사물인터넷융복합논문지 Vol. 8, No. 5, pp. 1-9, 2022 https://doi.org/10.20465/KIOTS.2022.8.5.001

SIEM 기반 사이버 침해사고 대응을 위한 데이터 보완 메커니즘 비교 분석

이형우*
한신대학교 컴퓨터공학부 교수

Analysis of Cyber Incident Artifact Data Enrichment Mechanism for SIEM

Hyung-Woo Lee*
Professor, Division of Computer Engineering, Hanshin University

요 약 최근 IoT 및 휴대용 통신 단말에 각종 서비스가 연동되면서 해당 디바이스의 보안 취약점을 악용한 사이버 공격이 급증하고 있다. 특히 지능형 지속위협(APT) 공격을 통해 대단위 네트워크 환경에서 이기종 형태의 디바이스를 대상으로 한 사이버 공격이 급증하고 있다. 따라서 침해사고 발생시 대응 체계의 유효성을 향상시키기 위해서는 위협 분석 및 탐지 성능이 향상되도록 수집된 아티팩트 데이터에 대한 데이터 보완(Data Enrichment) 메커니즘을 적용할 필요가 있다. 이에 본 연구에서는 침해사고 분석을 위해 수집된 아티팩트를 대상으로 기존의 사고관리 프레임워크에서 수행하는 데이터 보완 공통 요소를 분석하여 실제 시스템에 적용 가능한 특징 요소를 도출하고, 이를 토대로 개선된 사고분석 프레임워크 프로토타입 구조를 제시하였으며 도출된 데이터 보완 확장 요소의 적합도를 검증하였다. 이를 통해 이기종 디바이스로부터 수집된 아티팩트를 대상으로 사이버 침해사고 분석 시 탐지 성능을 향상시킬 수 있을 것으로 기대된다.

주제어 : 사물인터넷, 보안 정보 및 이벤트 관리(SIEM), 사이버 침해사고 대응, 데이터 보완 메커니즘.

Abstract As various services are linked to IoT(Internet of Things) and portable communication terminals, cyber attacks that exploit security vulnerabilities of the devices are rapidly increasing. In particular, cyber attacks targeting heterogeneous devices in large-scale network environments through advanced persistent threat (APT) attacks are on the rise. Therefore, in order to improve the effectiveness of the response system in the event of a breach, it is necessary to apply a data enrichment mechanism for the collected artifact data to improve threat analysis and detection performance. Therefore, in this study, by analyzing the data supplementation common elements performed in the existing incident management framework for the artifacts collected for the analysis of intrusion accidents, characteristic elements applicable to the actual system were derived, and based on this, an improved accident analysis framework The prototype structure was presented and the suitability of the derived data supplementary extension elements was verified. Through this, it is expected to improve the detection performance when analyzing cyber incidents targeting artifacts collected from heterogeneous devices.

Key Words : Internet of Things, Security Information and Event Management(SIEM), Cyber Incident Response, Data Enrichment Mechanism.

이 논문은 2022학년도 한신대학교 학술연구비의 지원에 의하여 연구되었음.
*교신저자 : 이형우(hwlee@hs.ac.kr)
접수일 2022년 7월 20일 수정일 2022년 9월 3일 심사완료일 2022년 9월 5일

1. 서론

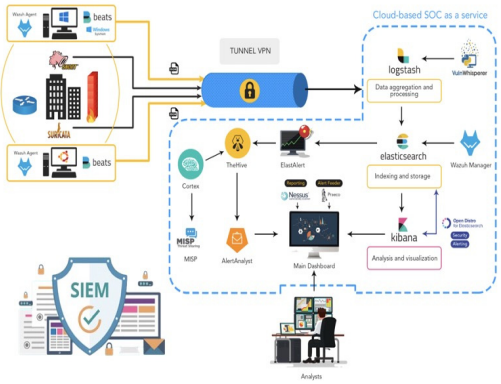
일상생활에서 사용하는 각종 통신 기기들이 대부분 IoT(Internet of Things)·휴대용 통신 수단 그리고 이와 결합한 다양한 형태의 서비스와 연동되면서 해당 기기의 보안 취약점을 악용한 사이버 공격이 급증하고 있으며 특히 지능형 지속위협(APT) 공격 등을 통해 고도화되고 지능화된 공격이 계속되고 있다[1,2,3]. 침투와 탐색 그리고 정보 수집 및 유출의 4단계로 수행되는 APT 공격에 대한 사고관리 분석 및 대응 체계 구축이 시급하다. 따라서 이와 같은 APT 공격에 능동적으로 대응하며 위협 탐지 및 대응을 위해 보안 정보 및 이벤트 관리(SIEM : Security Information and Event Management) 시스템[4]이 개발 및 배포되고 있다. 점차적으로 지능화/고도화되는 APT 공격 등에 능동적으로 대응하기 위해서는 각종 IoT 및 End Point 단말로부터 수집된 이기종 아티팩트 데이터에 대한 보완(Data Enrichment) 과정을 수행하여 메타데이터 정보를 획득/축적하여 사이버 침해사고 분석 과정에서 IoC(Indicator of Compromise)에 대한 정확도 및 유효성을 향상시킬 필요가 있다. 예를 들어, 사이버 침해사고 분석의 정확도를 향상시키기 위해 Domain/DNS info, Gep IP info, Whois lookup info 등 아티팩트에 포함된 속성 정보에 따라 추가적으로 데이터를 보강하는 과정을 수행할 필요가 있다. 만일 각 단말에 대한 OS 정보, Logged-in User 정보, GeoLocation 정보, 사용자 IP/Machine 및 Ownership 정보 등 아티팩트 내 추가적인 정보를 보강하는 과정을 수행한다면 사이버 침해사고 분석/대응 과정의 효율성/정확도를 향상시킬 수 있다.

