네트워크 침입 탐지 시스템을 위한 인-네트워크 지능 연구 동향

이정원, 한수빈, 방민지, 박고은, 백상헌
고려대학교

{jw_2, subin993, qkdalswl2, gopark, shpack}@korea.ac.kr

A Study on In-Network Intelligence for Network Intrusion Detection Systems

Jeongwon Lee, Subin Han, Minji Bang, Goeun Park, Sangheon Pack
Korea Univ.

요 약

네트워크 침입 탐지 시스템은 혼합된 사용자 트래픽 환경에서 공격 트래픽을 식별하는 핵심 보안 기술로, 최근 인공지능 모델을 활용해 복잡한 패턴을 효과적으로 학습하며 탐지 성능이 향상되었다. 그러나 모델의 높은 계산 복잡성으로 인해 트래픽을 제어 평면으로 복사하여 처리하는 방식이 사용되는데, 그 결과 지연과 대역폭 소모 문제가 발생한다. 이를 해결하기 위한 방안으로 데이터 평면에서 직접 추론을 수행하는 인-네트워크 지능을 적용한 방식이 활발히 연구되고 있다. 이에 본 논문에서는 데이터 평면에서 실시간으로 침입을 탐지하면서 모델을 점진적으로 업데이트하는 인-네트워크 지능 기반 연구를 소개한다.

Ⅰ. 서 론

네트워크 침입 탐지 시스템 (Network Intrusion Detection System, NIDS)은 이상 징후를 탐지하는 데 중요한 네트워크 보안 인프라로 필수적인 역할을 수행해왔다. 최근에는 머신러닝 및 딥러닝을 활용하여 복잡한 트래픽 패턴에 대한 탐지 정확도를 크게 향상시켰다. 그러나 모델의 높은 계산 복잡성으로 인해 실제 패킷이 이동하는 데이터 평면에서 직접 처리하지 못하고, 트래픽을 복사해서 제어 평면으로 보내어 처리하는 방식을 채택하고 있다. 하지만 이 과정에서 높은 지연시간과 네트워크 대역폭 소모 문제가 발생하며 결과적으로 실시간 탐지를 저해한다.

최근 Programming Protocol-Independent Packet Processors (P4) 와 같은 스위치 프로그래밍 언어로 동작을 정의할 수 있는 프로그래머블 스위치의 등장으로, 트래픽을 제어 평면으로 보내지 않고 데이터 평면에서 일부 네트워크 기능을 수행할 수 있게 되었다 [1]. 이로 인해 데이터 평면 내에서 직접 추론을 수행하는 인-네트워크 지능 (In-Network Intelligence)이 가능해졌고, 특히 인공지능 모델을 인-네트워크 지능에 접목시키는 연구들이 NIDS 분야에서 다수 등장하고 있다. 인-네트워크 지능을 구현하기 위해서는 모델의 파라미터나 결정 경계 (decision boundary)를 P4 데이터 평면에서 사용 가능한 매치-액션 테이블 형태로 변환하여 매핑하는 방식이 일반적으로 활용된다.

하지만 프로그래머블 스위치는 가용 메모리 자원이 극히 제한적이며 단순한 연산만 지원하기 때문에 복잡한 인공지능 모델을 데이터 평면에 구현하기 어렵다. 이를 극복하기 위해 지식 증류 (knowledge distillation)를 통한 모델 경량화 [2]나, 정규 표현식과 같은 스위치 친화적인 형태로의 근사 기법 [3]이 제안되었다. 그럼에도 불구하고 대부분의 기존 인-네트워크 지능 기반 방법은 초기 학습 이후의 지속적인 모델 업데이트를 고려하지 않으며, 모델의 미세한 변경에도 전체 하드웨어 테이블을 재구성해야 하는 운영상의 한계가 존재한다. 본 논문에서는 이러한 하드웨어적 제약을 극복하고, 필요한 부분만 점진적으로 업데이트하는 인-네트워크 NIDS 의 최신 연구 동향을 소개한다.

