Smart Media Journal / Vol.7, No.4 / ISSN:2287-1322
http://dx.doi.org/10.30693/SMJ.2018.7.4.70

네트워크 보안 관제를 위한 로그 시각화 방법

A log visualization method for network security monitoring

조우진*, 신효정**, 김형식*
(Woo-Jin Joe, Hyo-Jeong Shin, Hyong-Shik Kim)

요 약

정보시스템에서 정보 보안의 중요성이 강조됨에 따라 이에 대응하기 위해 많은 기업이 보안 솔루션을 도입하고 있다. 하지만 많은 예산을 들여도 이를 관리하는 보안 관제가 없으면 제대로 기능하지 못하게 된다. 보안 관제는 문제 발생 시 빠른 대처가 필수적이며, 관제 목적에 맞는 적절한 시각화 대시보드를 설계하여 필요한 정보를 빠르게 전달할 수 있도록 할 필요가 있다. 본 논문에서는 오픈소스 Elastic Stack을 이용하여 보안 로그를 시각화 하는 방법을 제시하고, 관제 목적에 적합한 대시보드로 구현함으로써 제시된 방법이 네트워크 보안 관제에 적합함을 보인다. 대시보드는 비정상적인 트래픽 증가와 공격 경로 분석 등의 목적으로 효과적으로 활용될 수 있음을 확인하였다.

■ 중심어 : 보안 관제 ; 시각화 ; 오픈소스 ; Elastic Stack

Abstract

Current trends in information system have led many companies to adopt security solutions. However, even with a large budget, they cannot function properly without proper security monitoring that manages them. Security monitoring necessitates a quick response in the event of a problem, and it is needed to design appropriate visualization dashboards for monitoring purposes so that necessary information can be delivered quickly. This paper shows how to visualize a security log using the open source program Elastic Stack and demonstrates that the proposed method is suitable for network security monitoring by implementing it as a appropriate dashboard for monitoring purposes. We confirmed that the dashboard was effectively exploited for the analysis of abnormal traffic growth and attack paths.

■ keywords : Security monitoring ; Visualization ; Open Source ; Elastic Stack

Ⅰ. 서 론

최근 기업별로 보안의 위험성이 인지됨에 따라 보안 관제의 필요성이 대두되고 있다. 기업에서 어렵게 개발한 핵심 기술이 경쟁 업체로 유출되는 경우, 유출된 기업은 회복하기 어려울 정도의 큰 피해를 볼 수밖에 없다. 그림 1에 따르면 최근 5년간 보안 규제가 강화되어 매년 유출 건수는 줄어들지만, 여전히 많은 수의 기업 기술이 유출되고 있다고 한다[1].

개인정보 유출도 큰 이슈 중 하나이다. 개인정보 유출은 주로 해킹, 내부 직원 및 협력사 직원을 통해 발생하게 되는데, 대표적으로 2014년 1월 국민, 롯데, 농협 카드에서 내부자를 통해 유출되었다. 약 1억 건이 넘는 개인 정보가 유출되자 3개월간의 영업정지라는 처벌을 받았으며, 이 결과 기업의 순이익이 감소하는 상황으로 이어졌다. 표 1은 연도별 개인정보 유출 현황이다[2]

그림 1. 연도별 기술유출 현황(2012 ~ 2016)

* 충남대학교 컴퓨터공학과
** 충남대학교 소프트웨어연구소
이 연구는 충남대학교 학술연구비에 의해 지원되었음.

접수일자 : 2018년 07월 31일
수정일자 : 1차 2018년 10월 05일, 2차 2018년 11월 19일
게재확정일 : 2018년 11월 20일
교신저자 : 신효정, e-mail : hyojeongshin@gmail.com

표 1. 연도별 개인정보 유출 현황(2012 ~ 2016)

년도	유출업체 수	피해 규모(단위: 만건)
2012	2	1,295
2013	5	19
2014	73	2,853
2015	13	325
2016	19	1,103
합계	112	5,595

기업들은 이에 대응하기 위해 정보 보안의 예산을 늘리는 등 역량 강화에 노력하고 있고, 여러 솔루션을 적용하고 있다. 그러나 단순히 보안 제품만 도입하고 사후 운영이 제대로 되지 않고 있어 정보 보안 사고는 여전히 발생한다.

정보 보안 사고를 예방하기 위해서는 보안 시스템을 구축한 이후에도 이를 통합 관리 및 분석할 필요가 있다. 많은 예산을 들여 다양한 보안 시스템을 구축하여도 이를 관제하지 않으면 속수무책으로 당할 수밖에 없다. 이를 위해서는 보안 관제 전담 인력을 두어 위협을 사전에 탐지할 수 있어야 하며, 적절한 관제 환경을 구성하여 침입이 발생하면 관제 담당자가 이를 빠르게 파악할 수 있어야 한다[3, 4].

보안 시스템에서는 로그가 생성되는데, 이를 수집하여 실시간으로 분석하는 것이 보안관제이다. 하지만 여러 보안 시스템 간에 많은 데이터를 수집할 경우, 기존의 데이터베이스로는 이를 수용하기 어렵고, 사람의 인지능력 또한 한계가 있어 실시간 분석이 어려워진다. 이에 많은 양의 데이터를 알기 쉬운 형태로 바꿔주는 시각화 기법이 주목받으나, 데이터 특성에 따른 시각화 방법에 대한 연구가 부족하여 분석가의 역량에 따라 적절하지 않은 시각화 방법이 도출될 수도 있다.

