USB 사용 이력 기반 이상 행위 탐지를 위한 시각화 대시보드 구현 : 기업 연구소의 폐쇄망 환경 관점

김민규*, 이용구**
* **한국수력원자력 중앙연구원
e-mail:kim.minkyu@khnp.co.kr* , lee.yongku@khnp.co.kr**

Design of a visualization dashboard for detecting anomalies based on USB usage history

Min-Kyu Kim*, Yong-Ku Lee**
* **Central Research Institue, Korea Hydro & Nuclear Power Co., Ltd.

요 약

폐쇄망 환경의 연구소는 외부와 연결이 차단되어 있지만, 업무 편의성과 데이터 이전을 위해 USB 사용이 빈번하여 보안 위험이 존재한다. 본 연구는 보안 솔루션에서 수집된 USB 사용 로그를 기반으로 사용자 행위의 이상 여부를 분석하고 시각화하는 대시보드를 설계한다. 제안된 시스템은 USB 사용 이력이 저장되는 보안솔루션의 데이터를 수집하고, 이를 시각화하여 이상 행위를 효과적으로 탐지하는 방법을 제시한다. 사용자별 USB 사용 타임라인, 시간대별 행위 밀도 히트맵, 매체 이동 상태별 파일 이동 현황 바차트, 출발지-목적지 연결 Sankey Diagram 등을 활용하여 직관적인 이상 행위 탐지가 가능하고, 이상 탐지 가상 시나리오를 통해 효과성을 검증하였다.

1. 서론

1.1 연구 배경 및 필요성

폐쇄망은 외부 네트워크와 물리적으로 분리된 내부 네트워크를 의미하며, 국방, 원자력, 핵심 기술 연구소 등 보안이 중요한 영역에서 주로 사용된다. 이러한 폐쇄망 환경에서도 업무 편의성과 데이터 이전을 위해 이동식 저장매체(USB)의 사용이 빈번하게 이루어지고 있으며, 이는 내부 정보 유출의 주요 경로가 될 수 있다. 본 연구는 폐쇄망이라는 제한된 환경 내에서 진행되었으며, 보안 솔루션을 통해 USB 사용을 엄격히 통제하여 사용자의 모든 파일 전송 기록을 저장할 수 있는 환경이다. 이러한 통제의 주요 목적은 연구소 내 기술자료의 유출을 예방하고, 불가피하게 발생할 수 있는 유출 사고에 대한 추적을 가능하게 하는 것이다.

폐쇄망 환경에서 로그 기반의 USB 사용 행위 분석을 통해 보안 위협을 탐지하는 것은 중요한 방안이 될 수 있다. 그러나 기존의 로그 기반 분석은 주로 텍스트 형태의 결과를 제공하여 대량의 데이터에서 이상 행위를 직관적으로 파악하기 어렵다는 한계가 있다. 특히 연구소 환경에서는 대량의 USB 사용 로그가 발생하지만, 이를 효과적으로 모니터링하고 분석할 수 있는 도구가 부족한 실정이다. 이에 USB 사용 로그 데이터를 효과적으로 시각화하여 이상 징후를 쉽게 포착할 수 있는 대시보드의 필요성이 대두되고 있다.

1.2 연구 목적

본 연구의 목적은 폐쇄망 환경의 연구소에서 보안 솔루션을 통해 수집된 USB 사용 로그를 기반으로, 파일 전송 이력에 대한 가시성을 확보하고 정보보안 비전문가의 시각으로도 위규 사항을 파악할 수 있도록 분석된 정보를 제공하는 시각화 대시보드를 설계하는 것이다. 구체적인 목표로는 폐쇄망 환경에서 수집되는 USB 사용 로그에 대한 효과적인 분석 방법론을 개발하고, 사용자 행위의 이상 여부를 쉽게 판단할 수 있는 직관적인 시각화 대시보드를 설계하며, 정보보안 비전문가도 이해하기 쉬운 분석 정보 제공을 위한 시각화 방법을 구현하고, 시각화 대시보드의 효과성 검증을 위한 실험 설계 및 평가를 수행하는 것이다.

