J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr

웹쉘 수집 및 분석을 통한 머신러닝기반 방어시스템 제안 연구

A study on machine learning-based defense system proposal through web shell collection and analysis

김 기 환[1] 신 용 태[2*]
Ki-hwan Kim Yong-tae Shin

요 약

최근 정보통신 인프라의 발달로 인터넷접속 디바이스가 급속하게 늘어나고 있는 실정이다. 스마트폰, 노트북, 컴퓨터, IoT디바이스까지 인터넷접속을 통하여 정보통신서비스를 받고 있는 것이다. 디바이스 운영환경이 대부분이 웹(WEB)으로 이루어져 있는 관계로 웹쉘을 이용한 웹사이버 공격에 취약하다. 웹쉘이 웹 서버에 업로드 될 경우 웹 서버의 제어가 손쉽게 이루어 질 수 있어서 공격빈도가 높은 것으로 확인된다. 웹쉘로 인한 피해가 많이 발생하면서 각 기업에서는 침입차단시스템, 방화벽, 웹방화벽등 다양한 보안장비로 공격에 대응하고 있지만, 현재 출시되는 대부분의 웹쉘 대응 장비는 패턴 기반으로 탐지가 이루어지기 때문에 웹쉘 변종에 있어서는 탐지가 어려우며 이런 특성으로 웹쉘 공격의 예방 및 대처하기 위해서는 기존의 체계와 보안소프트웨어만 가지고 대응 하기에는 힘든 상황이 현실이다. 이에 인공지능 머신러닝 과 딥러닝기법을 활용하여 알려지지 않은 웹쉘을 사전에 탐지하는 등 신규 사이버 공격에 대하여 대처 할 수 있는 인공지능 머신러닝 기반의 웹쉘 수집 및 분석을 통하여 자동화된 웹쉘 방어시스템에 대하여 제안하고자 한다. 본 논문에서 제안하는 머시러닝기반 웹쉘 방어시스템 모델은 웹환경에 대한 사이버공격중의 하나인 악성 웹쉘에 대하여 수집, 분석, 탐지를 빠르게 하여,안전한 인터넷환경구축 및 운영시 필수적으로 적용이 필요한 웹정보보안 시스템 설계,구축에 많은 도움이 될 것으로 생각한다.

☞ 주제어 : 웹서비스, 웹쉘공격, 머신러닝, 웹쉘수집 및 분석, 방어시스템

ABSTRACT

Recently, with the development of information and communication infrastructure, the number of Internet access devices is rapidly increasing. Smartphones, laptops, computers, and even IoT devices are receiving information and communication services through Internet access. Since most of the device operating environment consists of web (WEB), it is vulnerable to web cyber attacks using web shells. When the web shell is uploaded to the web server, it is confirmed that the attack frequency is high because the control of the web server can be easily performed. As the damage caused by the web shell occurs a lot, each company is responding to attacks with various security devices such as intrusion prevention systems, firewalls, and web firewalls. In this case, it is difficult to detect, and in order to prevent and cope with web shell attacks due to these characteristics, it is difficult to respond only with the existing system and security software. Therefore, it is an automated defense system through the collection and analysis of web shells based on artificial intelligence machine learning that can cope with new cyber attacks such as detecting unknown web shells in advance by using artificial intelligence machine learning and deep learning techniques in existing security software. We would like to propose about. The machine learning-based web shell defense system model proposed in this paper quickly collects, analyzes, and detects malicious web shells, one of the cyberattacks on the web environment. I think it will be very helpful in designing and building a security system.

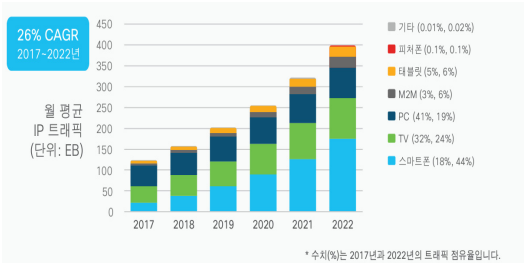
☞ keyword : Web service, WebShell attack, Machine learning, WebShell collection and analysis, Defense System

1. 서 론

최근 정보통신 인프라의 발달로 인터넷접속 디바이스가 급속하게 늘어나고 있는 실정이다. 스마트폰, 노트북, 컴퓨터, IoT디바이스까지 인터넷접속을 통하여 정보통신서비스를 받고 있는 것이다. 그림 1과 같이 전 세계 디바이스와 회선 수가 인구(CAGR 1%)와 인터넷 사용자 수(CAGR 7%)보다 빠른 속도로 증가(CAGR 10%)하고 있으며,이러한 추세가 가구당 그리고 일인당 디바이스 및 회선의 평균 개수 증가를 가속화하고 있다. 매년 향상된 기능과 지능을 갖춘 다양한 폼 팩터의 여러 가지 디바이스

