(Elasticsearch,Logstash,Kibana,Filebeat) 이용한

로그 3D 시각화, 침입 관제

2019.08.20.

팀명 ELK



Category	첨부파일 버전	문서 최종 수정일
Manual + Utility	없음	2019. 08. 22

ELK

# 개정이력

개정 번호	개정 내용 요약	추가/수정 항목	개정 일자
0.1	최초 제정 승인	목차 / 개요	2019.08.03
0.2		내용	2019.08.15
0.3		내용	2019.08.17
0.4	중간 제정 승인	내용 / 부록	2019.08.18
1.0	최종 제정 승인	결론	2019.08.20

# 문 서 규 칙

TATED *	제목			
KIIRI	Category	첨부파일 버전	문서 최종 수정일	팀명
	1	2	3	

- 작성 및 확인은 Microsoft Word 2016으로 작성되었으며, PDF로 읽는다.
- Category(①)에는 Manual, Utility, Tip, Analysis Report 로 구분하며, 기재된 정보가 Manual과 Utility가 혼합된 경우에는 "Manual + Utility" 라고 표기되며, 머리글의 Category에 해당 구분 정보를 표기된다.
- 첨부 파일 버전(②)은 첨부 파일이 존재하는 경우에 기재되며, 첨부 파일의 버전이 표기된다. (유틸리티의 경우 최종 버전은 날짜 표기 대신 버전으로 대체한다)
- 문서 최종 수정일(③)에는 문서의 최종 수정 날짜가 표기된다.



 Category
 첨부파일 버전
 문서 최종 수정일

 Manual + Utility
 없음
 2019. 08. 22

ELK

# 목 차

1.	서론		6
	1.1. 프	로젝트 개요	6
	1.1.1.	주제 선정 동기 및 목표	6
	1.1.2.	AS-IS, TO-BE	7
	1.1.3.	프로젝트 역할	7
	1.1.4.	프로젝트 일정	8
	1.2. 프	로젝트 환경	9
	1.2.1.	환경 구성	9
	1.2.2.	활용 도구	9
2.	본론		10
	2.1. 개설	2	10
	2.2. 프	로젝트 내용	10
	2.2.1.	ELK 의 작업 과정	10
	2.2.2.	centos의 데이터 선정 및 수집 방법	10
	2.2.3.	침입 판단의 기준	10
	2.2.4.	그래프 시각화(2D, 3D)	11
	2.2.5.	머신러닝 이용한 패턴 분석.(Single,multi)	13
3.	결론		15
	3.1. 결류	로, 추후 과제	15
	3.1.1.	결론	15
	3.1.2.	추후 과제	18
4.	참조, ·	출처	19



Category	첨부파일 버전	문서 최종 수정일
Manual + Utility	없음	2019. 08. 22

ELK

# 표 목 차

[표	1-1] AS-IS / TO BE	7
ſπ	1-2] 프로젝트 역할 분담	7
[丑	1-3] 프로젝트 일정	8
[丑	1-4] 환경 구성	9
ſπ	1-5) 환용 도구	q



Category	첨부파일 버전	문서 최종 수정일
Manual + Utility	없음	2019. 08. 22

ELK

# 그 림 목 차

그림	1-1 예: 2D 시간별 로그	11
그림	1-2 예: 3D 그래프	12
그림	2-1 머신러닝 로그 증강 테스트 화면 구성	13
그림	2-2 (그림 2-1)감지 부분 확대	13
그림	2-3 머신러닝 bash 특정 로그 감지 테스트 화면 구성	14
그림	2-4 감지 부분 확대	14
그림	3-1 대쉬보드 구성	15
그림	3-2 좌 상단 2D그래프 확대	16
그림	3-3 우 상단 2D그래프 확대	.16
그림	3-4 좌 하단 3D그래프 확대	17
그림	3-5 우 하단 3D그래프 확대	17



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종 수정일	ELK
Manual + Utility 없음 2019		2019. 08. 22	

# 1. 서론

#### 1.1. 프로젝트 개요

#### 1.1.1.주제 선정 동기 및 목표

IT 기술의 발전에 따라 취약점을 공격하는 바이러스가 다양해짐에 따라, Anti-Virus 프로그램을 회피하는 신종 해킹 방법을 이용한 침입이 늘어나고 있다. 결과적으로 스피어 피싱, APT 공격 등 공격 방식의 발전으로 인한 새로운 침입 방법이 늘어남에 따라 Endpoint를 보호 하는 새로운 방법이 추가 되어야 한다.

