



국민대학교
소프트웨어융합대학
소프트웨어학부

캡스톤 디자인 I

종합설계 프로젝트

프로젝트 명	SeMo(Security Monitoring Platform)
팀 명	Do Mo!(Do Monitoring!)
문서 제목	2차 중간보고서

Version	1.5
Date	2020-05-28

팀원	전하훈 (조장)
	김성은
	최운호
	최현인
	허윤서
지도교수	윤 명근 교수



CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 소프트웨어융합대학 소프트웨어학부 및 소프트웨어 학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 "SeMo"를 수행하는 팀 "Do Mo!"의 팀원들의 자산입니다. 국민대학교 소프트웨어학부 및 팀 "Do Mo!"의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

문서 정보 / 수정 내역

Filename	중간보고서-SeMo.doc
원안작성자	전하훈, 김성은, 최운호, 최현인, 허윤서
수정작업자	전하훈, 김성은, 최운호, 최현인, 허윤서

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2020-05-13	전하훈	1.0	최초 작성	최초 작성
2020-05-24	김성은	1.1	내용 추가	수정된 연구내용 추가
2020-05-25	허윤서	1.2	내용 추가	향후 추진 계획 수정
2020-05-26	최운호	1.3	내용 추가	수행내용 작성
2020-05-26	최현인	1.4	내용 수정	전반적인 내용 수정
2020-05-28	전원	1.5	내용 검토	전반적인 내용 검토

 국민대학교 소프트웨어학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	SEMO	
	팀 명	Do Mo!	
	Confidential Restricted	Version 1.5	2020-MAY-28

목 차

1	프로젝트 목표	4
2	수행 내용	5
2.1	시연 시나리오 구현	5
2.2	모델-ELK연동	6
2.3	ELK - Web 연동	7
2.4	대시보드 구성	8
2.5	ELK 가이드라인 작성	10
3	향후 추진계획	13
3.1	대시보드 가이드라인 제공	13
3.2	Web FrontEnd 콘텐츠 추가	13
4	개발 현황	14
5	참고문헌	15



1. 프로젝트 목표

Cisco 네트워크 트래픽 전망 조사에 따르면 앞으로 네트워크 트래픽의 양이 방대해 질 것이라고 전망하였고, 과학기술정보통신부에서 발표한 보안분야 매출 현황에서도 보안관제의 매출이 상승한 것으로 보아 앞으로의 네트워크 보안관제의 역할은 중요해진다. 그러나, 이러한 보안관제에 걸림돌인 "오탐(False Positive)"은 현저하게 많은 양의 데이터를 차지한다. KISTI (한국 과학 기술 정보 연구원)로부터 받은 IPS 로그데이터에서 정오탐 비율은 월 기준으로 많게는 1:9 적게는 3:7 (정탐:오탐) 비율이 나타나고 있다. 또한, 논문[1]에서 오탐의 존재는 보안관제사들이 정확한 분석을 하는데 방해를 주고, 오탐을 분석하는데 걸리는 시간이 대부분이라고 주장한다.

따라서 "DoMo!"는 이러한 오탐의 분류를 자동으로 할 수 있는 알고리즘을 개발하고, 이러한 정오탐 분석을 용이하게 해 줄 플랫폼을 구축한다. 해당 플랫폼은 보안관제 분야가 구축되지 않은 기업에 가이드라인이 될 수 있게 오픈소스 형태로 제공한다.

 <div> 국민대학교 소프트웨어학부 캡스톤 디자인 I </div>	중간보고서		
	프로젝트 명	SEMO	
	팀 명	Do Mo!	
	Confidential Restricted	Version 1.5	2020-MAY-28