이를 통해 이상 행위 탐지의 효율성과 직관성을 높이고, 초기 단계에서 내부 위협을 탐지하여 정보 유출과 같은 보안 사고를 예방하고자 한다. 또한 기술 자료 유출에 대한 예방 및 추적 기능을 강화함으로써 연구소 내 중요 정보 자산을 보호하는데 기여하고자 한다.

2. 관련 연구 및 기술 배경

2.1 로그 기반 이상행위 탐지 및 시각화 관련 선행 연구

로그 기반 이상행위 탐지 및 시각화에 관한 연구는 다양한 분야에서 진행되어 왔다. 웹 로그 데이터를 기반으로 한 침입 탐지 시각화 연구에서 웹 공격 패턴을 다양한 시각화 기법으로 표현하여 이상 행위를 효과적으로 탐지할 수 있음을 보인 선행연구가 있었고,[1] 오픈소스 ELK 스택을 활용한 보안 모니터링 시스템을 구현하여 네트워크 트래픽 로그, 방화벽 로그, 서버 로그 등 다양한 로그 데이터를 수집, 분석, 시각화하는 방법을 제시했다.[2] 모두 로그 데이터의 시각화를 통해 실시간으로 이상 행위를 탐지하고 대응하는 방법을 설명했으며, 특히 시계열 그래프, 지리적 위치 분석, 히트맵 등 다양한 시각화 도구를 활용하여 보안 위협을 직관적으로 파악할 수 있음을 확인하였다.

네트워크 보안 관제를 위한 로그 시각화 방법 연구[3]에서는 대시보드 설계의 중요성을 강조하고, 다양한 시각화 기법(바차트, 파이차트, 꺾은선그래프, 히스토그램, 히트맵 등)을 적용하여 신속한 정보 제공 및 이상 탐지가 가능함을 설명했다. 그들은 보안 모니터링의 목적에 따라 수집된 다양한 로그를 기반으로 각 로그의 특성에 맞는 시각화 도구를 선택하여 통합 대시보드를 구성하는 방법론을 제시했다. 또한 RGB 팔레트를 이용한 보안 로그 시각화 및 보안 위협 인식 연구[4]에서는 색상을 통한 시각적 패턴 인식의 효과성을 입증했다. 이 연구는 대량의 로그 데이터를 색상 코드로 변환하여 패턴화함으로써 이상 징후를 빠르게 감지할 수 있는 방법을 제안했다.

2.2 보안 솔루션 로그의 수집 및 처리 방식 개요

본 연구에서는 정보보안 솔루션에 의해 통제되는 USB를 통한 파일 전송 기록을 통합적으로 수집하고 분석하는 방식을 채택하였다. 구체적으로 MySQL 데이터베이스를 활용하여 [표 1]과 같은 정보를 수집하고 있다.

[표 1] 정보보안 솔루션에 저장되는 정보

정보명	설명
일시	파일 전송이 발생한 일자, 시간 정보
사용자 ID	파일 전송을 수행한 사용자 식별 정보
USB 시리얼번호	사용된 USB 장치의 고유 식별 번호
IP 주소	파일 전송이 발생한 시스템의 네트워크 주소
원본 파일	전송된 파일의 원래 경로 및 위치
목적지 폴더	파일이 전송된 대상 위치
파일 처리 방법	이동, 복사, 삭제 등 파일에 대한 처리 유형
파일 이름	전송된 파일의 이름
확장자	파일의 유형을 나타내는 확장자 정보
파일 크기	전송된 파일의 용량 정보

이런 데이터는 폐쇄망 환경 내에서 발생하는 모든 USB 파일 전송 활동을 추적할 수 있게 하며, 이상 행위 탐지를 위한 기초 자료로 활용된다. 데이터 수집은 자동화된 프로세스로 진행되며, 수집된 데이터는 정규화 과정을 거쳐 분석 및 시각화에 적합한 형태로 변환된다.

2.3 시각화 도구 및 기술 소개

본 연구에서는 파이썬을 기반으로 한 웹 형태의 시각화 대시보드를 구현하였으며, 사용된 패키지 목록은 [표 2]와 같다.