1 Corporate Strategy Team, ETRI, Daejeon, 34129, Korea.
2 Spartan SW Education Center, Soongsil University, Seoul, 06978, Korea.
* Corresponding author (shin@ssu.ac.kr)
[Received 30 June 2022, Reviewed 9 July 2022(R2 11 August 2022, Accepted 19 August 2022]

가 새로 출시되어 보급되고 있다. M2M 활용 분야(예: 스마트 미터, CCTV, 의료 모니터링, 수송 및 포장, 자산 추적)의 다양화가 디바이스와 회선 증가의 주요 원인으로 손꼽히며,2022년에 M2M 회선이 전체 디바이스 및 회선의 51%를 차지할 것으로 전망된다.



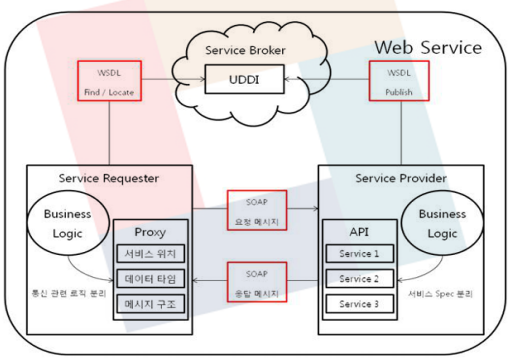
(그림 1) 전 세계 디바이스 및 회선 증가〔1〕
(Figure 1) Increase in Devices and Lines Worldwide〔1〕

M2M 회선은 2022년까지 2.4배(146억개) 가까이 증가하면서 예측 기간 동안 가장 급격한 상승 곡선(CAGR 19%)을 그릴 것으로 예상된다. 뒤를 이어 스마트폰이 9%의 CAGR로 1.6배의 증가율을 보일 전망이다. 그리고 (평면 TV, 셋톱 박스, 디지털 미디어 어댑터 (DMA), 블루레이 디스크 플레이어, 게임 콘솔로 구성되는) 스마트 TV는 7%의 CAGR를 기록하면서 32억 대까지 증가할 것으로 추산된다. 한편 PC는 예측 기간 동안 꾸준히 하락(2.5%)할 전망이다. 그러나 예측 기간 동안 태블릿이 PC를 앞지르는 일은 없을 것으로 보인다(PC 12억 대, 태블릿 7억 9천만 대).유선 및 모바일 디바이스를 포함한 모든 디바이스의 2022년 점유율은 소비자가 72%를, 비즈니스 부문은 나머지 28%를 차지할 것으로 예상된다. 소비자 점유율은 12.0%의 CAGR를 기록한 비즈니스 부문에 비해 다소 완만한 증가세(CAGR 8.8%)를 보일 것으로 예측된다.[1]

위와 같이 인터넷에 접속하는 디바이스의 운영환경 및 프로그램은 웹(WEB)프로그램과 웹(WEB)애플리케이션 환경에서 동작되고 있는 것이다.

웹 프로그램과 웹 어플리케이션은 웹(web) 기반의 응용 소프트웨어를 의미하는 용어로 이 두 가지 단어는 동일하다. 웹은 네트워크의 한 종류로HTTP 프로토콜에 의한 통신을 기반으로 한다. 미국의 웹 사이트 내용을 웹브라우저를 통해서 볼 수 있는 것은 요청한 페이지를 미국의 웹 서버로부터HTTP 통신을 통해 전달 받기 때문이다. 웹 어플리케이션은 웹 환경에서 이용할 수 있는 다양한 기능을 제공하는 프로그램으로 게시판, SNS, 인터넷쇼핑몰 등 매우 다양하다. 이러한 어플리케이션은 단독으로도 정상적인 실행이 가능하고, 다른 어플리케이션과 결합되어 좀 더 큰 프로그램이 될 수도있다.

웹 어플리케이션은 클라이언트(Client)의 요청과 이에 대한 서버(Server)의 응답으로 구성된다. 좀더 상세하게 알아보면 그림 2 와 같이 Service Broker는 서비스 등록 및 검색, 저장, 관리를 통해 Service Provider가 제공하는 서비스를 Service Consumer에게 연결한다. Service Requester는 Service Provider에게 해당되는 서비스를 검색하여 Service Provider의 서비스를 요청하는 주체이다. Service Provider는 웹 서비스를 구현하여 제공하는 주체이다.[2]



(그림 2) 웹 서비스 구성도〔2〕
(Figure 2) Web service configuration diagram〔2〕

2022년 상반기 사이버 위협 동향은 랜섬웨어 갱단의 활발한 활동과 가상자산의 공격 피해로 요약할 수 있다.