패턴 기반 네트워크 보안 관제의 한계를 극복하기 위하여 빅데이터 시각화를 이용한 위협 탐지 방법도 제안되었다. 현정훈 외 1인은 트래픽이 많이 발생하는 것과 같은 이상 현상을 탐지하기 위해 시각화 대시보드를 설계하고, 이 대시보드에서 위협으로 판단된 건수를 데이터베이스화 하여 날짜별로 침해 위협 정도를 측정하였다[5]. 그렇지만 제시된 방법은 이상 현상이라 판단되는 횟수를 기록하는데 초점을 맞춰서 이상 현상의 원인을 분석하지는 않았다.

우리는 이상 현상을 분석하는 것에 초점을 맞추어 다른 정보와의 연관 관계를 표현할 수 있도록 대시보드를 설계하는 방법을 연구한다. 우선 이상 현상을 효과적으로 탐지할 수 있도록 로그 데이터 유형에 따른 시각화 기법을 정리하고, 대시보드를 통해 정보 간의 유기적인 관계를 표현함으로써 실제 탐지된 위협에 대한 분석이 가능함을 보인다.

Ⅱ. 본 론

1. 네트워크 보안 관제

보안 관제란 침입으로부터 시스템 자원의 손상을 막기 위해 시스템이나 네트워크에서 발생하는 이벤트를 감시하고, 발생하는 문제에 대하여 실시간으로 대처하는 것을 말한다. 이를 위해 보안 관제 담당자는 보안 시스템에서 발생하는 이상 징후를 알아차리고, 이것이 위협인지 아닌지를 정확하게 판단하기 위한 전문적인 역량과 지식, 경험이 필요하다. 즉, 보안 관제는 보안에서 가장 먼저 고려할 대상이며, 보안을 종합적으로 다뤄야 하는 업무이다.

보안 관제는 1차 방어선, 최전방이라고도 불린다. 즉, 침입이 발생했을 때 빠른 대응을 하고 재발을 예방해야 한다. 이를 성공적으로 수행하기 위해서는 네트워크에서 발생하는 이벤트를 먼저 관리해야 한다. 개방형 시스템을 기반으로 하는 네트워크가 외부 공격에 가장 쉽게 노출되어 있기 때문이다.

네트워크에서 관리 대상은 스위치 등으로 인해 분류된 네트워크망이고, 각각의 망에 설치된 보안 시스템으로부터 수집되는 이벤트를 보안 관제망으로 전송할 필요가 있다. 그림 2는 네트워크 보안의 기본 흐름이다[6]. 1차로 외부 네트워크로부터 유입되는 트래픽은 많은 위협이 있을 수 있으므로 방화벽으로 적극적인 필터링을 적용해야 한다. 그리고 필터링 된 패킷에 대해서는 침입 탐지 시스템(IDS)을 통해 공격을 탐지해야 한다. IDS를 통해 모니터링 시 미러링 방식과 인라인 방식이 있는데, 그림 2는 미러링 방식으로 TAP 장비를 통해 네트워크 트래픽 사본을 받아 모니터링 하는 방식이다.

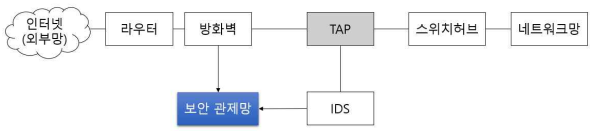

그림 2. 네트워크 보안 기본 흐름

그러나 방화벽이나 IDS와 같은 보안시스템 각각에 보안 전담 인력을 두면 관리 및 유지비용이 소요된다. 따라서 최소화된 보안 관리 인력으로 체계적인 유지 및 관리 목적으로 ESM(Enterpise security management, 기업 보안 관리)이라는 개념이 나오게 되었다. 아래 그림 3은 이기종 보안 시스템으로부터 로그를 수집하여 분석하는 ESM 시스템의 구성요소를 보여준다[7].

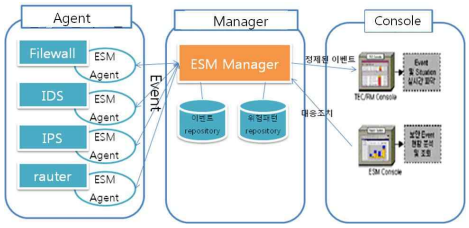

그림 3. ESM 구성 요소

방화벽이 IP 주소나 포트를 기준으로 비정상 트래픽을 차단하는 것이라면 IDS는 포트에 대한 정보뿐만 아니라 패킷의 데이터까지 분석하여 정상적인 트래픽 여부를 결정한다. 방화벽의 경우 열려있는 포트를 통해 들어오는 공격을 차단할 방법이 없으나, IDS에서는 이러한 공격을 인식할 수 있는 기능이 있다.

본 논문에서는 관제의 대상으로 오픈소스 IDS 도구인 Bro에서 발생하는 보안 로그를 수집하여 분석한다. Bro는 라우터 혹은 스위치로부터 트래픽을 수집 하여 로그 파일을 생성하며, 클러스터 형태로 트래픽을 분산함으로써 고대역 네트워크를 분석하기에 적절한 도구이다.

2. 보안 로그 관리

로그는 시스템에서 일어나는 동작에 대한 기본적인 추적 정보다. 대부분 시스템은 로그를 남기는데 이로부터 시스템 성능 관리, 시스템의 장애 원인 분석 등의 다양한 정보를 얻을 수 있다[8]. 또한 로그는 보안 목적으로 분석이 가능한데, 실제로 침입이 발생했을 경우 시스템에 남는 로그를 수집하여 외부 침입 탐지 및 추적을 할 수 있다.

