OCSF 기반 보안 로그 통합 분석 연구

문석환*, 박혜수**, 김미소**, 박상경*, 송지현**, 어영민**

*인하대학교 (학부생), **중부대학교 (학부생)

A Study on OCSF-based Integrated Analysis of Security Logs

Seokhwan Mun*, Hyesu Park**, Miso Kim**,
Sangkyoung Park*, Jihyun Song**, Youngmin Eo**

*Inha University (Undergraduate student)

**Joongbu University (Undergraduate student)

요 약

디지털 전환, 클라우드 확산, 사물 인터넷 기기 증가 등으로 사이버 환경이 복잡해지고 다양한 형태의 보안 로그가 발생하고 있다. 이렇게 복잡해진 시스템에서 단일 로그로 탐지할 수 없는 공격이 존재하기 때문에 로그의 상관 관계를 분석하여 이상 행위를 식별해야 한다. 하지만 서로 다른 포맷과 수집 방식으로 인해 로그 간 통합 분석이 어렵고, 이상 탐지 및 사고 대응이 지연되는 문제가 지속적으로 제기되는 상황이다. SIEM 등을 활용하는 기존 분석 방법은 로그의 형식과 필드명의 차이로 인해 자동화 및 재사용성의 효율이 떨어진다. 따라서 본 연구는 이러한 문제를 해결하기 위해 키워드 기반의 OCSF(Open Cybersecurity Schema Framework) 클래스 결정 알고리즘을 적용하여 로그의 분류 정확도를 향상시킨다. 이후 대규모 언어 모델을 활용하여 OCSF를 기반으로 로그를 정규화함으로써 로그 매핑 성능과 작업 편의성을 개선한다.

I. 서론

디지털 환경이 다양한 시스템으로 복잡하게 구성되면서, 기업이 처리해야 할 로그 데이터의 양과 종류가 급증하고 있다. 2022년 클라우드 보안 경고 피로 보고서(2022 Cloud Security Alert Fatigue Report)[1]에 따르면, 이러한 데이터 처리 부담으로 인해 보안 팀의 55%가 중요한 경고를 놓치고 있다고 응답했다.

이기종 로그 통합 문제는 보안 전문가들 사이에서 지속적으로 지적되었으나, 이 문제를 해결하기 위한 체계적인 연구는 아직 부족한 상황이다. 본 연구는 이러한 현실적인 한계를 인식하고, 로그 통합 분석 체계를 구축하는 것이 실제 환경에서 얼마나 효율적이고 실용적인지 입증하고자 한다.

본 연구의 주요 기여는 다음과 같다:

- 보안 로그를 통합하기 위한 핵심 스키마로서 OCSF의 필요성과 실효성을 검증한다.
- 정규화된 로그에 시그마 룰을 적용하여, 다양한 환경에서도 범용적으로 활용 가능한 탐지 로직을 수립한다.
- 제안하는 통합 프레임워크가 실제 환경에서 내부 보안을 강화할 수 있음을 논의하고, 향후 보안 체계의 개선 가능성을 제시한다.

II. 연구 방법론

OCSF 기반 보안 로그 통합 및 실시간 분석 시스템을 구현하고, 다음과 같은 단계로 연구를 진행 후 평가하였다.

1) 로그 수집

로그 수집기를 활용하여 다양한 시스템에서 발생하는 여러 종류의 로그를 수집한다.

2) OCSF 기반 로그 정규화

본 연구에서 제안하는 알고리즘을 통해 수집된 로그의 특징을 분석하여 적합한 OCSF 클래스를 결정한다. 이후 gpt-4.1-mini 모델을 사용

하여 로그 정보를 OCSF(v1.5.0)의 각 필드에 매핑한다.

3) 저장 및 시각화

OCSF 포맷으로 정규화된 로그를 ELK 스택을 활용하여 대시보드를 구성한다. 이를 통해 로그 정보를 시각화하여 보안 이벤트의 전반적인 흐름과 이상 징후를 분석한다.