랩서스(Lapsus$)는 가장 활발하게 활동한 랜섬웨어 갱단이다. 이들은 2021년 12월에 브라질 보건부 해킹을 시작으로 올해 마이크로소프트, 엔비디아 등 세계 유수의 기업뿐 아니라 옥타(Okta)와 같은 글로벌 보안전문기업을 해킹했다. 블랙캣(BlackCat)은 이탈리아 패션 브랜드 몽클레르(Moncler)와 스위스 항공서비스 기업 스위스포트(Swissport)를 공격하면서 2022년 상반기 새로 모습을 드러낸 랜섬웨어 갱단이다. 러시아에 기반을 둔 콘티(Conti)와 락빗(Lockbit)2.0 랜섬웨어 갱단은 작년에 이어 올해도 악명을 떨쳤다. 기업과 공공을 가리지 않고 공격하는 그들은 코스타리카와 페루의 정부기관, 캐나다 민간 군사훈련 기업 등으로 공격 대상을 확대했다. 2월에 시작한 러시아-우크라아나 전쟁은 물리 공간과 함께 사이버 공간에서도 전투가 벌어지는 하이브리드 전쟁이 됐다. 러시아는 침공 이전부

터 악성코드 배포, DDoS 공격 등 사이버 공격을 했고, 침공 이후에도 군사공격을 하기 전후로 사이버 공격을 적극 활용했다. 또한, 사이버전에 양쪽을 지지하는 해커그룹이 참여하면서 다른나라와 민간기업에 대한 공격으로 확산되는 양상을 보였다.[3]

사이버공격은 국내 및 국외를 가리지 않고 지속적으로 일어나고 있다. 그중에서 악성코드 공격을 통한 공격이 증가하고 있는 것이 현실이다. 특히, 웹사이트의 취약점을 이용한 공격이 증가하고 있다.

웹 기반 환경의 특성 및 업무의 특성을 고려하지 않고 설계를 진행하는 경우, 다양한 보안상 취약점이 실제 시스템 운영 단계에서 드러나게 되고, 이는 시스템 유지보수 비용의 증가를 초래하게 된다. 그러므로, 웹 기반 환경으로의 전환 시에는, 시스템 안전성 확보 및 보안 관련 유지보수 비용 절감을 위해 설계 단계에서부터 반드시 보안을 고려한 설계가 진행되어야 한다.[4]

웹기반 환경에서는 설계뿐만 아니라 구축후 운영시에도 정기적으로 보안취약점 점검이 필요하다.

자유게시판, 민원게시판 등과 같이 파일 업로드가 가능한 게시판의 파일첨부 기능에 대한 확장자 필터링(PHP, JSP, ASP 등 웹서버에서 실행 가능한 파일의 업로드를 제한)이 제대로 이루어지지 않았을 경우 공격자가 파일첨부 기능을 이용하여 해킹도구(웹쉘)를 업로드 하고 설치하여 해당 시스템을 제어할 수 있는 명령어의 실행, 내부 중요자료 접근 및 악성코드 등을 유포할 수 있는 취약점이 존재한다.[5]

웹쉘은 해커가 원격으로 웹 서버에 명령을 내릴 수 있도록 작성한 웹스크립트 파일이다. 해커는 웹쉘을 이용하여 보안시스템을 우회, 시스템에 접근하여 파일 수정, 복사, 삭제 등의 시스템 제어를 할 수 있고 웹 소스코드에 악성코드를 설치해 사용자들의 PC를 공격하거나 연결된 데이터베이스의 정보를 유출하는 등 큰 피해를 입힐 수있다.[6]

웹쉘은 웹서버 시스템에 업로드 되어 동작되며, 일반적으로 운영 되고 있는 네트워크 보안장비나 솔루션에서 탐지나 차단 등의 대응이 어렵다. 최근에는 웹쉘 탐지 전용 솔루션들이 개발 출시되어 운영되고 있으나, 탐지된 내역에 대한 분석이나 처리에 한계를 보이고 있다. 이로 인해 운영의 비효율성, 운영 안정화의 장기화 등의 문제점들이 발생하고 있다.[7]

웹쉘 방어를 위하여 어떠한 방안이 있는지 아래 와 같이 고려해 보아야 할 사안이 있다. 웹쉘 방어에 효과적인 대응을 위해서는 우선, 웹서버에 파일 업로드하는 것을 효과적으로 제어할 수 있어야 한다. 이를 위해 적절한 기능을 갖춘 웹 방화벽을 도입하고 웹서버에서 업로드 된 파일들이 실행권한이 없도록 웹서버 환경을 설정하는 것이 중요하다. 2차적으로는 웹쉘의 이용을 탐지하는 방법을 갖추는 것이 필요하다. 해커들은 흔히 시스템에 대한 제어권을 얻기 위해 웹쉘을 통하여 특정 시스템 커맨드를 이용하거나, 상위 디렉토리 또는 특정 디렉토리에 접근을 한다. 이때 위의 명령어들이 웹쉘에 전달되어야 하므로 이러한 명령어들을 탐지함으로써 웹쉘의 이용을 차단하고 웹쉘의 존재 여부를 판단할 수 있다.