전형적인 로그의 구조는 이벤트가 발생했던 시간을 기반으로 사용자에게 알리기 위한 메시지가 들어있거나 혹은 이벤트의 정보를 필드별로 구분한다. 많은 사람이 접속하여 사용할 수 있는 서버용 운영체제인 리눅스는 원격 접속과 관련하여 그림 4와 같은 로그를 남기며, 오픈소스 IDS인 Bro는 그림 5와 같은 로그를 남긴다.

```
Feb 3 12:36:45 localhost sshd[25922]: Failed password for invalid user test
        from ::ffff:61.237.15.202 port 47765 ssh2
Feb 9 22:56:31 localhost sshd[27922]: Accepted password for root
        from ::ffff:221.154.90.109 port 4763 ssh2
```

그림 4. 리눅스 로그 예시

```
1500609601.105317 CX1wlU2KMSVH09BEa5 150.183.146.130 45045 150.183.95.96 53
        udp dns 0.000001 0 210 SHR T T 0 ^d 0 0 2 266 (empty)
1500609601.898624 CBgVY61r4GnFZHxa7j 218.49.22.138 12278 150.183.158.227 443
        tcp ssl 0.014746 1169 0 S0 F T 0 SAD 4 1290 0 0 (empty)
```

그림 5. Bro 로그 예시

Bro는 패킷을 분석해 패킷에 대한 정보를 각 필드별로 남긴다. 표 2는 그림 5에 나온 필드에 대한 설명이다.

표 2. Bro 로그에 포함된 정보에 대한 예

필드	값	설 명
ts	1500609601.105317	첫 패킷이 도착했을 때의 UNIX 시간
orig_host	150.183.146.130	연결 요청한 IP 주소
proto	udp, tcp	프로토콜 정보
service	dns, ssl	이용하는 서비스
duration	0.000001	연결이 지속되는 시간

보안 관제 담당자는 이러한 로그를 기반으로 침입자가 어떠한 작업을 했는지, 어떠한 경로로 접속했는지를 추적해야 한다. 그러나 단일성 이벤트만으로는 APT와 같은 지능형 공격을 탐지해낼 수 없으며, 현재 구축된 시스템과는 상관없는 양성 오류(false positive)가 발생할 수 있다. 예를 들어 데이터베이스가 없는데 SQL 인젝션 공격이 시도되었다면 관제 담당자는 이를 신경 쓸 필요가 없다. 또한 웹 서버에서 404 에러가 뜨고 동시에 CGI 공격이 탐지되었다면 이는 주목할 만한 정보가 된다[9].

따라서 보안 영역뿐 아니라 응용 계층의 정보 및 이벤트를 수집하여 관리할 필요성이 생겨 SIEM(Security Information & Event Management, 보안 정보 및 이벤트 관리)이 주목받게 되었다. SIEM은 기존의 ESM 역할을 기업 전반으로 확대해 보안 영역뿐 아니라 기업 전반에서 로그를 수집하여 분석하는 모니터링 체계이다.

SANS가 2017년 보안관제센터(Security Operation Center)에 종사하는 직원들을 대상으로 한 설문조사에 따르면 77%의 응답자들이 SIEM도구를 사용한다고 했지만, 엄청난 양의 데이터를 처리하기에 어려움이 있다고 한다[10]. 이는 가공되지 않은 비정형화된 로그는 데이터의 양이 많아질수록 더욱 이해하기 어렵고 필요 없는 정보들이 생기기 때문이고, 보안 로그뿐만 아니라 응용 계층의 로그까지 수집할 경우 상호연관성까지 알아내야 하기 때문이다.

SIEM이 기업 전반에서 많은 양의 로그를 수집하기 때문에 기존의 관계형 데이터베이스 처리 방식이 아닌 빅 데이터 플랫폼이 필요하며, 그 목적이 보안 관제이기 때문에 실시간 분석이 가능해야 한다. 또한 단일성 이벤트만 분석하지 않고 여러 이기종 시스템으로부터 생성된 로그 간에 상호 연관성을 분석해야 한다. 즉, SIEM은 단순히 로그를 통합하고 분석하는 시스템이 아니다.

3. 빅 데이터 분석 도구

네트워크의 규모가 커짐에 따라 많은 양의 로그가 발생한다. 또한 앞에서 설명한 바와 같이 상호연관성 분석을 위해서는 여러 계층에서 데이터를 수집할 필요성이 있다. 이러한 대량의 데

이터를 저장하기 위해서는 빅 데이터 기술이 필요하다. 따라서 이번 절에서는 여러 빅 데이터 분석 도구들을 비교하여 보안 로그를 분석하기 위한 최적의 도구를 선별한다.

빅 데이터 분석 도구가 제공하는 기능은 데이터 통합, 상관관계분석, 알림, 시각화 대시보드 기능 등이 있다. 대부분의 빅 데이터 도구들은 해당 기능들을 지원하나, 각 도구마다 성능상 차이가 있을 수 있고, 상업용이 아닌 오픈소스 도구도 존재하므로 기업의 환경을 고려하여 선택해야 한다. 본 논문에서는 오픈소스 도구인 Elastic Stack을 이용하므로, Elastic Stack과 다른 도구들을 비교한다.

Elastic Stack은 원래 3가지 오픈소스(Elastic search, Logstash, Kibana)로 구성되어 ELK Stack이라 불렸지만, 5.0.0 버전부터 Beats가 포함되어 Elastic Stack이라 부른다[11]. Elasitc Stack은 단일 소프트웨어가 아닌 Elasticsearch(분산형 검색 엔진), Logstash(데이터 수집/변환), Kibana(데이터 시각화), Beats(데이터 전송) 4가지 도구로 구성되어 있다[12]. Beats가 로그 파일을 수집하면 Logstash에서 파싱하여 정형화된 형태로 변환하여 Elasticsearch에 저장한다. Kibana는 Elasticsearch에 저장된 데이터를 시각화한다

Logstash는 200개 이상의 플러그인으로 다양한 형태의 파일을 수집하고 변환할 수 있으며, 그림 6과 같이 파이프라인 형식으로 구성되어 필터 과정을 거친다[13]. Elasticsearch는 분산형 검색엔진으로써 확장성이 뛰어나 저장 공간이 부족할 경우 쉽게 대응할 수 있다. Kibana의 경우 히스토그램부터 지도까지 다양한 시각화 표현이 가능하고 여러 시스템의 로그를 대시보드에서 함께 보여줄 수 있어 연관분석이 가능해 빅 데이터 분석에 적합하다.