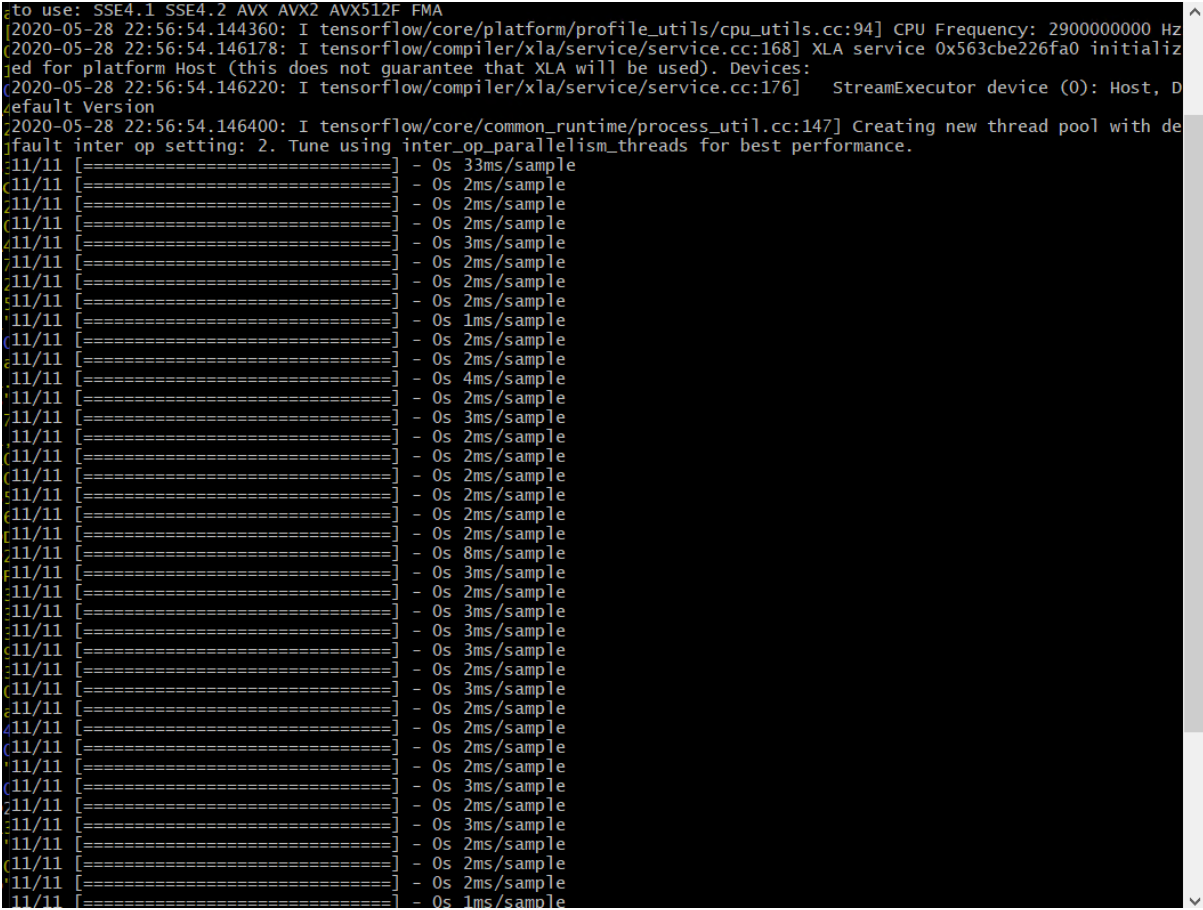
2. 수행 내용

2.1 시연 시나리오 준비

시나리오에 따라 백그라운드 상태에서 프로그램이 진행되도록 python daemon을 사용했으며 총 두 가지의 데몬 파일을 구현했다.

첫 번째 데몬 파일(IPS_daemon)은 실제 IPS 장비에서 데이터가 들어오는 것을 연출하기 위해 일정한 주기로 로그 파일에 데이터를 쌓아주는 역할을 한다.

두 번째 데몬파일은 2.2 에 이어서 설명하겠다.