[표 2] 대시보드 구현에 사용된 파이썬 패키지 설명

패키지명	설명
streamlit	파이썬 기반 마이크로 웹 프레임워크로, 대시보드의 백엔드 서버 구현에 사용
Pandas	데이터 처리 및 분석을 위한 라이브러리로, MySQL에서 추출한 데이터의 가공 및 분석에 활용
Plotly	인터랙티브한 차트 및 그래프 생성을 위한 라이브러리로, 특히 대시보드의 동적 시각화에 활용
Bootstrap	반응형 웹 디자인을 위한 CSS 프레임워크로, 대시보드의 UI 구현에 사용

3. 시스템 설계

3.1 로그 수집 구조

본 연구에서 제안하는 시스템은 폐쇄망 환경의 연구소에서 사용되는 USB 보안 솔루션으로부터 로그를 수집하고 분석하는 구조이다. 시스템의 구조는 크게 데이터 수집 단계, 데이터 처리 단계, 시각화 단계의 3단계로 구성된다.

데이터 수집 단계에서는 USB 보안 솔루션을 통해 모든 USB 사용 관련 로그를 MySQL 데이터베이스에 수집한다. 이 과정에서 일시, 사용자 ID, USB 시리얼번호, IP 주소, 원본 파일, 목적지 폴더, 파일 처리 방법, 파일 이름, 확장자, 파일 크기 등의 세부 정보가 자동으로 기록된다.

데이터 처리 단계에서는 수집된 로그 데이터를 정제하고 분석하는 작업을 수행한다. 이 단계에서는 파이썬의 Pandas 라이브러리를 활용하여 데이터를 가공하고, 분석에 필요한 추가 정보를 생성한다.

시각화 단계에서는 파이썬 웹 프레임워크를 기반으로 처리된 데이터를 대시보드 형태로 시각화한다. streamlit, Plotly 등의 라이브러리를 활용하여 사용자별 USB 사용 이력, 시간대별 행위 분석, 파일 이동 현황 등 다양한 관점에서 데이터를 분석하고 표현한다.

3.2 수집 항목

USB 사용 이력 기반 이상 행위 탐지를 위해 [표 1]의 데이터를 모두 동일하게 수집하여 분석한다. 해당 데이터는 USB를 통한 정보 유출 위험을 효과적으로 탐지하는데 필요한 정보를 포함하며, 특히 파일 처리 방법과 같은 세부 정보는 파일 복사 및 이동 패턴을 분석하여 이상 행위를 탐지하는데 중요한 역할을 한다.

3.3 로그 전처리 및 정규화

수집된 로그 데이터는 다양한 소스에서 제공되며 형식이 다를 수 있다. 효과적인 분석과 시각화를 위해 다음과 같은 전처리 및 정규화 과정을 거친다:

3.3.1 경로 단순화
파일 경로는 시스템마다 다를 수 있으며, 긴 경로명은 분석과 시각화를 복잡하게 만든다. 경로의 중요한 부분만 추출하거나 패턴화하여 분석을 용이하게 한다. 예를 들어, "C:\Users\username\Documents\confidential\project-x\report.docx"를 "Documents\confidential\project-x\report.docx"로 단순화할 수 있다.

3.3.2 시간대 일치화
모든 로그의 시간 정보를 한국 표준 시간대(UTC+9)로 변환하여 일관된 시간 기반 분석이 가능하도록 한다.

3.3.3 파일 유형 분류
파일 확장자를 기반으로 파일 유형을 분류한다. 예를 들어, ".docx", ".xlsx", ".pptx" 등은 "Office 문서"로, ".dwg", ".dxf" 등은 "CAD 파일"로 분류할 수 있다. 이런 분류는 특정 유형의 파일 이동에 대한 패턴 분석을 용이하게 한다.

3.3.4 데이터 보강
외부 데이터 소스를 활용하여 로그 데이터를 보강한다. 예를 들어, 조직의 근무 시간 정보를 추가하여 근무 시간 내/외 작업을 구분하거나, 부서별 사용 패턴 분석을 위한 이상 유무 판단의 기준이 되는 임계치 정보를 조정할 수 있다.