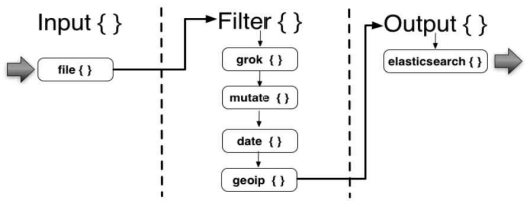

그림 6. Logstash 파이프라인

Elastic Stack과 자주 비교되는 대표적인 상용 솔루션에는 Splunk[14]가 있다 Splunk는 데이터 수집 시 full text를 저장하고, Elastic Stack은 Logstash에서 필터를 거친 다음 변환되어 저장된다는 차이점이 있다. Splunk는 상업용 소프트웨어라 고객 지원이 잘 되어 쉽게 이용할 수 있다는 장점이 있지만, 기업 환경 혹은 고객의 요구사항에 맞추는데 한계가 있다[15]. 반면, Elastic Stack은 환경에 맞춰 직접 플러그인을 설치하는 등 설정을 해야 하지만, 기업 환경에 맞춰 최적화가 가능하다는 장점이 있다.

본 논문은 네트워크 보안 관제를 대상으로 하기 때문에 실시간 분석이 중요하며, 이는 데이터 조회 속도가 분석에 많은 영향을 미치게 한다. 따라서 Elastic Stack의 저장 공간인 Elasticsearch와 다른 분석 도구들이 사용하는 저장 공간의 조회 성능 차이를 비교할 필요가 있다.

먼저, 상용 소프트웨어와의 성능 차이를 확인하기 위해 Splunk와 Elastic Stack의 조회 성능 비교가 필요하다. 로그의 개수별 조회 시간에 대한 실험 결과 로그의 개수가 적을 때는 차이가 많이 나지 않지만, 10억 개의 로그가 쌓일 경우 약 10초 정도의 차이가 발생했다[15]. 즉, 오픈소스 임에도 상용 소프트웨어와 성능 상 차이가 크지 않다는 것을 알 수 있다. 따라서 기업의 규모가 클 경우 Splunk를 선택하는 것이 바람직하고, 수집되는 데이터 규모가 작은 중소기업의 경우 Elastic Stack을 선택하는 것이 적절하다.

Elastic stack을 제외한 다른 오픈소스 도구들과의 성능 비교도 필요하다. Elastic stack이나 Splunk는 저장 공간을 따로 가지지만, Pentaho, Jaspersoft 등의 다른 분석 도구들은 Hbase와 같은 NoSQL 데이터베이스로부터 보고서를 생성한다[16]. 따라서 Hbase와 Elasticsearch의 조회 성능 비교가 필요하다. 로그 용량별 조회 시간을 실험한 결과 로그 용량이 커질수록 Elasticsearch의 조회 속도가 빨라지는 것을 보아[17] Hbase를 이용한 분석 도구들보다 Elasticsearch를 이용하는 Elastic stack이 분석에 더 적합함을 알 수 있다.

그림 7은 Google trends에서 Elastic Stack, splunk, jaspersoft, pentaho 분석 도구에 대한 검색 빈도를 측정한 것이다[18]. 14년도 이후부터 Elastic Stack이 다른 도구들 보다 많이 검색됨을 확인할 수 있다. 즉, 시간이 지날수록 Elastic Stack은 많은 주목을 받고 있음을 알 수 있으며, 이는 오픈소스이기에 비용적 부담이 없기 때문이라 추정된다. 또한 실시간 분석이 필요한 보안관제 특성상 데이터 조회 시간이 빨라야 하는데, Elasticsearch는 Hbase에 비해 조회 시간이 빠르고 Splunk와 성능 상 차이가 크지 않기 때문에 최적의 비용으로 보안 관제를 구축할 수 있다.

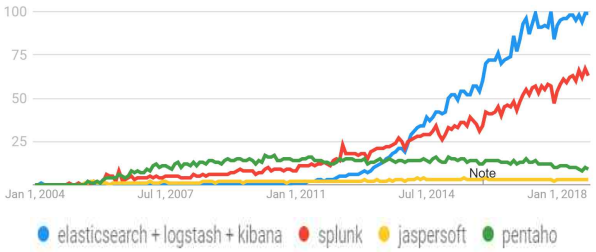

그림 7. 빅 데이터 도구별 구글 검색 빈도

빅 데이터 분석 도구의 중요한 기능중 하나는 시각화 기법이다. 시각화는 많은 양의 데이터들을 간단하고 이해하기 쉬운 형태로 만들어주는 효과적인 기법으로써, 기존의 방법으로는 알기 어려웠던 정보들을 제공한다. 예를 들어 시각화를 통해 비슷한 분포 또는 특정 패턴이 반복되는 것을 보고 의도적인 침입을 파

악할 수 있다. 즉, 기존의 로그 분석 방법이 침입 발생 후 시그니처를 생성해내는 사후 대처였다면, 시각화를 이용하면 예측 분석에 의한 선제적 대응이 가능해지고 이기종 로그 간의 상호 연관성을 제공할 수 있다[19].

