 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09


캡스톤 디자인 I

종합설계 프로젝트

프로젝트 명	Security Monitoring Platform Using Event Logs (SeMo)
팀 명	Do Mo! (Do Monitoring!)
문서 제목	결과보고서

Version	1.5
Date	2020-06-09

팀원	전 하훈 (조장)
	김 성은
	최 운호
	최 현인
	허 윤서
지도교수	윤 명근 교수

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

CONFIDENTIALITY/SECURITY WARNING

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

문서 정보 / 수정 내역


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

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2020-05-30	전하훈	1.0	최초 작성	최초 작성
2020-05-31	김성은	1.1	내용 추가	수행내용 작성 및 수정
2020-06-03	허윤서	1.2	내용 추가	수행내용 작성 및 수정
2020-06-05	최운호	1.3	내용 추가	수행내용 작성 및 수정
2020-06-07	최현인	1.4	내용수정	전반적인 내용 수정
2020-06-08	전원	1.5	내용검토	전반적인 내용 검토


 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

목 차

1	프로젝트 목표	3
1.1	프로젝트 개요	4
1.2	추진 배경 및 필요성	5
1.2.1	추진 배경	5
1.2.2	현재 기술 시장 현황	7
1.2.3	현재 기술 시장 문제점	10
2	개발 목표 및 내용	11
2.1	목표	11
2.2	연구/개발 내용 및 결과물	12
2.2.1	연구/개발 내용	12
2.2.1.1	데이터 Feature 필드 설명	12
2.2.1.2	딥러닝	13
2.2.1.2.1	성능 지표	14
2.2.1.2.2	전처리	15
2.2.1.2.3	모델 선정	16
2.2.1.2.4	Pre-Now Model	18
2.2.1.2.4.1	Pre-Now Model 전처리	19
2.2.1.2.4.2	Pre-Now Model 구조	19
2.2.1.3	ELK	22
2.2.1.4	웹	24
2.2.2	시스템 기능 요구사항	32
2.2.3	시스템 비기능(품질) 요구사항	33
2.2.4	시스템 구조 및 설계도	34
2.2.5	활용/개발 된 기술	35
2.2.6	현실적 제한 요소 및 해결 방안	36

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.2.6.1	하드웨어	36
2.2.6.2	소프트웨어	36
2.2.7	결과물 목록	36
2.3	기대효과 및 활용방안	37
2.3.1	기대효과	37
2.3.2	활용방안	38
3	자기평가	39
4	참고문헌	40
5	부록	43
5.1	사용자 매뉴얼	43
5.2	설치 매뉴얼	44
5.3	테스트 케이스	48

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

1 개요

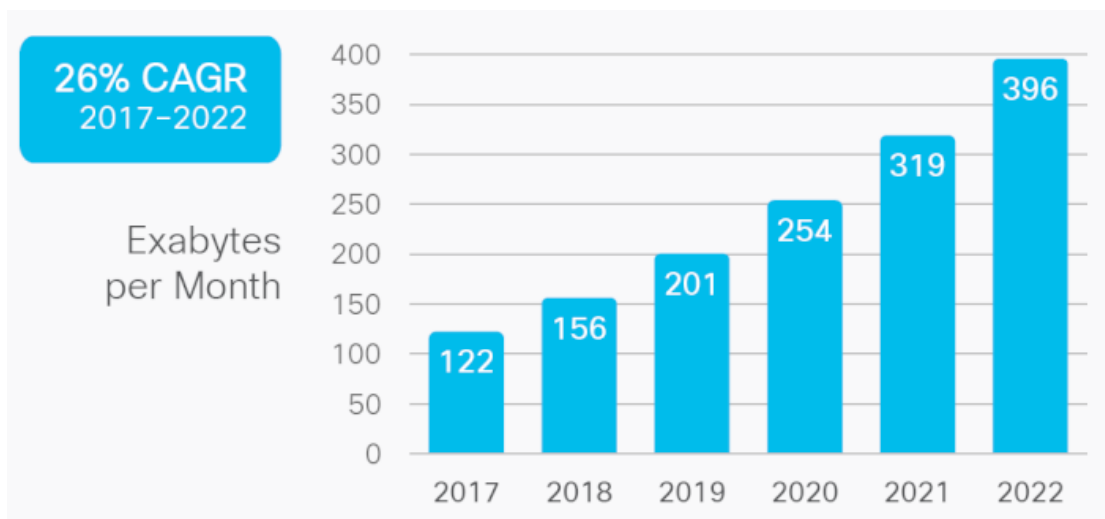
1.1 프로젝트 개요

정보화 시대를 맞아 네트워크 트래픽의 양이 방대해지면서 보안분야에서의 보안 관제의 역할이 더욱 중요해지고 있다. SeMo는 이러한 문제를 해결하고자 보안관제사들이 자동 처리 규칙들을 업데이트하는 데 도움을 주기 위해 고안된 프로젝트이다. 수집된 로그데이터를 받아서 딥러닝을 통해 정탐과 오탐 여부를 판별한 후 분석하여 사용자에게 분석결과를 시각화 하여 웹을 통해 보여준다.

1.2 추진 배경 및 필요성

1.2.1 추진 배경


네트워크 장비회사 CISCO에 네트워크 트래픽 전망 조사에 따르면 앞으로 네트워크 트래픽



양이 방대해질 것이라고 전망했다[그림 1].

[그림 1] 전세계 월별 IP 트래픽 전망

(출처 : Cisco VNI Global IP Traffic Forecast, 2017-2022)

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

이처럼 증가한 네트워크 트래픽 양으로 인해 보안의 중요성이 커지면서 보안관제에 대한 중요성도 커지고 있다. 과학기술정보통신부와 '한국 정보보호 산업 협회'가 발표한 '2019 국내 정보보호산업 실태 조사'에 따르면, 국내 정보보안 서비스 부분에서 보안관제 서비스 매출규모는 2018년대비 4.7%의 매출 규모 상승률을 보이고 있으며, 보안관제 서비스의 수출 성장세는 40.8%(220억원)를 기록했다.

[표 3-10] 정보보안산업 중분류 매출 현황

(단위 : 백만원, %)


구분		2018년	2019년(E)	증감율(%)
정보보안 시스템 개발 및 공급	네트워크보안 시스템 개발	729,393	771,656	5.8
	시스템보안 솔루션 개발	488,402	523,115	7.1
	정보유출방지 시스템 개발	426,128	456,251	7.1
	암호/인증 시스템 개발	151,879	161,760	6.5
	보안관리 시스템 개발	297,920	327,790	10.0
	소계	2,093,723	2,240,572	7.0
정보보안 관련 서비스	보안컨설팅 서비스	302,099	321,478	6.4
	보안시스템 유지관리/ 보안성 지속 서비스	351,942	359,645	2.2
	보안관제 서비스	273,927	286,880	4.7
	보안교육 및 훈련 서비스	1,740	2,990	71.8
	공인/사실 인증서	59,496	66,122	11.1
	소계	989,203	1,037,115	4.8
합계		3,082,926	3,277,687	6.3

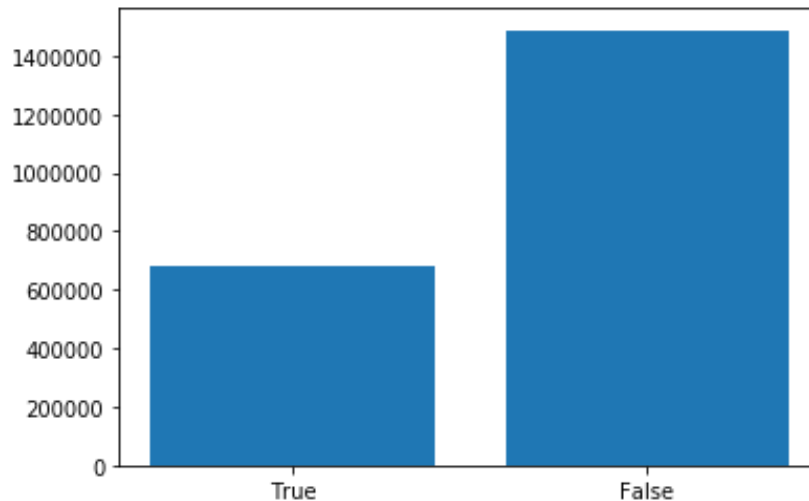
[그림 2] 2019 국내 정보보호산업 실태조사 보고서

(출처: <https://www.dailysecu.com/news/articleView.html?idxno=106428>)

현재, 대부분의 보안관제센터에는 오랜 기간 관제요원들의 업무 경험을 축약시켜서 만든 침입탐지 이벤트 자동 처리 규칙들(Signatures)이 침입 방지 시스템(Intrusion Prevention System) 장비에 집합되어 실행된다.

그러나 보안관제 서비스가 이와 같은 과정을 거침에도 불구하고 오탐(False Positive)이 지속적으로 발생하고 있다는 문제점이 존재한다. 이러한 오탐은 빅데이터 기반 보안관제에 큰 부담을 주고 있으며, 대량으로 발생하는 경보를 처리할 관제 인력의 부족과 경보 내 다수의 오탐은 보안관제사의 업무 효율성을 떨어뜨리는 요소가 되었다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09



[그림 3] KISTI IPS Event Log 데이터 정탐/오탐 비율
(출처: KISTI 사이버보안 안전센터)

1.2.2 현재 기술 시장 현황


1) SK 인포섹



[그림 4] SK infosec 로고

(출처: <http://www.skinfosec.com/>)

SK 인포섹은 보유하고 있는 위협정보, 소속 보안전문가의 분석논리, 글로벌 보안기업과 공유하는 위협인텔리전스 등 양질의 정보를 학습하는 머신러닝 분석 알고리즘을 개발하여 보안관제 시스템 '시큐디움'에 적용시켰고, 국내외 2,000여 곳에 8,000대 이상의 보안 시스템에서 탐지한 이상징후를 수집한 후 다양한 분석규칙을 통해 공격 여부를 판별하고 있다. 또한 머신러닝을 활용해서 탐지 결과에 대한 효과 검증이 가능한 분야인 정탐과 오탐을 자동판정하는 모델을 우선 적용하였다. 그 결과, 머신러닝 적용 전보다 리소스를 70%나 감소시켰고 그만큼 줄

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

어든 리소스를 위협 가능성이 높은 탐지이벤트 분석에 집중시켜 품질을 향상시켰으며, 공격여부를 판별하는 사람의 노력과 시간을 매우 단축시켰다.