Ⅱ. 모델 업데이트 기반 인-네트워크 NIDS 동향

인-네트워크 NIDS 는 학습 방식에 따라 지도 학습 방식과 비지도 학습 방식으로 나눌 수 있다. 지도 학습 방식은 여러 공격과 정상 패턴에 대한 라벨링된 데이터를 학습하여 모델을 구축한다. 이 방식은 라벨 정보를 직접 활용하므로 다중 클래스 분류에 강점을 가지지만 대량의 라벨링된 데이터 확보가 필수적이라는 단점이 있다. 반면 비지도 학습 방식은 라벨 없이 데이터 자체의 특징을 분석하여 정상 패턴과 다른 이상 징후를 찾아내는 방식으로 동작한다. 본 장에서는 각 방식에서 점진적 모델 업데이트를 구현한 대표적인 연구인 Helios [4], Genos [5]를 소개한다.

2.1 지도 학습 방식

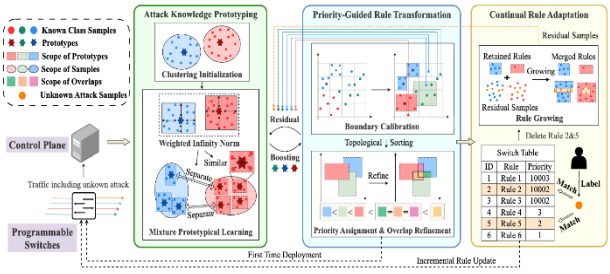

그림 1. Helios 의 워크플로우

Helios 는 지도 학습 기반의 인-네트워크 공격 탐지 기법으로, 새로운 공격 패턴이 점진적으로 추가되는

시나리오에서 모델의 성능을 유지하며 지속적으로 학습하는 기법을 제안한다. 해당 기법은 특징 공간 상에서 각 클래스를 대표하는 프로토타입을 여러 개 설정한 다음 유입되는 트래픽의 특징 값이 공간상에서 어떤 프로토타입과 가장 가까운지를 측정함으로써 해당 트래픽의 클래스를 판별하는 다중 클래스 분류를 수행한다. 하지만 프로그래머블 스위치는 소수점 연산이 불가능하므로 복잡한 거리 계산식을 처리할 수 없기 때문에 Helios 는 유클리드 거리 대신 정수 비교만으로 처리 가능한 weighted infinity norm 기반의 거리 계산 방식을 활용한다. 또한, 프로토타입의 범위를 결정하는 결정 경계가 중첩되는 구간에 대해서는 우선순위 기반으로 규칙을 변환한 다음 매치-액션 테이블에 배포함으로써 오분류 가능성을 낮추고 고속 처리를 보장한다. 만약 배포된 규칙들에 해당하지 않는 트래픽이 유입되면, 이를 잠재적 공격으로 간주하여 제어 평면으로 보고하고 분석을 요청한다. 새로운 공격에 대한 라벨링이 완료되면, Helios 는 전체 모델을 재학습하는 대신 신규 클래스에 대한 프로토타입만을 생성하여 기존 규칙 세트에 통합하는 방식을 사용해 하드웨어를 재구성하는 비용을 절감하고 효율적인 점진적 학습을 수행한다. 하지만 Helios 는 새로운 공격 패턴을 학습시키기 위해서는 반드시 데이터를 직접 분류하고 라벨링하는 과정이 선행되어야 하며, 충분한 양의 데이터셋이 확보될 때까지는 새로운 공격에 대한 분류가 지연될 수 있다는 한계점이 존재한다.