4. 보안 로그 시각화 방법

데이터 특성에 따라 다양한 형태의 시각화가 가능하다. 유용한 정보를 얻기 위해서는 적절한 시각화를 선택하는 것이 중요하다. 여러 대상의 단순히 양의 크고 작음을 비교한다거나 비율을 비교하고 싶을 때는 바 차트와 파이 차트를 사용하며, 혹은 한 변수의 시간에 따른 변화를 보고 싶을 때는 꺾은선 그래프를 이용한다. 다음은 시각화 방법별로 적절한 보안 로그 분석 사례들에 대해 설명한다.

가. 바 차트

바 차트는 가장 기본적인 차트로써, 하나의 축을 기준으로 데이터를 시각화한 차트이다. 데이터 수치를 바의 길이로 표현함으로써 상대적인 크기 차이를 한 눈에 알아볼 수 있다는 장점이 있으며, Kibana에서는 정렬 기능이 있어 데이터가 많이 기록된 순서대로 혹은 가장 적게 기록된 순서대로 정렬 할 수 있다.

네트워크 통신을 할 때 패킷에 IP주소가 남게 되는데, Bro IDS는 로그를 남길 때 이러한 IP주소를 포함한다. 따라서 IP 주소별로 몇 회의 접속이 있었던 지 로그를 보면 알 수 있는데, 바 차트로 표현할 경우 그림 8과 같이 다른 IP 주소와의 상대적인 차이를 확인할 수 있다. 즉, 가장 많이 접속한 IP 주소를 얻을 수 있고, 해당 IP 주소가 다른 IP 주소들에 비해 얼마나 많이 기록되었는지를 확인할 수 있다.

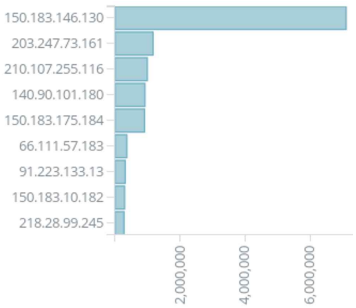

그림 8. IP 주소별 접속 횟수

나. 파이 차트

파이 차트는 비율을 알아보는 데 주로 사용되는 차트로써, 수치를 각도 혹은 면적으로 표현한 차트이며, 최대한 구성요소를 제한하는 것이 가독성에 좋다. Kibana에서는 구성요소의 수를 사용자 임의로 설정할 수 있으며, 불필요한 요소는 필터로 제외시킬 수 있다. 예를 들어 그림 9의 서비스에서 식별이 되지 않는 정보는 '-'로 표시되었는데 이 비율이 절반 이상을 차지한다. 실제 분석가가 파이차트를 구현 할 때 분석가의 판단으로 이를 제외시키고 분석하는 것이 가능하다.

원격 접속을 시도할 경우 이용한 서비스와 프로토콜이 로그로 남게 된다. 이용 가능한 서비스는 정해져 있기 때문에 파이차트가 적합하며, 그림 9와 같이 전체 비율 중 얼마나 차지하는지에 대한 정보를 얻을 수 있다. 이를 이용하면 가장 많이 이용한 서비스가 어느 정도의 비율을 차지하는 지와 다른 서비스들에 비해 얼마나 많이 차지하는지를 알 수 있다.

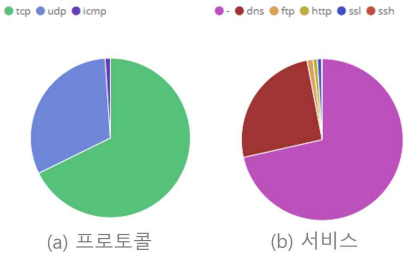

그림 9. 프로토콜 및 이용 서비스

다. 꺾은선 그래프

꺾은선 그래프는 수량을 점으로 표시하고 그 점들을 선분으로 이어 그린 그래프이다. 시간에 따라 지속적으로 변화하는 연속적 데이터는 숫자로만 적으면 흐름을 파악하기 어렵기 때문에 꺾은선 그래프가 유용하다. Kibana에서는 시간 간격을 사용자 임의로 설정할 수 있다. 전체 기간 동안 이상 현상을 관찰하기 위해서는 주 단위 혹은 월 단위로 설정할 수 있으며, 특정일을 기준으로 관찰 할 경우 시간 단위로 설정할 수 있다.

로그에는 항상 이벤트가 발생한 시간(timestamp)이 기록된다. 따라서 로그를 시간 축을 기준으로 이벤트 발생 빈도수를 표시하게 되면 특정 기간 몇 번의 이벤트가 발생 했는지 알 수 있다. 이를 이용하면 시간의 흐름에 따른 이벤트 발생량을 쉽게 파악할 수 있어 이상 현상 분석이 가능하다. 예를 들어 그림 10은 꺾은선 그래프를 이용해 접속 빈도수를 나타낸 것인데, 특정 시간대에 많은 접속이 발생한 것을 확인할 수 있었다.

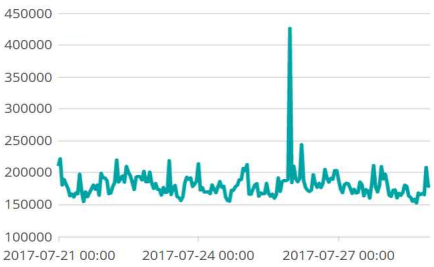

그림 10. 시간대별 접속량

라. 히스토그램

히스토그램은 어떠한 변수에 대해 구간별 빈도수를 나타낸 그래프다. 이를 통해 해당 변수 값의 분포를 알 수 있는데, 히스토그램 모양이 어느 방향으로 치우쳤느냐에 따라 데이터가 어느 값에 몰렸는지를 알 수 있다. Kibana에서는 이러한 변수 값의 범위를 사용자가 지정할 수 있다. 예를 들어 그림 11은 연결 지속 시간을 0.1초의 동일한 구간으로 만든 히스토그램이다.