기존/신종 침입 탐지를 개선한 추가 방법으로 침입을 로그의 변화로 침입을 탐지한다. 침입 초기, 크래커는 터미널을 이용하여 해킹을 시작하는 데 이 점을 중심으로 정상 사용자와 침입자의 입력을 직관적 구별되도록 만든다. 추가로 머신 러닝을 이용하여 사용자의 기존 패턴과 차이를 감지하여 시각적으로 침입을 판단한다.

첫번째로 로그 데이터로 침입을 구별 할 수 있는 기준을 세우는 것이다. 두번째로 로그 같은 종류의 문자 데이터를 직관적인 시각적 정보로(2D, 3D그래프와 머신러닝) 변환 하여 관제의 편의성을 높이는 것. 세번째로 변환된 정보로 침입의 유무를 분석하는 것이 목표이다,



Endpoint log기반 탐지		
Category	첨부파일 버전	문서 최종 수정일
Manual + Utility	연 음	2019. 08. 22

### 1.1.2.AS-IS, TO-BE

AS – IS	TO – BE
기존 백신 기반 탐지의 문제점	로그 기반 탐지의 장점
✓ 알려진 악성코드 위주 진단	✔ 로그 기반 침입 진단
✓ 공격자도 테스트 가능	✔ 에러로 인한 악영향 차단
✓ 에러로 인한 장애	✔ 시각적, 직관적 정보 제공
✓ 텍스트 정보로 인한 불편함	✔ 신종 악성 침입 파악, 추측
✓ 신종 악성 코드 제한적 진단	✔ 백신을 피한 침입 감지
✓ 침입 전 예방 및 차단	

[표 1-1] AS-IS / TO BE

### 1.1.3. 프로젝트 역할

이름	직책	역할
홍기선	PM	1. 자료 수집, 분석, 정리 2. 설계 계획/일정 3. VMware / ELK 환경 구성 및 기능/옵션 테스트 4. 로그 데이터 확인, 특징 및 구조 파악 5. 로그 선별, 전송 설정, 로그 시각화 3D 그래프 6. 머신 러닝 작업 7. 문서 작성 (일일/착수/최종 보고서)

[표 1-2] 프로젝트 역할 분담

페이지 7 / 21



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종 수정일	
Manual + Utility	없음	2019. 08. 22	

# 1.1.4.프로젝트 일정

일정	7월 1주	2주	3주	4주	8월 1주	2주	3주	4주
주제 선정								
자료 수집								
환경 구축								
로그 분석								
로그 설정/								
전송								
3D 시각화								
머신 러닝								
오류 수정								
보고서								

[표 1-3] 프로젝트 일정



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종 수정일	EL
Manual + Utility	없음	2019. 08. 22	

# 1.2. 프로젝트 환경

# 1.2.1.환경 구성

CentOS	터미널 입력 로그 기록 및 전송 위한 테스트 환경 Version – Centos7
<b>∫ vm</b> ware <sup>™</sup>	centos 환경을 구축 위해 설치한 가상화 소프트웨어 Version – Workstation pro 14

[표 1-4] 환경 구성

# 1.2.2.활용 도구

Elasticsearch	전송된 정보 검색 및 분석 엔진 도구 ELK Version – 6.8.1
Logstash	비트로 수집 된 로그 변환 후 전송 도구
Kibana	데이터 시각화 및 유저 인터페이스 확장 가능한 도구
Beats	경량형 문자 로그 수집 및 옵션 추가 하는 파밍 도구

[표 1-5] 활용 도구



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종 수정일	
Manual + Utility	없음	2019. 08. 22	

### 2. 본론

#### 2.1. 개요

로그를 수집하여 각 단말의 데이터를 정보화 즉, 시각화하여 인간이 관측하기 편한 직관적인 정보로 만든다. 이후 시각 정보를 기반으로 관측자가 침입을 감지한다. 머신 러닝을 이용하여 단말의 기존 패턴과 차이를 감지하여 관측자에게 시각적으로 알리는 것까지 이번 프로젝트이다.