[그림 1] 실시간 inference 화면

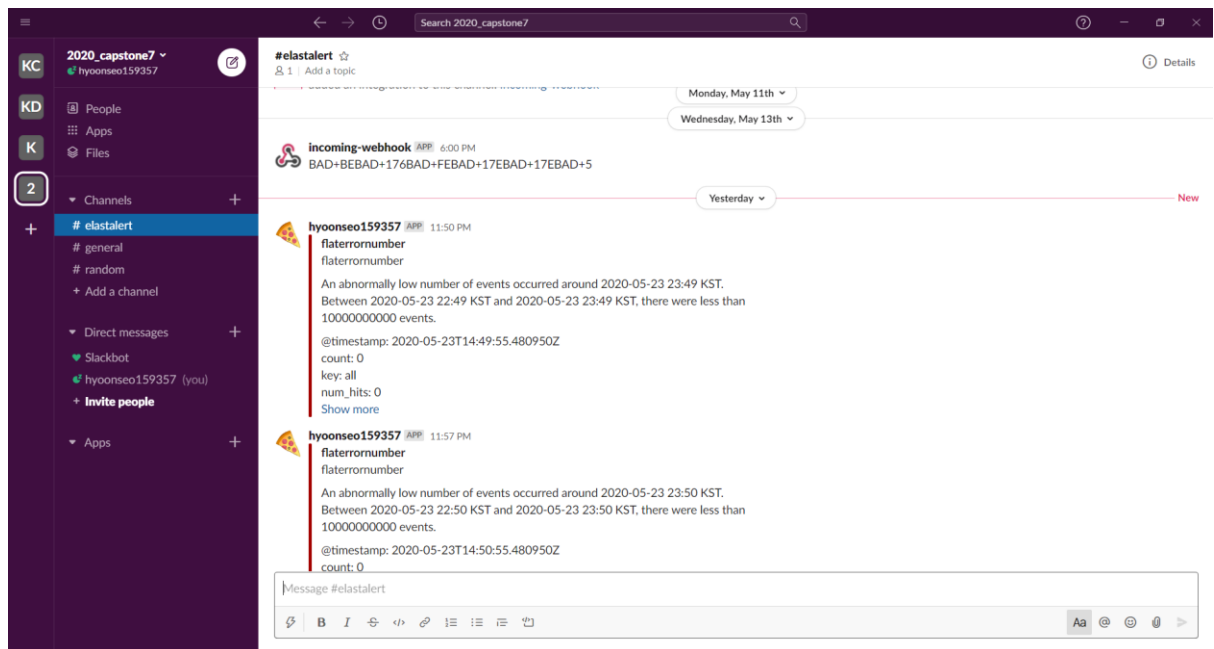
 <div> 국민대학교 소프트웨어학부 캡스톤 디자인 I </div>	중간보고서		
	프로젝트 명	SEMO	
	팀 명	Do Mo!	
	Confidential Restricted	Version 1.5	2020-MAY-28

2.3 ELK - Web 연동

사용자 편의를 위해 Django로 웹프레임을 구축하였다. ELK의 시각화 대시보드를 웹에 임베딩하여 여러가지 대시보드를 활용할 수 있도록 한다.

우선 Kibana에서 대시보드를 구성한 다음 share - Embed code 에서 short url을 카피해 Django의 templates에 적용했다. 보안을 위해 Kibana 데이터에 접근할수 없도록 대시보드만 보이도록 설정하였다. Kibana 대시보드 설정에서 auto refresh를 3초로 설정해주면 Django에서 따로 auto refresh를 설정해주지 않아도 최신 데이터가 계속 업데이트된다.

또한 실행중 에러가 발생할경우 Elastalert가 감지하여 Slack으로 error 알람을 전송한다.



[그림 3] Slack으로 전송된 error 알람 예시

 <div> 국민대학교 소프트웨어학부 캡스톤 디자인 I </div>	중간보고서		
	프로젝트 명	SEMO	
	팀 명	Do Mo!	
	Confidential Restricted	Version 1.5	2020-MAY-28

