AI 복합 학습을 통한 네트워크 침입 탐지 시스템 연구

황득빈*, 이용준
극동대학교
binhhoangex2000@gmail.com

Network Intrusion Detection System via Hybrid AI Learning

Hoang Duc Binh, Lee Yong Joon*
Far East University

요 약

최근 인공지능(AI)의 발전과 함께 기존 네트워크 침입 탐지 시스템(NIDS)의 한계를 극복하려는 연구가 활발하다. 서명 기반 또는 단일 딥러닝 모델은 제로데이·신종 공격에 취약하고, 네트워크 플로우 간 구조적 관계를 충분히 반영하지 못한다. 본 연구는 SAGEConv-GNN(국지 구조 학습)과 Transformer Encoder(전역 시퀀스 문맥)를 Gated Fusion으로 결합한 하이브리드 NIDS를 제안한다. 그래프는 UNSW-NB15 플로우를 노드로 하여 동일 SrcIP 그룹 내 KNN(k=6)과 temporal chain으로 에지를 구성하고, 학습형/사인·코사인/차수 기반 다중 위치 인코딩을 적용하였다. 학습 안정성을 위해 early stopping, label smoothing, class weight 등을 사용하였다. UNSW-NB15(≈50,000 flows) 실험에서 제안 모델은 GCN, GAT, GraphSAGE, Transformer 단일 모델을 상회하여 정확도 99.70%, 정밀도 0.9644, 재현율 0.9908, F1-score 0.9772를 달성하였다. 이는 불균형 데이터 환경에서도 높은 탐지 성능과 낮은 오탐지율을 유지함을 보여 주며, AI 기반 하이브리드 NIDS의 실용적 적용 가능성을 제시한다.

Ⅰ. 서 론

IoT, 5G, 클라우드, ICS의 확산으로 네트워크 공격 표면이 확대되면서 NIDS의 고도화 요구가 커지고 있다. 서명 기반은 알려진 공격에 강점이 있으나 제로데이에 취약하고, 이상 탐지는 미지 공격에 유리하나 오탐률이 높다는 근본 한계가 있다. 이에 본 연구는 Transformer Encoder의 전역 시퀀스 학습 능력[1]과 SAGEConv-GNN의 관계적 구조 학습 능력[2]을 결합한 하이브리드 NIDS를 제안한다. 데이터는 UNSW-NB15를 사용하여 현대적 공격 시나리오를 반영하고, 단일 GCN, GAT, GraphSAGE, Transformer와 공정 비교를 수행한다[3]. 목표는 정확도·재현율·F1 향상을 실험적으로 입증하고, 최근 하이브리드 추세와의 정합성을 확인하는 것이다[5][6].

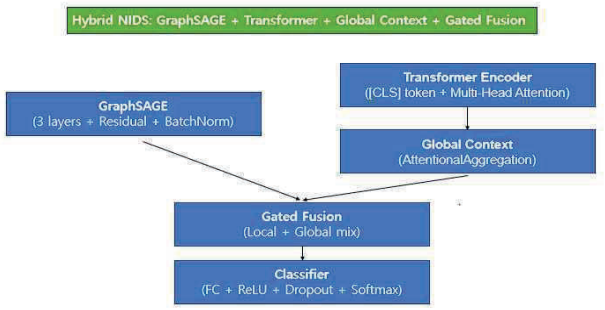
Ⅱ. 관련 이론 및 선행연구

Transformer는 Self-Attention으로 전 구간의 전역 의존성과 장기 종속성을 병렬적으로 학습하여 네트워크 트래픽의 시간적 문맥을 포착하는 데 적합하다[1][4]. 반면 GraphSAGE는 이웃 샘플링과 메시지 패싱으로 1-K-hop의 국지 구조를 효율적으로 학습하며 대규모 그래프에 확장성이 높다[2]. 최근 연구는 구조(Local)와 시퀀스(Global)를 함께 모델링하는 하이브리드 방향으로 수렴하고 있으며, 여러 벤치마크에서 97-99%대의 높은 성능이 보고된다. 다만 아키텍처 복잡도, 구식 데이터셋 의존, 클래스 불균형, 실시간 적용성/확장성은 여전히 해결 과제로 남아 있다[5]. 이에 본 연구는 표 2의 한계—아키텍처 복잡도, 구식 데이터셋 의존, 클래스 불균형, 실시간 적용성/확장성—를 고려하여, Gated Fusion으로 두 표현이에 본 연구는 표 1의 한계—아키텍처 복잡도, 구식 데이터셋 의존, 클래스 불균형, 실시간 적용성/확장성—를 고려하여, Gated Fusion으로 두 표현을 통합하고, 데이터 누수 억제를 위한 그래프 구성·분할 및 동일 조건 비교를 채택한다.