구성: 로그의 수집, 전송, 시각화, 침입 기준 설정, 분석, 머신 러닝 - 패턴 분석

테스트 과정: 각 그래프의 필터 값 설정과 결과 확인, 수정

#### 2.2. 프로젝트 내용

#### 2.2.1.ELK의 작업 과정

ELK의 환경 구성으로 VMware 가상머신에 centos를 설치하여 진행한다. 진행 변화에 따른 내용은 스냅샷으로 저장한다. ELK 로그 수집, 전송, 데이터 분석하는 모든 과정 설명한다면, filebeat 경량 로그 수집기로 얻은 데이터를 logstash로 통합 수집하여 elasticsearch와 kibana로 전송한다. 여기서 kibana는 시각화, elasticsearch(머신 러닝 포함)는 분석하는 프로그램이다.

#### 2.2.2.centos의 데이터 선정 및 수집 방법

Utmp, Wtmp, Lastlog, Btmp, History, Secure, Messages, Dmesg, Boot.log, Xferlog, Cron, Mail, Maillog등 centos에서 추출 할 수 있는 데이터를 선정하여 내부 데이터 파악 작업 한다. 그 뒤 파일의 데이터를 로그화 작업 (예: history -> bash.log) 끝내고 침입 표현에 적절한 시각화, 직관적인 데이터 선정해서 진행하였다. 이번 프로젝트의 경우 history.log 선정하였다.

#### 2.2.3.침입 판단의 기준

Linux Unix BSD Post-Exploitation List(침입 후 입력하는 명령어 리스트)와 권한 관련 명령어를 기준과 로그의 변화로 시각적인 정보 생성한다. 그리고 생성한 정보를 토대로 알 수 있는 3 가지 대조 방법으로 정보 침입 판단을 내린다. 첫번째로 사용자 명령어 입력 시간 기록과 대조하는 것이다. 평소와 다른 사용자 패턴을 보이는 것에 중점을 둔다. 두번째로 로그의 증감 변화를 확인한다. 침입 하는 경우, 로그를 삭제하거나 악성 프로그램의 기능으로 로그가 폭발적으로 증가하는 변화를 보인다. 세번째는 터미널 명령어 Linux Unix BSD Post-Exploitation List 일부 (reverse shell code)에 중점을 둔다. 일반적인 사용자가 쓸 일이 없는 침입에 쓰이는 코드들로 명령어 사용이 확인된다면 침입이라 판단 가능하다.



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종 수정일	
Manual + Utility	없음	2019. 08. 22	

#### 2.2.4.그래프 시각화(2D, 3D)

history log에 있는 명령어를 키워드화하여 2가지 방식으로 분석 및 시각화 한다. 키워드의 필터 구성으로는 Linux Unix BSD Post-Exploitation를 참조하며, 일반 사용자 명령어 중 해커가 침입 후 사용할 가능성이 높은 명령어로 필터를 선정 하였다.

