빅데이터 처리를 위한 보안관제 시각화 구현과 평가

윤성열, 김정호, 전상준
한밭대학교 컴퓨터공학과

Design and Evaluation Security Control Iconology for Big Data Processing

Yun Seong Yeol, Kim Jeong Ho, Jeon Sang Jun
Dept. of Computer Science, HanBat National University

요 약

본 연구에서는 민간기업들이 전체적인 보안관제 인프라를 구축 할 수 있도록 오픈소스 빅데이터 솔루션을 이용하여 보안관제 체계를 구축하는 방법을 기술한다. 특히, 보안관제 시스템을 구축할 때 비용·개발시간을 단축 할 수 있는 하나의 방법으로 무료 오픈소스 빅데이터 분석 솔루션 중 하나인 Elastic Stack을 활용하여 인프라를 구축했으며, 산업에 많이 도입되는 제품인 Splunk와 비교실험을 진행했다. Elastic Stack을 활용해 보안로그를 단계별로 수집-분석-시각화 하여 대시보드를 만들고 대용량 로그를 입력 후 검색속도를 측정하였다. 이를 통해 Elastic Stack이 Splunk를 대체 할 수 있는 빅데이터 분석 솔루션으로서의 가능성을 발견했다.

1. 서론

1.1 연구배경

G-PRIVACY 2019에서 한국인터넷진흥원(KISA)의 발표자료에 의하면 사이버 침해사고는 대부분 취약한 솔루션을 사용하는 기업의 웹사이트 및 인터넷 기반 신서비스가 해킹에 취약해 저부유출 되는 경우가 80%를 차지한다고 말했다. 이제는 정부, 공공기관 뿐만 아니라 일반 기업체들도 정보보호에 대한 중요성의 인식이 퍼지며 서비스 구축시 다양한 국가 지원사업 및 자체 예산을 통해 네트워크 장비, 서버, 보안 장비 등을 포함한 전산자원을 필수로 구축하고 있지만 문제는 침해사고 발생 및 전산 장애 등을 실시간으로 한눈에 볼 수 있는 시스템의 구축은 일반 기업에서는 비용 및 개발인력부족 등의 이유로 도입을 망설이고 있는 실정이다.[1]

1.2 연구목적

본 논문에서는 빅데이터 분석시스템 도입 시 비용·개발시간을 단축 할 수 있는 하나의 방법으로 무료 오픈소스 솔루션 중 하나인 Elastic Stack에 대해 기술하고, 빅데이터 분석 솔루션중 산업계에서 주로 사용하는 제품인 Splunk와 검색 성능 비교실험을 진행하였다. 이를 통해 Elastic Stack이 Splunk를 대체 할 수 있는 솔루션인지 실험하였다. 오픈 소스코드 솔루션 Elastic Stack을 활용해 로그 분석시스템을 구축했으며, 대용량 보안 로그 분석은 유료 솔루션과 비슷한 성능을 발휘한다는 것을 확인하였다. 또한, 단순한 문자열 분석뿐만 아니라 시각화 분석과 대시보드를 통해 실시간 보안이벤트 처리가 가능한 SIEM(Security information and event management) 솔루션으로의 발전 가능성도 확인할 수 있었다.

2. 관련 연구

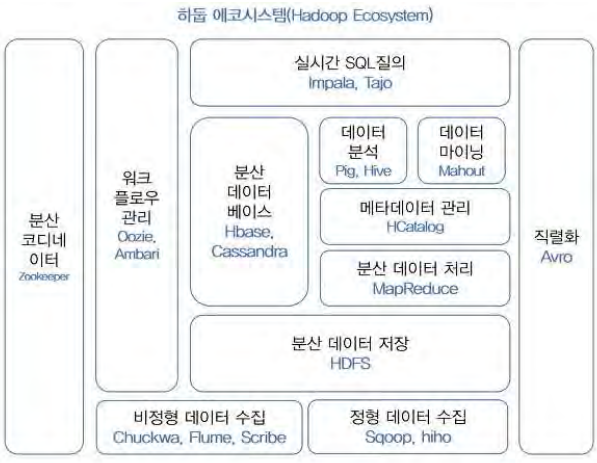
2.1 Splunk

Splunk는 빅데이터 분석솔루션으로 현재 산업계에서 가장 많이 사용되는 솔루션이다. Splunk는 강력한 UI를 지원하고 사용자가 원하는 UI로 변경 가능하며 데이터 즉시 분석이 가능하다. Fortune 100대

