하이브리드 클라우드 환경에서 시계열 로그 분석 기반 예방적 보안 모델 구축 연구

김건민[1], 김예진[2], 김태림[3], 김경백[1]

전남대학교 인공지능융합학과[1], 전남대학교 소프트웨어공학과[2], 전남대학교 인공지능학부[3]

geonminkim@jnu.ac.kr, ye031010@jnu.ac.kr, ktr0706@jnu.ac.kr,kyungbaekkim@jnu.ac.kr

A Time-Series Log Analysis Approach for Preventive Security in Hybrid Cloud Environments

Geonmin Kim[1], Yejin Kim[2], Taerim Kim[3], Kyungbaek Kim[1]

[1]Dept. of AI Convergence, Chonnam National Univ.,
[2]Dept. of Software Enginnering, Chonnam National Univ.,
[3]Dept. of Artificial Intelligence, Chonnam National Univ.

요 약

하이브리드 클라우드 환경은 온프레미스, 퍼블릭, 프라이빗 클라우드를 통합하여 유연성과 확장성을 제공하나, 이질적 구조와 모호한 네트워크 경계로 인해 보안 위협이 증가한다. 기존의 룰 기반 보안이나 이상 탐지 기법은 발생한 공격을 식별하는 데 초점을 두어 선제적 대응이 어렵다는 한계가 존재한다. 이에 본 연구는 탐지 중심 보안에서 예방 중심 보안으로의 전환을 목표로, 하이브리드 클라우드 환경에서 발생하는 로그 데이터를 수집 및 분석하여 공격 발생을 사전에 예측하는 모델을 제안한다. 실제 89만 건의 로그 데이터를 활용한 실험 결과, 제안 모델은 96% 이상의 정확도와, 0.94 이상의 F1-Score를 달성하였다. 특히, 정상 및 공격 클래스 모두에서 균형잡힌 성능을 보이며, 위협 및 탐지 누락을 최소화하면서 불필요한 차단도 줄일 수 있음을 입증하였다.

Ⅰ. 서 론

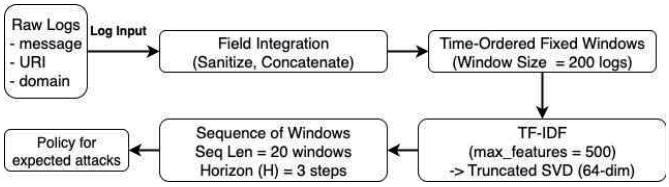
클라우드 컴퓨팅은 현대 기업의 핵심 IT 인프라로 자리매김하며, 비용 효율성과 확장성을 동시에 제공한다. 특히 온프레미스 자원과 퍼블릭, 프라이빗 클라우드를 통합하는 하이브리드 클라우드 환경은 다양한 비즈니스 요구를 충족할 수 있는 유연성을 제공한다[1]. 그러나 이러한 이점과 동시에 보안 측면에서는 도전 과제를 안고 있다[2-3]. 서로 다른 플랫폼과 관리 체계가 얽혀 있는 이질적인 구조는 네트워크 경계를 불명확하게 만들고 공격 표면을 크게 확대한다. 결과적으로 관리자는 복잡해진 로그와 이벤트 속에서 위협을 식별해야하며, 이는 전통적인 보안 기법의 한계를 불러온다. 이에 최근에는 머신러닝과 인공지능을 활용한 이상 탐지 기법이 보안 대응에 적극 활용되고 있다[4]. 로그 데이터를 분석해 공격을 탐지하고, 그 결과를 정책 엔진에 반영하는 방식은 실시간 대응 능력을 강화하는 효과를 보이나, 이 역시 공격이 발생한 후에 대응이 가능하다는 한계를 갖는다[5]. 공격 발생과 탐지, 정책 반영 사이에는 불가피하게 시간 지연이 존재하며, 시스템은 노출 상태에 놓인다. 본 연구는 이러한 한계를 극복하기 위하여 위협 예측 개념을 도입하여 탐지 중심 보안에서 예방 중심 보안으로의 전환을 제안한다. 시계열 로그 분석과 딥러닝 모델을 통해 미래 공격 가능성을 예측하고, 이를 정책 생성과 연계하여 사전 차단을 실현한다.