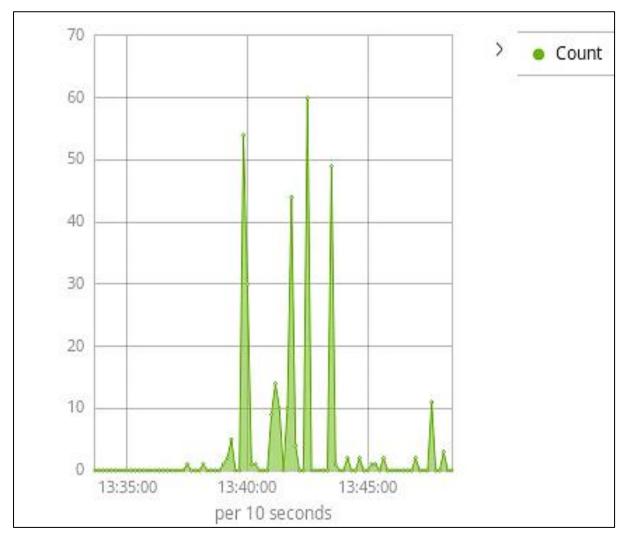


그림 1-1 예: 2D 시간별 로그

① 2D 그래프 이용해 통해 시각화하는 것으로 로그의 변화를 볼 수 있도록 한다.

Kibana - Visualize 카테고리 선택 그래프 옵션을 이용하여 2D 그래프 표현하며, 시간별 로그의 생성양 및 일반적인 명령어 사용 빈도수 확인한다.



Category	첨부파일 버전	문서 최종 수정일
Manual + Utility	없음	2019. 08. 22

ELK

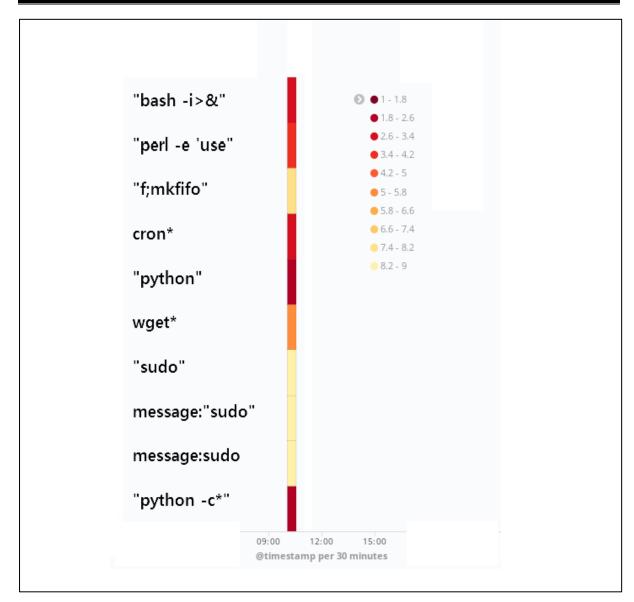


그림 1-2 예: 3D 그래프

② 3D 그래프의 경우 x축 시간대, y축으로 키워드 필터링 하여 표현한다.

Kibana - Visualize 카테고리 선택 Heatmap 그래프 옵션을 이용하여 3D 그래프를 표현한다. x 축 @Timestamp 옵션으로 설정. 시간 별 키워드가 입력되는 시간대를 파악하여 사용자 이용시간 나타낸다. y축 filters 옵션으로 키워드 리스트를 입력한다. z축은 횟수를 색의 진함으로나타낸다. 본 프로젝트에서는 붉은 색의 스펙트럼으로 표현한다.



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종 수정일	
Manual + Utility	없음	2019. 08. 22	

### 2.2.5.머신러닝 이용한 패턴 분석. (Single,multi)

① Single Metric 머신러닝

그림 2-1은 인위적으로 로그를 제거, 생성 했을 때 급변하는 로그 패턴을 잡아낸 것이다.

그림 하단에 시간, 수치 별 리스트를 나타내며 이로 인해 크래커가 인위적으로 로그 제거 시침입을 감지 가능하다.(예: 그림 2-1, 2-2)