2) 이글루 시큐리티(IGLOO SECURITY)



[그림 5] IGLOO SECURITY 로고

(출처: <http://www.igloosec.co.kr/index.do>)

이글루시큐리티는 한국 정보보안 업체 중 하나로, 종합보안관제 서비스를 제공하고 있다. 공공, 금융, 교육, 기업, 해외를 대상으로 솔루션사업과 보안관제 서비스를 제공하고 있으며, 보안컨설팅 사업도 하고 있다. 이글루 시큐리티의 SPiDER TM AI Edition은 ELK Stack을 사용하여 빅데이터를 기반으로 위협 상황의 실시간 식별 및 분석을 수행하고, 머신러닝 기반의 지도학습과 비지도학습을 통해 관제의 효율성을 극대화했다. 또한 보안진단 자동화 솔루션과의 연동으로 사이버 위협 요인을 사전에 해소하고 사이버 위협 정보 공유 시스템을 통해 글로벌 위협 인텔리전스 정보를 제공한다


3) 로그프레스소(LOGPRESSO)



[그림 6] LOGPRESSO 로고

(출처: <https://ko.logpresso.com/>)

로그프레스소는 빅데이터 기반의 통합보안관제와 AI기반의 이상거래차단, 빅데이터기반의 통합로그 관리 등을 제공하는 업체이다. 로그프레스소의 통합보안관제는 실시간 이벤트 연관분석

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

(CEP)와 배치연관분석을 지원하고 있고 비지도 학습 머신러닝 모델을 통해 이상징후를 탐지한다. 또한 드래그&드롭을 통해 분석을 간편화했으며 IP가시성과 CTI연계를 지원한다.


4) 시만텍(Symantec)

시만텍은 미국의 보안소프트웨어 회사로, 시만텍의 MSS-ATP는 네트워크 및 엔드포인트에서 탐지된 내용을 자동으로 비교분석하여 고객사의 환경을 위협하는 요소들을 탐지해 우선적으로 처리하고, 자동 역추적을 통해 공격받은 실제 엔드포인트를 파악한다. 또한 엔드포인트 보안기술과 네트워크 기반의 지능적 보안 위협 탐지 기술을 양방향으로 통합해 신속한 대응이 가능하며 지속적으로 정보를 수집하는 시만텍의 GIN(Global Intelligence Network)과 평판 기반 보안기술인 Insight를 사용해 잠재적 악성파일을 평가하면서 보안 위협 조사의 효율을 높였다.



[그림 7] 시만텍 보안관제 서비스 특징

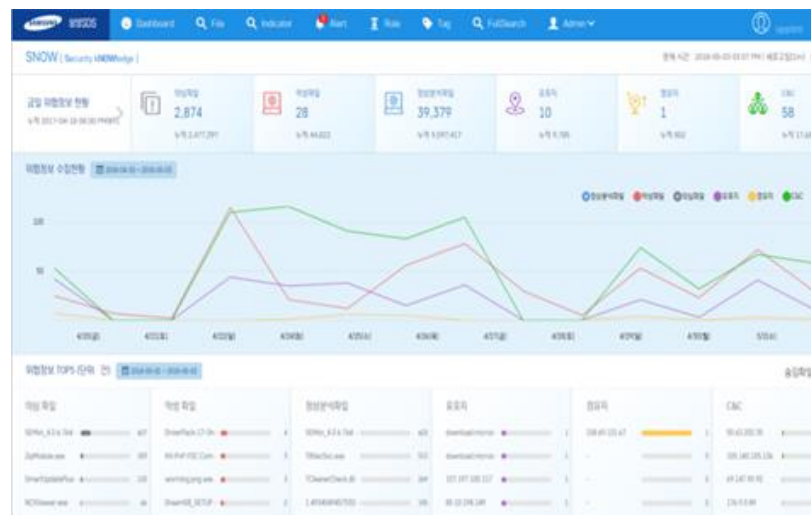
(출처: http://143.127.10.117/content/ko/kr/enterprise/fact_sheets/b-overview-solution-mss-advanced-threat-protection-21332713_KR.pdf)

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

5) 삼성SDS

삼성SDS는 삼성의 ICT를 담당하는 회사로, AI와 클라우드, 블록체인 등의 IT 기술력을 기반으로 다양한 영역에서 서비스를 제공하고 있다.

삼성SDS의 보안관제는 로그, 트래픽, 파일 등에 대해 AI/시나리오 기반 탐지정보를 제공하는 STORM과, STORM에서 제공하는 정보와 TI 정보 및 탐지 결과를 기반으로 분석하는 SNOW를 사용한다.




[그림 8] SNOW 실행화면

(출처: <https://image.samsungsds.com/>)

1.2.3 현재 기술 시장 문제점

현재 기술시장의 정오탐 판별 문제의 현황을 살펴보면, 정오탐의 비율이 약 3:7로 오탐의 비율이 많은 것을 알 수 있다. 보안관제사들이 이러한 오탐에 분석하는 시간을 들이는 것은 능률하락으로 이어진다[7].

이러한 정오탐 분류에 머신러닝을 도입해 판별하는 업체(SK 인포섹, 이글루 시큐리티 등)들은 피쳐 추출을 사용한다. 피쳐 추출 같은 경우에는 전문적인 네트워크 지식이 필요하기 때문에

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

전문적인 인력을 필요로 한다. 또한, 현재 시장에 존재하는 기업들은 보안관제를 용역을 맡겨서 진행한다.

본 프로젝트는 이러한 문제를 해결하기 위해 오탐을 자동으로 분류하고 정오탐 분석을 용이하게 해 줄 플랫폼을 구축한다. 해당 플랫폼은 보안관제 시스템이 구축되어 있지 않은 기업에 가이드라인을 제공하게 된다.

2 개발 목표 및 내용


2.1 목표

IPS 장비는 정해진 규율에 따라 악성 이벤트를 수집한다. 이 때 정오탐 여부를 고려하지 않고 수집하기 때문에 보안관제사들이 수동으로 정오탐을 분류해야 한다.

이 프로젝트는 앞서 언급한 IPS장비에서 수집된 이벤트 로그데이터의 정오탐 여부를 자동으로 분류해주는 것을 목표로 한다. 또한 결과를 시각화하여 사용자에게 보여준다. 이로써 보안관제사들이 정오탐을 수동 분석하고 라벨링하는 수고를 덜어내 능력을 높인다.

세부 목표는 다음과 같다.

1. 정오탐 분류를 수행할 모델을 설계한다. 모델은 KISTI 보안관제사들이 분석한 라벨이 붙은 데이터들로 지도학습 시킨다.
2. 실시간으로 들어오는 데이터들은 전처리를 거친 후 모델에서 정탐 또는 오탐으로 분류된다. 분류된 데이터들은 데이터베이스(Elasticsearch)에 올라갈 수 있도록 후처리하여 CSV 파일로 저장된다.
3. CSV 파일을 Logstash Pipeline을 통해 Elasticsearch에 저장하고 Kibana 대시보드를 구성하여 저장된 데이터들을 시각화한다.
4. 최종적으로 Kibana 대시보드를 웹에 임베딩하여 사용자에게 보여준다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.2 연구/개발 내용 및 결과물

2.2.1 연구/개발 내용


2.2.1.1 데이터 Feature 필드 설명

데이터는 KISTI 사이버보안 안전센터에서 IPS 로그로부터 2018년 8월부터 2019년 3월까지 수집한 이벤트 로그 빅데이터 2,166,757개를 사용한다.

필드명	설명
atdate	DB에 데이터가 입력된 시간
payload	헤더를 포함한 HEX 페이로드
orgIDX	대상기관 인덱스

[표 1] 데이터 Feature 필드 설명

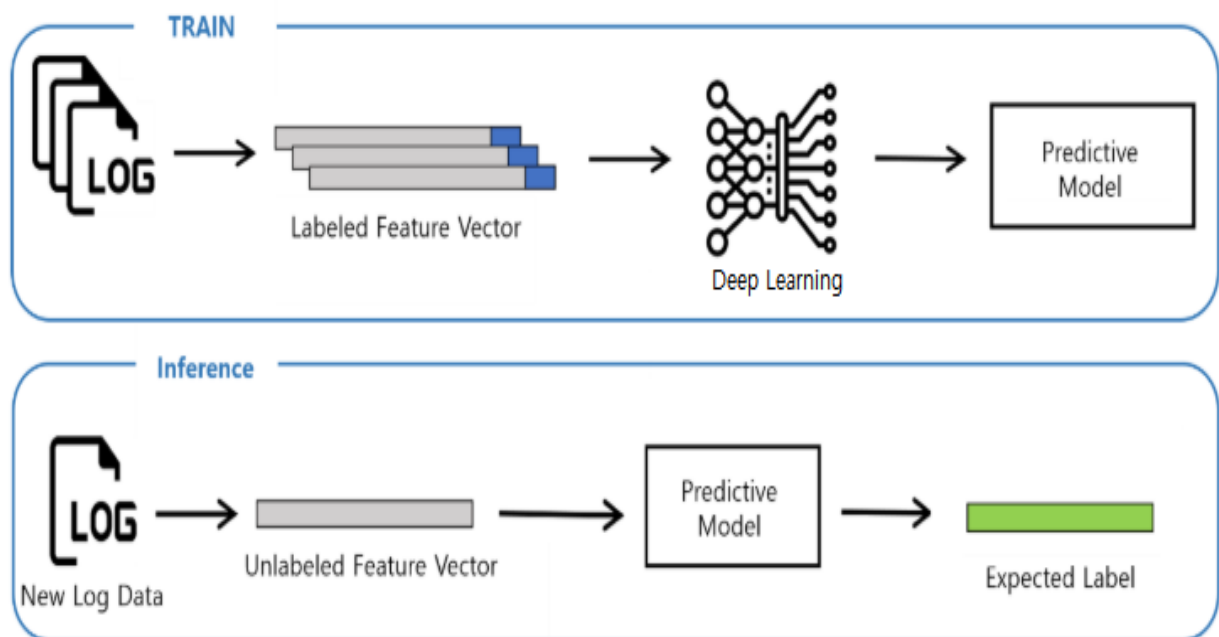
(출처: KISTI 사이버보안 안전센터)

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09


2.2.1.2 딥러닝

IPS 장비 특성상 대상 기관에 최적화 되어 규칙이 적용되고, IPS 장비가 다양하게 있으므로 대다수의 IPS장비에서 적용할 수 있는 Payload, Source IP, Destination IP와 같은 일반적인(general) 필드들을 딥러닝을 통해 학습한다.