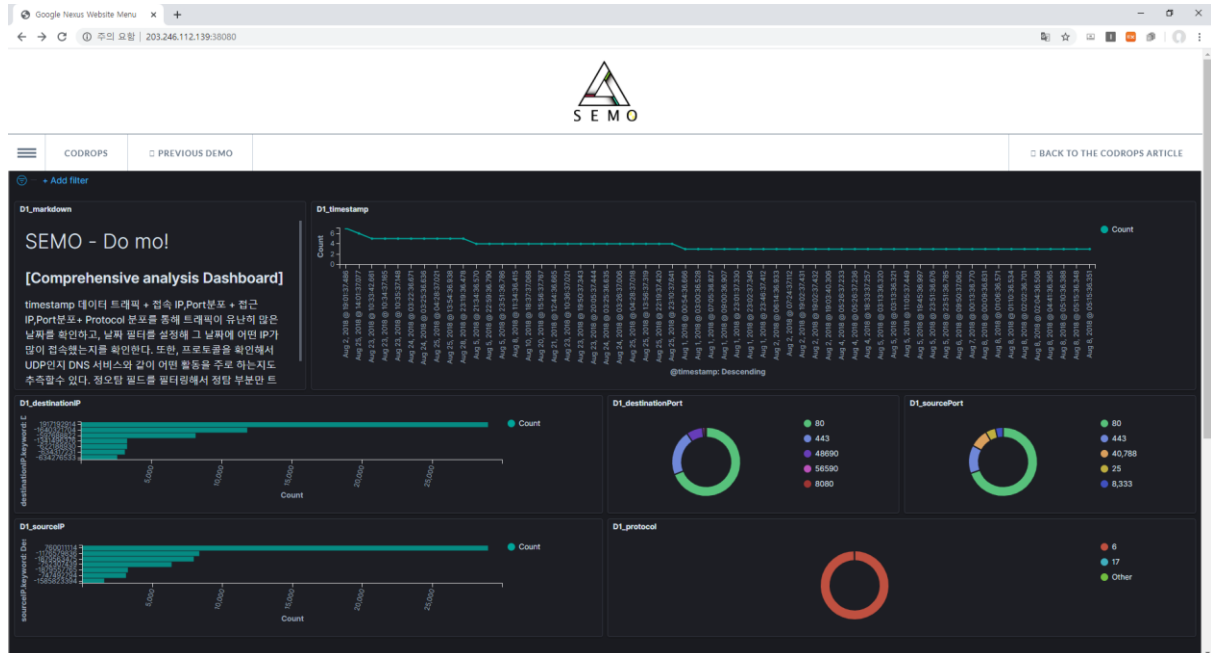
2.4 대시보드 구성

대시보드 구성은 보안관제사들이 분석하기 쉽도록, 분석하려는 목적에 맞게 빠지는 필드 없이 한 눈에 볼 수 있도록 구성해야 한다.

대시보드는 크게 종합분석 대시보드와 위협탐지 대시보드로 기능이 나뉜다.

1) 종합분석 대시보드

예를 들면, 타임스탬프 데이터 트래픽 + 특정 IP 접속횟수 + IP 분포 + 접속 Protocol 분포를 통해 트래픽이 유난히 많은 날짜를 확인하고, 날짜 필터를 설정해 그 날짜에 어떤 IP가 많이 접속했는지를 확인한다. 또한, 프로토콜을 확인해서 UDP인지 DNS 서비스와 같이 어떤 활동을 주로 하는지도 추측할 수 있다. 정오탐 필드를 필터링해서 정탐 부분만 트래픽 분석 대시보드로 필터링 해주면 분석해야 할 양이 훨씬 줄어들 것이다.