위와 같은 웹쉘 공격 자체를 차단하는 방법 외에 정적 분석도구(Static Analysis Tools)을 이용하는 방법도 효과적이다. 웹쉘로 활용될 수 있는 웹소스를 검색하거나 악의적으로 수정된 웹페이지 등을 주기적 또는 실시간으로 모니터링 함으로써 웹쉘 공격을 차단할 수 있다.[8]

전체 침해사고 유형 중 65%가 웹쉘로 인한 피해로 파악되고 있으며, 피해를 입은 웹서버에서 웹쉘이 발견되는 경우는 90%에 달한다. 이에 보다 효과적인 웹쉘 탐지를 위해서는 인공지능 기술을 활용하여 웹쉘을 분석하고 탐지하는 시스템의 연구가 필요하다. 이를 활용할 경우 웹쉘 자체의 특징을 파악하여 이를 기준으로 웹쉘에 대한 분석을 수행함으로 해당 작업에 대한 인력 및 자원의 효율성을 향상시킬 수 있을 것으로 사료된다. 웹쉘이 갖는 잠재적인 위협과 막대한 피해규모에도 불구하고 웹쉘 방지에 탐지가 쉽지않은 상황이다. 이러한 위협에 대한 실제적인 방안으로는 현존하는 데이터에 대한 특징 뿐 아니라, 웹쉘 자체의 대한 특징을 학습함으로 웹쉘을 탐지해낼 수 있는 인공지능 기반의 접근법을 활용하는 것이며, 이를 통해 기존 수동으로 탐지해내던 웹쉘에 대한 탐지를 자동화 할 경우 웹쉘 탐지에 대한 성능을 대폭 향상될 수 있다고 판단된다.[9]

본 연구에서는 웹쉘탐지를 자동화 하여 웹사이트를 통한 웹쉘 공격을 방어 할 수있도록 웹쉘수집 및 분석을 머신러닝기법을 통하여 방어하는 방안을 제시한다.

본 논문의 구성은 2장에서 관련 연구를 소개하고, 3장에서 AI 머신러닝기반 자동환된 웹쉘 수집 및 분석을 통한 웹쉘 방어시스템 모델을 설명한다. 마지막으로 4장에서는 전반적인 결론 및 향후연구과제에 대하여 요약한다.

2. 인공지능(AI) 머신러닝 연구 사례

2.1 해외 사례

인공지능 머신러닝의 학습 및 추론 기술은 데이터에

내제된 패턴, 규칙, 의미 등을 알고리즘 기반으로 스스로 학습하게 하여 새롭게 입력되는 데이터에 대한 결과를 예측 가능하도 록 하는 기술이다., 이러한 인공지능의 학습 및 추론 기술을 활용하여 각 분야에서 기술혁신을 통하여 , 기업의 생산성 및 공공부문의 서비스를 획기적으로 바꿀 수 있는 기술이므로, 국내뿐만 아니라 해외 IT 기업들은 인공지능 기술 개발에 사활 을 걸고 있는 실정이다. 인공지능 기술 분야에 대해 WIPSON 특허 데이터베이스를 사용하여 2017년 3월 까지의 한국, 미국, 일본, 유럽, 중국 특허청에서 등록 및 공개된 특허를 조사 및 분석대상으로 연구한 논문에 따르면, 특허출원 현황은 표 1과 같다.[10]

(표 1) 국가별 인공지능 관련 특허출원〔10〕
(Table 1) Patent application related to artificial intelligence by country〔10〕

나라	건수	%
미국	4,860건	46%
일본	2,386건	23%
한국	1,398건	13%
중국	1,342건	13%
유럽	564건	5%

아래 표2 인공지능 머신러닝 알고리즘 기술을 활용하여 정보보호분야 중 정보보안관제의 경우, 상관관계 분석이 탐지에 있어 가장 중요한 역할을 담당하게 된다. 수집되는 로그를 얼마나 빠르게 실시간으로 분석하는지, 얼마나 다양한 인텔리전스 정보를 통해 탐지하는지가 탐지 효율성을 결정하게 된다.[11] 추가적으로, 악성코드탐지및분석. 웹쉘 탐지 및 분석등에 적용 될 수 있도록 연구가 진행되고 있는 실정이다.