그 후, 딥러닝으로 학습된 모델을 바탕으로 새로운 로그 데이터에 대해 예측 과정을 거친 후 라벨링 한다.



[그림 10] 딥러닝

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.2.1.2.1 성능 지표

검증에 사용한 성능 지표는 다음과 같다.

1) Confusion Matrix

		실제 정답	
		True	False
분류 결과	True	True Positive	False Positive
	False	False Negative	True Negative

[그림 11] Confusion Matrix

- a) True Positive(TP) : 실제 정답을 정답이라고 예측 (정답)
- b) False Positive(FP) : 실제 오답을 정답이라고 예측 (오답)
- c) False Negative(FN) : 실제 정답을 오답이라고 예측 (오답)
- d) True Negative(TN) : 실제 오답을 오답이라고 예측 (정답)

2) 정확도 (Accuracy)

예측된 값이 실제 값과 얼마나 가까운지 나타내는 척도이다.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

3) 재현율 (Recall)


실제 정답 중에 예측한 정답을 나타내는 지표이다.

$$Recall = \frac{TP}{TP + FN}$$

4) 정밀도 (Precision)

예측한 정답 중에 실제 정답을 나타내는 지표이다.

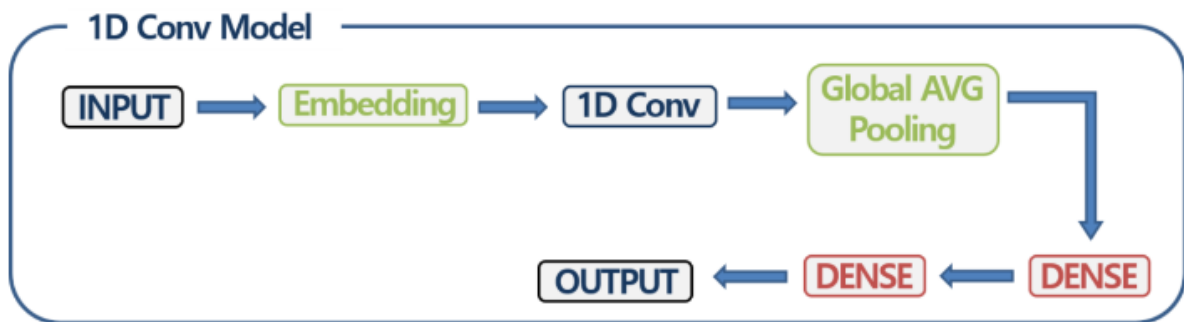
$$Precision = \frac{TP}{TP + FP}$$

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

패딩은 페이로드의 최대 길이인 1600으로 패딩을 진행하고, 해당 값은 등장하지 않는 값인 256으로 패딩시킨다.

2.2.1.2.3 모델선택

1) 1D Conv



[그림 13] 1D Conv 모델 구조

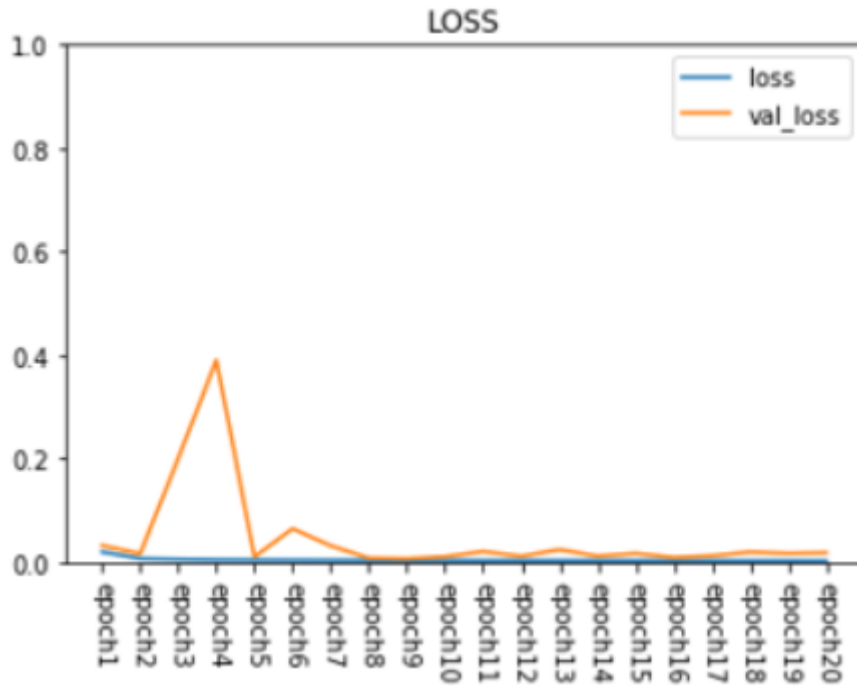
Keras의 1D Conv 레이어를 사용한 모델과 LSTM을 사용한 모델의 성능을 비교하여, Convolution과 RNN 모델 중에 적합한 모델을 선정하기 위해 실험을 하였다. 그 결과, 1D Conv의 성능이 더 좋다고 판단하여 1D Conv를 사용하였다.

2) Keras-embedding layer

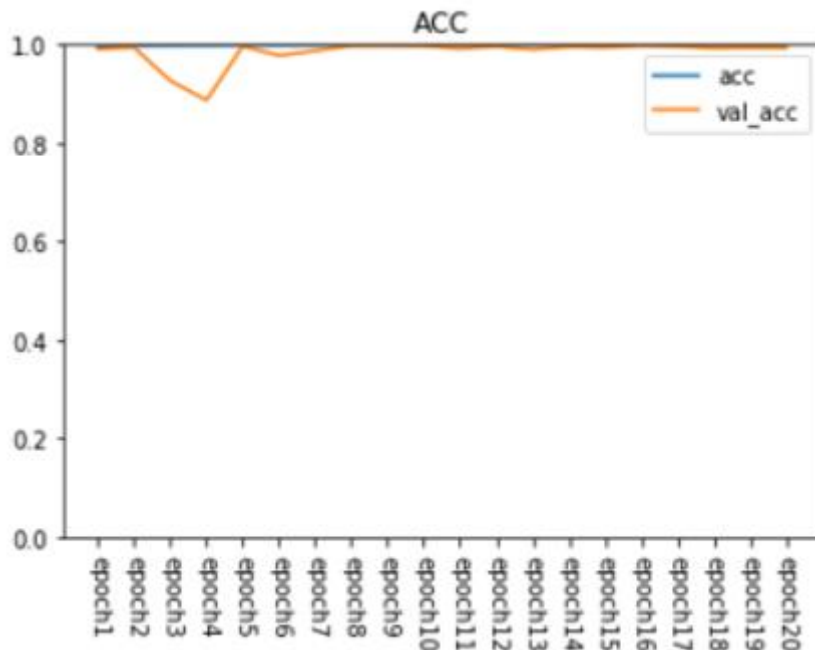
Keras의 임베딩 레이어는 각 값을 의미있는 값으로 임베딩시켜주는 레이어이다. 따라서, 각 페이로드를 1600 x 8 차원으로 변환한 뒤 학습을 진행하였다. 해당 모델은 다음과 같은 그래프를 갖는다.




결과보고서		
프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
팀 명	Do Mo!(Do Monitoring)	
Confidential Restricted	Version 1.5	2020-JUN-09



[그림 14] Keras embedding 결과 LOSS 그래프



[그림 15] Keras embedding 결과 ACCURACY 그래프

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

Accuracy	Precision	Recall
0.9980	0.9839	0.9960

[표 2] Keras embedding 결과표

결과는 [표 2]와 같다.


	Accuracy	Precision	Recall
Raw	0.8781	0.4360	0.9230
One-Hot	0.9967	0.9756	0.9912
Keras	0.9980	0.9839	0.9960

[표 3] Raw, One-Hot, Keras-embedding 비교표

결론적으로, Raw(0~1 정규화), One-Hot, Keras-embedding를 비교하였을 때 가장 성능이 좋은 Keras-embedding을 임베딩 기법으로 사용하였다.

2.2.1.2.4 Pre-Now Model

IPS 특성상 각 기관에 대한 데이터의 양, 특성, 룰셋의 최적화 여부 등 차이점을 보이기 때문에 데이터를 각 기관별로 모은 뒤 실험을 진행하고, 기관 별로 데이터를 모을 경우 직전 페이로드가 현재 페이로드와 연관성이 있다고 생각하여, 모델 구조를 직전 데이터가 현재 데이터에 영향을 주도록 바꾸고 그에 맞게 전처리 방법도 변경하였다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.2.1.2.4.1 Pre-Now Model 전처리


전처리에서는, orgIDX(대상기관 인덱스) 별로 그룹화하여 시간순으로 정렬하고, 정렬한 데이터의 직전 페이로드를 Previous 데이터로 만들었다. 만들 때, 제일 앞에 나온 페이로드의 경우 256으로 패딩한 값이 Previous로 들어갔다[그림 15].