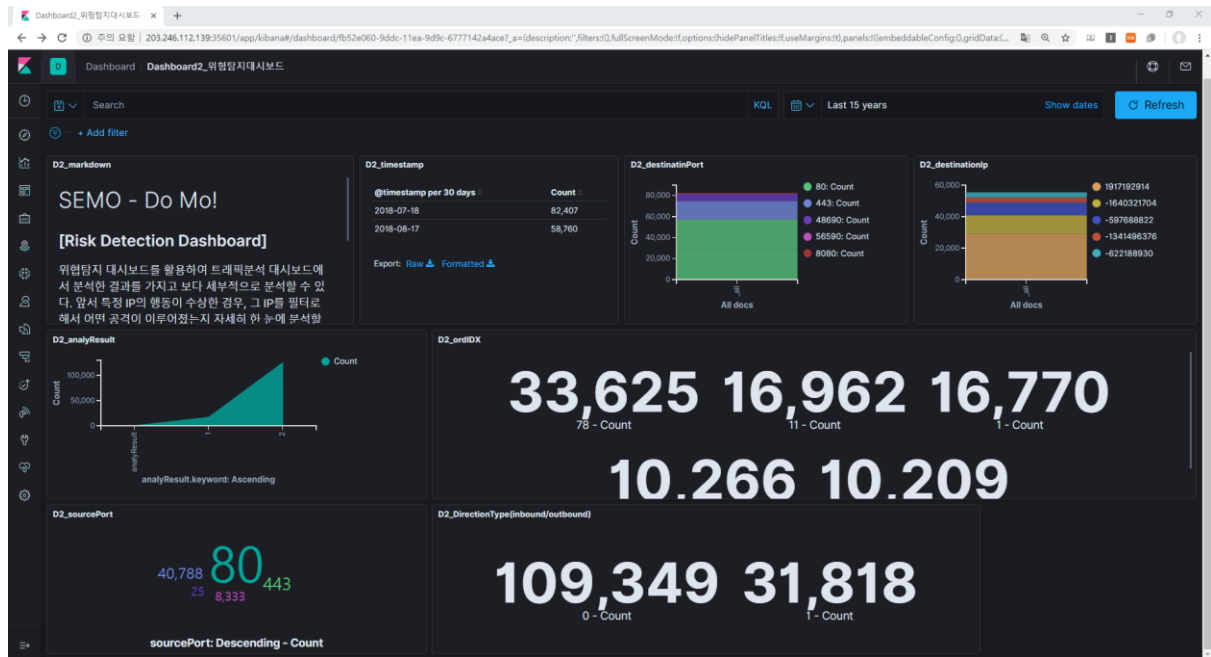


[그림 4] 종합 분석 대시 보드 데모


 국민대학교 소프트웨어학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	SEMO	
	팀 명	Do Mo!	
	Confidential Restricted	Version 1.5	2020-MAY-28

2) 위험탐지 대시보드

위험탐지 대시보드를 활용하여 트래픽분석 대시보드에서 분석한 결과를 가지고 보다 세부적으로 분석할 수 있다. 앞서 특정 IP의 행동이 수상한 경우, 그 IP를 필터로 해서 어떤 공격이 이루어졌는지 자세히 한 눈에 분석할 수 있는 시각화 대시보드를 제공한다. 특정 IP의 공격탐지 대시보드에서 목적지 IP, 포트 번호, 총 이벤트 발생 수, 탐지 장비, 정오탐 필드, 인바운드/아웃바운드 유무 필드들을 포함하여 해당 IP의 행동이 수상한 경우 필터링하여 어떤 특정한 IP나 포트번호에 과한 시도가 있었는지, 이벤트 수가 지나치게 많은지, 정오탐 필드로 수상한 시도를 많이 하였는지를 추측할 수 있다. 이처럼 특정 공격 별로 분석이 용이할 수 있도록 실제 보안관제 대시보드들과 정보보안 공격유형들을 참고하여 대시보드를 구성하였다.



[그림 5] 위험 탐지 대시 보드 데모

 국민대학교 소프트웨어학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	SEMO	
	팀 명	Do Mo!	
	Confidential Restricted	Version 1.5	2020-MAY-28

2.5 ELK 가이드라인 작성

SEMO가 제공하는 best practice를 이용하기 위해선 ELK 스택을 필수로 구축하여야 한다. 보다 쉽게 사용자가 ELK 스택을 구축 할 수 있도록 가이드라인을 작성하였다. 상세한 내용은 다음과 같다.

ELK

해당 가이드라인은 ELK 스택 7.7 버전을 사용하였으며, 설치 및 사용에 관련된 내용은 버전별로 상이할 수 있습니다.

설치 순서

ELK Stack 설치 순서는 다음과 같습니다.

1. Elasticsearch
2. Kibana
3. Logstash (여기까지만 설치하시면 됩니다. 아래는 추가적으로 필요한 경우에 설치하세요)
4. Beats
5. APM Server
6. Elasticsearch Hadoop

이 순서대로 설치하면 각 제품이 의존하는 구성 요소가 제 위치에 있게 됩니다.