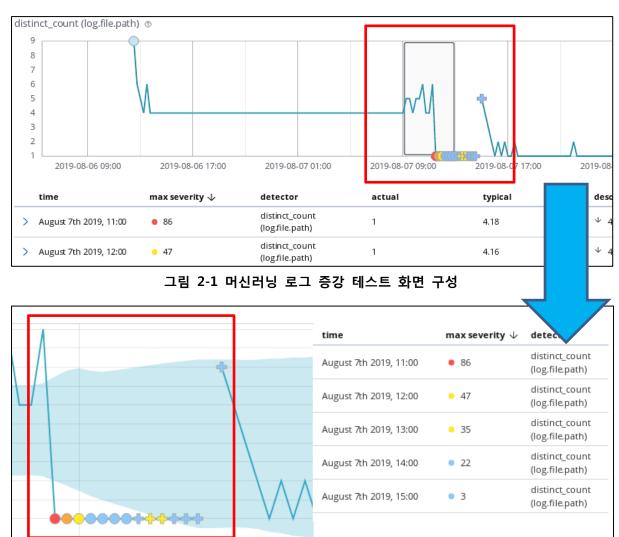


그림 2-2 (그림 2-1)감지 부분 확대

그림 2-2 로그의 삭제로 인한 변화를 나타낸 것이다. 각 포인트를 원으로 나타내며, 변화가 큰 부분을 색깔로 표현하여 붉은 색으로 나타낸다.

페이지 13 / 21



Endpoint	log기빈	: 탐지
----------	-------	------

	, ,	-
Category	첨부파일 버전	문서 최종 수정일
Manual + Utility	없음	2019. 08. 22

#### ② Multi Metric 머신러닝

머신러닝 설정에서 필드 값 log.file.path 선택 한 뒤 Create Job눌러 분석 결과 생성한다.

(예: 그림 2-3)

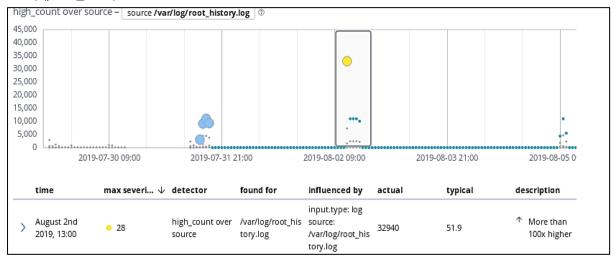


그림 2-3 머신러닝 bash 특정 로그 감지 테스트 화면 구성

그림 2-3은 history.log에서 추출한 것으로 기존 패턴과 다른 변화를 나타낸다. 그래프 x축에 가까운 파란 원이 기존 패턴을 나타내며 큰 원형으로 나타낸 파란 원과 노란 원이 다른 패턴을 나타낸다.

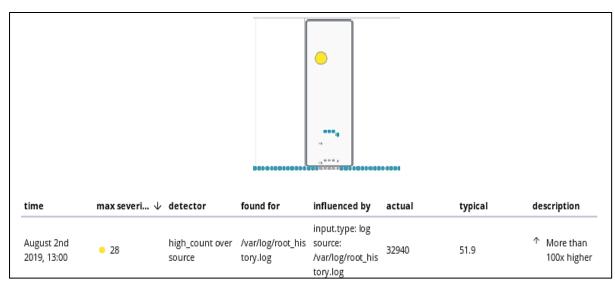


그림 2-4 감지 부분 확대

그림 2-4 인위적인 방법으로 생성된 로그의 생성을 잡아낸 부분이다. 반복 입력된 코드가 이용되어 급증한 로그의 양을 나타낸 것이다. 파란색 작은 원이 정상적인 로그의 생성양을 나타내며 노란색 원이 급증한 부분을 나타낸다.

페이지 14 / 21



Endpoint log기반 탐지		
Category	첨부파일 버전	문서 최종 수정일
Manual + Utility	없음	2019. 08. 22

### 3. 결론

#### 3.1. 결론, 추후 과제

#### 3.1.1. 결론

머신러닝(그림 2-1~2-4) 사용자 패턴, 대쉬보드(그림 3-1~3-5) 결과로 2D, 3D 그래프와 머신 러닝으로 터미널에 입력되는 코드와 날짜 별 로그의 양 등 실시간으로 쌓이는 로그데이터를 시각적인 정보로 변환 한 것을 보여준다.