4) 시그마 룰 작성 및 확장성 평가

시그마 룰을 OCSF 포맷에 맞게 작성하고, pySigma 라이브러리를 활용하여 Lucene 쿼리로 변환한다. 변환된 쿼리로 단일 로그 기반의 공격 탐지를 진행하고, 로그의 상관 관계를 고려한 패턴 기반 탐지를 진행한다.

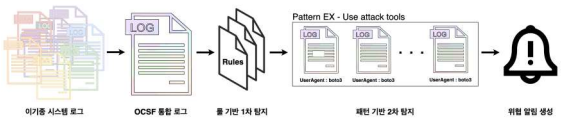
5) 공격 테스트

다양한 공격 시나리오를 구상하여 제안하는 프레임워크가 이상 행위를 효과적으로 탐지하고 분석할 수 있는지 평가한다. 또한, 새로운 공격이 식별되었을 때 시그마 룰을 신속하게 정의하고 확장하여 즉각적인 대응이 가능한지 검증한다.

III. 시스템 설계 및 평가

본 장에서는 제안하는 OCSF 기반 보안 로그 통합 및 분석 시스템의 아키텍처를 설명하고, 구축된 환경에서 공격 시나리오를 바탕으로 시스템 성능을 평가한다.

3.1 시스템 아키텍처 설계



[그림 1] 전체 시스템 아키텍처. 본 연구에서 제안하는 프레임워크의 구성과 흐름이다.

제안하는 시스템은 로그 수집 계층, OCSF 정규화, 그리고 데이터 저장 및 분석 계층으로 구성된다.

3.1.1 로그 수집 계층

다양한 이기종 시스템에서 발생하는 원시 보안 로그를 수집하는 역할을 담당한다.

시스템	버전	시스템	버전
Windows	Windows Server 2022	AWS S3	-
Ubuntu	24.04	AWS EC2	-
MySQL	8.0.42	AWS System Manager	-
Apache	2.4.58	AWS VPC	-
Suricata	8.0.0	AWS VPC Flow Log	-
Nginx	1.24.0	AWS CloudTrail	-
PHP	8.3.6	AWS GuardDuty	-
ModSecurity	2.9.7		

[표 1] 윈도우, 리눅스, AWS(2025.7.15. 기준) 등 다양한 로그 소스를 구성한다.

3.1.2 OCSF 정규화

Algorithm 1 Keyword-Based OCSF Class Determination Algorithm

```
function PREDICT_CLASS(Log_Entry)
    if Log_Entry is a structured log then
        Check specific fields
        if a direct match is found in ALL_KEYWORDS then
            return corresponding Class_UID
        end if
    else
        Initialize Match_Scores
        for each Class_UID in ALL_KEYWORDS do
            Calculate a Total_Score based on:
              - Presence of associated keywords in Log_Text
              - Position of keywords in Log_Text
              - Rank/importance of keywords in the Class_UID's keyword list
            if Total_Score then
                Add Total_Score to Match_Scores for Class_UID
            end if
        end for
        if Match_Scores is not empty then
            return Class_UID with the highest Total_Score
        end if
    end if
    return None
end function
```

제안하는 알고리즘은 원시 로그를 사전에 정의된 키워드 테이블을 참조하여 분석하고, 로그 내의 키워드 매칭 및 점수 부여를 통해 가장 적합한 OCSF class_uid를 결정한다. 이 키워드 테이블은 각 OCSF 클래스와 관련된 핵심 키워드 및 패턴을 포함한다. 이는 LLM의 추론에만 의존하는 것이 아니라, 명시적인 규칙 기반의 초기 분류를 통해 OCSF 정규화의 투명성과 일관성 보장에 기여한다.

해당 접근 방법은 LLM을 단독으로 사용할 때보다 정규화 정확도가 83.4% 증가했고, 이후 시그마 룰 작성과 패턴 기반 탐지를 효율적으로 수행할 수 있게 한다.

3.1.3 데이터 저장 및 분석 계층

OCSF로 정규화된 데이터를 기반으로 실제 위협을 탐지하고 분석하는 기능을 제공한다.