이에 본 연구에서는 (1) 기존의 SIEM[4] 및 CTI(Cyber Threat Intelligence)[5] 등 사이버 침해사고 대응기술에 대한 기술 동향을 조사하여, (2) 기존 프레임워크에서 수행하는 데이터 보완(Data Enrichment) 프로세스 과정 및 기술을 분석하고, (3) 기존 SIEM 프레임워크에 적용된 데이터 보완 과정에 대해 비교 분석하여 (4) 향후 개선된 사고분석 프레임워크 프로토타입을 설계 및 구현 과정에 활용토록 하여, 대단위 이기종 네트워크 기반 지능형 사이버 공격 발생 시 디지털 포렌식(Digital foresics)[6,7,8] 기반 사이버 침해사고 분석 과정의 신뢰성을 더욱 향상시킬 수 있을 것으로 기대된다.

2. SIEM 기술 현황 분석

2.1 SIEM 기반 사이버 공격 정보 수집 및 공유

대단위 네트워크에서의 보안 위협이 더욱 급속하게 확산되고 있을 뿐만 아니라 지속적으로 새로운 위협이 등장하고 있다. 모바일 사용자, 원격지 및 네트워크 액세스 디바이스 수가 늘어남에 따라 End Point 네트워크 진입 지점(Edge Node)도 급격히 증가하고 있다. 특히 IoT 기기 등 새로운 디바이스 연결과 신규 애플리케이션 연동으로 해당 대단위 네트워크 내에서 보안 취약점이 증가하여 새로운 형태의 공격이 발생하는 근본적인 원인을 제공하기도 한다. 따라서 이와 같은 문제점에 대응하기 위해 개발된 SIEM(Security Information and Event Management) 소프트웨어 및 프레임워크는 아래 그림 1과 같이 Edge Node 부터 최종 사용자까지 전체 네트워크 등을 대상으로 각종 로그를 수집, 저장 및 분석하는 기능을 제공하며, 종합적인 보안 보고 및 규제 준수 관리와 함께 신속한 공격 탐지, 차단 및 응답을 통해 보안 위협을 실시간으로 모니터링 하는 기능을 제공한다. 그러므로 결국 SIEM 소프트웨어는 조직에서 네트워크에 대한 최신 보안 위협을 탐지할 수 있는 강력한 도구이며, SIEM에서 제공하는 보안 이벤트에 대한 분석 기능과 실시간 보고 기능을 통해 조직의 IT 보안에 대한 포괄적 방어 체계를 향상시킬 수 있다.



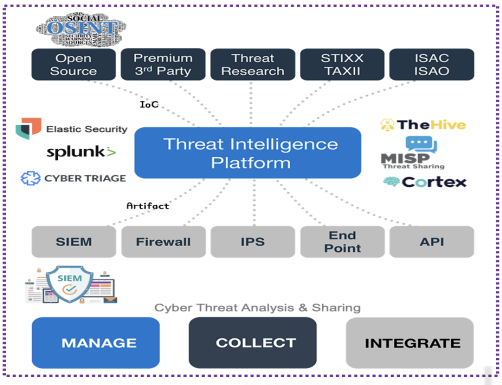
[Fig. 1] SIEM-based incident management and response system diagram

SIEM 소프트웨어 및 프레임워크는 네트워크 전반에서 생성되는 이벤트 레코드/로그 정보를 저장/수집하는 기능을 제공한다. 수집된 로그들은 추후 디지털 포렌식 도구와 연계되어 침해사고 발생에 대한 증거물을 확보하는 과정에도 연계된다. 내부적으로 구문 분석 및 표준화 과정을 통해 각기 서로 다른 SIEM 시스템을 통해 수집된

로그 메시지를 공통 데이터 모델로 매핑/변환하고 다양한 소스 형식으로 로깅된 관련 이벤트를 통합 분석이 가능하다. 또한 SIEM 내부에서 제공되는 상관분석 기능을 이용하여 개별 시스템 또는 애플리케이션의 로그 이벤트를 서로 연결시킴으로써 보안 위협의 탐지 및 대응 속도와 정확도를 향상시켜준다. 이 과정에서 SIEM 내부에서는 수집된 데이터에 대한 중복 최소화, 이벤트 레코드 통합, 데이터 보완(Data Enrichment) 및 상관분석 과정을 진행하게 된다.

2.2 TIP 기반 사이버 공격 정보 수집 및 분석

보안사고에 대한 대응을 위해 SIEM 소프트웨어 및 프레임워크는 일차적으로 유입된 공격에 대한 감지(탐지) 과정을 수행한다. 탐지 과정에서 정확도 및 유용성을 증대하기 위해 다양한 형태의 Data Enrichment 과정이 수반된다. MISP[10] 시스템 등에서는 각종 데이터를 수집하며 각기 정해진 데이터 포맷에 따라 정보를 저장/축적하게 된다. 그리고 만일 보안사고를 감지하였다면 해당 로그 필드 또는 주요 Alert 공격 이벤트 정보에 대한 데이터 강화 과정을 수행하게 된다. 이와 같은 초동대응 과정을 진행한 후 복구 및 사후 대응 과정으로 전환된다. 결국 사이버 공격에 대한 감지(탐지) 및 대응을 수행할 때 SIEM 또는 CTI를 이용하여 아래 그림 2와 같은 사이버 위협 지능형 대응 플랫폼(TIP : Threat Intelligent Platform)을 구축하게 된다. IntelMQ[11], The Hive[12], Cortex[13], Splunk[14] 등 오픈소스 형태로 제공되는 각종 TIP를 이용하여 OSINT 또는 3rd Party 데이터를 수집하고, STIXX/TAXII 등의 포맷으로 전송되는 각종 정보를 축적하게 된다. 이때 네트워크 내에 Firewall/IPS 및 End Point 등으로부터 생성되는 각종 이벤트 정보를 수집하게 된다.

