특징 편차 기반 이상 트래픽의 행동 단위 해석 방법

박서현[1], 서유정[2], 이현우[3]
[1]한국에너지공과대학교 에너지공학과 학부생
[2]한국에너지공과대학교 에너지AI트랙 석사과정
[3]한국에너지공과대학교 에너지AI트랙 교수
parksh5295@kentech.ac.kr, yujeongseo@kentech.ac.kr, hwlee@kentech.ac.kr

Behavior-Level Interpretation of Anomalous Network Traffic Based on Feature Deviation

Seohyeon Park[1], Yujeong Seo[2], Hyunwoo Lee[3]
[1]Undergraduate, Dept. of Energy Engineering, Korea Institute of Energy Technology
[2]Graduate, Dept. of Energy Engineering, Korea Institute of Energy Technology
[3]Professor, Dept. of Energy Engineering, Korea Institute of Energy Technology

요 약

최근 네트워크 이상 탐지 기술은 높은 성능을 보이고 있으나, 탐지 결과가 주로 개별 특징 수준의 차이로 제시되어 이를 직관적인 행동 의미로 해석하기 어렵다. 본 연구는 이상 트래픽의 특징 수준 편차를 행동 단위로 집계하여 해석하는 방법을 제안한다. 제안 방법은 정상 트래픽으로부터 특징별 기준 벡터를 구성하고, 각 샘플의 특징 편차를 계산한 뒤, 의미적으로 관련된 특징들을 행동 그룹으로 묶어 행동 수준의 편차 점수로 변환한다. 또한 정상 트래픽의 행동 점수 분포를 기준으로 이상 정도를 평가함으로써 추가 라벨 없이도 행동 중심 설명이 가능하도록 설계하였다. NSL-KDD와 NetML 데이터셋을 이용한 실험 결과, 특징 수준 엔트로피 중앙값은 각각 0.96 및 2.66이었으나, 행동 수준 엔트로피는 최대 0.29 및 0.95로 낮아져 이상 설명이 소수의 의미적 행동 축에 집중되는 경향을 확인하였다. 또한 모든 그룹 구성에서 엔트로피 변화량이 양수로 나타나, 행동 단위 설명이 특징 수준 설명보다 더 집중적인 구조를 가진다는 경향이 일관되게 유지되었다. 이러한 결과는 제안 방법이 서로 다른 특징 구성을 갖는 데이터셋에서도 이상 탐지 결과를 더 이해하기 쉬운 행동 중심 설명으로 변환하는 실용적 사후 해석 방법으로 활용될 수 있음을 시사한다.

1. 서론

최근의 네트워크 환경에서는 방대한 양의 트래픽 데이터가 지속적으로 생성되고 있어 자동화된 이상 탐지가 보안 모니터링 시스템의 필수 요소로 자리잡았다. 이상 탐지를 위해 통계 기반, 규칙 기반, 데이터 기반 분석 방법 등 다양한 접근이 제안되어 왔으며, 실제 환경에서도 이상 여부를 자동으로 판별하는 기술은 상당 수준 발전하였다. 그러나 실제 보안 분석 과정에서는 단순히 이상 여부를 탐지하는 것뿐만 아니라, 탐지된 트래픽이 어떤 특성에서 정상과 다르게 나타나는지를 이해하는 과정이 요구된다. 기존의 이상 탐지 결과의 설명 가능한 해석은 주로 개별 특징 수준의 차이를 나열하는 방식으로 제공되지만, 이러한 정보는 수십 개 이상의 특징이 동시에 존재하는 네트워크 데이터에서 직관적인 해석으로 이어지기 어렵다는 한계를 가진다.

