웹 공격 탐지를 위한 ELK Stack 시스템 구축에 관한 연구

김진수*, 오원재*, 이승재*, 한지호*, 김민수**

*중부대학교 (대학생), **중부대학교 (교수)

Research of Building ELK Stack system for web attack detection

Jin Su Kim*, Won Jae Oh*, Seung Jae Lee*, Ji Ho Han*, Min Su Kim**

*JoongBu University(Undergraduate student)

**JoongBu University(Professor)

요 약

정보통신 기술의 발전과 최근 팬데믹(pandemic) 상황으로 인하여 사회 전반적으로 비대면 환경의 일상화가 가속화되면서, 보안 위협의 이슈 또한 고도화된 형태로 증가하고 있다. 보안 공격의 형태는 홈페이지 변조, 침해사고 등 웹 서버에 대한 공격으로 보안 영역에서는 웹 로그 데이터를 이용한 실시간 탐지를 통해 대응을 위한 웹 서버 공격 탐지 시스템을 관리·운영하게 된다. 이에 본 연구에서는 발생 된 로그 중 필터링 알고리즘을 통해 웹 공격 관련 로그를 탐지 및 분류하고, ELK Stack 시스템을 구축하여 탐지된 로그들을 실시간으로 분석 및 시각화한 웹 서버 공격 탐지 시스템을 제안한다.

Ⅰ. 서 론

정보통신 기술의 발전과 최근 팬데믹(pandemic) 상황으로 인하여 사회 전반적으로 비대면 환경의 일상화가 가속화되면서, 보안 위협의 이슈 또한 고도화된 형태로 증가하고 있다. 2020년 KISA에서 코로나19로 인해 비대면 일상화에 따른 인터넷 이용이 증가한 만큼 해킹 공격 기법의 지능화·다양화로 인한 침해사고가 증가하고 있다고 발표하였다.[1]

보안 공격의 형태는 홈페이지 변조, 침해사고 등 웹 서버에 대한 공격으로 보안 영역에서는 웹 로그 데이터를 이용한 실시간 탐지를 통해 대응을 위한 웹 서버 공격 탐지 시스템을 관리 및 운영하게 된다. 과학기술정보통신부와 한국정보보호산업협회의 기업과 개인의 정보보호 인식 및 침해사고 예방/대응 활동 등에 대한 '2020년 정보보호 실태조사'에 의하면 작은 비중(IT 예산 대비 1% 미만)이더라도 정보보호 예산을 편성한 기업이 약 61.8% 상당히 증가했다. 또한, 기업 부문에서도 정보보호의 중요성에 대한 인식이 작년 대비 약 4.5% 증가한 것을 확인할 수 있다.[2] 이처럼 기업 차원에서도 중요한 정보의 유출을 막기 위해 다양한 방법을 고안하고 있으나, 발전하는 공격 기술 혹은 관리 소홀과 같은 이유로 여전히 위험에 노출된 실정이다. 이처럼 웹 공격이 발전함에 따라 공격 경향을 빠르게 파악하고 대처하는 것이 중요하다.

OWASP Top 10에서 항상 빠지지 않고 나오는 주요 공격들이 있다. 웹 애플리케이션으로 비정상적인 명령어, 쿼리 등을 보내 시스템에 접근하는 취약점인 Injection, 공격자가 피해자의 브라우저에 악의적인 스크립트를 넣는 XSS 등이 대표적이다. 또한 웹 서버에서는 사용자의 행위에 따라 매 순간 많은 양의 로그가 발생한다.

이에 본 연구에서는 서버 운영자뿐만 아니라 사용자들의 가독성을 위해 발생 된 로그 중 필터링 알고리즘을 통해 웹 공격 관련 로그를 분류한다. 또한, ELK Stack을 활용하여 분류된 로그들을 실시간으로 분석 및 시각화함으로써 기존의 웹 로그 분석 연구에서 실시간 분석 및 시각화가 어렵다는 단점을 보완했다. 결론적으로, 본 연구에서는 발생 된 로그 중 필터링 알고리즘을 통해 웹 공격 관련 로그를 탐지 및 분류하고, ELK Sta

ck 시스템을 구축하여 탐지된 로그들을 실시간으로 분석 및 시각화한 웹 서버 공격 탐지 시스템을 제안한다.