2.2 비지도 학습 방식

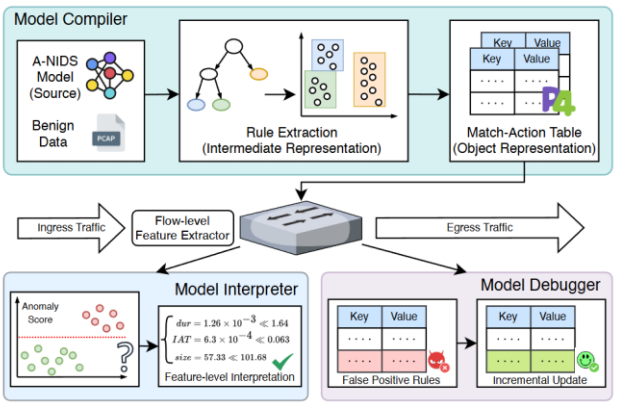

그림 2. Genos 의 프레임워크

Genos 는 비지도 학습 기반 접근 방식을 활용한 연구로, 특정 NIDS 모델에 구애받지 않고 인-네트워크 환경에 배포할 수 있는 범용 프레임워크를 제안한다. Genos 는 먼저 정상 트래픽의 데이터 분포가 여러 그룹으로 나뉘는 특성을 고려하여, 모델이 산출한 이상 점수를 기반으로 특징 공간을 다수의 부분 공간으로 분할한다. 이후 각 부분 공간에서 정상 트래픽이 밀집된 영역을 초기 정상 영역으로 정의한다. 이후 이상 점수의 기울기를 추적하여 정상과 공격을 구분하는 경계면을 향해 탐색 지점을 반복적으로 이동시켜 복잡한 결정 경계를 정밀하게 추정한다. 추정된 경계는 복잡한 형태인 경우가 많으므로 Genos 는 이를 P4 스위치에서 처리 가능한 axis-aligned rules 형태로 근사하여 최종 추출한 다음 스위치에 배포한다. 오탐 (false positive)이 발생하는 경우 전체 모델을 재학습하는 대신 영향을 받는 부분 공간의 규칙만 미세 조정하여 하드웨어 재구성 비용을 절감한다. 하지만 Genos 는 정상과 공격을 구분하는 이진 분류만을 수행하며 세부적인 공격 유형으로 세분화하는 다중 클래스 분류 기능은 결여되어 있다는 한계점이 있다.

Ⅲ. 결 론

본 논문에서는 진화하는 공격 패턴에 대한 실시간 탐지의 중요성이 증가함에 따라, 인-네트워크 지능을 활용해 이상 징후를 탐지하는 연구 동향을 소개하였다. 구체적으로는 서로 다른 학습 방식을 활용하면서도 공통적으로 점진적 업데이트를 하는 연구들의 특징과 한계점을 살펴보았다. 향후에는 두 방식의 장점을 통합하여, 라벨이 없는 환경에서 미지의 공격을 탐지함과 동시에 발견된 공격을 다중 클래스로 분류하는 연구를 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학 ICT 연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2025-RS-2022-00156353)

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2024-00341965)

참 고 문 헌

[1] P. Bosshart *et al.*, "P4: programming protocol-independent packet processors," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 3, pp. 87-95, July 2014.

[2] G. Xie, Q. Li, Y. Dong, G. Duan, Y. Jiang, and J. Duan, "Mousika: Enable general in-network intelligence in programmable switches by knowledge distillation," in *Proc. IEEE Conference on Computer Communications (INFOCOM) 2022*, Virtual Conference, May 2022.

[3] Z. Zhang, Y. Huang, G. Duan, Q. Li, D. Zhao, Y. Jiang, L. Ma, X. Xiao, and H. Xu, "Metis: understanding and enhancing in-network regular expressions," in *Proc. Advances in Neural Information Processing Systems (NeurIPS) 2023,* New Orleans, Louisiana, USA, Dec. 2023.

[4] Z. Shi, D. Zhao, Y. Zhu, G. Xie, Q. Li and Y. Jiang, "Helios: Learning and Adaptation of Matching Rules for Continual In-Network Malicious Traffic Detection," in *Proc. ACM The Web Conference (WWW) 2025*, Sydney, NSW, Australia, April 2025.

[5] R. Li, Q. Li, Y. Zhang, D. Zhao, X. Xiao and Y. Jiang, "Genos: General In-Network Unsupervised Intrusion Detection by Rule Extraction," in *Proc. IEEE Conference on Computer Communications (INFOCOM) 2024*, Vancouver, BC, Canada, August 2024.