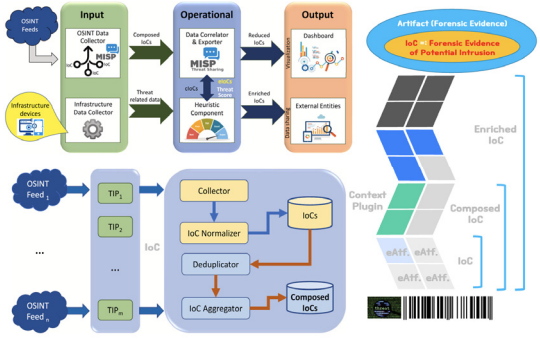


[Fig. 2] TIP-based Data Enrichment Mechanism

3. SIEM 기반 데이터 보완 기법

3.1 SIEM 기반 데이터 보완 메커니즘

아래 그림 3과 같이 SIEM은 Log 데이터에 대한 수집/분석 과정에서의 효율성을 향상시키는 것을 주요 목적으로 하고 있다. 중앙 집중화된 저장소에 Log 데이터를 수집한 다음에 자동화된 Alert를 발생시키는 것을 주요 목적으로 한다. 이를 위해 SIEM에서는 CyberTriage[15], Google GRR[16] 및 Elastic Security[17] 등과 같은 프레임워크와 연동하며 수집된 로그 및 이벤트 정보에 대한 통합, 시각화 및 데이터 강화 과정을 진행하게 된다.

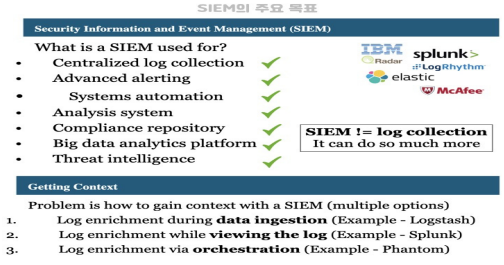


[Fig. 3] SIEM-based Data Enrichment Mechanism

SIEM에서 URL 정보에 대해 Enrichment가 수행되는 경우를 살펴보면 URL Query에 대해서 서브 도메인, 상위 도메인 및 생성 날짜, Geo.ASN 정보 등을 검색한 다음 해당 정보를 추가하는 과정을 수행하게 된다. 추가적인 로그 정보 등을 수집한 다음 부가적인 정보에 대한 조합 과정을 통해 Data Enrichment 과정을 진행한다. 해당 시스템이 윈도우즈 시스템일 경우 이벤트 정보를 수집한 후 Syslog 정보 등과 연계하여 최종적으로 강화된 Log 파일을 생성하게 된다. 결국 SIEM 기반 Log Data Enrichment(eLog) 방식은 공격 탐지와 관련된 핵심적인 로그 이벤트 필드에 대한 데이터 강화 과정을 수행하여 풍부한 정보를 포함한 아티팩트 데이터를 구축하는 것을 주요한 목적으로 하고 있다는 것을 확인할 수 있다.

3.2 SIEM 기반 데이터 보완 특징 분석

결국 SIEM 기반 Data Enrichment 과정의 주요 특징을 분석해 보면 아래 그림 4와 같이 Log 데이터를 대상으로 Context 정보를 추가하는 과정이라 정의할 수 있다. 육하원칙에 따라 로그 내 관련 정보를 추가하는 과정이며 이와 같은 과정의 주요 목적은 사이버 침해사고에 대한 분석 과정에서 탐지 효율성 및 유용성을 향상시키는 것이 핵심 요구 사항이라 할 수 있다. 따라서 SIEM 기반 Data Enrichment 과정은 Low-Value 형태의 아티팩트 데이터에 Context Plug-in 방식을 적용하여 사이버 침해사고 분석/탐지/대응에 유용한 데이터 생성/축적을 목적으로 하는 것이라 정의할 수 있다. 따라서, 아래 그림과 같이 SIEM 기반 수집된 데이터에 대한 순환처리 과정을 통해 원본 Raw 데이터 형태의 Log 정보(Low-Value Artifact)에 대한 Data Enrichment 과정을 통해 High-Value 형태의 eLog 아티팩트를 생성하는 과정이라 할 수 있다.