위 그래프와 머신러닝의 사용자 패턴 분석으로 사용자 외 침입에 의한 로그의 종류와 패턴을 파악 할 수 있다. 결과적으로 백신과는 다른 분석 방법으로 침입을 파악 할 수 있다.

또한 이 그래프의 경우 bash 로그만 분석한 내용으로 추가로 message 로그, utmp 로그, wtmp 로그, secure 로그 기타 등등 로그를 복합적으로 분석하게 된다면 더욱더 정확한 침입 감지가 가능하다고 생각한다.

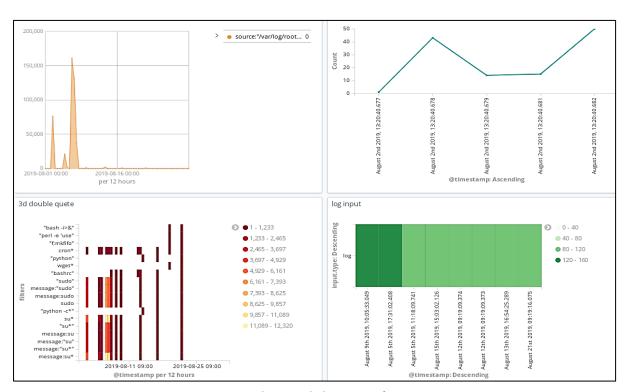


그림 3-1 대쉬보드 구성

페이지 15 / 21



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종 수정일	
Manual + Utility	없음	2019. 08. 22	

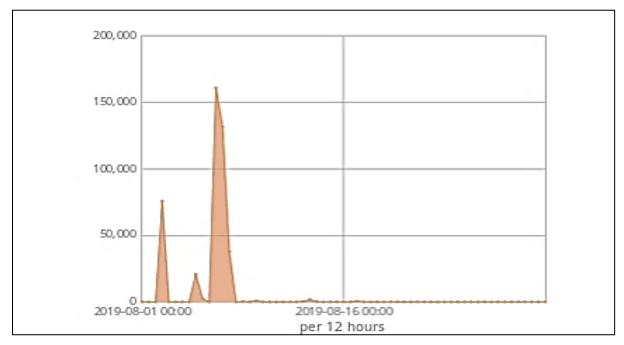


그림 3-2 좌 상단 2D그래프 확대

전체 로그의 생성양을 표현한 그래프로 12시간 간격으로 변화를 나타낸다.

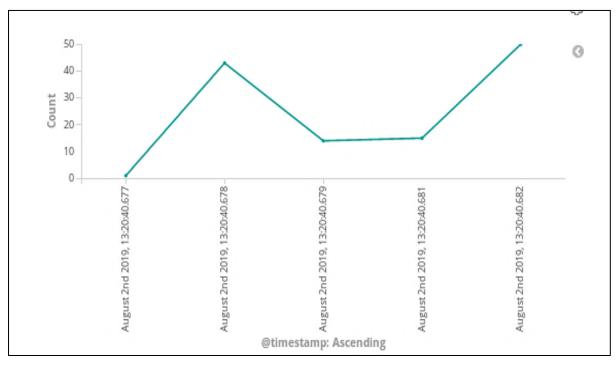


그림 3-3 우 상단 2D그래프 확대

로그의 변화를 그린 그래프로 급격한 변화를 나타낸 날짜가 기준이다.

페이지 16 / 21



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종	수정일

Manual + Utility 없음 2019. 08. 22

**ELK** 

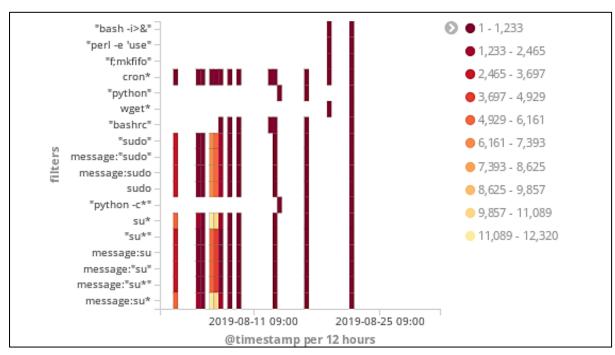


그림 3-4 좌 하단 3D그래프 확대

x 축은 시간, y 축은 Reverse Shell code 와 권한 명령어, z 축은 입력 횟수를 색으로 표현한 3D 그래프이다.