3.4 이상행위 탐지 기준 설계

이상 행위 탐지를 위해 네 가지 주요 항목을 기준으로 설정하였다. 이상 행위로 분류되는 항목에는 근무 시간 외 USB 사용, 대용량 파일 전송, 단시간 내 다수 파일 복사 행위, 그리고 민감한 확장자 파일 이동이 포함된다.

구체적으로, 근무 시간은 오전 9시부터 오후 6시까지로 정의하여 이 시간대를 벗어난 USB 활동을 이상 행위로 탐지한다. 대용량 파일 전송과 관련해서는 100MB를 초과하는 파일 전송을 이상 행위의 임계값으로 설정하였으며, 이 기준은 단시간 내 다수 파일 복사 행위의 총량에도 동일하게 적용된다. 또한, 민감한 확장자 파일의 경우 연구소 환경에서 특히 중요한 CAD 도면 파일(.dwg)과 대용량 PDF 문서와 같은 설계도면 관련 파일을 중점적으로 모니터링하도록 탐지 기준을 설계하였다.

3.5 시각화 구성요소

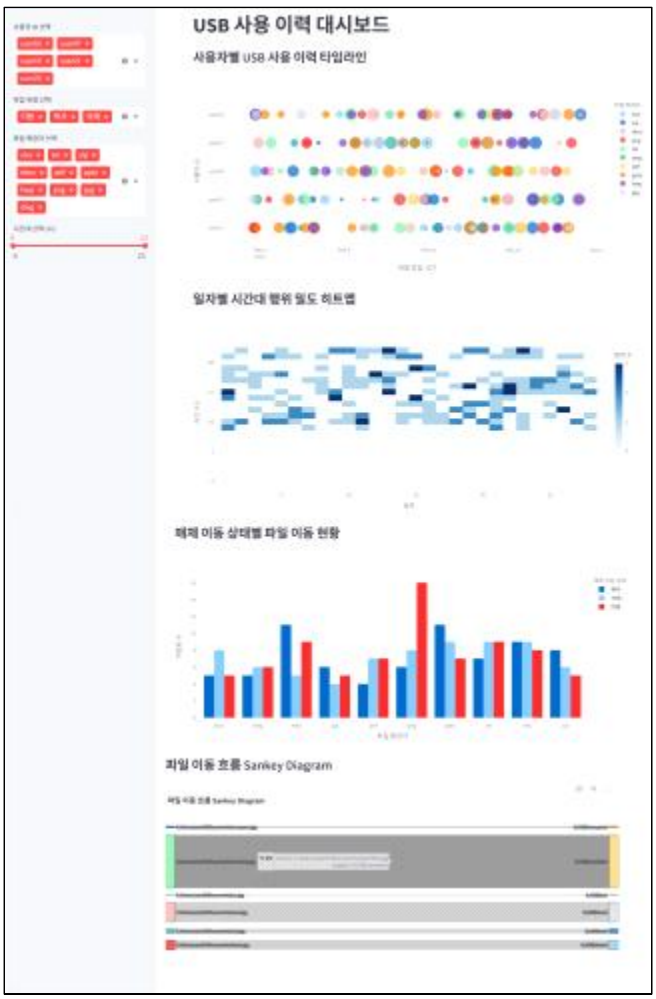
3.5.1 사용자별 USB 사용 이력 타임라인
사용자별 USB 사용 이력 타임라인은 수평축의 시간을 기준으로 각 사용자의 USB 활동을 시간 순서대로 표시하는 시각화 방식이다. 주요 이벤트(USB 연결, 파일 작업, USB 해제)를 아이콘으로 표시하고, 파일 크기에 따라 마커 크기를 조절하며, 파일 유형별로 색상을 구분하여 직관적인 식별이 가능하다. 시간대 구분과 드릴다운 기능을 통해 비정상 시간대 사용이나 특이 패턴을 빠르게 파악할 수 있다.

3.5.2 시간대별 행위 밀도 히트맵
시간대별 행위 밀도 히트맵은 수평축에 24시간, 수직축에 요일을 표시하여 USB 사용 활동의 집중도를 색상 강도로 표현한다. 셀의 색상 강도는 해당 시간대의 USB 활동 빈도를 나타내며, 파일 작업 유형별 분리 표시, 부서별/개인별 필터링, 이상치 강조 기능을 제공한다. 이 히트맵을 통해 조직의 전반적인 USB 사용 패턴을 파악하고 비정상 시간대 활동을 식별할 수 있다.