Ⅱ. 이론적 배경

2.1 apache 웹 서버 로그

웹 로그는 서버에서 이루어지는 모든 일을 구체적으로 기록하여 보관한 데이터이다. 웹 로그는 시스템의 문제를 찾거나, 성능 현황을 분석하기 위해서 사용된다. 시스템에는 많은 로그가 기록되지만, 이 중 apache server의 access log는 서버가 처리하는 모든 요청을 기록한다.[3] apache access log의 구조는 Fig. 1, Fig. 2, Table. 1과 같다.

```
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}i\"" combined
```

Fig. 1. The structure of access log

```
127.0.0.1 - - [15/Jun/2021:19:08:19 +0900] "GET /dvwa/DVWA/ HTTP/1.1"
①         ②                                ③   ④           ⑤
200 6340 "http://localhost/dvwa/" "Mozilla/5.0 ...."
⑥   ⑦    ⑧                       ⑨
```

Fig. 2. Example for access log

Table 1. Description table of access log

No	Description
①	원격 호스트 IP주소
②	요청 시간
③	method(메서드)
④	요청 URL
⑤	프로토콜 버전
⑥	HTTP 상태코드
⑦	HTTP 헤더를 제외한 전송 크기
⑧	referrer(리퍼러)
⑨	User-Agent

기록되는 정보 중 데이터 전송 크기, Referrer, User-Agent 등의 정보는 요청 패킷을 만들 때 정상과 동일하게 조작이 가능하며, Method, URL, Protocol 버전의 경우 클라이언트에서 자동으로 생성되는 정보이다. 반면, 원격 호스트 IP, 요청날짜/시간의 경우 사용자와 프로그램이 접속할 때 위변조가 불가능한 기록이다. 본 연구에서는 웹 공격 탐지를 위해 access log를 활용한다.

2.2 ELK Stack

ELK Stack이란 Elasticsearch, Logstash, Kibana, 이 오픈 소스 프로젝트 세 개의 머리글자와 Beats를 추가하여 ELK Stack이라 한다. 또한, ELK Stack은 데이터 수집, 분석, 시각화 등 기타 유용한 기능들을 제공하는 빅데이터 플랫폼으로 구성되어 있다.[4]

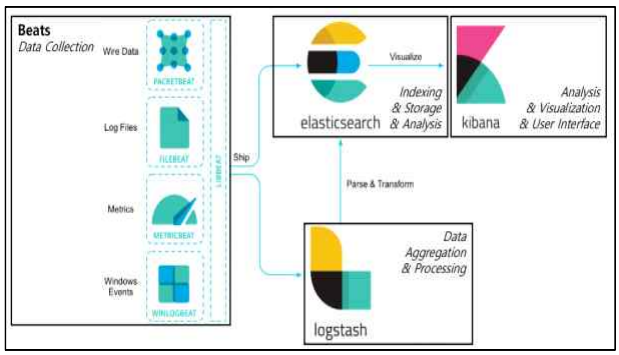

Fig. 3. The structure of ELK Stack

Elasticsearch는 ELK Stack 구조의 핵심 기능을 담당하는 JSON 기반의 분산 검색 데이터베이스 엔진으로써 데이터 분석 및 저장 기능을 담당한다.

Logstash는 수집할 로그를 선정해서 지정된 대상 서버(Elasticsearch)에 인덱싱하여 전송하는 역할을 담당하는 소프트웨어로써 데이터 전처리 기능을 담당한다.