본 연구는 이러한 문제를 해결하기 위해 특징 편차를 행동 단위 수준에서 해석하는 방법을 제안한다. 제안 방법은 정상 트래픽 분포를 기준으로 각 특징의 편차를 계산한 뒤, 의미적으로 관련된 특징들을 행동 그룹으로 구성하여 행동 수준의 편차 점수로 집계한다. 이를 통해 이상 트래픽을 개별 수치의 이상 여부가 아닌, 정상 대비 어떤 행동 특성이 변화하였는지의 관점에서 설명할 수 있도록 한다. 또한 정상 분포 기반의 분리 점수를 활용함으로써 공격, 세부 유형과 같은 추가 라벨 없이도 해석이 가능하도록 설계하였다.

NetML 및 NSL-KDD 데이터셋을 이용한 실험을 통해 제안 방법은 특징 편차 정보를 행동 단위의 요약된 설명으로 변환할 수 있음을 확인하였다. 이는 서로 다른 특징 구성을 갖는 데이터셋에서도 동일한 해석 절차를 적용하여 이상 트래픽의 행동적 특성을 일관된 방식으로 기술할 수 있음을 보여준다. 또한 행동 그룹의 구성은 분석 목적에 따라 명시적으로 정의되어야 하는 설계 요소임을 확인함으로써, 이상 탐지 결과 해

석 과정에서 행동 단위 설계의 중요성을 시사한다.

2. 이상 탐지 설명 가능성 관련 연구 동향

네트워크 트래픽 이상 탐지는 다양한 접근을 통해 지속적으로 발전해 왔으며, 이상 여부를 자동으로 식별하는 기술은 높은 수준의 성능을 달성하고 있다. 이에 따라 단순 탐지를 넘어, 탐지 결과를 해석하고 분석에 활용하려는 요구 또한 증가하고 있다[1].

최근 설명 가능한 이상 탐지 및 네트워크 분석 분야에서는 특징 중요도나 입력 기여도 기반의 설명 방법이 주로 활용되며, 개별 특징이 탐지 결과에 미치는 영향을 정량적으로 제시하는 접근이 널리 연구되었다. 그러나 이러한 방법들은 대부분 개별 특징 수준의 설명에 머무르며[2], 여러 특징 간의 의미적 관계를 통합하여 상위 수준의 행동 단위로 해석하는 접근은 상대적으로 제한적으로 이루어져 왔다. 특히 이상 트래픽이 정상과 어떤 행동적 특성에서 구별되는지를 구조적으로 설명하는 연구는 아직 충분히 탐색되지 않은 영역으로 남아 있다[3].

3. 연구 방법

본 연구는 이상 트래픽의 특징 수준 편차를 행동 단위로 해석하기 위해, 정상 분포 기반의 편차 계산과 행동 그룹 집계를 결합한 방법을 제안한다. 전체 절차는 (1) 정상 기준 구축, (2) 특징 편차 계산, (3) 행동 단위 집계, (4) 정상 분포 기반 평가의 네 단계로 구성된다.

(1) 데이터의 정상 트래픽 집합을 N, 각 트래픽 샘플을 i, 각 특징 j에 대한 정상 평균을 다음과 같이 정의하며,

$$\mu_j = \frac{1}{|N|} \sum_{i \in N} x_{ij}$$

이를 통해 정상 활동을 대표하는 벡터 μ를 구성한다.

(2) 각 트래픽 샘플에 대한 특징의 편차는 정상 평균과의 절대적 차이로 정의한다.

(3) 특징을 상위 수준의 행동 단위로 해석하기 위해, 각 행동 그룹에 대해 특징 집합 G_k를 정의한다. 실제 계산에서는 데이터에 존재하는 수치형 특징만을 사용하여 $\widetilde{G_k} \subseteq G_k$를 구성했다. 샘플 i의 행동 단위 편차 점수는 동일한 행동을 반영하는 특징들의 편차를 평균하여, 개별 특징 수준의 편차를 행동 수준의 편차로 변환한다.