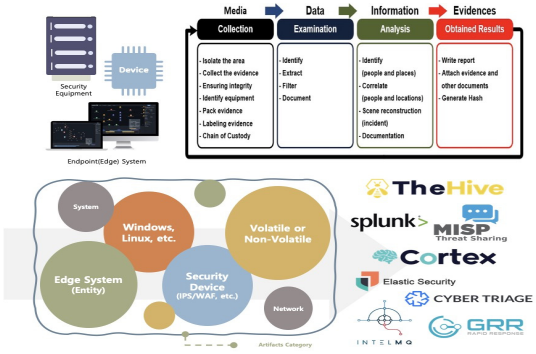
[Fig. 4] Data Enrichment for SIEM

상기에 분석한 내용을 토대로 본 연구에서는 CTI와 SIEM 기반 Data Enrichment 과정의 주요 특징과 세부 메커니즘 분석 내용을 기반으로 사고관리 프레임워크 기반 Data Enrichment 세부 메커니즘에 대한 분석 과정을 수행하였다. 각 디바이스 또는 플랫폼/프레임워크로부터 수집된 데이터에 대한 특성 정보를 추출한 후 각종 위협 분석 프레임워크를 기반으로 데이터 보완 과정을 분석하였다.

4. SIEM 기반 데이터 보완 기술 비교 분석

4.1 주요 프레임워크 별 데이터 보완 방법 분석

아티팩트 데이터를 대상으로 사고관리 프레임워크에서 수집된 Low-Value 형태의 원본 이벤트 정보(Log 데이터 또는 이벤트 데이터 등 대상)에 대해 공격 탐지 및 대응 과정의 효율성을 향상시키기 위해 각종 프레임워크에서 데이터 보완 과정을 진행하는 과정에서 주요 특징과 내부 메커니즘, 그리고 각종 TIP 등에서 발견된 공통점, 차이점 및 특이사항 등에 대한 분석 과정을 진행하였다. 이와 같은 목적을 달성하기 위해 아래 그림 5와 같이 아티팩트에 대한 Data Enrichment 대상을 선별하고 각각에 대해 어떠한 범주 및 방법을 통해 강화 과정을 수행하는 것이 합리적인지에 대한 분석 과정을 진행하였다.



[Fig. 5] Artifact Data Enrichment Extension Feature Analysis Target

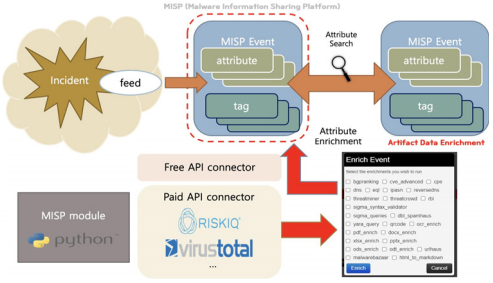
본 연구에서는 SIEM 개별 시스템 내에서 진행되고 있는 데이터 보완 기법에 대해 각각 분석하고 공통점 및 주요 기능을 비교 분석하였다. MISP[10], IntelMQ[11], The Hive[12], Cortex[13], Splunk[14], CyberTriage[15], Google GRR[16] 및 Elastic Security[17] 등 SIEM 관련 시스템 주요 8개 솔루션을 대상으로 각각 어떠한 데이터 보완 기법을 사용하고 있는지에 대해서 분석하였다. 각각에 대해 개별적인 특징을 살펴보면 다음과 같다.

4.2 MISP 및 IntelMQ 데이터 보완 방법 분석

우선 MISP 시스템과 IntelMQ, 그리고 CyberTriage와 Google GRR 및 Cortex 시스템에서 사용하고 있는 데이터 보완 기법의 세부 속성/대상 정보를 분석하였다. MISP 시스템은 오픈 소스 기반 사이버 위협 공유 플랫폼으로, 아래 그림 6과 같이 운영자에게 위협 관련 인텔리전스 피드(feed) 정보를 제공하며, 유료 또는 커뮤니티에서 제공하는 정보를 토대로 공격 징후, 취약 정보, 대테러 정보를 저장 및 공유한다. 특히 cyber security indicators를 사용하여 악성코드 분석을 공유, 저장, 협동할 뿐 아니라 Indicator of Compromise(IoC) 정보

를 이용해 정보통신 인프라, 단체를 겨냥한 공격, 사기, 위협을 감지하고 방지한다.

MISP은 network 및 system indicator를 속성 정보(attribute)로 표현하고 관련된 속성 정보를 하나로 묶어 이벤트 정보로 표현한다. MISP module은 Python으로 작성됐으며, expansion module은 VirusTotal이나 X-force Exchange 등과 같은 외부 API를 사용하여 IP, hash, URL, domain 등의 속성 정보에 대한 데이터 보완 과정을 수행한다. 특히 개별 expansion module을 이용하여 DNS domain 정보와 같은 특정 속성 정보에 대해 연관성이 있는 부가적인 정보를 제공한다.



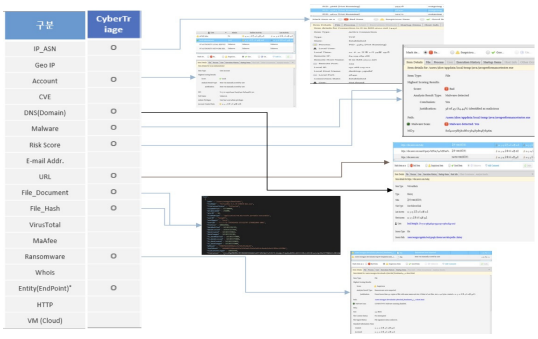
[Fig. 6] MISP Data Enrichment Mechanism

또한 IntelMQ 시스템은 메시지 큐잉 프로토콜(Message Queuing Protocol)을 이용하여 각종 보안 feed를 수집하고 처리하는 솔루션으로 collector, parser, expert 및 output 봇으로 구성된다. Collector bot은 메일 박스, 로컬 파일, MISP, 각종 서버 등에서 정보를 가져온다. parser bot은 특정 collector bot마다 존재하는 bot으로, 정보를 파싱하며, Expert bot은 각종 이벤트 정보에 대한 데이터 강화 기능을 제공한다.