Kibana는 Elasticsearch 내에 있는 데이터를 시각화해주며, 사용자가 신속하게 데이터를 이해할 수 있도록 다양한 다이어그램 및 그래프를 기본적으로 제공하고 있다. 또한, 데이터를 시각적으로 탐색하고 실시간으로 분석할 수 있다.

Beats는 단일 목적의 데이터 수집기 플랫폼으로써 데이터 수집 기능을 담당한다. 서버의 에이전트로 설치하여 다양한 유형의 데이터를 Elasticsaerch 또는 Logstash에 전송하는 오픈 소스 데이터 발송자이다.

Ⅲ. 관련 연구

3.1 Webalizer

Webalizer는 무료로 제공되며, C언어로 작성되어 매우 빠른 속도와 이식성이 뛰어나 일반적으로 사용되는 로그 파일 분석 프로그램 중 하나이다. 표준 웹 브라우저에서 볼 수 있도록 매우 상세하고 쉽게 구성할 수 있는 사용 보고서를 HTML 형식으로 생성한다. 하지만 Webalizer는 지속적인 실시간 모니터링이 불가능하며 웹 공격을 탐지해 주지 않는 단점이 있다.

3.2 AWStats

AWStats는 Webalizer에서 지원하지 않는 인터페이스를 포함하고 있다. 또한, CGI 또는 명령줄에서 작동하며 몇 가지 그래픽 웹페이지에서 로그에 포함된 가능한 모든 정보를 보여주는 로그 분석 도구이다. 무료로 설치할 수 있으며 대용량 로그 파일을 빠르게 처리할 수 있다. 하지만, AWStats 역시 실시간 모니터링이 불가능한 단점이 있다.

Ⅳ. 제안 방법

4.1 시스템 목표

웹 공격 탐지 시스템은 크게 '실시간 모니터링'과 '웹 공격 탐지'라는 주제로 다음과 같이 구축 목표를 정했다.

첫째, access log 파일을 실시간으로 불러올 수 있도록 한다.

둘째, 필터링 알고리즘을 통해 각 공격 기법의 패턴을 판별하여 공격 기법별로 분류하고, 탐지된 로그는 해당 공격 기법의 로그 파일로 저장한다.

셋째, ELK Stack 시스템에서 탐지된 로그를 실시간 모니터링으로 확인 후, 해당 로그를 분석한다.

해당 시스템의 구조도는 Fig. 4와 같다. 공격자가 웹 서버에 공격을 시도하였을 때, access log 파일에 로그가 기록된다. 각 공격 기법들의 패턴을 통해 access log에 저장된 로그 중 공격 로그를 판별한 후, 각 공격 기법별로 로그를 저장한다. 그 후, 탐지된 공격 로그를 Elasticsaerch에 저장 후 분석하고 Kibana를 이용한 실시간 모니터링을 통해 공격 로그를 탐지한다.

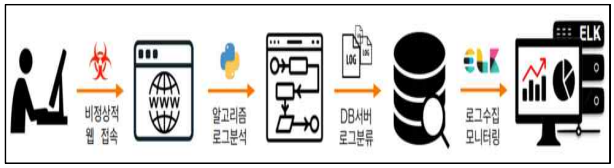

Fig. 4. The structure of detection system

4.2 시스템 구현

본 연구에서는 OWASP TOP 10 기반의 주요 공격 기법인 Injection과 XSS 공격을 중점으로 진행하였다. 알고리즘의 구성에 대한 설명은 Fig. 5와 같다. 알고리즘의 시작은 로그 파일을 불러온 후, 해당 로그가 일반 로그인지 공격 관련 로그인지 판단한다. 판단 기준은 각 공격 기법별 패턴을 활용하였다. 공격 패턴과 일치하는 로그가 있을 시, 해당 공격 로그로 따로 저장한다. 알고리즘 구현 방법은 python으로 제작하였다. 메인 함수에서 sleep 함수를 사용하여 사용자가 지정한 시간 동안 로그가 생기지 않을 시 시스템을 종료한다.