$$D_{ik} = \frac{1}{|\widetilde{G_k}|} \sum_{j \in \widetilde{G_k}} d_{ij}$$

행동 그룹 G_k의 구성은, NSL-KDD와 같이 고정된 특징들을 가지는 경우에는 사전에 정의된 특징 구성과 의미에 따라 행동 그룹을 직접 지정하고, NetML과 같이 특징 이름이 가변적인 데이터셋에서는 열 이름의 키워드 규칙을 기반으로 특징을 자동으로 할당한다. 본 연구에서 기본 그룹 구성(baseline)은 주 실험에 사용한 기본 행동 그룹 체계를 의미하며, NSL-KDD에서는 연결 강도, 데이터 볼륨, 호스트 접근 및 셀, 프로토콜 및 오류율, 시간 패턴 등의 사전 정의 그룹을, NetML에서는 바이트 및 패킷, 프로토콜 및 플래그, 시간 관련 키워드 등 자동생성된 그룹을 사용한다. 또한 그룹 구성 수준에 따른 해석 변화 분석을 위해 여러 행동 그룹을 통합한 통합 구성(coarse) 및 기존 그룹을 세분화하여 상세한 행동 단위로 분리한 세분화 구성(fine) 또한 고려한다.

(4) 정상 트래픽에 대해 계산된 D_{ij}값을 이용하여 각 행동 그룹에 대한 정상 분포의 분위수를 구하고, 이상 트래픽에 대해 계산된 D_{ij}값과 정상 트래픽의 분포를 비교하여 행동 수준의 이상 정도를 구분한다.

4. 실험 결과

4.1. 특징 수준과 행동 수준 엔트로피 비교

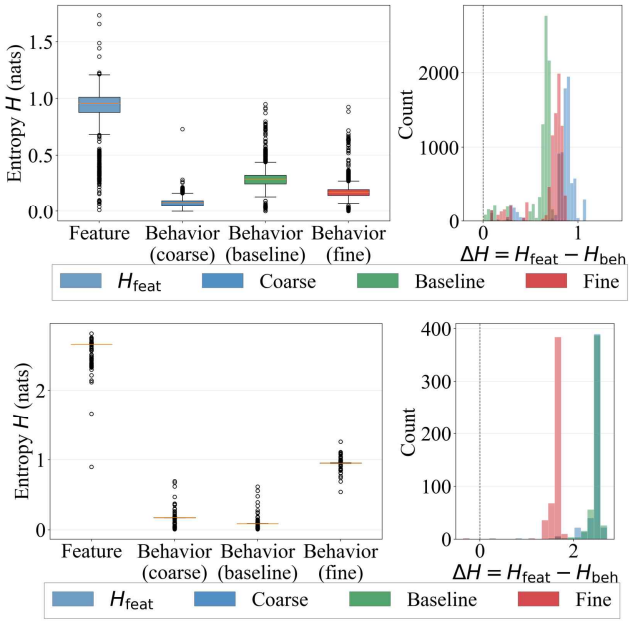

그림 1 각 그룹화 방법별 엔트로피 비교

그림 1은 NSL-KDD와 NetML 데이터셋의 이상 샘플 전체에 대해 계산한 엔트로피 분포를 나타낸다. 두 데이터셋 모두에서 특징 수준 엔트로피는 행동 수준 엔트로피보다 전반적으로 크게 나타났다. 중앙값 기준으로 NSL-KDD의 특징 수준 엔트로피는 약 0.96이었으나, 행동 수준에서는 그룹 구성에 따라 약 0.08–0.29 수준으로 낮아졌다. NetML에서도 특징 수준 엔트로피는 약 2.66인 반면, 행동 수준 엔트로피는 약

0.09–0.95 수준에 머물렀다. 이는 특징 수준에서는 이상 설명이 여러 변수에 분산되는 반면, 행동 수준에서는 더 적은 수의 의미적 축에 집중됨을 의미한다. 즉, 제안 방법은 특징 편차를 행동 단위로 집계함으로써 이상 설명을 보다 집중적이고 요약된 형태로 재구성한다. 이러한 경향이 서로 다른 특징 스키마를 갖는 NSL-KDD와 NetML 모두에서 확인된다는 점은, 제안 방법이 특정 데이터셋에만 의존하지 않는 해석 절차로 활용될 수 있음을 시사한다.