<표 1> 하이브리드 NIDS 연구 동향 요약표

논문 / 저자 / 연도	모델 구조	데이터셋	성능 지표	주요 장점	한계점 / 향후 연구
Farhan et al., 2025	DNN + Extra Trees	UNSW-NB15	ACC 97.9%	특징 차원 축소: 해석 가능성	이진 분류 한정: 앙상블 + 실시간 연구 제안
Sun et al., 2024	GNN (GCN, GAT)	합성 데이터, CIC-IDS2017	정밀도/재현율 높음	그래프 해석 가능성: 실시간 적용 가능	합성 데이터셋 한정: 시계열 모델링 없음
Alashjaee, 2025	CNN + LSTM + Attention	NSL-KDD, BoT-IoT	ACC 95-97%	공간적 + 시간적 + Attention 결합: 낮은 지연(latency)	구식 데이터셋: 확장성 문제
Dash et al., 2025	LSTM + 메타휴리스틱	NSL-KDD, CIC-IDS2017, BoT-IoT	ACC 97-99%(NSL/CIC): ~81-86%(BoT-IoT)	최적화를 통한 정확도 향상	BoT-IoT 성능 약함: 향후 IoT 하이브리드 필요

Ⅲ. 방법 및 모델 개요



(그림 1) 제안한 하이브리드 NIDS 아키텍처

본 연구의 입력 그래프는 UNSW-NB15에서 추출한 플로우를 노드로 하여, 동일 Source IP 그룹 내부에서 KNN(k=6)과 temporal chain을 결합해 에지를 정의하였다. 데이터 누수 방지를 위해 srcip 단위의 group-aware split을 적용하였고, 모델 입력 표현을 강화하기 위해 학습형, 사인/코사인, 노드 차수 기반의 다중 위치 인코딩을 결합하였다. 모델은 잔차 연결과 배치 정규화를 포함한 3층 GraphSAGE로 국지 구조 임베딩을 얻은 뒤, Transformer Encoder를 통해 [CLS] 토큰 중심의 전역 문맥을 학습한다. 두 경로의 출력은 Gated Fusion으로 동적으로 결합되어 구조적·시계열적 신호의 중요도를 조정하며, 최종적으로 MLP-Softmax를 통해 이진 분류를 수행한다. 학습은 AdamW와 OneCycleLR을 사용하고, class weight, label smoothing, early stopping, AMP를 적용하여 안정성과 일반화 성능을 높였다.

IV. 실험 및 결과

실험은 약 50,000개의 플로우를 사용해 동일 전처리·하이퍼파라미터 조건에서 단일 GCN, GAT, GraphSAGE, Transformer와 공정 비교로 진행되었다. 제안 모델은 정확도 99.70%, 정밀도 0.9644, 재현율 0.9908, F1 0.9772, AUC ≈ 0.99를 기록하여 모든 기준에서 단일 모델을 상회하였다. 혼동 행렬 요약(TN 18,355, FP 47, FN 10, TP 618)은 오탐과 미탐이 모두 낮음을 보여 주며, 특히 공격 클래스에서 높은 재현율을 유지하면서도 정밀도를 크게 희생하지 않았다. 이는 구조(Local)와 전역(Global) 문맥을 동시에 반영한 융합이 희귀 공격 탐지의 강건성 향상에 기여함을 시사한다.

<표 2> 기존 모델과 HybridSAGETransformerGlobal의 성능 비교

모델명	정확도	정밀도	재현율	F1-score
GCN	0.4601	0.5287	0.7200	0.3607
GAT	0.5620	0.5344	0.7689	0.4186
GraphSAGE	0.9934	0.9164	0.9966	0.9527
TransformerEncoderOnly	0.9932	0.9152	0.9957	0.9516
HybridSAGETransformerGlobal	0.9970	0.9644	0.9908	0.9772

V. 결론

본 연구는 SAGEConv-GNN의 구조 학습과 Transformer Encoder의 전역 시퀀스 학습을 Gated Fusion으로 결합한 하이브리드 NIDS를 제안하였고, UNSW-NB15 기반 실험에서 단일 모델 대비 재현율과 F1의 우위를 입증하였다. 제안 방식은 불균형 데이터에서도 낮은 오탐·미탐을 동시에 달성하여 실사용 환경에서의 신뢰성을 높인다. 향후 연구는 CIC-IoT2023과 ToN-IoT 등 최신 데이터셋 검증, 경량화·양자화에 기반한 실시간 배포, 그리고 설명 가능한 AI(XAI) 모듈의 통합에 중점을 둘 것이다.

참 고 문 헌

[1] A. Vaswani, N. Shazeer, N. Parmar, et al., “Attention Is All You Need,” Advances in Neural Information Processing Systems (NeurIPS), pp. 5998 - 6008, 2017.

[2] W. L. Hamilton, R. Ying, and J. Leskovec, “Inductive Representation Learning on Large Graphs,” NeurIPS, pp. 1024 - 1034, 2017.

[3] N. Moustafa and J. Slay, “UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems,” in Proc. MilCIS, IEEE, pp. 1 - 6, 2015.

[4] T. Lin, Y. Wang, X. Liu, and X. Qiu, “A Survey of Transformers,” arXiv preprint arXiv:2106.04554, 2021.

[5] L. Sun, et al., “GNN-IDS: Graph Neural Network based Intrusion Detection System,” Computer Communications, vol. 213, pp. 1 - 13, 2024.

[6] S. Ullah, et al., “TNN-IDS: Transformer Neural Network-based Intrusion Detection for MQTT-enabled IoT,” IoT Security Journal, vol. 8, no. 1, pp. 33 - 45, 2023.