기업 중 92명이 이용하고 있으며 그 외에 9,000여 개 이상의 기업, 서비스공급자와 정부가 Splunk 솔루션을 이용하고 있다. 2020년에 Splunk는 7년 연속 '가트너 매직 쿼드런트'의 SIEM(Security Information & Event Management) 부분에서 리더로 선정될 정도로 보안업계에서도 많이 활용되고 있다.

2.2 Hadoop

Hadoop은 아파치 재단의 프로젝트로 빅데이터 처리 플랫폼으로 많이 알려져 있으며 Hadoop은 아파치 재단의 다양한 서브 프로젝트들과 융합되어 하둡 에코 시스템이라는 명칭으로 불린다.



<그림 1> Hadoop Ecosystem

Hadoop은 HDFS(Hadoop Distributed File System) 등을 이용하여 대용량 데이터를 분산저장하고, 저장된 데이터를 빠르게 처리 할 수 있으며 저사양 서버를 이용한 스토리지 구성도 가능하여 가격대비 뛰어난 효율을 보인다. 허나 Hadoop은 다양한 프로젝트들과 융합되어 시너지 높은 모델이 될 수 있지만 다양한 프로젝트 간의 호환성과 보안관제에서 중요한 시각화에 대한 약점이 있어 보안관제 솔루션으로는 적합하지 않다.

2.3 Spark

Spark는 2009년 미국 버클리 대학에서 개발한 오픈소스 스프트웨어로 메모리를 활용하여 빅데이터를 저장하고 처리하기 때문에 Hadoop의 처리 성능에 비해 약 30배 이상 차이난다. 하지만 빅데이터를 분석하기 위하여 원천 데이터를 RDD로 변경하여 메모리로 데이터를 처리하기 때문에 구축 비용이 매우 비싸기 때문에 서브 분석용도로만 사용하고 있다.

2.4 Elastic Stack

<그림 2> Elastic Stack 구성

Elastic Stack은 <그림 2-3>처럼 Elasticsearch+Logstash+Kibana+filebeats로 구성되어 있다. Elastic Stack 중 Elasticsearch는 Apache Lucene를 바탕으로 개발된 검색엔진 솔루션이며. Logstash는 beats 등을 이용하여 수집한 각종 로그를 JSON 형태로 만들어 Elasticsearch로 전송하는 역할을 하고 Kibana는 Elasticsearch에 저장된 Data를 사용자에게 그래프, 테이블 등 시각화 형태로 보여주는 솔루션이다.

2.4.1 Elasticsearch

Elasticsearch는 Lucene을 기반으로 만들어진 분산 검색 엔진이다. Lucene은 손쉽게 검색 기능을 추가할 수 있게 도와주는 자바 형태의 검색 라이브러리이다. 로그 데이터를 검색하고, 분석하는데 사용되는 오픈소스이다. 현재 깃허브, 위키피디아, 넷플릭스 등에서 구축하여 운영 중이다. 본 연구에서는 보안솔루션의 로그를 분석하고 검색, 통계등을 내는 SIEM(Securty Information & Enterprise Management)의 검색엔진으로 사용 하였다.[3]