Ⅱ. 하이브리드 클라우드 보안 취약점

하이브리드 클라우드 보안의 기존 연구는 주로 탐지 기법에 집중되어 왔다[2-3]. 가장 전통적인 접근은 룰 기반 탐지로, 시그니처를 활용하여 알려진 공격을 빠르게 식별하는 방식이다. 이는 IDS/IPS 시스템에 널리 적용되어 왔으나, 변종 공격이나 제로데이 공격에는 취약하다. 이러한 한계를 보완하기 위해 머신러닝 기반 이상 탐지 연구가 진행되고 있다. 네트워크 트래픽과 로그를 분석하여 정상 및 비정상 패턴을 분류하는 방식은 기존 룰 기반 탐지에 비해 일반화 성능이 우수하다. 그러나 대부분 단일 이벤트 단위로 동작하며, 특히 공격이 발생한 이후의 대응에 머무른다는 한계를 갖는다. 이에 본 연구는 기존에 발생한 공격 유형들을 시계열로 분석하여 추후 발생할 수 있는 공격 패턴을 예측하는 예방 중심 보안 기법을 제안한다.

Ⅲ. 제안 모델

본 연구는 하이브리드 클라우드 환경에서 수집되는 로그를 시계열 단위로 분석하여 공격 발생을 사전에 예측하는 방법을 제안한다. 먼저 로그 데이터를 수집하고 메시지, URI, 도메인, User-Agent 등 다양한 필드를 통합하여 하나의 텍스트 표현으로 변환한다. 각 로그는 시간 순서대로 정렬되며, 고정 크기 윈도우로 분할되어 시퀀스 분석의 기본 단위가 된다. 윈도우별 라벨은 단순 이벤트 기반이 아닌, 해당 구간에서 공격이 발생한 비율을 계산하여 부여한다. 이를 통해 데이터 특성에 따라 동적으로 기준이 조정되며, 불균형한 데이터에도 유연하게 대응할 수 있다. 윈도우는 이후 TF-IDF 임베딩과 차원 축소 과정을 거쳐 잠재 표현으로 전환된다. 제안 모델은 일정 길이의 윈도우 시퀀스를 입력으로 받아 H-step 미래의 공격 여부를 예측하도록 설계되었다. 이를 통해 단순한 탐지가 아닌 예방적 대응이 가능하도록 하였다. 시계열 분석에는 LSTM을 사용하였다.



(그림 1) 시계열 로그 분석을 통한 공격 예측 및 대응 정책 생성

Ⅳ. 실험 설계

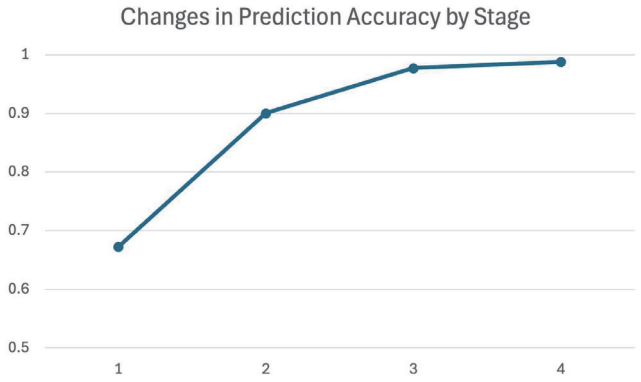
실험에는 실제 하이브리드 클라우드 환경에서 수집된 로그를 활용하였다. 로그는 Elasticsearch를 통해 적재되었으며, 약 89만 건을 확보하였다. 데이터는 메시지, URI, 도메인, User-Agent 등 다양한 속성을 포함하고 있으며, 모든 로그는 시간 순으로 정렬된 뒤, 윈도우 단위 구조로 변환되었다. 로그 텍스트는 TF-IDF 임베딩 후 차원 축소를 진행하였다. 모델 입력은 길이 20의 윈도우 시퀀스로 구성되었으며, 예측 대상은 3-Step 미래의 공격 여부로 설정하였다.

성능 평가는 정확도, 정밀도, 재현율, F1-Score를 기준으로 수행하였으며, 정상 클래스(0) 및 공격 클래스(1) 지표를 중심으로 결과를 분석하였다. 이를 통해 제안 모델이 두 클래스에서 모두 균형 잡힌 성능을 달성하는지를 검증하였다.

Ⅴ. 실험 결과

	precision	recall	f1-score	support
0	0.9835	0.9613	0.9273	310
1	0.9273	0.9684	**0.9474**	158
macro avg	0.9554	0.9648	0.9598	468
weighted avg	0.9645	0.9637	**0.9639**	468

(표 1) 공격 예측 모델 실험 결과



(그림 2) 단계 별 공격 예측 정확도 변화 추이

본 연구에서 제안한 시계열 예측 모델은 초기 학습 단계에서는 정확도가 67% 수준에 머물렀으나, 반복적인 학습을 통해 모델은 로그 패턴을 빠르게 습득하였다. 특히 2회차 실험에서는 정확도가 90%를 상회하고, 3회차에는 검증 정확도가 97%, 4회차 98%로 상승하였다. 이는 제안 모델이 로그 시퀀스 내 시간적 상관성을 효과적으로 학습할 수 있음을 보여준다.