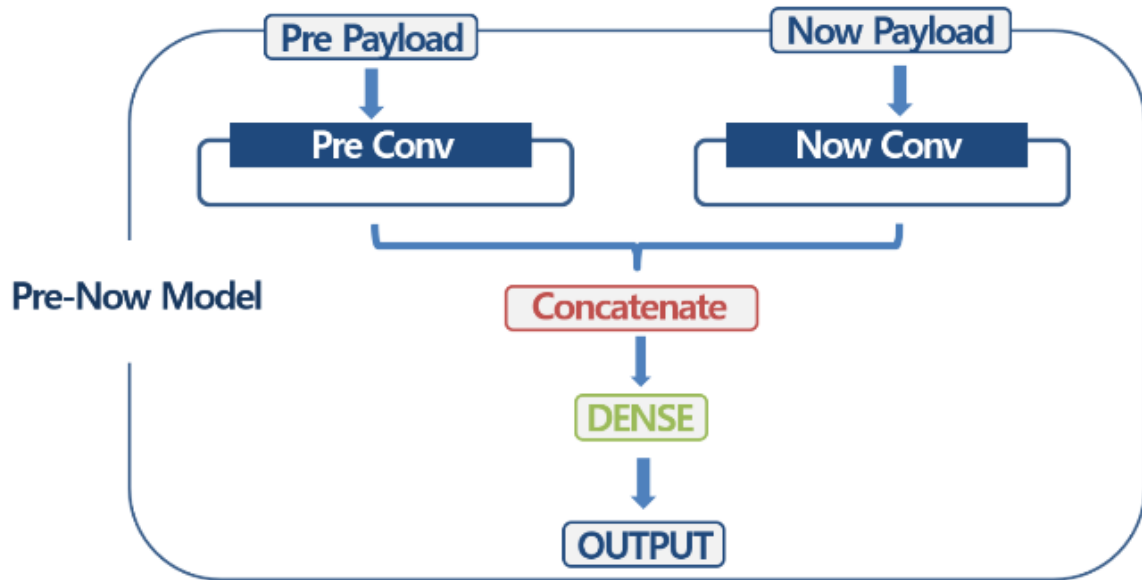


[그림 15] Pre-Now Model 전처리 방법

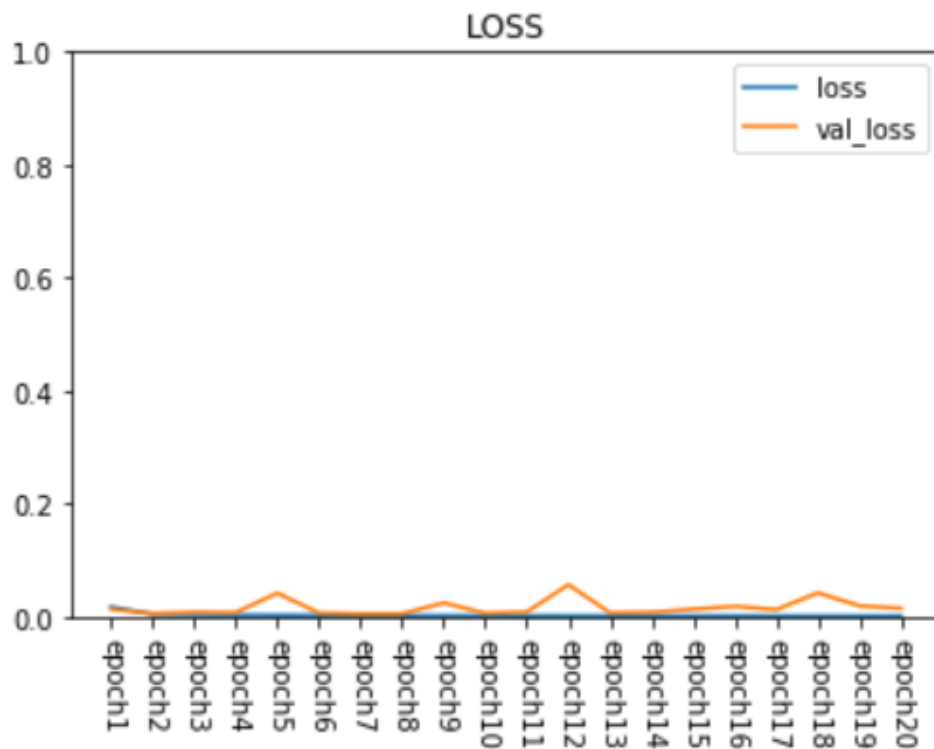
2.2.1.2.4.2 Pre-Now Model 구조

앞서 보여준 Keras-embedding을 사용한 1D Conv를 사용하여, 직전 페이로드와 현재 페이로드를 데이터 인풋으로 넣어주었다. 각각 Conv 레이어가 끝난 후 Concatenate되고 마지막에 Dense 레이어를 거쳐 결과값을 출력하였다[그림 16].


 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

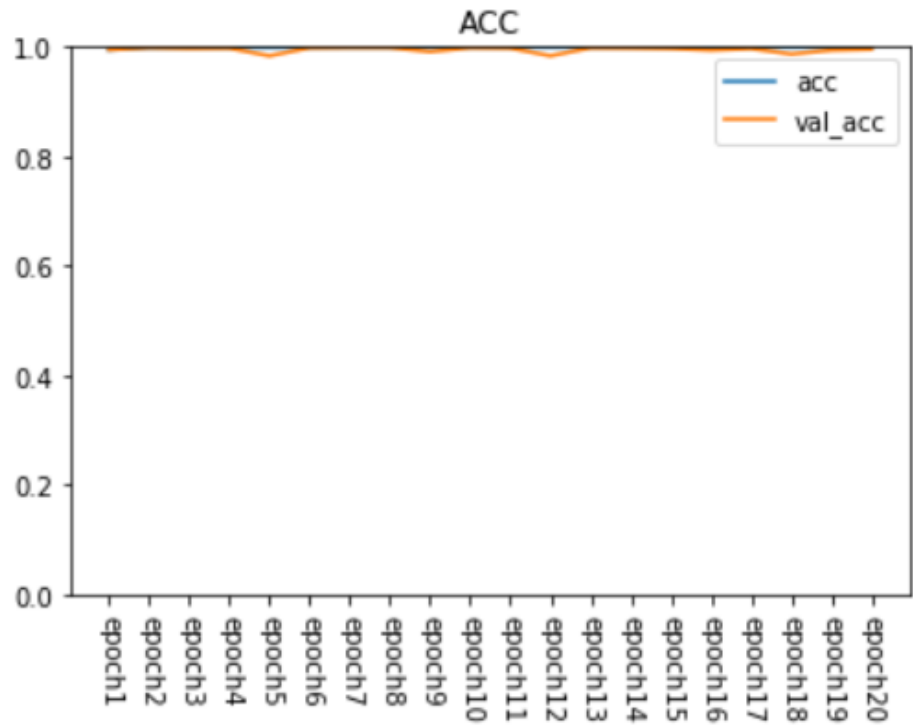


[그림 16] Pre-Now Model 구조



[그림 17] Pre-Now 모델 결과 LOSS 그래프

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09




[그림 18] Pre-Now 모델 결과 ACCURACY 그래프

Accuracy	Precision	Recall
0.9987	0.9904	0.9958

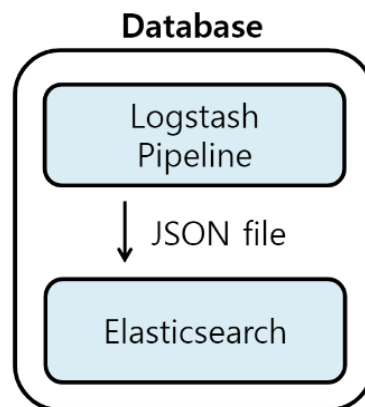
[표 4] Pre-Now 모델 결과표

[그림 17], [그림 18], [표 4]를 통해 알 수 있듯이 가장 좋은 성능을 보였고, Loss와 Accuracy가 보다 빠르게 수렴하였다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.2.1.3 ELK


대용량 데이터를 실시간으로 처리하고 학습된 모델로 판별한 데이터를 시각화하기 위해 ELK를 구축하였다.



[그림 19] Logstash에서 Elasticsearch까지의 흐름

Logstash는 로그 파일로부터 지속적으로 입력을 받아서 Elasticsearch 서버에 데이터를 저장한다. 모델을 거치고 나온 데이터는 판별값으로 라벨링이 되어 나온다. 이 때, 판별값이 정답인 데이터의 'id' 컬럼을 매칭 인덱스로 사용하여 Elasticsearch 데이터에 라벨링 컬럼을 추가한다. 최종적으로 악성 로그를 분석하기 위해 Kibana는 정답으로 판별된 데이터만 시각화한다.

ELK는 7.6.2 버전을 사용하였고, ELK 스택에 필요한 jvm은 Java SE Development Kit 11을 사용하였다. 개발 과정에서는 서버 컴퓨터를 SSH로 작업하고 있기때문에 서버의 IP 주소를 사용하여 Elasticsearch에 데이터를 올리지만 배포되는 버전에서는 localhost로 작업가능할 수 있도록 Logstash configuration 스켈레톤 파일을 제공하였다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09


```

1  input {
2    file {
3      path => "file path"
4      start_position => "beginning"
5      sincedb_path => "/dev/null"
6    }
7  }
8  }
9  filter {
10   csv {
11     separator => ",",
12     columns => ["id","uid","stdrPort","atdate","autoFlag","sourcePort",
13     "payload","eventType","accidentType","destinationIP","packetSize",
14     "vfnStatus","detectEnd","attackType","detectStart","sourceIP",
15     "vfnUpdate","protocol","eventCount","destinationPort","batchID",
16     "detectName","metaType","detailResult","directionType","orgIDX",
17     "autoEmailSendFlag", "jumboPayloadFlag", "analyResult",
18     "doubtFlag","etcInfo"]
19   }
20   date{
21     match => ["atdate", "yyyy-MM-dd HH:mm:ss.SSS"]
22   }
23   mutate {remove_field => ["detectName"]}
24   mutate {convert => ["stdrPort", "float"]}
25   mutate {convert => ["sourcePort", "float"]}
26   mutate {convert => ["eventType", "float"]}
27   mutate {convert => ["accidentType", "float"]}
28   mutate {convert => ["attackType", "float"]}
29   mutate {convert => ["etcInfo", "float"]}
30 }
31 output {
32   elasticsearch {
33     hosts => ["http://localhost:9200"]
34     index => "IndexToElasticsearch"
35   }
36   stdout{}
37 }

```

[그림 20] logstash.conf

[그림 20]과 같이, 데이터는 Logstash Pipeline을 거쳐서 Elasticsearch에 올라가는데, Pipeline의 경우 사용자가 자신의 데이터에 맞춰 변경할 수 있도록 하였다.