(표 2) 인공지능 머신러닝 알고리즘 종류〔11〕
(Table 2) Types of artificial intelligence machine learning algorithms〔11〕

	자율 학습 (Unsupervised)	지도 학습 (Supervised)
연속 (Continuous)	밀도 추정 (Density estimation) • Expectation Maximization • Parzen Window	회귀 (Regression) • 결정 트리 • Random Forests • Boosting Trees • Neural Networks • Support Vector Regression
범주 (Clustering)	군집화 (Clustering) • K-Means • DBSCAN	분류 (Classification) • Naive-Bayes • K-Nearest Neighbors • Logistic Regression • Support Vector Machine • Trees

2.2 국내 사례

표 3과 같이 정보보안부문에서 웹쉘 탐지 및 방어 시스템을 연구개발하고 있는 국내 사례 중 첫 번째 사례는 에이쓰리시큐리티社의이지스쉘모니터(EGISShell Monitor)이다. 이지스쉘모니터는 최근 웹 서버 해킹에 사용되는 악성 프로그램인 웹쉘을 실시간 모니터링 하고 탐지하는 신개념 웹 보안 관제서비스로 웹쉘을 방어하는 웹서버 단의 최종 방어수단이다. 두번째 사례는 SecureAT社의 쉘캅(shelcop)이다. SecureAT사의 ShelCop은 해커가 웹서버의 취약점을 이용해 웹서버 관리자 계정을 획득, 저장된 내부 자료의 유출 및 백도어 프로그램 설치 등 악의적인 목적으로 작성 후 웹서버에 설치한 웹쉘 프로그램을 탐지하고 발견 시, 실행을 방지하는 솔루션이다. 쉘캅은 웹서버의 해킹을 방지하고 해킹발생 후, 정보 유출과 해킹의 거점으로 사용되는 2차 피해를 방지한다. 현재 SGA와 시큐브社에서 공급하고 있다. 세번째 사례는 하로스社의 쉘 가드(Shell Guard)이다. 하로스社의 쉘가드는 웹쉘탐지솔루션인 '휘슬(WHISTL)'을 한국인터넷진흥원과 공동 개발했으며 현재 약 4,000여 기업에서 사용 중이다. 쉘가드는 하나 이상의 웹서버에 대해 각각 웹쉘 프로그램을 탐지하고 관리자는 중앙관리 및 조치가 가능하다. 특히, 업로드 파일에 대한 웹쉘 프로그램을 실시간으로 탐지하고 원격관리를 통한 즉각적인 대응이 가능하다. 또한, 주기적인 웹쉘 패턴을 업데이트 및 자체 패턴 규칙 등을 정의해 신속한 대응이 가능하고 각종 탐지 및 검역 내용에 대한 각종 통계를 제공한다.[12]

(표 3) 국내 웹쉘탐지 솔루션〔12〕
(Table 3) Domestic webshell detection solution〔12〕

제조사	제품명	주요 기능
A3 시큐리티	AEGIS Shell Monitor	• 실시간 웹 보안관제 • 웹쉘실시간 모니터링 탐지
SecureAT	ShellCop	• 웹쉘 프로그램 탐지 실행 방지
하로스	Shell Guard	• 업로드파일 웹쉘탐지

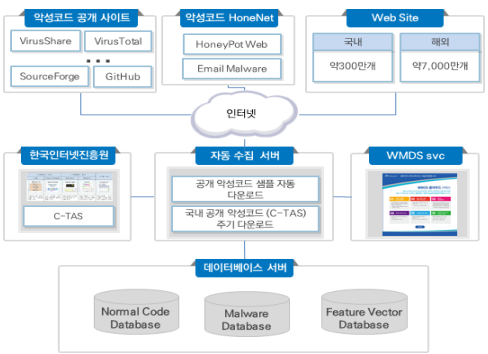
3. 머신러닝(Machine Learning)기반 웹쉘 수집 및 분석을 통한 방어시스템

3.1 머신러닝(Machine Learning)기반 웹쉘 수집 및 분석

최근 악성코드 은닉 여부를 탐지하기 위한 정적분석 기법에 관한 연구가 활발하게 이루어지고 있지만, 순차 알고리즘에 의존하고 분산 컴퓨팅을 지원하지 않아서 막대한 런타임 오버헤드가 발생한다. 그리고 순수 메모리 기반 알고리즘은 제한된 메모리 환경에서 비효율적이다. 이와 같이 정적 분석 기법은 모델의 탐지율이 낮고, 탐지 대상이 난독화된 경우에는 탐지하는데 오래 걸린다.[13]

악성코드를 빠른시간 안에 수집, 분석, 탐지를 위해서는 머신러닝기반의 자동화 기술이 필요하다.