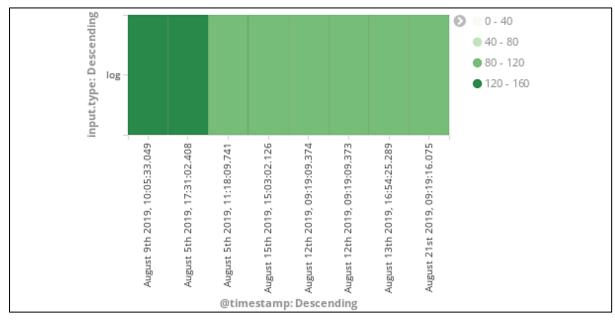


그림 3-5 우 하단 3D그래프 확대

로그 값 변화를 3D 그래프 버전으로 나타냈다. 로그가 많은 날짜 순으로 정렬된 그래프이다.

페이지 17 / 21



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종 수정일	
Manual + Utility	없음	2019. 08. 22	

#### 3.1.2.추후 과제

추후 과제는 여러 코드와 연관성 및 영향을 분석 하여 추가하는 것이다. 위 프로젝트의 한계점은 터미널 로그를 기반으로 침입을 감지하는 것이라 한계가 명확하며, 한계를 벗어날 방법으로 history 로그이외 (예:utmp,wtmp,message..) 데이터로 남는 파일들을 로그로 만들어 분석하는 방법이 있다. 다른 하나는 위 결론 란에 서술한 로그들 간의 연관성 및 내부에 나타나는 정보의 맵(기록 간 영향력을 표현)침입 시 발생하는 패턴을 분석한다면 침입 탐지를 더욱 확실하게 해줄 것이다 본다.

또한 개선 시 주의사항이 있다. 그래프 생성 조건에 필터 값 입력하는 경우, 각 코드와 명령어 설정에 따른 오류가 존재한다. 예를 들어 리버스 쉘 코드의 경우 ""를 쓰는 방법이 아닌 경우, Elastic 내부 검색 시스템이 명령어로 인식하여 실행된다. 이로 인한 필터링이 안되며 시각화가 무의미 해지는 상황이 발생된다.



Category	첨부파일 버전	문서 최종 수정일
Manual + Utility	없음	2019. 08. 22

ELK

### 4.참조, 출처

논문 - 클라우드 기반 데이터 처리기술을 이용한 대용량 보안로그 연관분석

학위 논문 검색 사이트 RISS - 저자: 최대수

날짜: 2012 년 2월. 사이트 주소:

http://www.riss.kr/search/detail/DetailView.do?p\_mat\_type=be54d9b8bc7cdb09&control\_no=416e560167c72f57ffe0bdc3ef48d419

논문 - 시각화 기법을 이용한 악성코드 분석 및 분류 연구학위 논문 검색 사이트 NDSL - 송인수, 이동휘, 김귀남 (경기대학교산업기술보호특화센터)

날짜: 2019 년 9월 24일. (채택일 기준) 사이트 주소:

http://www.ndsl.kr/ndsl/search/detail/article/articleSearchResultDetail.do?cn=JAKO20101 1949347080

제목 - 아파치 웹서버 로그 파일 상세 분석

웹페이지 운영자 : 신광식

날짜: (표기 안됨.) 사이트 주소: http://www.linuxlab.co.kr/docs/00-09-6.htm

제목 - [번역] 엘라스틱서치와 키바나 실용적인 소개서

웹페이지 운영자 : 서진규

날짜: 2019 년 7월 11일. 사이트 주소:

https://velog.io/@jakeseo\_me/%EB%B2%88%EC%97%AD-%EC%97%98%EB%9D%BC%EC%8A%A4%ED%8B%B1%EC%84%9C%EC%B9%98%EC%99%80-%ED%82%A4%EB%B0%94 %EB%82%98-%EC%8B%A4%EC%9A%A9%EC%A0%81%EC%9D%B8-%EC%86%8C%EA%B 0%9C%EC%84%9C