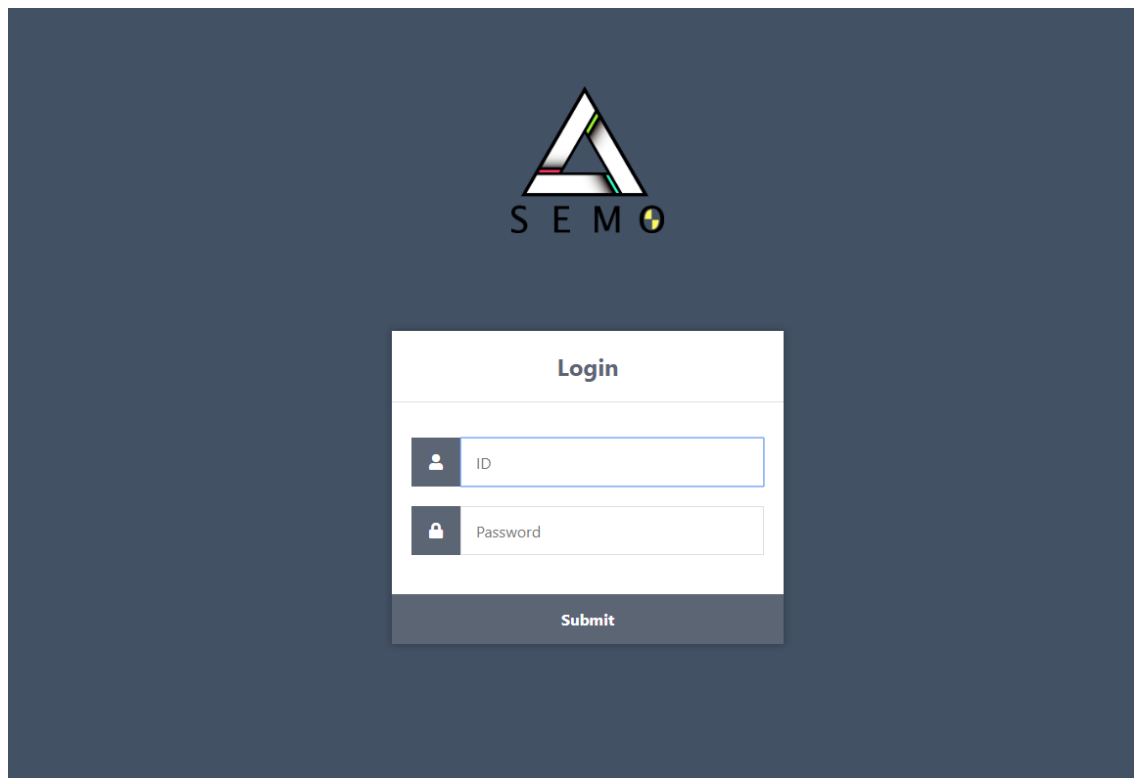
 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.2.1.4 웹


웹은 분석, 시각화 기능을 제공하는 플랫폼을 Django를 기반으로 구현한다. ELK로부터 얻어진 데이터는 로컬 환경에 저장되고, 저장된 파일은 딥러닝을 통해 정오탐 판별과 분석이 된다. 분석결과는 브라우저에 나타난다.

● 웹 UI/UX 구성

로그인 페이지를 구성하여 Django Admin에 관리자권한으로 아이디와 비밀번호를 부여받은 사람만이 해당 플랫폼을 사용할 수 있다. 만약 로그인을 건너뛰고 url로 다른페이지에 접속을 시도했을 경우, 해당페이지에 접근할수 없는 대신 로그인페이지로 이동하도록 하였다.




[그림 21] 로그인 화면

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

● ELK-Web연동

사용자 편의를 위해 Django로 웹 프레임을 구축하고, ELK의 시각화 대시보드를 웹에 임베딩하여 여러가지 대시보드를 활용할 수 있도록 하였다.

우선 Kibana에서 두가지 대시보드를 구성한 다음 share - Embed code 에서 saved object url을 Django의 templates에 적용했다. 키바나에서 오직 대시보드 구성만 보여주기 때문에 데이터 검색이나 삭제등을 수행할 수 없고 필터링만 가능하여 데이터 보안에 효율적이다.

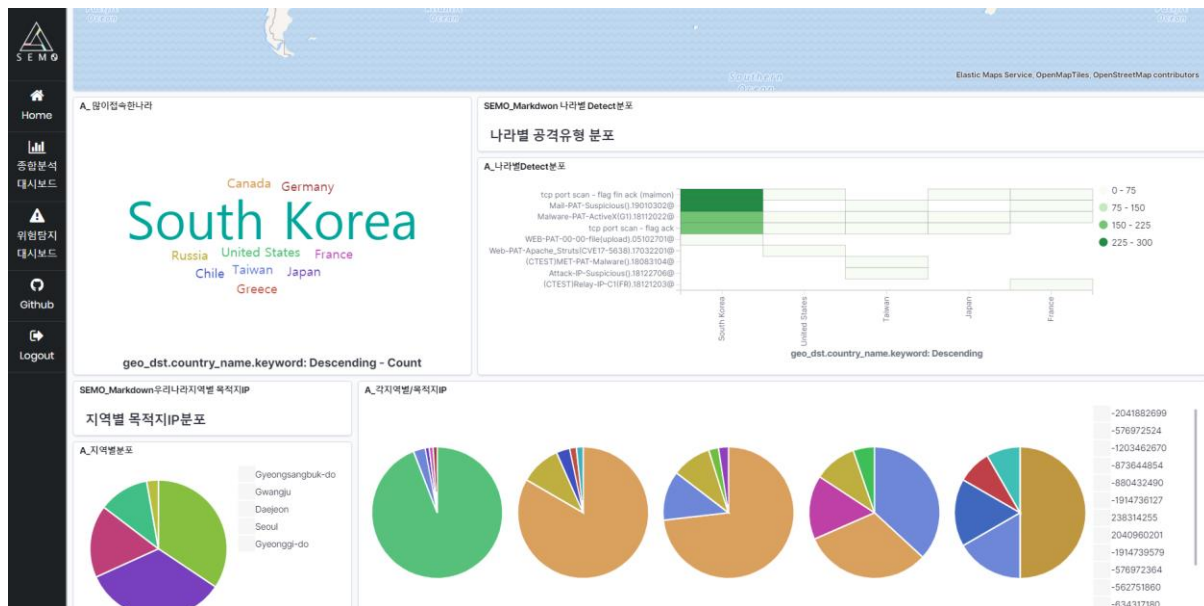
 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

● 대시보드 구성


1) 종합분석 대시보드

기존에 존재하는 대시보드를 벤치마킹하여 데이터 플로우 지도,총 데이터 개수(정오탐 별 개수로 분류), 위험분포 발생건수, DNS request(프로토콜 분포로 대체), 분단위 데이터 흐름을 만들어 주었고, 추가한 항목으로는 Top9 침입별 나라분포, accidentType별 나라 분포, 대상기관 인덱스별 데이터 분포를 추가하였다.

종합탐지 대시보드를 통해 시간별 데이터 흐름과 정오탐별 데이터분포를 확인할수 있으며, 지역별, 대상기관별, 나라별 등등 각 항목별로 데이터 흐름을 확인하여 전체적인 흐름을 체크하면서 튜는 데이터를 확인할 수 있다. 정오탐 필터를 적용하면 확인해야할 데이터수가 줄어 효율적이다. 만약 이상을 감지했을 경우 위험 탐지 대시보드에서 특정 필터를 설정하여 더 자세히 확인해 볼 수 있다.




[그림 22] 종합분석 대시보드

 <div> <p>국민대학교</p> <p>컴퓨터공학부</p> <p>캡스톤 디자인 I</p> </div>	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09



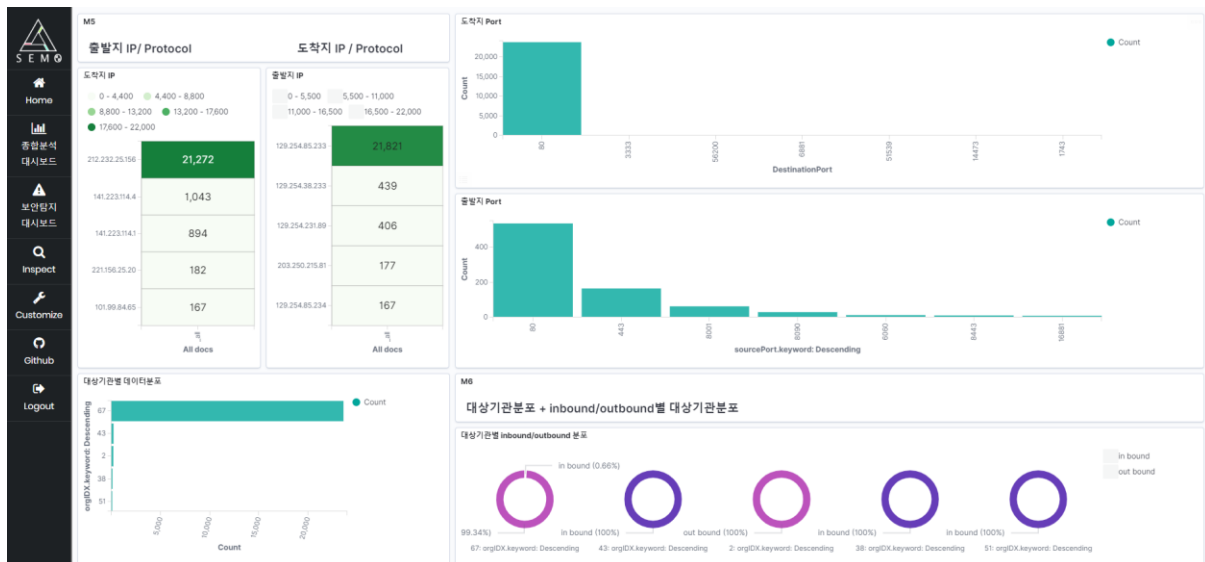
[그림 23] 종합분석 대시보드

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09


2) 보안탐지 대시보드

보안관제 2.0 프로토타입[12]을 벤치마킹하여 보안탐지 대시보드를 구성하였다. 보안관제 2.0 프로토타입에서 따온 항목으로는 출발지 IP, 목적지 IP, 출발 IP별 도착 IP분포, 출발 IP별 도착 Port번호 분포이며, 추가한 사항으로는 출발지Port, 목적지Port, 프로토콜 분포, 대상기관 데이터 분포, 대상기관 분포별 인바운드, 아웃바운드 분포 항목이다.