3.5.3 매체 이동 상태별 파일 이동 현황 바차트
매체 이동 상태별 파일 이동 현황 바차트는 파일의 이동, 복사, 삭제 등 작업 유형별로 이벤트 수와 데이터 용량을 시각화한다. 파일 확장자별로 색상을 구분하고, 시간축 슬라이더를 통한 기간 필터링, 사용자/부서별 필터, 정렬 기능을 제공하여 특정 파일 유형의 비정상적 이동이나 대용량 파일 전송과 같은 잠재적 위험 행위를 시각적으로 식별할 수 있다.

3.5.4 Sankey Diagram 기반 흐름 시각화

Sankey Diagram은 왼쪽의 출발지와 오른쪽의 목적지 간 파일 이동의 흐름을 시각화하는 도구로, 연결선의 두께는 이동된 데이터 용량이나 파일 수에 비례한다. 파일 유형이나 작업 유형에 따라 색상을 구분하고, 상호작용 기능과 시간 필터를 제공하여 특정 경로의 상세 정보 확인과 기간별 데이터 흐름 변화 추적이 가능하다. 이 방식은 특정 출발지에서 목적지로의 비정상적 데이터 이동 패턴을 효과적으로 파악할 수 있게 한다.



[그림 1] 완성된 대시보드 웹 화면

4. 시나리오 검증 및 결론

4.1 실험 설계 및 이상행위 시나리오 검증

30일간의 가상 데이터를 기본 데이터셋으로 활용하여 실험을 진행하였다. 총 4가지 이상행위 시나리오(근무 시간 외 사용, 대용량 파일 전송, 다수 파일 복사, 민감 확장자 파일 이동)를 특정 사용자 로그에 설계하고 시각화 대시보드를 통해 검증하였다. 실험 결과, 설계된 시각화 요소들은 각 이상 행위 유형을 직관적으로 표현하여 기존 텍스트 기반 로그 분석 방식에 비해 이상 행위 탐지 효율이 크게 향상되었다. 특히 Sankey Diagram을 통한 데이터 흐름 시각화가 파일 이동 패턴 식별에 매우 효과적임을 확인하였다.

4.2 결론 및 향후 과제

폐쇄망 환경 연구소의 USB 사용 이력 기반 이상 행위 탐지를 위한 시각화 대시보드를 설계하고 효과를 검증하였다. 정보보안 비전문가도 쉽게 이해할 수 있는 직관적인 시각화 방식을 통해 기술 자료 유출 위험에 대한 가시성을 확보하고 빠른 이상 징후 탐지가 가능해졌다. 향후 과제로는 머신러닝 기반 이상 탐지 자동화, 폐쇄망 환경에 최적화된 경량 대시보드 설계, 그리고 실시간 알림 체계 구현을 통해 시스템의 효율성과 정확성을 더욱 향상시키는 방안을 연구할 계획이다. 이를 통해 기술 자료 보호 및 정보 유출 방지 역량을 지속적으로 강화할 수 있을 것으로 기대된다.

참고문헌

[1] 이수용, 구본화, 조준일, 조금환, "웹 로그 데이터셋을 활용한 침입탐지 시각화 연구", 한국정보과학회 학술발표논문집, 제34권 2호, pp. 267-271, 6월, 2007년.
[2] 현종현, 김형진, "오픈소스 ELK 스택을 활용한 빅데이터 분석 기반의 보안 운영 구현", 디지털콘텐츠학회 논문지, 제19권 1호, pp. 181-192, 1월, 2018년.
[3] 조우진, 신효정, 김형식, "네트워크 보안 관제를 위한 로그 시각화 방법", 스마트미디어저널, 제7권 4호, pp. 72-78, 12월, 2018년.
[4] 이동건, 김휘강, 김은진, "RGB Palette를 이용한 보안 로그 시각화 및 보안 위협 인식", 정보보호학회논문지, 제25권 1호, pp. 61-73, 2월, 2015년.