웹쉘 수집 및 분석을 자동화하기 위해서는 그림 3과 같이 국내외 웹사이트를 통한 악성코드 자동 수집 모듈 개발과 국내외 웹사이트로부터 악성코드 탐지 수집이 필요하다. 즉, 본 연구개발을 위해서는 국내 약 300만개의 사이트, 해외 약 7,000만개의 사이트를 확보하여 탐지 수행하여, 탐지 결과 발견된 악성코드를 악성코드 데이터베이스에 메타정보와 함께 보관하는 것이 필요하다. 추가로 공개 악성코드 자동 수집 모듈 개발을 위하여 공개 악성코드 사이트로부터 Virusshare, VirusTotal, GitHub, SourceForge 등의 공개 악성코드 사이트를 발굴하여 악성코드 주기적 자동 수집 연계 모듈 개발과 한국인터넷진흥원의 C-TAS 시스템 연계 모듈 개발과 함께 악성코드 URL 시그니처 정보를 주기적으로 수집하여 웹쉘 정보와 악성코드 정보를 연동한 시스템을 제안한다.



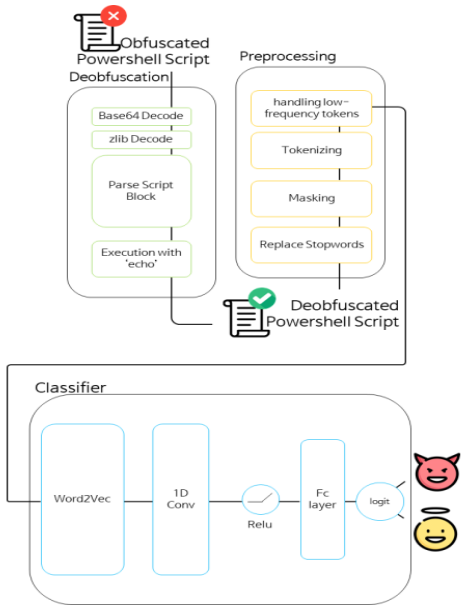
(그림 3) 국내외 웹사이트악성코드 자동수집 모델
(Figure 3) Malicious code automatic collection model through domestic and foreign website

3.2 딥러닝(Deep Learning)기반 웹쉘 탐지 시스템

해커는 웹쉘공격시 탐지를 우회하기 위하여 난독화된 웹쉘을 통하여 공격하는 경우가 많아지고 있다. 이에 빠른 웹쉘탐지를 위하여 딥러닝기반 아래의 방안에 대하여 검토 및 적용이 필요하다.

간단하고 빠른 역난독화 처리과정, Word2Vec과 CNN (Convolutional Neural Network)으로 구성되어 스크립트의 의미를 학습하고 특징을 추출해 악성 여부를 판단할 수 있는 딥러닝 기반의 분류 모델이 필요하다.딥러닝 기반 악성코드 탐지 방법은 그림 4 와 같이 크게 3가지 부분인 역난독화, 전처리, 분류기로 구성되어있다.

역난독화 단계에서는 난독화 되어있는 파워쉘 스크립트에 역난독화를 적용하여 의미를 지니는 원래의 코드로 복구하는 기법을, 전처리에서는 분류기의 성능 향상을 위해 텍스트 데이터에 적용할 수 있는 전처리 방법을, 분류기에서는 전처리된 데이터에서 의미를 학습하고 특징을 추출해 악성 여부를 판단할 수있는 분류기 모델을 활용한다. 2021 사이버보안 AI, 빅데이터 활용 경진대회의 AI 기반 파워쉘 악성 스크립트 탐지 트랙에서 제공된 1400



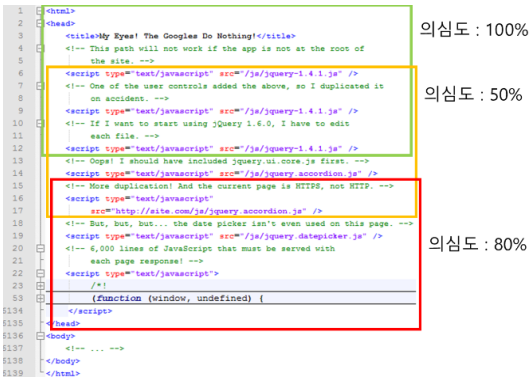
(그림 4) 딥러닝 기반 웹쉘 탐지 시스템 [14]
(Figure 4) Overall Structure of Malware Detection Method[14]

개의 악성코드와 8600개의 정상 스크립트를 이용하여 제안한 모델을 테스트한 결과 기존보다 5.04배 빠른 역난독화 실행시간, 100%의 역난독화 성공률, 0.01의 FPR(False Positve Rate), 0.965의 TPR(True Positive Rate)로 악성코드를 빠르고 효과적으로 탐지함을 보인다.[14]

3.3 슬라이딩 윈도우 기반 웹쉘 탐지 시스템

웹쉘을 탐지하는 방안으로 코드 위험도를 분석하여 웹쉘탐지를 하는 슬라이딩 윈도우 기반 기술이 있다.