보안탐지 대시보드에서 정오탐 필터를 적용한다음, 목적지 IP + 이벤트 수 + 탐지장비(대상 기관) 항목조합으로 행위적 위협 패턴을 감지할 수 있고, 아웃바운드 트래픽을 확인함으로써 엔드포인트를 확인할 수 있다. 또한 출발지 IP별 도착지 IP, Port를 확인하여 특정 IP에서 트래픽이 몰렸는지도 확인 가능하다.




[그림 24] 종합 탐지 대시보드

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

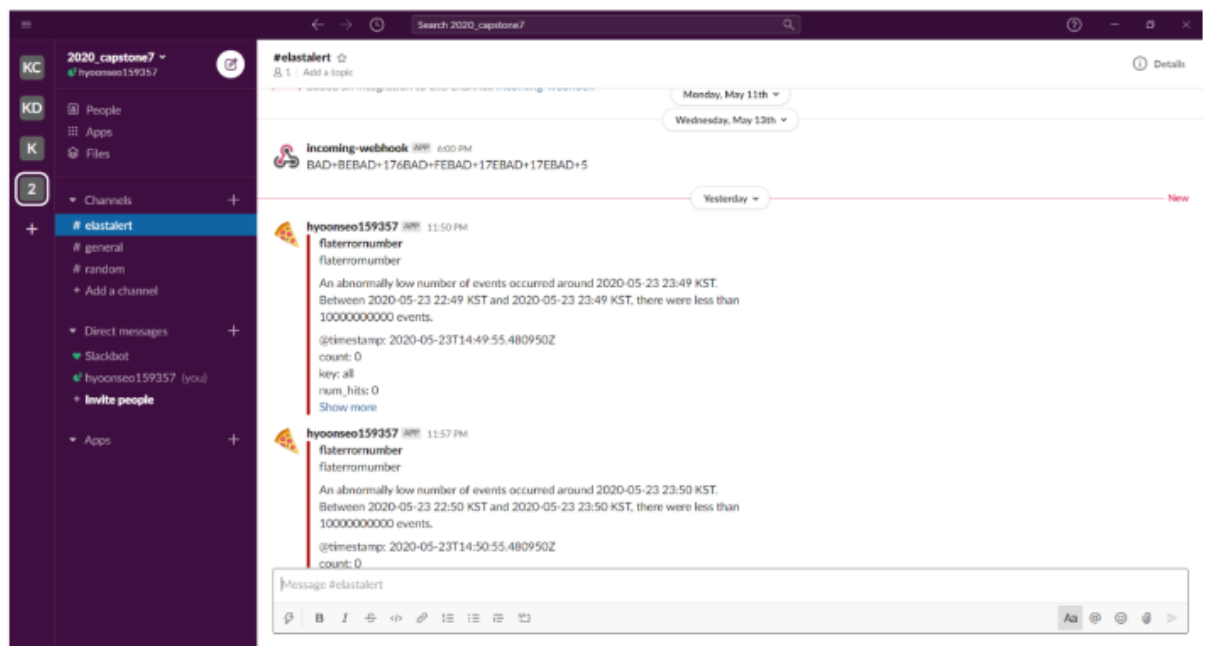


[그림 25] 보안 탐지 대시보드

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

● Elastalert

갑자기 많은 양의 데이터가 들어온다거나, 특정 필드값이 변하였다던가 등 알림을 받기 원하는 룰이 있을 경우, example_rules에서 선택하거나 새로 yaml파일로 생성해준다. 실행 중 해당 조건을 만족시켰을 경우 Slack으로 지정해 둔 work branch로 알림이 전송된다.




[그림 25] Elastalert

● 대시보드 가이드라인 작성

Kibana 대시보드를 구성하는 큰순서와, 대시보드를 구성하는 차트들을 종류별로 Visualize 하는 방법을 가이드라인으로 만들어 활용할 수 있도록 하였다.

● Elastalert 가이드라인 작성

설치방법과 rule 종류 및 작성 예시를 제공하여 활용할 수 있도록 하였다.

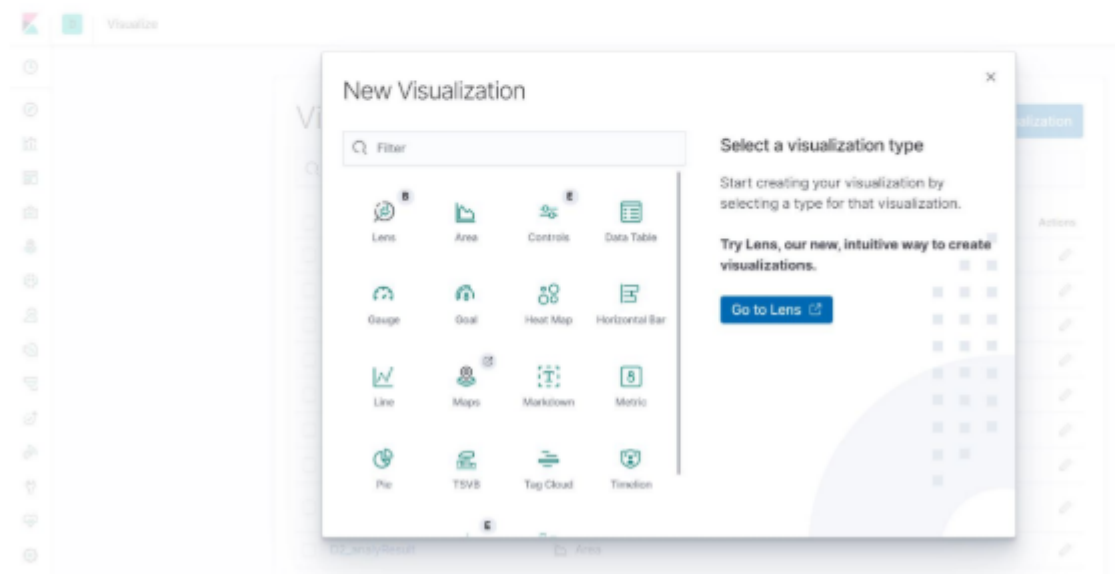
 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

키바나 대시보드 만드는 큰순서


1. Discover -> change index pattern에서 대시보드를 만들 index설정 -> show dates에서 적절한 date범위 조절
2. Visualization -> create visualization -> 시각화하고싶은 방법선택(ex:수직바차트,파이차트) -> data index 를 선택한다> 보이고 싶은 조건에 맞추어 구성
3. Dashboard -> create dashboard -> create new버튼으로 visualization 새로 생성 or add 버튼눌러서 만들어둔 visualization들 중 선택

Visualization 구성방법

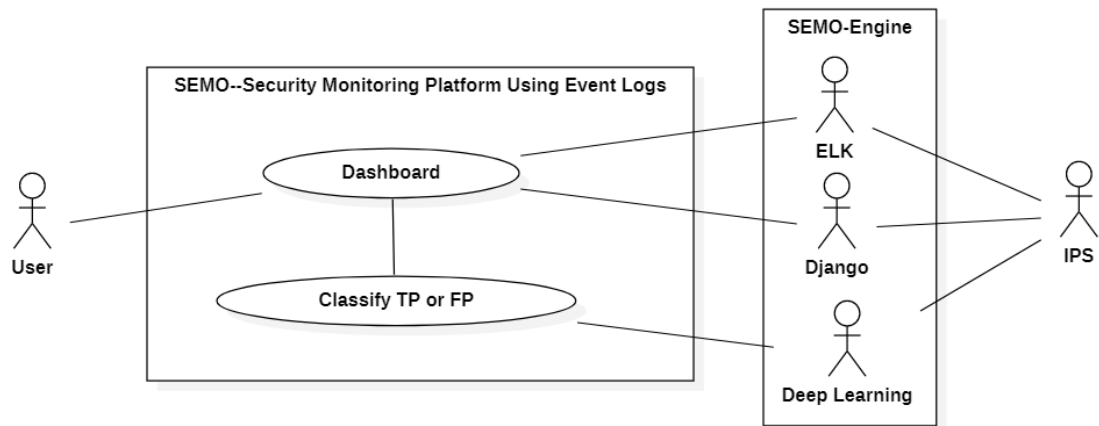
데이터 시각화를 시작하려면 사이드 탐색 메뉴에서 시각화 를 클릭합니다.



[그림 26] Kibana 대시보드 가이드라인

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.2.2) 시스템 기능 요구 사항



[그림 27] Use Case Diagram

4 실시간 데이터 업로드 - 완료

실시간으로 데이터를 업로드한다.

→ IPS장비에서 수집된 실시간 로그 데이터를 Logstash로 받아와 업로드한다.

5 데이터 시각화 (종합분석 대시보드 / 보안탐지 대시보드) - 완료

Kibana 대시보드를 통해 사용자가 원하는 정보를 시각화한다.

→ 실제 쓰이고 있는 보안관제 대시보드들을 벤치마킹하였으며[12][13][14] 전체적인 데이터 흐름 속에서 이상징후를 탐지할수있는 종합탐지 대시보드와, 이상징후를 보다 자세히 살펴볼 수 있는 위험탐지 대시보드로 구성하였다.


6 Login / Logout - 완료

로그인과 로그아웃 기능을 구현한다.

→ Django의 admin 데이터를 활용해 로그인, 로그아웃 기능을 구현하였으며, 비로그인 상태에서 url로 페이지에 접근시도를 했을경우 해당 페이지에 들어갈수없는데 로그인페이지로 이동시켜준다.

7 데이터분석 -완료

사용자들은 시각화된 종합탐지 대시보드와 위험탐지 대시보드를 통해 이상징후를 탐지할수

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

있다.

8 모델을 통한 정오탐 분류 - 완료

IPS장비에서 수집된 이벤트 로그 데이터에 대한 정오탐을 자동으로 분류한다.

2.2.3 시스템 비기능(품질) 요구사항

1) 데이터는 사용자만 접근 가능하다. - 달성


로그인과 로그아웃 기능을 구현하여 사용자만 데이터에 접근할 수 있도록 하였다.