Bro의 conn.log 파일에는 연결 지속 시간이 기록된다. 해당 로그를 히스토그램으로 표현하면 네트워크 연결 지속 시간이 어느 구간에 몰려있는 지를 파악할 수 있다. 이를 통해 서비스가 정상적으로 운영 되는지를 확인할 수 있는데, 예를 들어 평소와 다른 모양의 분포를 띄고 있다거나, 평소에는 없던 지속 시간이 기록되어 있다면 특정 목적을 위해 의도된 행위임을 추측할 수 있다. 그림 11을 보면 대부분의 연결 지속시간이 0.1초 전에 종료되는 것을 알 수 있으며, 1.0초 이상이 다음으로 많은 것을 알 수 있다. 만약 0.1초와 1.0초 사이의 지속 시간이 많이 기록되었다면 이는 의심해볼 수 있다.

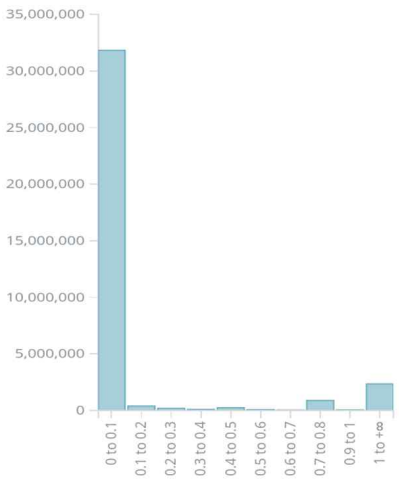

그림 11. 연결 지속시간

마. 히트 맵

히트 맵은 열(heat)과 지도(map)를 결합시킨 단어로, 색상으로 표현할 수 있는 다양한 정보를 지도위에 열 분포 형태의 그래픽으로 시각화한 것이다. 이를 통해 각 지역별로 어느 정도의 접속을 시도했는지 직관적으로 파악할 수 있다. Kibana에서는 기준 값(Threshold)에 따라 색상에 변화를 줄 수 있다. 예를 들어 기준 값을 100으로 잡으면 빈도수가 100보다 많아질수록 빨강에 가까워지고 반대의 경우 보라색에 가까워진다. 기준 값을 너무 높게 잡으면 모든 지역이 보라색에 가깝게 나오기 때문에 적절한 값이 필요하다.

Elastic stack의 Logstash는 IP 주소를 지리 정보로 변환해주는 플러그인이 있어 IP 주소의 빈도수를 지역별 빈도수로 변환할 수 있다. 이를 통해 가장 많이 접속한 IP 주소가 어느 나라에 속한지를 알 수 있을 뿐만 아니라 특정 지역을 기준으로 연관성을 추측할 수 있다. 그림 12는 그림 8의 IP 주소를 변환하여 히트 맵으로 나타낸 것으로 미국에서 가장 많이 접속했음을 알 수 있다.

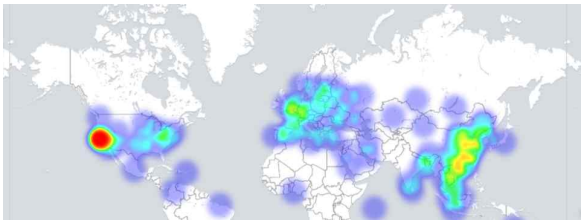

그림 12. 지리 정보

5. 관제 적용 사례

대시보드란 목적에 맞게 필요한 정보를 최대한 빠르게 얻기 위한 모니터링 화면이다. 필요에 따라 적절한 대시보드를 설계하는 것이 중요한데, Kibana는 서로 다른 로그 파일을 하나의 시각화 대시보드에서 보여줄 수 있어 다른 이벤트와의 상호 연관성 분석을 가능하게 한다.

Bro는 패킷을 분석해 여러 종류의 로그 파일을 남긴다. TCP/UDP/ICMP 프로토콜에 대해서는 conn.log라는 파일에 남겨 이용하는 서비스 및 포트 번호 등을 기록하고, 위협이 존재하는 것으로 의심되면 notice.log에 기록하며, 그 외의 정보는 weird.log 파일에 쓴다[20].

가. 대시보드 설계

그림 13은 트래픽 및 행위 분석 목적으로 conn.log 파일만을 이용해 대시보드를 구성한 것이다. 시간대별 접속량, IP 주소, 포트 번호, 연결 상태 등에 관한 정보를 얻을 수 있도록 했다.

그림 14는 위협 탐지 목적으로 conn.log, weird.log, notice.log 파일을 이용해 만든 대시보드이다. Bro가 기록하는 경고 메시지와 트래픽 정보를 한 대시보드에 보임으로써 수상한 패킷이 탐지될 경우 해당 패킷에 대한 다양한 정보를 쉽게 얻을 수 있고, 비정상적인 트래픽 상승과의 상호연관성을 분석할 수 있다.