제목 - Linux 기초 (Shell Script)

웹페이지 운영자: 박원영

날짜: 2018 년 9월 17일. 사이트 주소:

 $\frac{https://blog.naver.com/PostView.nhn?blogId=pk3152\&logNo=221360352090\&from=search\&redirect=Log\&widgetTypeCall=true\&directAccess=false$ 



Category	첨부파잌 버전	문서 최종 수정일
97	0 1 1	
Manual + Utility	없음	2019, 08, 22

ELK

제목 - [Linux] 리눅스 주요 로그 파일 정리

웹페이지 운영자 : 공대냥이 (실명 안나옴.)

날짜: 2018 년 8월 21일.(약 12개월 전 - 블로그에 12개월 전이라 적힘.)

사이트 주소: https://itragdoll.tistory.com/13

제목 - ELK Stack 을 이용한 로그 관제 시스템 만들기

웹페이지 운영자: modolee

날짜: 2019 년 3 월 26 일. 사이트 주소: https://medium.com/day34/elk-

stack%EC%9D%84-%EC%9D%B4%EC%9A%A9%ED%95%9C-%EB%A1%9C%EA%B7%B8-%EA%B4%80%EC%A0%9C-%EC%8B%9C%EC%8A%A4%ED%85%9C-%EB%A7%8C%EB%9

3%A4%EA%B8%B0-ca199502beab

제목 - [CentOS7] CentOS 7 에 ELK, Filebeat 설치 방법

웹페이지 운영자: 프로그웍스

날짜: 2019 년 3월 21일. 사이트 주소:

https://progworks.tistory.com/78?category=822251

제목 - 백억 개의 로그를 모아 검색하고 분석하고 학습도 시켜보자:

로기스 현동석 / 김광림 / 양은숙

웹페이지 작성자: NAVER D2

날짜: 2017 년 8월 17일. 사이트 주소:

https://www.slideshare.net/deview/ss-80885724

제목 - 한시간에 만드는 대용량 로그 수집 시스템

웹페이지 운영자 : 조대협 날짜: 2017 년 1 월 24 일.

사이트 주소: https://bcho.tistory.com/1158

제목 - 머신러닝으로 시스템 이상 징후를 자동으로 분석하는 예제 따라하기

웹페이지 운영자 : 올빼미 날짜: 2017 년 8 월 7 일.

사이트 주소: https://olpaemi.blog.me/221068974781



Endpoint log기반 탐지			
Category	첨부파일 버전	문서 최종 수정일	ELK
Manual + Utility	없음	2019. 08. 22	

제목 - ELK 와 Event Log Explorer 를 이용한 대용량 이벤트 로그 분석

웹페이지 운영자 : 이글루 시큐리티

날짜: 2018 년 10 월 1 일.

사이트 주소:

http://www.igloosec.co.kr/BLOG\_ELK%EC%99%80%20Event%20Log%20Explorer%EB%A5 %BC%20%EC%9D%B4%EC%9A%A9%ED%95%9C%20%EB%8C%80%EC%9A%A9%EB%9F %89%20%EC%9D%B4%EB%B2%A4%ED%8A%B8%20%EB%A1%9C%EA%B7%B8%20%EB %B6%84%EC%84%9D?searchItem=&searchWord=&bbsCateId=1&gotoPage=3

제목 - Linux/Unix/BSDPost-ExploitationCommandList.

웹페이지: http://www.handgrep.se

날짜: 게시 날짜 없음

사이트 주소:

http://www.handgrep.se/repository/cheatsheets/postexploitation/LinuxUnixBSDPost-Exploitation.pdf