4.3 CyberTriage 및 GRR 데이터 보완 방법 분석

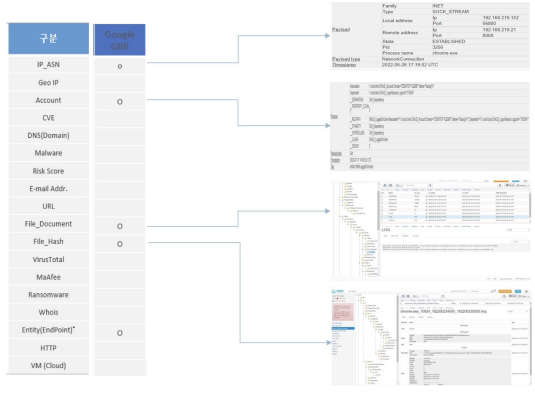
CyberTriage는 라이브 디지털 포렌식을 목적으로 하는 침해 사고 대응 프레임워크로, 관련된 침해사고 데이터에 빠르게 수집할 수 있으며, 40여개 이상의 malware/ransomware 탐지 엔진으로 이용하여 실행 파일에 대한 검사 과정 및 탐지 기능을 제공하기 위해서 Endpoint에 Collection tool을 설치하여 데이터를 수집한 후 각각의 수집한 항목들을 대상으로 데이터 보완 과정을 수행한다. 그리고 최종적으로 CyberTriage GUI 형태로 제공되는 Dashboard 내에서 확인하거나 Html, CSV, JSON 파일로 추출도 가능하다.

CyberTriage를 이용한 데이터 보완 과정을 살펴보면 아래 그림 7과 같이 IP_ASN, DNS, URL, Account 정보, Risk Score, Ransomware 등에 대한 판단 정보, 시스템 내 의심되는 파일에 대한 Hash 정보를 토대로 침해사고 여부를 판별할 수 있는 부가적인 데이터를 보완해 주는 기능을 제공한다.



[Fig. 7] CyberTriage Data Enrichment Extension Feature

Google GRR 시스템 역시 CyberTriage와 유사하게 원격 시스템 환경을 토대로 라이브 디지털 포렌식 기능 제공을 목적으로 하는 사고 대응 프레임워크이다. GRR 클라이언트는 조사하려는 시스템에 배포/설치되며 GRR 서버는 웹 기반 GUI를 사용할 수 있다. GRR은 다수의 클라이언트에서 많은 양의 데이터를 쉽고 빠르게 가져오기 위하여 Google의 protobuf 방식을 사용하여 수집한 침해사고 이벤트 정보를 수집하게 된다.



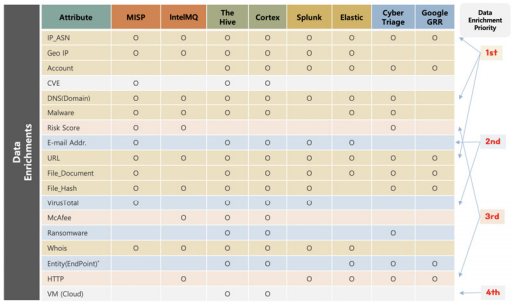
[Fig. 8] CyberTriage Data Enrichment Extension Feature

Google GRR은 수집된 데이터 중에서 IP_ASN 데이터 보완 과정을 통해 각각의 호스트/클라이언트 IP 주소에 대한 보완 과정 및 Netstat 정보를 수집하고, Account

보완 기능과 File_Hash 그리고 File_Document에 대한 보완 기능을 통해 Virtual Filesystem에 수집된 파일에 대한 그림 8과 같은 데이터 강화 기능을 수행한다.

4.4 데이터 보완 기능 비교 분석 결과

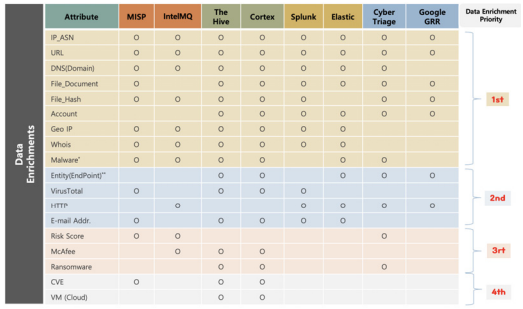
앞서 제시한 내용 이외에도 본 연구에서는 TheHive, Splunk 및 Elastic security 시스템에서 각기 수행하고 있는 데이터 보완 과정에 대해 공통점 및 특이점을 각각 분석하였으며, 8개의 프레임워크를 대상으로 각각 데이터 보완 속성에 대해 종합적으로 분석한 결과는 아래 그림 9와 같다. 그림에서 제시된 바와 같이 가장 많은 데이터 보완 기능을 제공하고 있는 것은 TheHive 및 Cortex 시스템인 것으로 확인되었다.