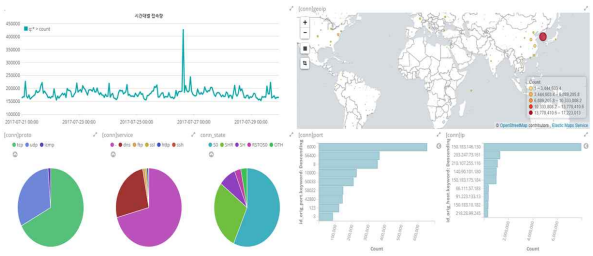

그림 13. 트래픽 분석 대시보드

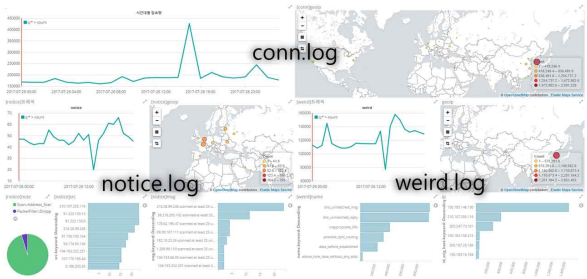

그림 14. 위협 탐지 대시보드

나. 트래픽 분석 사례

트래픽 분석 대시보드를 보면 특정 기간에 트래픽이 많은 것을 확인할 수 있다. 이에 대한 원인을 파악하기 위해 위협 탐지 대시보드에 해당 날짜로 시간 필터를 설정하면 그림 15와 같이 weird.log 파일의 name 필드에서 “dns_unmatched”를 확인할 수 있다.

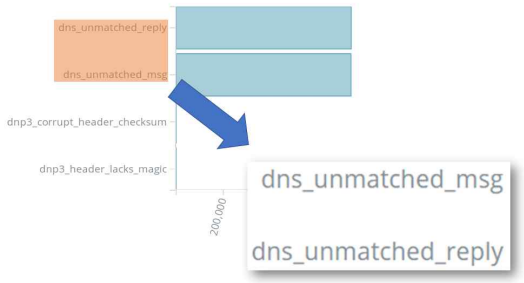

그림 15. weird.log의 name 필드

비정상적으로 많은 DNS 서비스 오류가 기록되었음을 확인할 수 있었으나, 정확한 트래픽 상승의 원인을 알기위해 어떤 IP 주소가 DNS 요청을 했는지 알아낼 필요가 있다. 그림 16은 트래픽이 가장 많이 발생한 기간으로 필터를 설정한 트래픽 분석 대시보드이다. 바 차트를 통해 표시한 IP 주소별 접속 횟수를 살펴보면 “150.183.146.130”이 가장 많이 접속했다는 것을 알 수 있다. 이제 해당 IP 주소가 어떤 서비스를 이용했는지를 알아내야 한다.

그림 16. 특정 기간 필터를 적용한 트래픽 분석 대시보드

트래픽 분석 대시보드는 특정 IP 주소를 설정함으로써 어떤 활동을 했는지 쉽게 파악할 수 있다. 그림 17은 IP 주소 “150.183.146.130”를 필터로 설정한 대시보드로써, 해당 IP 주소가 지도상에서 한국에 위치했음을 알 수 있고, UDP 프로토콜을 사용했으며, DNS 서비스가 대부분임을 알 수 있다. 또한 꺾은선 그래프를 통해 나타낸 시간대별 접속량을 살펴보면 비정상적인 트래픽 상승을 확인할 수 있다. 따라서 트래픽 상승의 원인은 해당 IP 주소가 DNS 서비스를 요청했는데 정상 작동하지 않아서 그림 13에서 남은 것처럼 반복해서 요청했음을 추측할 수 있다.

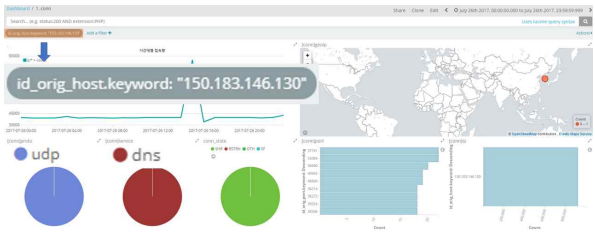

그림 17. IP 주소 필터를 적용한 트래픽 분석 대시보드

다. 침입 경로 분석 사례

트래픽이 많았던 기간에 위협탐지 대시보드를 살펴보면 그림 18과 같이 notice.log 파일의 note필드에서 “Scan::Address_Scan” 이벤트가 남은 것을 확인할 수 있다. 이는 하나의 포트에 대해 여러 번의 Address Scan을 시도했음을 의미하는데, Bro 정책 설정에서 Scan::port_scan_threshold에 설정된 값보다 많은 횟수가 탐지되면 기록된다.

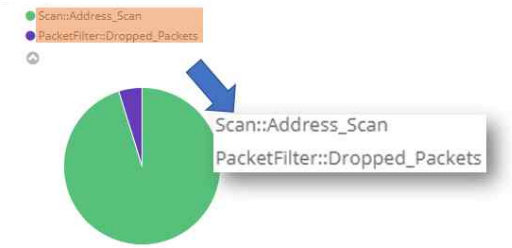

그림 18. notice.log의 note 필드

트래픽 상승 원인은 아니었지만 누가 Address Scanning을 시도했는지 분석해야한다. 먼저 notice.log에서 해당 이벤트를 남긴 IP 주소를 특정해 낸다. 그림 19와 같이 notice.log에 남은 msg 필드를 살펴보면 IP 주소 218.28.99.245가 스캔했다는 기록이 남아있다.

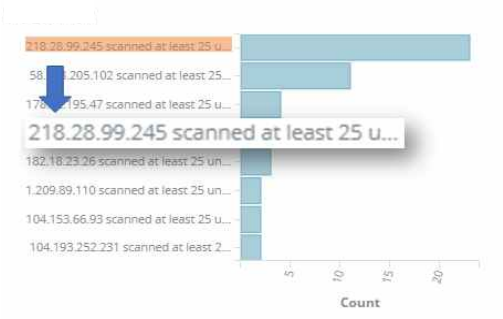

그림 19. notice.log의 msg 정보

이를 이용해 해당 IP 주소를 필터에 설정하고 geoip정보를 살펴보면 중국에서 시도했음을 대시보드만 보고 쉽게 파악할 수 있었다.