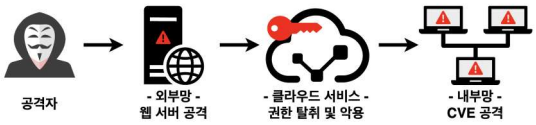
• 시그마 룰 적용: 정규화된 로그에 시그마 룰을 적용하여 공격을 탐지한다. pySigma 라이브러리를 활용하여 시그마 룰을 Lucene 쿼리로 변환하고, Elasticsearch에 적용한다.

• 패턴 기반 탐지: 공격을 특정할 수 있는 연관된 로그들이 일정 시간 내에 수집된다면 이를 하나의 패턴으로 간주하여 공격을 탐지한다.

• 실시간 경고 및 시각화: Slack API를 활용하여 탐지된 위협을 실시간으로 알림을 보내고, Kibana 대시보드로 모니터링하여 위협에 신속히 대응할 수 있는 기반을 마련한다.

3.2 실험 및 평가

3.2.1 공격 실험



[그림 2] CVE-2016-6662, CVE-2021-3156, CVE-2022-0847 등을 활용한 공격 시나리오.

해당 시스템의 실효성을 검증하기 위해 다양한 시스템 내외부 공격 시나리오를 구상하고 탐지 성능을 평가한다. 공격 탐지 시, 작성한 시그마 룰을 기반으로 Slack 알림이 전송되는지 확인하고, 해당 공격 정보가 대시보드에 실시간으로 시각화되어 분석 가능한지 확인한다.

3.2.2 평가

평가 항목	평가 내용	결과 및 의의
로그 통합 분석 실효성	이기종 로그의 상관관계 분석을 통한 공격 탐지 능력 검증	단일 로그로 탐지할 수 없는 공격 식별 & 정규화된 로그 기반 통합 분석 효율성 향상
시그마 룰 및 정의된 패턴 기반의 공격 탐지율 측정	새로운 위협 패턴에 대한 시스템의 적응성 및 통합 속도 평가	신규 공격에 대한 시그마 룰 작성 및 적용 시간을 기존 SIEM 대비 약 5분 단축. 탐지 정확도와 효율성은 보안 전문가의 역량에 의존
공격 탐지 정확도	시그마 룰 및 정의된 패턴 기반 공격 탐지율 측정	• 로그 탐지 정확도: 94.8% (474/500) • 패턴 탐지 정확도: 86% (43/50) • 탐지 실패 및 한계점 발견
키워드 기반 OCSF 클래스 결정 알고리즘 성능	'키워드 기반 알고리즘'과 'LLM 단독 사용'의 OCSF 클래스 분류 정확도 비교	• 제안 알고리즘: 96.1% (98/102) • LLM 단독 사용: 12.7% (13/102)

Ⅳ. 결론

본 논문에서는 이기종 시스템에서 다양한 형식의 보안 로그를 OCSF라는 공통된 포맷으로 정규화하고 통합 분석하는 실효적인 방안을 제시한다. 키워드 기반 분류 알고리즘으로 OCSF 클래스를 결정하고, 이를 바탕으로 대규모 언어 모델에 추가 정보를 제공함으로써 로그 정규화의 정확도를 향상시키는 방식을 제안한다. 또한, 로그 발생 위치에 관계없이 일관된 정규화 처리를 보장하므로, 하이브리드 환경에서 높은 유연성을 갖는다. 결론적으로, OCSF 기반의 로그 정규화는 여러 환경에서 시그마 룰을 효과적으로 관리·적용하여 다양한 공격 표면에 대해 탐지 범위를 확장시킬 수 있다.

[참고문헌]

[1] 2022 Cloud Security Alert Fatigue Report, ORCA Security, [Online]. Available: https://orca.security/wp-content/uploads/2022/03/Orca-2022-Cloud-Security-Alert-Fatigue-Report.pdf

[2] Julien Piet, Vivian Fang, Rishi Khare, Vern Paxson, Raluca Ada Popa and David Wagner, Semantic-Aware Parsing for Security Logs, June, 2025.

[3] Xiaojian Liu, IEEE Student Member, Yi Zhu, Shengwei Ji, Web Log Analysis in Genealogy System, IEEE International Conference on Knowledge Graph (ICKG), 2020.