웹 공격에 많이 활용되는 웹쉘의 탐지를 위하여 제안하는 슬라이딩윈도우 기반의 탐지 기법은 시간이 지남에 따라 발전해가는 웹쉘 탐지 우회 기술에 대응하여 보다 정확한 탐지를 제공하는 기술이며, 이를 기반으로 웹쉘의 다양한 변종 또한 탐지할 수 있다. 슬라이딩윈도우기법은 전통적으로 데이터 통신 및 처리분야에서 사용되던 방법이며, 신호 및 데이터를 다룸에 있어 보다 섬세한 접근이 필요할 때 활용되는 방법이다. 이러한 기법을 활용하여 스크립트에 부분적으로 존재하는 코드 위험도를 분석하여 웹쉘을 탐지하기 위한 방법을 제안하고 있으며, 이에 대한 접근은 그림 5와 같다.[15]

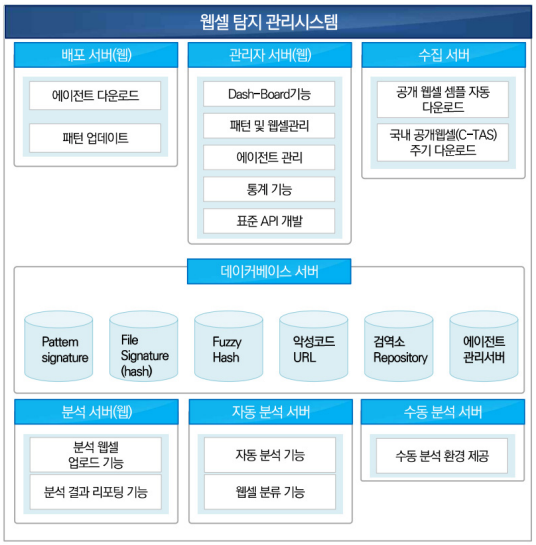


(그림 5) 슬라이딩윈도우 기반 코드 위험도측정〔15〕
(Figure 5) Sliding window-based code risk measurement for each part〔15〕

3.4 머신러닝기반 웹쉘 방어시스템

본 논문에서 제안한 인공지능 머신러닝 기반의 웹쉘 방어시스템을 상세하게 살펴보면, 인공지능 머신러닝을 적용한 인터넷 웹쉘 자동수집 기능, 웹쉘 분류 기능, 웹쉘 자동분석 기능, 분석결과 리포팅 기능이 있다.

머신러닝기반 웹쉘 방어시스템은 그림 6과 같이 웹사이트에 서비스할 탐지 에이전트와 탐지 관리시스템으로 구성되며, 탐지 관리시스템은 관리자 웹서버를 중심으로 수집서버, 배포서버, 데이터베이스 서버와 함께 분석서버군을 포함한다. 분석서버는 분석 웹서보와 머신러닝을 적용한 자동 분석서버, 수동분석서버로 구분되며, 내외부연계를 통하여 공개 웹쉘샘플을 획득할수 있는 GitHub와 SouceForge와 연계하여 머신러닝 기반 웹쉘방어시스템을 제안한다.



(그림 6) 머신러닝기반 웹쉘 방어시스템 모델
(Figure 6) Machine learning-based webshell defense system model

4. 결론 및 향후 연구과제

웹쉘(Web Shell)은 웹스크립트(asp, php, jsp, cgi)파일을 의미하므로, 자체가 악성코드가 아니기 때문에 WAF(웹방화벽),방화벽, IDS, IPS와 같은 보안장비나 바이러스나 악성코드를 탐지하는 기존의 백신, 스파이웨어 등에서 탐지가 쉽지 않다.

최근에는, 공격자가 탐지를 우회하기 위하여 암호화, 난독화된 웹쉘이 주를 이룸에 따라 더욱더 탐지가 안되고 있으며, 한번 침투하게 되면 주변에 파일서버, DB서버, 그룹웨어서버 등의 인접 시스템으로 전파가 용이하다. 이와 같이, 웹쉘이 갖는 잠재적인 위협과 막대한 피해

규모에도 불구하고 공개되어 있는 수많은 웹사이트와 허용된 프로세스에 의하여 침투 하므로 웹쉘 탐지가 쉽지 않은 상황이다. 이러한 위협에 대한 실제적인 방안으로는 현존하는 악성코드 와 웹쉘 대한 특징 뿐 아니라, 웹쉘 자체의 대한 특징을 학습함으로 웹쉘을 빠른시간안에 탐지해낼 수 있는 인공지능 머신러닝 기반의 웹쉘 방어시스템기술을 활용하는 것이며, 이를 통해 기존 수동으로 탐지하던 웹쉘에 대한 기술을 자동화 할 경우 웹쉘 탐지에 대한 시간, 탐지율을 대폭 향상 될 수 있다고 판단된다.