2) 서비스는 24시간 제공된다. - 달성

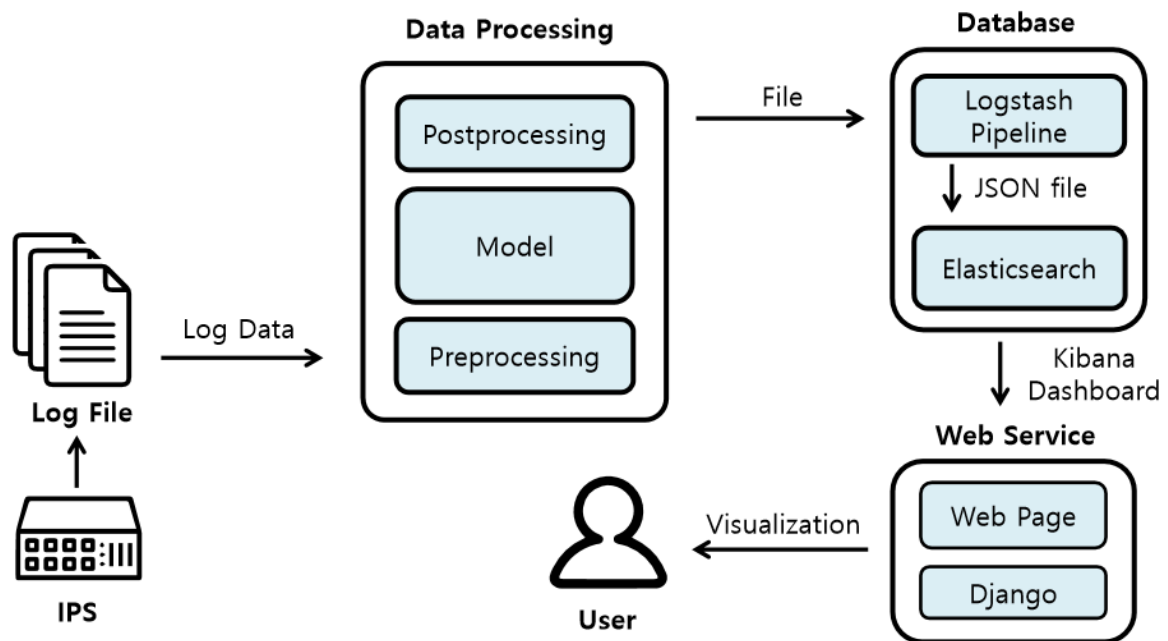
24시간 서비스를 제공함으로써 사용자가 원할 때 언제든지 이용할 수 있도록 하였다.

3) 온라인 서비스(오픈소스)로 가이드라인을 제공한다. - 달성

설치 매뉴얼을 본 프로젝트의 깃허브에 공개하여, 가이드라인을 보고 참고할 수 있도록 하였다. 제공된 가이드라인은 Elasticsearch, Logstash, Kibana, Elastalert 설치와 Kibana 대시보드 구성방법에 대해 다루고 있다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09


2.2.4 시스템 구조 및 설계도



[그림 28] 시스템 구조

시스템의 전체적인 흐름은 다음과 같다.

1. 우선 IPS 장비에서 수집된 이벤트 로그 데이터를 받는다.
2. 이 데이터들은 모델을 적용시키기 위한 전처리를 거친 후
3. 기존의 학습된 모델을 사용해 정오답 분류를 한다.
4. 분류된 데이터들을 Elasticsearch에 올라갈 수 있도록 후처리 한 후 csv파일로 저장한다.
5. Logstash는 이 파일을 읽어서 Pipeline을 통해 Elasticsearch로 전달한다.
6. Elasticsearch의 데이터들은 구현된 대시보드를 통해 시각화한다.
7. 최종적으로 Kibana 대시보드를 웹에 임베딩하여 사용자에게 보여준다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.2.5 활용/개발된 기술

1) Convolutional Neural Network

정오탐 분류 모델을 구축하기 위해 Convolutional Neural Network(CNN)을 사용하였다.

CNN은 심층 신경망(DNN)의 한 종류로, 하나 또는 여러개의 convolutional layer와 pooling layer, fully connected layer들로 구성된 신경망이다. 직전 페이로드와 현재 페이로드가 유의미한 관계라고 생각하여 현재 페이로드를 학습할 때 직전 페이로드를 고려하여 학습하도록 하였다. 현재 페이로드와 직전 페이로드를 데이터의 인풋으로 넣어주고, 각각 Conv 레이어가 끝나고 난 후 Concatenate된다. 마지막으로 Dense 레이어를 거쳐 정탐과 오탐을 분류한 결과값을 출력한다.

2) ELK Stack

모델을 거쳐 정탐과 오탐으로 분류된 데이터들을 시각화하기 위해 ELK Stack을 사용하였다.

ELK Stack은 검색과 분석 엔진인 Elasticsearch, 데이터를 수집 및 변환 후 Elasticsearch로 전송하는 Logstash, 차트와 그래프를 이용해 데이터를 시각화하게 해주는 Kibana를 말한다.

대용량의 데이터를 실시간으로 처리하기 위해 Elasticsearch와 Logstash를 사용하였다.

Logstash는 로그 파일로부터 지속적으로 입력을 받아서 Elasticsearch 서버에 데이터를 저장한다. 이때, 데이터는 Logstash Pipeline을 거쳐서 Elasticsearch에 올라가고 Kibana의 대시보드를 통해 시각화된다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.2.6 현실적 제한 요소 및 해결 방안

2.2.6.1 하드웨어

웹 서버에 엔진이 정상적으로 확장될 수 있는 서버 사양이 되어야 한다. 딥러닝 모델을 학습할 때 많은 양의 데이터를 학습할 경우 오랜 시간이 소요된다.

이는 GPU를 사용함으로써 모델이 학습되는데 걸리는 시간을 단축할 수 있다.

2.2.6.2 소프트웨어


실시간으로 로그를 분석하는 것이 주목적이므로 웹 서버와 엔진 간의 통신이 원활하게 되어야 한다.

또한, IPS 장비마다 로그 파일의 형식이 다르기 때문에 범용적인 소프트웨어를 만들어야 한다.

따라서 raw 데이터에서 수집 가능한 가장 일반적인(general)한 feature를 추출한다.

2.2.7 결과물 목록

1. Semo 정오탐 분류 모델
2. Semo 보안관제 플랫폼
3. 플랫폼 사용 가이드라인

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.3 기대효과 및 활용 방안


2.3.1 기대효과

네트워크 트래픽의 양이 방대해지면서 보안분야에서는 보안 관제의 역할이 더욱 중요해지고 있다. 본 프로젝트는 보안관제사들의 자동 처리 규칙 업데이트를 도와주기 위해 고안됐다. 이 플랫폼으로 인해 보안관제사들은 효율적으로 트리거를 업데이트 할 수 있다. 뿐만 아니라 수작업으로 이루어지던 정오탐 판별의 시간적 낭비가 사라진다.

우선적으로, 매일 새롭게 생성되는 방대한 보안 이벤트 분석을 자동화함으로써, 보안 업무의 효율성을 높일 수 있다. 우선 처리해야 할 고위험 이벤트를 선별함으로써 방대한 보안 이벤트 분석에서 소요되는 시간을 단축하여 보다 빠른 대응이 가능해진다. AI 알고리즘에 적용할 학습 데이터를 생성하고 AI 시스템에 내린 결과에 피드백을 주는 과정을 반복함으로써 예측의 정확성을 끌어올릴 수 있다. 예를 들어, 기존 상태에서 약간의 변화만 있는 신·변종 악성코드의 경우, 이전에는 보안 관리자가 일일이 보안 이벤트를 분석해 처리해야 했다. 하지만, 방대한 데이터를 학습한 AI 시스템을 구축해 단순한 공격은 자동 처리하게 하고 보안 담당자는 고위험군 이벤트 분석에 집중한다면, 이벤트 처리 효율성을 비약적으로 높일 수 있게 될 것이다.


더불어, 장기간 축적된 보안 데이터를 AI 알고리즘을 통해 분석함으로써, 날로 진화하는 고도화된 보안 위협을 보다 빠르고 정확하게 탐지하고 위협 대응에 소요되는 시간을 단축시킬 수 있다. 공격자들이 장기간에 걸쳐 기업 내부 시스템들을 교묘히 옮겨 다니며 공격하는 만큼, 단기간 수집된 보안 데이터 분석만으로는 공격자의 행위를 정확하고 빠르게 탐지하기 어려웠던 것이 사실이다.

딥러닝 기반의 AI 알고리즘을 통해 보안 데이터, 최신 위협 정보, 취약점 등 관련된 정보를 신속하게 연관 분석함으로써, 기업 전반을 아우르는 폭넓은 가시성을 확보할 수 있게 된다. 또한 악의적 행위·공격자 특성 등이 담긴 양질의 학습 데이터에 대한 비지도 학습을 통해 심각한 위협으로 발전할 수 있는 알려지지 않은 변칙 활동 및 이상행위를 보다 빨리 식별할 수 있게 된다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

2.3.2 활용 방안


보안관제 플랫폼 가이드라인을 제시함으로써 기존의 자체적인 보안관제 시스템이 구축되어 있지 않은 회사들은 자체적인 보안관제 시스템을 구축하여 활용할 수 있다. 또한 인공지능 보안관제 기술을 개발함으로써 정탐과 오탐의 자동 분석이 가능해져 보안관제사들의 업무 능력을 향상시킬 수 있다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

3 자기평가


현 기술시장에서 정오탐 판별 비율 중 오탐의 비율이 절대적으로 높다. 보안관제사들이 이러한 오탐에 분석하는 시간을 들이는것은 능률 하락으로 이어진다. 본 프로젝트에서는 IPS장비에서 수집된 이벤트 로그데이터의 정오탐 여부를 딥러닝을 활용해 자동으로 분류해주고, 이를 시각화하여 사용자에게 보여준다. 이로써 보안관제사들이 정오탐을 수동 분석하고 라벨링하는 수고를 덜어내 능률을 높이는것을 목표로 하였다.