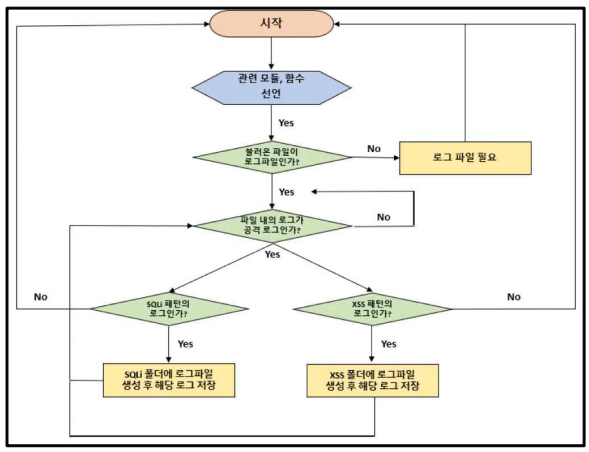

Fig. 5. Classification/detection algorithm

Ⅴ. 결론 및 향후 계획

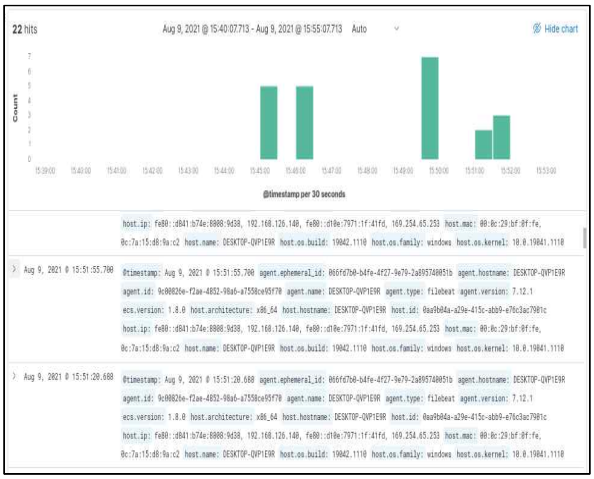

Fig. 6. Execution results

Fig. 6은 구현한 공격 탐지 시스템을 통해 각 주요 기법별로 해당 로그를 실시간 분석 및 시각화한 화면이다. Elasticsearch를 통해 로그를 분석하였고, kibana 웹 사이트에서 탐지된 로그의 상세 내용이 보여지고 있는 것을 확인할 수 있다.

본 연구에서는 SQL Injection, Cross-Site Scripting 공격이 이루어졌을 때 로그를 분류하고 실시간 모니터링을 통해 탐지를 진행하였다.

향후 연구에서는 클라우드 DB를 이용하여 탐지된 로그를 수집할 계획이다. 클라우드 DB는 기존 DB에 비해 확장성 및 과부하 문제점을 해결할 수 있다. 클라우드 DB를 통해 수집된 로그는 알고리즘을 통해 공격 로그가 분류되며, 로그 내 데이터 중 IP 부분을 클라우드 DB에 저장되어 있는 IP와 매핑한 뒤 차단할 수 있도록 할 예정이다. 또한, 공격자가 인가된 IP가 아닌 변조된 IP를 사용할 수도 있기 때문에, 제로트러스트 모델 보안 방식을 연구 및 적용할 계획이다.

[참고문헌]

[1] https://index.go.kr/smart/chart_view.jsp?idx_cd=1363&bbs=INDX_001&clas_div=C&rootKey=1.48.0

[2] https://m.korea.kr/news/pressReleaseView.do?newsId=156441395#pressRelease

[3] 이화성, 김기수. 2019. 웹서버 로그 데이터의 이상상태 탐지 기법. 한국정보통신학회논문지 23: 1311-1319.

[4] https://www.elastic.co/kr/what-is/elk-stack

[5] 전보국, 최경택, 박규순 (2021). 보안은 불신으로부터, 제로트러스트 모델의 군 적용 필요성 제고. 국방과 기술, (511), 116-129.