[Fig. 9] Extraction Result of Common Data Enrichment Extension Features

분석 결과 8개의 프레임워크에서는 IP_ASN 및 Geo_IP 등을 포함하여 대략 18개의 데이터 보완 속성을 제공하고 있다는 것을 확인할 수 있었으며, 공격자 IP 주소와 해당 Account 정보 그리고 네트워크 Domain 관련 상세 정보를 대부분의 시스템에서 공통적으로 적용하고 있는 것을 확인할 수 있었다. 이밖에도 해당 아티팩트 또는 이벤트 관련하여 로그 등에 기록되어 있는 URL 상세 정보를 Enrichment하는 기능을 사용하고 있었으며, 만일 각종 이벤트 또는 로그 정보 내에 첨부된 파일, 그리고 랜섬웨어 등 악성 행위 관련된 이벤트와 관련해서는 해당 파일에 대한 상세 정보와 이에 대한 해시 정보 등을 Enrichment하는 기능을 사용하고 있었다. 이와 함께 Whois, EndPoint에 대한 상제 정보를 Enrichment하는 기능을 대부분의 사고관리 프레임워크에서 적용하고 있었다.

8개의 프레임워크에서 공통적으로 제공하는 데이터 보완 기법인 경우 중요도가 높고 일반적으로 많이 사용되는 기법이라는 것을 의미한다. 따라서 본 연구에서는 데이터 보완 속성의 중요도를 우선순위 기준에 따라 정렬하였고 그 결과는 다음 그림 10과 같다. 분석 결과 IP_ASN, Geo IP, DNS, File Hash, Account, Whois, Malware 등 9개의 속성 정보에 대해서는 우선순위가 가장 높은 형태로 데이터 보완 과정이 수행되고 있음을 확인할 수 있었다. 그리고 CVE, VM cloud 기반 데이터에 대한 보완 과정은 상대적으로 우선순위가 낮다는 것을 확인할 수 있었다.



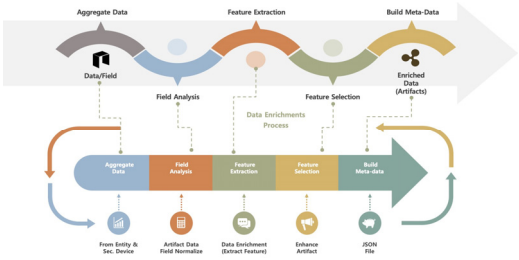
[Fig. 10] Extraction Result of Priority based Common Data Enrichment Extension Features

5. 도출된 데이터 보완 기술의 유효성 분석 기반 실제 시스템 적용 방안

5.1 데이터 보완 실제 사례 분석

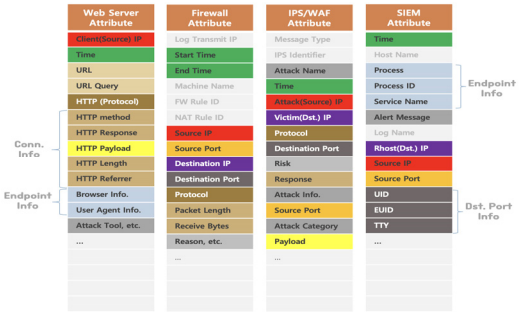
앞서 제시한 내용과 같이 기존의 프레임워크에 대한 분석 결과를 통해 도출된 Data Enrichment 세부 항목에 대한 적용 가능성 및 적합도를 실제 시스템 환경에 적용하여 검증하였다.

실제 시스템을 대상으로 침해사고 발생시 효율적인 대응을 위해서는 아래 그림 11과 같이 데이터 수집(Aggregate data), 특징 추출(Feature Extraction) 및 메타데이터 구축(Build Meta-data)와 같이 크게 세 가지 단계를 수행하게 된다. 그리고 이와 같은 3단계 과정은 다시 특징 분석(Feature Analysis) 및 특징 선택(Feature Selection)의 두 가지 단계와 결합하여 최종적으로 다섯 단계를 수행하게 된다. 구체적으로 각각의 보안장비로부터 획득되는 아티팩트 및 이벤트를 중심으로 각각의 필드 정보에 대한 유사도 및 공통 필드 도출 과정을 수행하였고, 이를 토대로 데이터 보완 대상이 되는 특징 필드를 선정할 수 있었으며, 해당 특징 필드 정보에 따라 앞서 제시한 데이터 보완 과정에서의 정합도를 측정할 수 있었다.



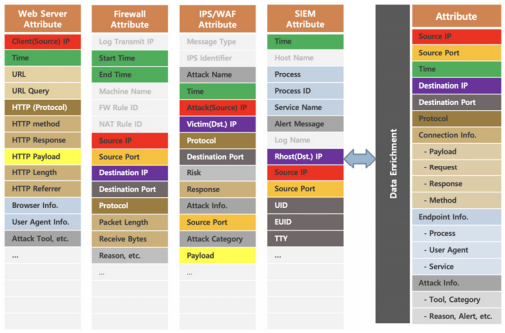
[Fig. 11] Detailed Procedure Diagram of Data Enrichment Process

구체적으로 아래 그림 12와 같이 실제 4종의 보안장비에서 수집 및 저장되는 로그 정보에 대한 분석 과정을 수행하였다. 웹 서버로 로그 정보(Web Server Log Data)와 방화벽 로그(Firewall Log), 침입차단 시스템/웹방화벽 로그(IPS/WAF Log) 및 SIEM 로그(SIEM Log)와 같이 네 가지 보안 장비로부터 획득되는 정보 내 필드 구성 요소를 각각 분석하고 앞서 제시한 데이터 보완 방식과의 연관성 및 적합도를 측정/검증하는 과정을 수행하였다.