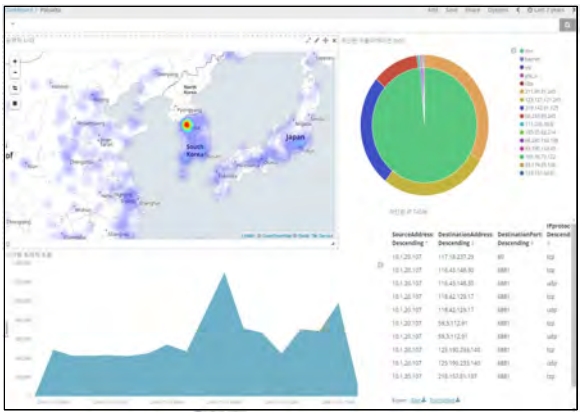
2.4.2 Logstash

Logstash는 보안 솔루션의 로그 데이터를 수집하여 Elasticsearch로 보내주는 역할을 한다. JRuby로 만들어졌으며, 로그를 수집해서 Elasticsearch로 보내주면 JSON형태로 로그 수집 및 가공한다.

2.4.3 KIBANA

Kibana는 Elasticsearch에서 가공한 데이터를 기반

으로 시각화를 해주는 오픈소스 솔루션이다. 시계열 분석등 다양한 방법을 이용해서 로그데이터를 가시화 한다. 지도상 표현이나, 그래프등을 통해 과거의 로그 데이터와 손쉽게 비교를 가능하게 해주어 데이터의 추이를 살펴 볼 수 있다.



<그림3> Kibana를 활용한 시각화 분석 예시

3. 비교실험

3.1 기능

Elastic Stack은 Splunk모두 사용자 편의성에 맞춰 커스터마이징 가능하며 검색쿼리, 데이터 시각화(표,차트, 대시보드 등)를 표현할 때 거의 동일한 기능이 가능하다.

Elasitc Stack **Splunk**

<그림 4> 시각화 Dashboard 비교

3.2 사용의 용이성

주관적인 판단 이지만 Splunk는 완성형 솔루션인만큼 데이터 수집 및 분석과 분석데이터를 시각화 하는 기능이 Elasticsearch에 비해 접근하기 쉬운 편이다. 하지만 Elasticsearch는 AWS 등을 활용하는 클라우드 환경에서 서비스를 배포할 경우 강점을 보이고 있다.

3.3 서비스 지원

빅데이터 솔루션을 활용하여 보안관제센터 구축시에 발생하는 문제를 해결 할 수 있는 커뮤니티는 필요하다. 개발사가 아닌 이를 이용하는 사용자간에도 자유롭게 의견을 나눌 수 있는 커뮤니티 또한 중요한 요소이다. Elasticsearch는 오픈소스 솔루션인 만큼 페이스북 한국그룹에 7400여명의 회원들이 소통하고 있으며 공식 사이트에서 커뮤니티를 운영하고 있어, Splunk에 비해 사용자 참여가 많은 편이다.

3.4 가격 및 지원

오픈소스 프로젝트인 Elastic Stack은 기본적인 구축에는 서버 비용 및 구축인건비 정도의 비용이 들어가며, 자체인력이 구축할 경우 서버 비용만 들어간다. 인공지능 및 클라우드 서비스를 이용할 경우 이에따른 제반 비용이 있지만 이는 Splunk도 동일하며 Splunk는 로그 용량별로 가격을 책정한다.

Index Volume	Perpetual License (per GB)	Annual Term License (per GB)	Volume Purchase Discount
1GB Per Day	$4,500	$1,800	0%
10GB Per Day	$2,500	$1,000	44%
50GB Per Day	$1,900	$760	58%
100GB Per Day	$1,500	$600	67%
>100GB Per Day	Contact sales for custom pricing with additional volume discounts		

<그림 5> Splunk 가격정책

3.5 기술지원

Splunk는 개발시에 자주 사용되는 언어용 SDK를 제작하여 배포할 뿐만 아니라 200개 이상의 RESTful API를 제공하며 문서화 또한 훌륭 하다. Elasticsearch또한 RESTful API를 제공하며 개발언어에 대하여 SDK를 제공한다. 하지만 Splunk는 API를 자체적으로 개발하여 확장하기에는 완성형 솔루션이라 어려운 편이지만 Elasticsearch는 오픈소스 솔루션이므로 다양한 맞춤형 APP을 개발할 수 있다.