Elasticsearch

```
sudo wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.7.0-linux-x86_64.tar.gz
```

[그림 6] ELK 가이드라인

2.5.1 ELK 설치 방법

처음엔 Debian package를 이용한 설치 방법을 제공하려 했으나, 사용자들이 보다 쉽게 설정할 수 있도록 ELK 아카이브를 직접 다운받는 방식을 제공하였다. 이 방식의 장점은 각 제품별로 아카이브안에 모든 설정파일 및 실행파일이 존재하여 사용자가 보다 쉽게 제품별 configuration을 수정할 수 있다는 점이다. 단점으로는 systemd 설정이 불가능하기에 백그라운드에서 실행하기가 쉽지 않다. 구축이 민감한 ELK라 판단하여 단점보다는 장점을 높게 평가했다.

또 ELK 스택은 설치 순서도 중요하기 때문에 이와 관련된 사항들을 문서화하였다.



Elasticsearch

```
sudo wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.7.0-linux-x86_64.tar.gz
sudo wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.7.0-linux-x86_64.tar.gz.sha512
shasum -a 512 -c elasticsearch-7.7.0-linux-x86_64.tar.gz.sha512
tar -xzf elasticsearch-7.7.0-linux-x86_64.tar.gz
```

refer by : <https://www.elastic.co/guide/en/elasticsearch/reference/current/targz.html>

Kibana

```
sudo curl -O https://artifacts.elastic.co/downloads/kibana/kibana-7.7.0-linux-x86_64.tar.gz
sudo curl https://artifacts.elastic.co/downloads/kibana/kibana-7.7.0-linux-x86_64.tar.gz.sha512 | shasum -a 512 -c -
tar -xzf kibana-7.7.0-linux-x86_64.tar.gz
```

Logstash

```
sudo curl -O https://artifacts.elastic.co/downloads/logstash/logstash-7.7.0.tar.gz
sudo curl https://artifacts.elastic.co/downloads/logstash/logstash-7.7.0.tar.gz.sha512 | shasum -a 512 -c -
tar -xvf logstash-7.7.0.tar.gz
```

Configuration

Elasticsearch와 Kibana는 yml 추가 설정이 필요합니다.

Elasticsearch Configuraiton : <https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk/elasticsearch>

Kibana Configuration : <https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk/kibana>

버전이 상이할 경우 각 제품별 설치 방법은 아래 링크 참고

<https://www.elastic.co/guide/index.html>

[그림 7] ELK 설치 가이드라인

2.5.2 Elasticsearch Configuration

Elasticsearch 설치를 마치면 elasticsearch.yml 파일을 통해 Elasticsearch Configuration을 진행 하여야 한다. ELK를 처음 접하는 사람이라면 이 부분에서 가장 힘들어 할 것이다. 왜냐하면 설정이 굉장히 민감하고, Warning이 많이 발생하기 때문이다. 개발과정에서 겪었던 경험을 바탕으로 완성된 elasticsearch.yml 예제 파일을 제공하여 처음 접하는 사람이 이러한 어려움을 피할 수 있도록 하였다.



Elasticsearch

elasticsearch.yml

- elasticsearch.yml guideline

실행 방법

<https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk>

위 설치 가이드라인을 따라 설치를 완료하였다면 실행방법은 다음과 같습니다.

```
[elasticsearch 아카이브가 설치된 경로]/elasticsearch-7.7.0/bin/elasticsearch
```

[그림 8] ElasticSearch Configuration 가이드라인

2.5.3 Kibana Configuration

Kibana 역시 kibana.yml 파일을 통해 Configuration을 진행 한다. 이 또한 예제를 제공하여 Configuration의 어려움을 덜어준다.

kibana

kibana.yml

- kibana.yml example

실행 방법

<https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk>

위 설치 가이드라인을 따라 설치를 완료하였다면 실행방법은 다음과 같습니다.

```
[kibana 아카이브가 설치된 경로]/kibana-7.7.0-linux-x86_64/bin/kibana
```

[그림 9] kibana configuration 가이드라인

logstash

logstash.conf

- logstash configuration example

실행 방법

<https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk>

위 설치 가이드라인을 따라 설치를 완료하였다면 실행방법은 다음과 같습니다.

```
[logstash가 설치된 폴더의 경로]/bin/logstash -f [작성한 logstash configuration 파일 경로]
```

[그림 10] logstash configuration example

 국민대학교 소프트웨어학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	SEMO	
	팀 명	Do Mo!	
	Confidential Restricted	Version 1.5	2020-MAY-28

2.5.4 ELK 실행 방법

사용자가 설치 및 설정을 마치고 바로 실행해 볼 수 있도록 실행 순서와 방법을 제공하였다.