- 1) 사용자가 간편하게 웹페이지에 접속하여 사용할 수 있어 접근성이 좋다.
- 2) 정오탐여부를 분류해주기 때문에 보안관제사들이 수동분석 후 라벨링하는 수고를 덜어주어 능률을 높일 수 있다.
- 3) 다양한 visualization을 제공함으로써 사용자가 필요로 하는 목적에 맞게 선택하여 활용할 수 있다.
- 4) 커스텀 가능한 대시보드를 제공하여 상황에 맞게 자체 구성할 수 있다.
- 5) 보안관제 플랫폼 구축을 위한 best practice 을 제공한다.


 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

4 참고문헌


번호	종류	제목	출처	발행 년도	저자	기타
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 2천700억원...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=&isShowPopularPosts=false&from=postView	2018		

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

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 산업매출 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		
11	논문	네트워크 보안 관제를 위한 로그 시각화 방법	https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artId=ART0024	2018		

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

			28329			
12	기사	보안관제 2.0 프로토타입- 인공 지능의 궁극적인 혜택 (위험탐지대시보드 참고)	9 http://m.comworld.co.kr/news/articleView.html?idxno=49740	2019		
13	참고 사이트	시야인사이트 - 시야 MilkyWay 구축실적사례 (종합분석대시보드 참고)	http://www.iseeya.co.kr/solution_milkyway.html	2020		
14	참고 사이트	any.run (종합분석대시보드 참고)	https://any.run/	2020		

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

5 부록

5.1 사용자 매뉴얼

서비스 대상자는 보안관제사이며, 보안관제사가 본 프로젝트를 사용하는데 어려움이 없도록 깃허브에 사용자 매뉴얼을 공개하였다.

● 대시보드 visualization 가이드라인

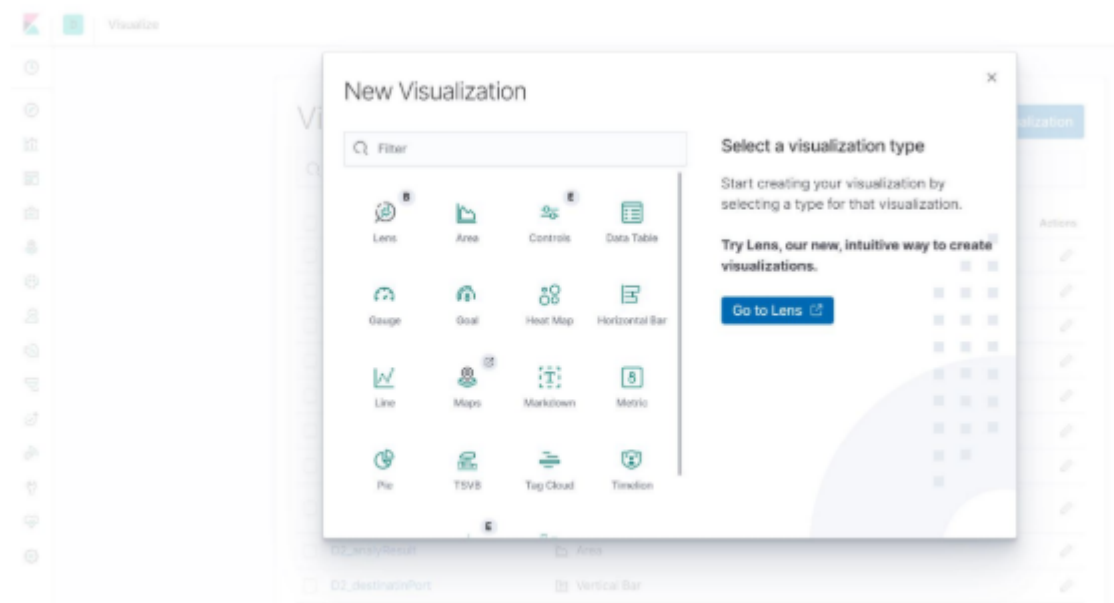
<https://github.com/kookmin-sw/capstone-2020-7/issues/59>

키바나 대시보드 만드는 큰순서


1. Discover -> change index pattern에서 대시보드를 만들 index설정 ->show dates에서 적절한 date범위 조절
2. Visualization -> create visualization ->시각화하고싶은 방법선택(ex:수직바차트,파이차트) -> data index 를 선택한다> 보이고 싶은 조건에 맞추어 구성
3. Dashboard -> create dashboard -> create new버튼으로 visualization 새로 생성 or add 버튼눌러서 만들어진 visualization들 중 선택

Visualization 구성방법

데이터 시각화를 시작하려면 사이드 탐색 메뉴에서 시각화 를 클릭합니다.



[그림 28] Kibana 대시보드 가이드라인

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

5.2 설치 매뉴얼

사용자가 원하는 기능을 추가함으로써, 시스템을 확장하여 이용할 경우 아래와 같은 방법으로 설치한다. 각각의 설치 매뉴얼은 본 프로젝트의 깃허브에 공개하였다.

1) ELK 설치 가이드라인:

<https://github.com/kookmin-sw/capstone-2020->

[7/commit/e1fe783d5385387e642ad1a3da3d0f9b92841219](https://github.com/kookmin-sw/capstone-2020-/commit/e1fe783d5385387e642ad1a3da3d0f9b92841219)

ELK

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


설치 순서

@@ -16,5 +15,41 @@ ELK Stack 설치 순서는 다음과 같습니다.

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
+ 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
```

[그림 29] ELK 설치 가이드라인1

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

```

+ ### 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.<br>
+ Elasticsearch Configuraiton : https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk/elasticsearch
+ <br>
+ Kibana Configuration : https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk/kibana <br>

```

[그림 30] ELK 설치 가이드라인2


2) Elasticsearch.yml 가이드라인: <https://github.com/kookmin-sw/capstone-2020-7/commit/771ee97a4f4594d9de4e7df3f9f49bb696e0fe61>

```

+ # Elasticsearch
+
+ elasticsearch.yml
+ - elasticsearch.yml guideline
+
+ ### 실행 방법
+
+ https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk <br>
+ 위 설치 가이드라인을 따라 설치를 완료하였다면 실행방법은 다음과 같습니다.
+
+ ```
+ [elasticsearch 아카이브가 설치된 경로]/elasticsearch-7.7.0/bin/elasticsearch
+ ```

```

[그림 31] Elasticsearch.yml 가이드라인

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

3) logstash configuration 가이드라인:

<https://github.com/kookmin-sw/capstone-2020-7/commit/6d7329bcc9eaa656d459bedbe05f341b5b6b1905>

```
+ # logstash
+
+ logstash.conf
+ - logstash configuration example
+
+ ### 실행 방법
+
+ https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk <br>
+ 위 설치 가이드라인을 따라 설치를 완료하였다면 실행방법은 다음과 같습니다.
+
+ ...
+ [logstash가 설치된 폴더의 경로]/bin/logstash -f [작성한 logstash configuration 파일 경로]
+ ...
```


[그림 32] logstash configuration 가이드라인

4) kibana.yml 가이드라인

<https://github.com/kookmin-sw/capstone-2020-7/commit/39f047e62344ef99cd038150c3a6159630ec1088>

```
+ # kibana
+
+ kibana.yml
+ - kibana.yml example
+
+ ### 실행 방법
+
+ https://github.com/kookmin-sw/capstone-2020-7/tree/feature/ELK/elk <br>
+ 위 설치 가이드라인을 따라 설치를 완료하였다면 실행방법은 다음과 같습니다.
+
+ ...
+ [kibana 아카이브가 설치된 경로]/kibana-7.7.0-linux-x86_64/bin/kibana
+ ...
```

[그림 33] kibana.yml 가이드라인

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

5) Elastalert 설치 가이드라인

<https://github.com/kookmin-sw/capstone-2020-7/issues/62>

[Elastalert 설치 가이드라인]

해당 가이드라인은 ELK 스택 7.7버전을 사용하였습니다.

```
$ pip install elastalert
$ git clone https://github.com/Yelp/elastalert.git
$ pip install "setuptools>=11.3"
$ python setup.py install
$ elastalert-create-index
```

[Elastalert 실행 가이드라인]

- 테스트용으로 정상적으로 실행되는지만 확인

```
$ elastalert-test-rule example_rules/d.yaml
```
- 실행

```
$ elastalert --verbose --rule example_rules/d.yaml
```


(d.yaml대신 실제로 실행할 rule을 선택.)

[Elastalert Example_rules 가이드라인]

Rule타입


- **Any** : The any rule will match everything. Every hit that the query returns will generate an alert.
- **Blacklist** : The blacklist rule will check a certain field against a blacklist, and match if it is in the blacklist.
- **Whitelist** : Blacklist와 흡사, this rule will compare a certain field to a whitelist, and match if the list does not contain the term.
- **Change** : This rule will monitor a certain field and match if that field changes.
- **Frequency** : This rule matches when there are at least a certain number of events in a given time frame.

[그림 34] Elastlaert 설치 및 실행 가이드라인

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

5.3 테스트 케이스

대분류	소분류	기능	테스트 방법	기대 결과	테스트 결과
모델	정오탐 분류	정탐과 오탐을 분류한다.	KISTI 데이터를 통해 정오탐 라벨링이 되는지 확인한다.	정탐과 오탐이 라벨링되어 분류된다.	성공
ELK	파일 전송	Logstash에서 Elasticsearch로 데이터를 전송한다.	Logstash에서 Elasticsearch로 데이터가 올라가는지 확인한다.	Logstash에 실시간으로 데이터가 올라간다.	성공
			2. Kibana 대시보드 설정에서 auto refresh를 3초로 설정한다. 3. 최신데이터가 계속 업데이트 되는지 확인한다.	auto refresh를 통해 최신데이터가 계속 업데이트 된다.	
			1. 종합분석 대시보드를 클릭한 경우 종합분석 대시보드를 보여준다. 2. 위험탐지 대시보드를 클릭	종합분석 대시보드와 위험탐지 대시보드가 각각 잘 나타난다.	

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	결과보고서		
	프로젝트 명	SeMo(Security Monitoring Platform Using Event Logs)	
	팀 명	Do Mo!(Do Monitoring)	
	Confidential Restricted	Version 1.5	2020-JUN-09

			한 경우 위험 탐지 대시보드를 보여준다.		
--	--	--	---------------------------	--	--