그림 20. geoip 정보(218.28.99.245)

III. 결 론

본 논문에서는 보안 관제에서 빠른 정보 전달을 위해 시각화 기법을 제시했다. 네트워크에서 생성되는 로그를 오픈소스 Elastic Stack으로 수집 및 시각화함으로써 최적의 비용으로 보안 관제 기능을 구현할 수 있고, 데이터의 특성에 따라 파이 차트, 바 차트, 지도 등 다양한 방식으로 시각화함으로써 최대한 직관적인 이해를 도울 수 있음을 보였다. 또한, 여러 개의 연관된 그래프를 하나의 대시보드에 종합하여 나타냄으로써 문제 상황에서 필요한 정보를 빠르게 전달할 수 있음을 확인하였다.

본 논문은 네트워크 보안 관제를 문제로 정의하여 Bro IDS에서 생성되는 로그로 한정했지만, 향후 다른 보안 로그에도 적용하여 침입 탐지를 위한 다양한 정보를 얻을 수 있도록 확장할 예정이다.

REFERENCES

[1] 연도별 기술유출 현황(2012~2016), http://www.smba.go.kr (accessed Sep., 27, 2018).
[2] 주요 개인정보 유출 현황(2012~2016), http://www.kcc.go.kr (accessed Sep., 27, 2018).
[3] 남승수, 서창호, 이주영, 김종현, 김익균, "통합 사이버 보안 상황분석을 통한 관제 상황인지 기술," *스마트미디어저널*, 제4권, 제4호, 80-85쪽, 2015년 12월
[4] 차준섭, "통합 사이버 보안 상황분석을 통한 관제 상황인지 기술," *스마트미디어저널*, 제4권, 제4호, 86-92쪽, 2015년 12월
[5] 현정훈, 김현중, "오픈소스 ELK Stack 활용 정보보호 빅데이터 분석을 통한 보안관제 구현," *디지털콘텐츠학회논문지*, 제19권, 제1호, 181-191쪽, 2018년 1월
[6] 장상근, *"네트워크 보안 시스템 구축과 보안 관제,"* 한빛미디어, 2016
[7] ESM의 구성도 및 구성요소, http://www.jidum.com/jidums/view.do?jidumId=608 (accessed Sep., 27, 2018).
[8] Akshaya H L, "A Basic Introduction to DevOps Tools," *International Journal of Computer Science and Information Technologies,* pp. 2349-2353, 2015.
[9] 김성락, "상호연관성 분석을 이용한 웹서버 보안관리 시스템", *한국컴퓨터정보학회논문지*, 제9권, 제4호, 157-165쪽, 2004년 12월
[10] SANS 2017 Security Operations Center Survey, https://pages.endgame.com/rs/627-YBU-612/images/SOC%20Survey%202017.pdf (accessed Sep., 27, 2018).
[11] Heya, Elastic Stack and X-Pacx, https://www.elastic.co/blog/heya-elastic-stack-and-x-pack (accessed Sep., 27, 2018).
[12] Getting started with the Elastic Stack, https://www.elastic.co/guide/en/elastic-stack-get-started/6.4/get-started-elastic-stack.html#install-elasticsearch (accessed Sep., 27, 2018).
[13] Lahmadi, Abdelkader, and Frédéric Beck. "Powering monitoring analytics with ELK stack," *9th International Conference on Autonomous Infrastructure*, Management and Security (AIMS 2015). 2015
[14] Splunk Quick Reference Guide, https://www.splunk.com/pdfs/solution-guides/splunk-quick-reference-guide.pdf (accessed Sep., 27, 2018).
[15] Sung Jun Son, "Performance of ELK Stack and Commercial System in Security Log Analysis," *Malaysia International Conference on Communications*, pp. 28-30, Nov., 2017.
[16] S. Vidhya, S. Sarumathi, and N. Shanthi.

"Comparative analysis of diverse collection of big data analytics tools," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 8, no. 9, 2014.

[17] 이봉환, "아파치 엘라스틱서치 기반 로그스태시를 이용한 보안로그 분석시스템," *한국정보통신학회논문지,* 제22권, 제2호, 382-389쪽, 2018년 2월

[18] Google Trends(2018), https://trends.google.com/trends/explore?date=all&q=elasticsearch%20%2B%20logstash%20%2B%20kibana,splunk,jaspersoft,pentaho&hl=en-US, (accessed Sep., 27, 2018).

[19] 이상준, 이동훈, "빅 데이터 로그를 이용한 실시간 예측 분석시스템 설계 및 구현," *정보보호학회논문지,* 제25권, 제6호, 1399-1410쪽, 2015년 12월

[20] Bro log file Document(2018), https://www.bro.org/sphinx-git/script-reference/log-files.html. (accessed Sep., 27, 2018).

저 자 소 개

조우진

2018년 충남대학교 컴퓨터공학과 학사 졸업.
2018년 충남대학교 컴퓨터공학과 석사과정 재학중

<주관심분야 : 악성코드 분석, 네트워크 보안, 빅 데이터 로그 분석>

신효정

1991년 서울대학교 컴퓨터공학과 학사 졸업.
1994년 서울대학교 컴퓨터공학과 석사 졸업.
1994년~2014년 KT 네트워크연구소 연구원.
2014년~현재 충남대학교 소프트웨어연구소 연구원, 컴퓨터공학과 강사.

<주관심분야 : 네트워크 보안, 네트워크 구조>

김형식

1988년 서울대학교 컴퓨터공학과 학사 졸업.
1997년 서울대학교 컴퓨터공학과 박사 졸업.
1999년~현재 충남대학교 컴퓨터공학과 교수.

<주관심분야 : 시스템 보안, 네트워크 보안, 컴퓨터시스템구조>