3. 향후 추진 계획

3.1 대시보드 가이드라인 제공

사용자가 원하는 방향으로 대시보드 설정을 할 수 있도록 가이드라인 제공한다. 특정 조건일 때의 그래프와 대시보드 구성 예제를 제공하여 사용자가 참고할 수 있도록 한다.

3.2 Web FrontEnd 콘텐츠 추가

사용자의 요구사항을 고려하며 콘텐츠를 고안하고 요청에 따른 최적화 서비스를 제공하기 위해 적절한 배치와 일관된 디자인 고려한다.



4. 개발 현황

파란색은 완료 상태를 의미하고 노란색은 진행 중인 상태를 의미한다.

항목	세부내용	1월	2월	3월	4월	5월	6월	비고
요구사항분석	요구 분석							
	정보 수집							
관련분야연구	주요 기술 연구							
	관련 시스템 분석							
설계	시스템 설계							
구현	코딩 및 모듈 테스트							
테스트	시스템 테스트							




중간보고서		
프로젝트 명	SEMO	
팀 명	Do Mo!	
Confidential Restricted	Version 1.5	2020-MAY-28

5. 참고 문헌

번호	종류	제목	출처	발행년 도	저 자	기 타
1	논문	ADELE: Anomaly Detection from Event Log Empiricism	https://ieeexplore.ieee.org/document/8486257	2018		
2	논문	DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning	https://dl.acm.org/doi/10.1145/3133956.3134015	2017		
3	논문	ATTACK2VEC: Leveraging Temporal Word Embeddings to Understand the Evolution of Cyberattacks	https://arxiv.org/abs/1905.12590	2019		
4	논문	Tiresias: Predicting Security Events Through Deep Learning	https://dl.acm.org/doi/10.1145/3243734.3243811	2018		
5	기사	2019국내 정보보안 산업매출 3조 2천 700억원...수출액은 1천80억원 기록	https://www.dailysecu.com/news/articleView.html?idxno=106428	2020		
6	보고서	이글루시큐리티 SPiDER TM AI Edition - AI 기반 보안관제시스템 기대 효과	http://www.igloosec.co.kr/brochure/kr/SPiDERTM_AIedition(kr).pdf	2019		
7	참고 사이트	직무인터뷰-sk인포섹	https://blog.naver.com/PostView.nhn?blogId=skinfosec2000&logNo=221407240489&parentCategoryNo=&categoryNo=&viewDate=&isSho	2018		



			wPopularPosts=false&from=postView			
8	참고 사이트	다가오는 인공지능 기반의 보안관제, 그 전에 준비해야할 것은?	http://www.igloosec.co.kr/BLOG_%EB%8B%A4%EA%B0%80%EC%98%A4%EB%8A%94%20%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5%20%EA%B8%B0%EB%B0%98%EC%9D%98%20%EB%B3%B4%EC%95%88%EA%B4%80%EC%A0%9C,%20%EA%B7%B8%20%EC%A0%84%EC%97%90%20%EC%A4%80%EB%B9%84%ED%95%B4%EC%95%BC%20%ED%95%A0%20%EA%B2%83%EC%9D%80[qs]?searchItem=&searchWord=&bbsCatId=1&gotoPage=1	2019		
9	기사	지난해 국내 정보보안·물리보안 산업 매출 규모, 10조5000억원	https://byline.network/2020/02/5-58/	2020		
10	논문	CNN and RNN based payload classification methods for attack detection	https://www.sciencedirect.com/science/article/pii/S0950705118304325#b14	2018		

 국민대학교 소프트웨어학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	SEMO	
	팀 명	Do Mo!	
	Confidential Restricted	Version 1.5	2020-MAY-28

11	논문	네트워크 보안 관제를 위한 로그 시각화 방법	https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artid=ART002428329	2018		
----	----	--------------------------	---	------	--	--