[Fig. 12] Analysis of Common Feature Elements Extracted from Security Equipment

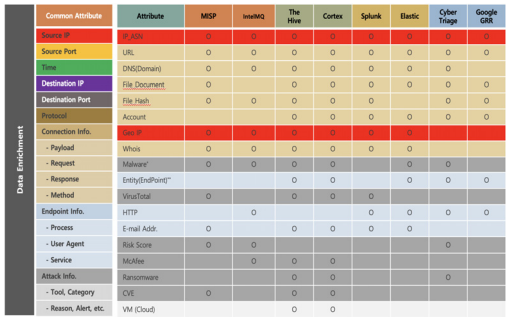
위 그림에서 확인할 수 있듯이 기존의 4가지 보안장비에서 수집되는 로그 정보 내에는 Source IP 정보, Time 정보가 공통적으로 포함(색깔로 표현)되어 있으며, 각 장비별로 수집 되는 정보가 조금씩 다르게 구성되어 있는 것을 확인할 수 있었다. 하지만 이를 다시 분석해 보면 일반적으로 Source Port, Destination IP에 관한 필드 정보를 포함하고 있으며, URL, HTTP method/response/referrer 등과 같이 Packet Connection 관련 정보가 포함되어 있으며, Browser Info/User Agent Info/Process ID 등과 같이 End Point에 대한 정보가 포함되어 있는 것을 확인할 수 있었다. 따라서 아래 그림 13과 같이 앞에서 공통 필드 요소를 추출하면 Source IP/Port, Time, Destination IP/Port, Protocol 및 Connection Info.(Payload, Request, Response, Method), EndPoint Info.(Process, User Agent, Service) 그리고 Attack Info.(Tool, Category, Reason, Alert, etc.)로 그룹지을 수 있었다.



[Fig. 13] Correlation Analysis of Common Feature Elements Extracted from Security Equipment

5.2 제안한 데이터 보완 방법의 적합도 검증

따라서 실제 4종의 보안장비에서 수집/저장되는 공통 필드 정보를 대상으로 앞서 제시한 데이터 보완 방법과의 적합도를 분석하면 다음과 같다. 각 장비에서 공통적으로 수집되는 특징 필드 항목 6개(Source IP, Source Port, Time, Destination IP, Destination Port, Protocol)와 3개 그룹(Connection Info., Endpoint Info., Attack Info.) 정보와 본 연구에서 제시한 데이터 보완 항목 18개(IP_ASN, URL, DNS, File/Hash, Geo IP, Whois, Malware 등) 중에서 VM(Cloud)를 제외하고 17개의 항목이 일치(94.4%)하는 것으로 나타났다.

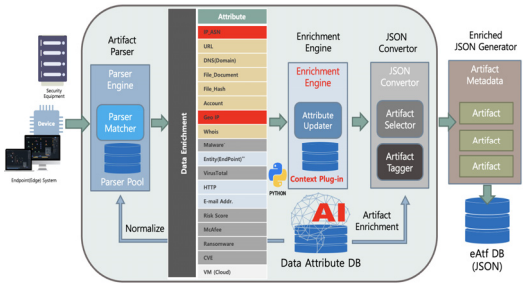


[Fig. 14] Compatibility Analysis of Common Data Enrichment Feature Elements Related from Security Equipment

따라서 본 연구에서 제시한 데이터 보완 항목을 실제 보안 장비에 적용하였을 경우 사이버 공격 침해사고 대응을 위한 데이터 보완 대상 항목으로 적합하다는 것을 확인할 수 있었다.

5.3 데이터 보완 시스템 실제 적용 방안

각종 EndPoint 및 Device로부터 수집된 아티팩트를 대상(Artifact Collection Process)으로, 이에 대한 전처리 과정으로 식별/필터링 및 추출 과정을 수행(Data Examination Process)한 후 아티팩트에 대한 분석 과정(Artifact Information Analysis Process)을 진행하는데 있어서 중요한 부분을 차지하는 Data Enrichment 과정을 수행하여 그림 15와 같이 디지털 포렌식 분석을 위한 메타데이터를 자동 생성할 수 있다.



[Fig. 15] System Diagram of Data Enrichment Extension Feature Element for Cyber Incident Response

아티팩트 수집 대상을 선별한 후 각종 디바이스와 Endpoint로부터 수집된 Artifact(비휘발성 아티팩트 중심)에 대한 Data Enrichment 과정을 수행하며, Feature Engineering을 토대로 각각의 데이터 필드에 대한 전처리/정규화 과정을 수행한다. 그리고 Data Enrichment 대상이 되는 주요 필드 등을 추출하는 Feature Extraction 과정을 수행한 다음에 Data Enrichment 모델링 과정을 진행하며, 최종적으로 사고관리 프레임워크에서 필요로 하는 확장요소(Extended Elements)를 JSON 기반 Meta-Data 형태로 추출/생성 및 축적할 수 있도록 프로토타입을 설계하였다.

5.4 기존 연구와의 비교 분석 및 평가

본 연구에서 제시한 방법은 기존 연구와 차별성을 제공한다. 본 연구에서 제시한 방법은 실제 침해사고 대응 프레임워크로 사용되고 있는 8개 솔루션 내 데이터 보완 메커니즘 현황 분석을 토대로 각각 공통적인 확장 요소를 도출한 것이다.

