석을 함으로써 침입에 대한 대응 및 예측이 가능하게 된다.

표 1. 전통적인 보안 시스템과 머신 러닝 알고리즘 활용한 보안 시스템 비교

Table 1. Comparing traditional and machine learning algorithm security systems

구분	전통적인 방식	머신 러닝 방식
정의	시그니처 기반 오용 침입탐지 기술로 사이버공격에 대한 축적된 지식을 바탕으로 패턴을 설정하여 가공된 데이터와 비교하는 방식	패턴 인식과 비정상 탐지 속성 집합을 정의하여 알고리즘 학습과정을 통해 데이터 속성 인식
핵심 기술	시그니처 기반 오용 침입탐지 기술	머신 러닝 학습 알고리즘, 빅데이터 분석
필요 시스템	방화벽, IDS	머신 러닝 알고리즘 활용한 IDS
시스템 성능	시스템 log을 취합할 수 있는 기본적인 시스템 성능	빅데이터 분석 및 알고리즘 실행을 위한 고성능 및 가용성 시스템
탐지 능력	사이버 공격에 대해 지속적 탐지 불가	빅데이터 분석, 머신 러닝 알고리즘을 통한 실시간 탐지 가능

5. 결 론

사이버 공격이 점점 지능화되고 새로운 공격 기법을 접목함에 따라 전통적인 보안 시스템은 실시간 탐지 및 분석에 한계가 있다. 머신 러닝 알고리즘은 과거의 공격 및 지속적으로 증가하는 사이버공격에 대한 다양한 정보의 빅데이터 분석과 학습을 통하여 실시간 탐지 및 분석 그리고 대응 및 예측으로 보안 인프라에서 필요로 하는 가시성을 제공하여 효과적으로 보안을 향상 시킬 수 있을 것으로 전망한다. 추후 본 연구에서 제시한 머신 러닝을 활용한 IDS 보안 시스템 구축을 토대로 실제 구체적인 구현 방법론과 구현이 보완되어야 할 것이며, 이를 통해 사이버공격에 대해 능동적인 대응에 활용될 것을 기대한다.

이 논문은 2019년도 남서울대학교 학술연구비 지원에 의해 연구되었음.

참 고 문 헌

[1] Korea Internet & Security Agency (KISA), “Cyber Security Threat Trend Report”, 2019, https://www.boho.or.kr

[2] Shai Ben-David, Shai Shalev-Shwartz, “Understanding Machine Learning: From Theory to Algorithms”, Cambridge University Press, 2014. https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf

[3] Andreas C. Müller, Sarah Guido, “Introduction to Machine Learning with Python: A Guide for Data Scientists”, O'Reilly Media, Inc., 2016. http://index-of.es/Varios-2/Introduction%20to%20Machine%20Learning%20with%20Python.pdf

[4] Giuseppe Bonaccorso, “Machine Learning Algorithms: A reference guide to popular algorithms for data science and machine learning”, 2007. https://dl.acm.org/citation.cfm?id=3165154

[5] Yoshua Bengio, Yves Grandvalet, “No Unbiased Estimator of the Variance of K-Fold Cross-Validation”, The Journal of Machine Learning Research, vol.5, pp.1089 - 1105, Dec. 2004. https://dl.acm.org/citation.cfm?id=1044695

[6] Leo Breiman, “Heuristics of Instability and Stabilization in Model Selection”, The Annals of Statistics, vol.24, no.6, pp.2350-2383, 1996. https://projecteuclid.org/download/pdf_1/euclid.aos/1032181158

[7] Ji-Hyun Kim, “Estimating Classification Error Rate: Repeated Cross-Validation, Repeated Hold-out and Bootstrap”, Computational Statistics & Data Analysis, vol.53, pp.3735 - 3745, Sep. 2009. https://www.sciencedirect.com/science/article/pii/S0167947309001601

[8] Gartner, Analyst(s): Carlton E. Sapp, “Preparing and Architecting for Machine Learning”, Jan. 2017. https://www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/preparing_and_architecting_for_machine_learning.pdf

[9] Hawkins, Douglas M., Subhash C. Basak, Denise Mills, “Assessing Model Fit by Cross-Validation.”, Journal of Chemical Information and Computer Sciences, pp.579 - 586, 2003. https://www.ncbi.nlm.nih.gov/pubmed/12653524

[10] Kohavi, Ron, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection.”, International Joint Conference on Artificial Intelligence, pp.1137 - 1143, 1995. https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf

저 자 소 개

강현선(Hyun-Sun Kang)

2007.2 단국대학교 전자계산학과 박사
2007.3-2009.2 단국대학교 강의전임교수
2010.9-현재 : 남서울대학교 교수
<주관심분야> 통신프로토콜, 물리적 보안