제품명	Elastic Stack	Splunk
평가자수	40명	82명
기능	4.5	4.7
사용의용이성	4.3	4.6
서비스지원	4.3	4.5
가격 및 지원	4.4	4.2
기술지원	4.1	4.5

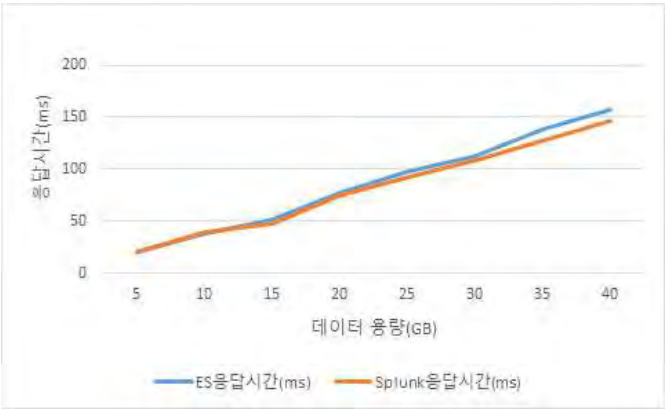
<표 1> Elastic Stack vs Splunk 비교설문

3.6 검색 성능 비교

빅데이터 기반 로그 솔루션 2개의 검색 성능을 비교하였다. 빅데이터라고 하기엔 데이터가 적지만 최대 1억건(40GB) 정도의 방화벽 로그데이터 Elasticsearch, Splunk 각각 index 작업을 거친후 특정IP를 Full Text검색 하였다. 검색 결과 미세하지만 Splunk의 검색속도가 우세 하였다.

ES응답시간(ms)	Splunk응답시간(ms)	용량(GB)
20	20	5
38	40	10
52	48	15
77	75	20
98	92	25
112	108	30
139	128	35
157	146	40

<표 2> 검색 성능 비교



<그림 6> 검색 성능 비교

4. 결론

본 논문에서는 오픈소스 기반의 빅데이터 솔루션인 Elastic Stack을 활용해 로그 분석시스템을 구축하는 방안을 기술하였다. 보안 솔루션의 로그를 수집하여 분석하였고, 이를 대시보드 형태로 시각화 하였다. 기존의 RDBMS(Relational Database Management System) 방식으로는 속도이슈가 있었던 대용량 로그에 대한 분석을 NOSQL인 Elasticsearch를 사용함으로써 검색속도가 개선되었고, 시각화 툴인 Kibana를 활용하여 실시간 보안이벤트 처리가 가능한 대시보드 형태로 구현하였다. 사이버 침해사고 발생시 가장 중요한 사항이 빠른 대응인 만큼 Elastic Stack을 활용한 대용량 로그 분석이 유료 솔루션과 비슷한 성능을 발휘한다는 것을 확인하였다.

참고문헌

[1] 한국인터넷진흥원, “2019년 개인정보 실태점검 이슈와 계획”

[2] 데일리시큐, Splunk 7년 연속 SIEM부분 리더 선정, https://www.dailysecu.com/news/articleView.html?idxno=107058, 2020

[3] 위키북스, 시작하세요! 하둡 프로그래밍, 2014

[4] Grartner , https://www.gartner.com/reviews/market/security-information-event-management/compare/elasticsearch-vs-splunk, 2019

[5] 한빛미디어, 네트워크 보안 시스템 구축과 보안 관제, 2016

[6] 인포더북스, 차세대 정보보호 인재 양성을 위한 보안관제 실무가이드, 2017

[7] 현정훈, “ 오픈소스 ELK Stack 활용 정보보호 빅데이터 분석을 통한 보안관제 구현”, 2017

[8] 한빛미디어, “데이터 시각화의 구현과 분석”, 2016