최종 테스트셋을 활용한 클래스별 성능 분석에서도 유사한 결과가 확인되었다. 정상 클래스(0)의 경우, 정밀도 98.35%, 재현율 96.13%, F1-Score 97.23%를 기록하였으며, 공격 클래스(1)의 경우 정밀도 92.73%, 재현율 96.84%, F1-Score 94.74%를 기록하였다. 정상 클래스에서는 높은 정밀도를 통해 정상적인 트래픽이 불필요하게 차단되는 상황을 최소화할 수 있었으며, 공격 클래스에서는 높은 재현율을 통해 실제 위협이 효과적으로 탐지되었다. 이와 같은 결과는 보안 체계에서 중요한 위협 누락의 최소화와 불필요한 차단 방지를 동시에 충족했음을 의미한다.

Ⅵ. 결론

본 연구는 하이브리드 클라우드 환경에서 발생하는 대규모 로그 데이터를 시계열 단위로 분석하여 미래 공격 발생 가능성을 예측하는 모델을 제안하였다. 기존의 보안 기법은 주로 발생한 공격을 식별하는 탐지 중심의 구조에 머물러 있었으나, 본 연구는 이를 넘어 예방적 보안의 가능성을 제시하였다. 실험 결과, 제안 모델은 96% 이상의 정확도와 0.94 이상의 F1-Score를 달성하며 높은 수준의 성능을 입증하였다. 특히, 정상 및 공격 클래스 모두에서 안정적인 성능을 보여, 탐지와 예방이라는 두 가지 요구를 동시에 충족하였다.

본 연구의 기여점은 다음과 같다. 첫째, 하이브리드 클라우드 환경에서 로그를 단순 이벤트 단위가 아니라 시계열 단위로 분석하여 예방적 보안으로 확장하였다. 이를 통해 공격 가능성을 사전에 파악하여 보안 정책의 선제적 배포로 이어질 수 있는 기반을 마련하였다. 둘째, 동적 임계치 설정을 통해 불균형한 데이터 분포에 유연하게 대응함으로써 실험 결과에서 정밀도와 재현율을 동시에 높이는 성과를 보였다. 향후 연구에서는 다양한 공격 유형별 예측 모델 확장, 멀티스텝 예측 지평 확대, 그리고 실시간 정책 엔진과의 통합 실험을 통해 실제 클라우드 운영 환경에 적용 가능한 수준으로 발전시킬 계획이다. 이를 통해 하이브리드 클라우드 보안 체계를 탐지 중심에서 예방 중심으로 전환하는 구체적 실행 방안을 제시하고자 한다.

ACKNOWLEDGMENT

본 연구는 한국인터넷진흥원(KISA)-정보보안 특성화대학 지원사업의 지원을 받아 수행된 연구임(50%). 본 연구는 2025년도 과학기술정보통신부 및 정보통신기획평가원의 소프트웨어중심대학사업의 연구결과로 수행되었습니다.(2021-0-01409)(50%).

참 고 문 헌

[1] A. Leff and J. T. Rayfield, "Integrator: An Architecture for an Integrated Cloud/On-Premise Data-Service," 2015 IEEE International Conference on Web Services, New York, NY, USA, 2015, pp. 98-104

[2] G. Raktate, K. Shelar, P. Parjane, S. Pangavhane, S. More and S. R. Deshmukh, "A Survey on Security Issues and Challenges in Cloud Computing," 2024 International Conference on Decision Aid Sciences and Applications (DASA), Manama, Bahrain, 2024, pp. 1-5

[3] S. B. Mallisetty, G. A. Tripuramallu, K. Kamada, P. Devineni, S. Kavitha and A. V. P. Krishna, "A Review on Cloud Security and Its Challenges," 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2023, pp. 798-804,

[4] S. J, K. Kanagasabapathi, K. Mahajan, S. Ahamad, E. Soumya and S. Barthwal, "AI-Enhanced Multi-Cloud Security Management: Ensuring Robust Cybersecurity in Hybrid Cloud Environments," 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2023, pp. 1-6

[5] G. Kim, Y. Kim, E. Lee, H. Jang and K. Kim, "Edge-Based Policy Caching for Low Latency Security Enforcement in Hybrid Clouds," 2025 25th Asia-Pacific Network Operations and Management Symposium (APNOMS), Kaohsiung, Taiwan, 2025, pp. 1-6