최근 IoT 기기를 대상으로 한 대단위 DoS 공격 또는 이와 유사한 악성코드 감염을 통한 악의적인 침해사고/공격 등이 급증하고 있다. 그리고 지능화된 APT 공격 등이 발생하였을 경우 능동적으로 대응하기 위한 데이터 보완 기술이 필요한데, 본 연구에서 제시한 우선순위 기반 데이터 보완 확장 요소를 토대로 실세계에 적용 가능할 것으로 기대된다. 특히 4가지 실제 보안 장비에서 수집/저장되는 로그 정보를 대상으로 적합도를 분석한 결과 본 연구에서 도출한 데이터 보완 항목은 94.4% 유사도를 갖는 것으로 확인되기에 향후 IoT 기기에 대한 보안성 강화 및 침해사고 분석을 위한 시스템 구축 및 연구개발 과정에서 매우 유용하게 사용될 수 있을 것으로 기대된다.

6. 결론

각종 통신 기기들이 대부분 IoT·휴대용 통신 수단과 결합된 상태로 다양한 서비스와 연동되면서 해당 기기의 보안 취약점을 악용한 사이버 공격이 급증하고 있다. 특히 최근에는 지능형 지속위협(APT) 공격을 통해 대단위 네트워크 환경에서 이기종 디바이스를 대상으로 한 사이버 공격이 계속되고 있다. 따라서 사이버 환경에서 사고관리 대응 체계의 유효성을 향상시키기 위해서는 위협분석 및 탐지 성능을 향상시킬 수 있도록 수집된 아티팩트 데이터에 대한 데이터 보완(Data Enrichment) 프로세스를 개발할 필요가 있다. 이에 본 연구에서는 침해사고 분석을 위해 수집된 아티팩트를 대상으로 기존의 사고관리 프레임워크에서 수행하는 데이터 보완 공통 요소를 분석하여 실제 시스템에 적용 가능한 특징 요소를 도출하고, 이를 토대로 개선된 사고분석 프레임워크 프로토타입 구조를 제시하였으며 도출된 데이터 보완 확장 요소의 적합도를 검증하였다. 이를 통해 이기종 디바이스로부터 수집된 아티팩트를 대상으로 사이버 침해사고 분석시 탐지 성능을 향상시킬 수 있을 것으로 기대된다.

REFERENCES

[1] S.N.Swamy and S.R.Kota, "An Empirical Study on System Level Aspects of Internet of Things (IoT)," IEEE Access, Vol.8, pp.188082-188134, 2020.

[2] Hassannataj Joloudari, J., Haderbadi, M., Mashmool, A., GhasemiGol, M., Shahab, S., and Mosavi, A., "Early detection of the advanced persistent threat attack using performance analysis of deep learning", arXiv e-prints, 2020.

[3] Chen, P., Desmet, L., Huygens, C., "A Study on Advanced Persistent Threats," Communications and Multimedia Security. CMS 2014, Lecture Notes in Computer Science, Vol.8735. Springer.

[4] Gustavo Gonzalez-Granadillo, Susana Gonzalez-Zarzosa, Rodrigo Diaz, "Security Information and Event Managment (SIEM): Analysis, Trends, and Usage in Critical Infrastructures," Sensors, Vol.21, No.14, 2021,

[5] Md Sahrom Abu, Siti Rahayu Selamat, Aswami Ariffin, Robiah Yusof, "Cyber Threat Intelligence – Issue and Challenges," Indonesian Journal of Electrical Emgineering and Computer Science, Vol.10, No.1, April 2018, pp.371-379.

[6] Hussam Mohammed, Hathan Clarke, Fudong Li, "An Automated Approach for Digital Forensic Analysis of Heterogeneous Big Data," Journal of Digital Forensics, Security and Law, Vol.11, No.2, 2016, pp.137-152.

[7] A. Alenezi, H. Atlam, R. Alsagri, M. Alassafi, and G. Wills, "IoT Forensics: A State-of-the-Art Review, Challenges and Future Directions," Proceedings of the 4th International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2019), pp.106-115.

[8] H.Lee, "Intrusion Artifact Acquisition Method based on IoT Botnet Malware," Journal of The Korea Internet of Things Society, Vol.7, No.3, pp.1-8, 2021.

[9] Maria Stoyanova, Yannis Nikoloudakis, Spyridon Panagiotakis, Evangelos Pallis, and Evangelos K. Markakis, "A Survey on the Internet of Things (IoT) Forensics: Challenges, Approaches, and Open Issues," IEEE COMMUNICATIONS SURVEYS & TUTORIALS, Vol.22, No.2, pp.1191-1221, SECOND QUARTER 2020.

[10] MISP, Open Source Threat Intelligence and Sharing Platform, "https://www.misp-project.org".

[11] IntelMQ, "https://intelmq.readthedocs.io".

[12] TheHive, "https://thehive-project.org".

[13] Cortex, "https://github.com/TheHive-Project/Cortex".

[14] Splunk, "https://www.splunk.com".

[15] CyberTriage, "https://www.cybertriage.com"

[16] Google GRR, "https://github.com/google/grr"

[17] Elastic Security, "https://www.elastic.com/security"

이 형 우(Hyung-Woo Lee) [종신회원]

■ 1994년 2월 : 고려대학교 컴퓨터학과 (학사)
■ 1996년 2월 : 고려대학교 컴퓨터학과 (석사)
■ 1999년 2월 : 고려대학교 컴퓨터학과 (박사)
■ 1999년 3월 ~ 2003년 2월 : 백석대학교 정보통신학부 교수
■ 2003년 3월 ~ 현재 : 한신대학교 컴퓨터공학부 교수

〈관심분야〉
사물인터넷, 정보보호, 모바일 보안 및 디지털 포렌식