현재 인공지능 머신러닝기반으로 웹쉘탐지 및 분석에 적용하기 위해서는 추가적으로 많은 연구가 필요한 실정이다. 인공지능 머신러닝을 활용하여 빠른 시간안에 웹쉘탐지 와 정확한 분석 체계를 수립하여, 사이버침해에 효과적으로 대응하여 안전한 인터넷환경과 사용자의 편의성을 고려한 연구도 필요하다.

본 논문은 웹환경에서 웹쉘 탐지부문에 대하여 머신러닝기반 방어시스템 모델을 제안하였다. 그러나, 웹환경에서 다양한 사이버공격이 많이지고 있는 관계로, 웹쉘뿐만 아니라, OWASP TOP10 기준으로 웹에 대한 다양한 사이버공격에 대해서도 머신러닝기술을 활용하여 방어할수 있는 웹통합방어시스템 모델에 관한 연구를 진행하고자 한다.

참고문헌(Reference)

[1] Cisco Visual Networking Index 2017-2022 Outlook and Trends", 2018.
https://www.cisco.com/c/dam/global/ko_kr/solutions/service-provider/visual-networking-index-vni/pdfs/white-paper-c11-741490-kr.pdf

[2] K. A. Kim, "A Study on Design of Improved Security Vulnerability of Web Application", Hanbat University GraduateSchoolofInformationand Communication, 2016.

[3] Korea Internet & Security Agency, "Cyber Threat Trend Report(thefirsthalf2022)", accessed July.12, 2022. https://www.krcert.or.kr/data/reportView.do?bulletin_writing_sequence=66820

[4] U. Chung, J.S. Moon, "Study on security requirements for the web based operation system of a shipping company", Journal of Korean Society for Internet Information, vol. 23, no. 1, p. 49 - 68, 2022. https://doi.org/10.7472/jksii.2022.23.1.49

[5] Jaehong Yoo, "A Study on the Improvement of Website Security Vulnerabilities", Domestic Master's Dankook University, 2021.

[6] S.H Hong, "Study on defense countermeasures against Webshell attacks of the Industrial Information System", Journal of Industrial Convergence, Vol. 16, No. 4, 47-52, 2018.
https://doi.org/10.22678/JIC.2018.16.4.047

[7] J.B. Lee, "A Study on the improvement of countermeasures for webshell hacking", Dongguk University, 2019.

[8] Gil Min-kwon, "Web Hacking Starts from 'Web Shell' ", dailysecu, 2011.
https://www.dailysecu.com/news/articleView.html?idxno=15,2011.06.02.

[9] K.H. Kim, "A Study on Detection Method of Malicious Code Based on Artificial Intelligence Machine Learning", ICONI, 2017.

[10] O.Y. Han, "Artificial Intelligence Trend Research and Technology Trend Analysis in the Era of the 4th Industrial Revolution", Proceedings of the Korea Internet and Information Society, Vol. 18, No. 2, 2017.

[11] J.U. Park, "Deep Learning-based Malicious Code Detection Using API Features", Graduate School of Konkuk University, 2017.

[12] K.H. Kim, S.S. Choi, Y.T Shin, "Development of artificial intelligence application of web shell collection and analysis system Trend Analysis", Korean Internet and Information Society, Vol. 19, No. 2, 2018.

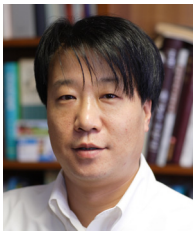
[13] C.R. Han, S.H Yun, M.J Han, I.G Lee., "Machine Learning-Based Malicious URL Detection Technique," Journal of the Korea Institute of Information Security & Cryptology, 32(3), 555-564, 2022.

[14] H.J Jung, H.Y. Ryu, "Deobfuscation Processing and DeepLearning-Based Detection Method for PowerShell-Based Malware," Journal of the Korea Institute of Information Security& Cryptology, 32(3), 501-511, 2022

[15] Kihwan Kim, Lee DongGeun, Hyoung Yi, Yongtae Shin, "A Study on Sliding Window based Machine Learning for Web Shell Detection," Proceedings of the Korean Society of Computer Information Conference, 27(2), 121-122, 2019.
https://koreascience.kr/article/CFKO201920461758039.page

◐ 저 자 소 개 ◑

김 기 환(Ki-Hwan Kim)
2007년 숭실대학교 대학원 보안학과(공학석사)
2010년 숭실대학교 대학원 컴퓨터학과(공학박사수료)
현재 ETRI 책임연구원
관심분야 : 정보보호, AI보안, ISMS-P,ISO27001인증, IoT보안, etc.
E-mail : itconsult@hanmail.net



신 용 태(Yong-tae Shin)
1994년 Unviersity of Iowa Computer Science 공학박사
현재 숭실대학교 스파르탄SW교육원 원장
관심분야 : 정보통신, 정보보호, etc.
E-mail : shin@ssu.ac.kr