행동 수준 엔트로피의 구체적 분포는 그룹 구성 방식에 따라 달라졌다. NSL-KDD에서는 통합 그룹 구성이 가장 낮은 엔트로피를 보였고, 기본 및 세분화 그룹 구성은 그보다 높은 값을 나타냈다. NetML에서도 세분화 그룹 구성의 엔트로피가 상대적으로 크게 나타나, 행동 축을 세분할수록 설명이 다시 분산될 수 있음을 확인할 수 있다. 그러나 그룹 구성 방식을 달리하더라도 특징 수준 엔트로피가 각 행동 수준 엔트로피보다 높고, $\Delta H = H_{feat} - H_{beh}$가 대부분 양수로 유지된다는 점은 일관되게 관찰되었다. 중앙값 기준 ΔH는 NSL-KDD에서 약 0.66–0.87, NetML에서 약 1.71–2.57로 나타났다. 이는 제안 방법의 효과가 특정 그룹 구성에만 의존하는 것이 아니라, 특징 편차를 행동 단위로 집계할 때 전반적으로 나타나는 구조적 특성임을 보여준다.

4.2. 개별 샘플을 통한 예시

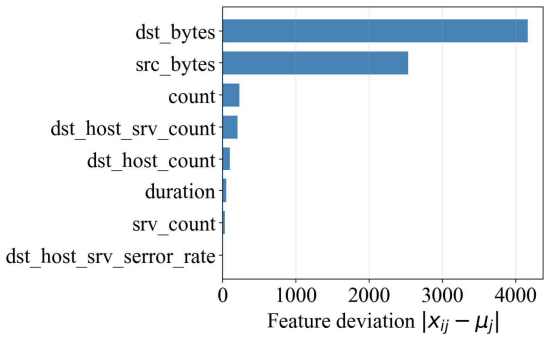

그림 2 NSL-KDD 샘플의 특징 수준 편차

그림 2, 3은 NSL-KDD의 이상 샘플 예시를 각각 특징 수준과 행동 수준에서 나타낸 결과이다. 그림 2에서는 dst_bytes와 src_bytes가 다른 특징들에 비해 압도적으로 큰 편차를 보여, 특징 수준에서는 소수의 볼륨 관련 특징이 이상 설명을 주도하는 것으로 해석된다.

그러나 행동 편차를 집계한 그림 3에서는 해석의 강조점이 달라진다. Protocol, error rates는 가장 강한 행동 이상으로 나타나며, Timing pattern 역시 비교적 큰 차이를 보인다. 반면 특징 수준에서 가장 큰 편차를 보였던 src_bytes와 dst_bytes는 행동 수준

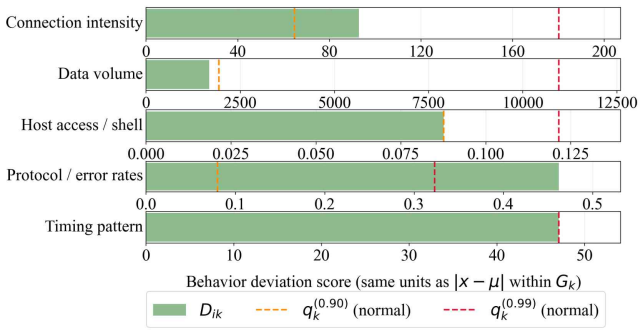

그림 3 NSL-KDD 샘플의 행동 수준 편차

Data volume으로 집계되었을 때 다른 그룹들보다 비교적 약한 행동 이상으로 드러난다. 이는 개별 특징의 절대 편차가 크더라도, 해당 행동이 정상 분포에서 크게 변하는 축이라면 행동 수준에서는 상대적 이상 정도가 높지 않을 수 있음을 보여준다.

5. 연구 결론 및 향후 연구 방향

본 연구는 이상 트래픽의 특징 수준 편차를 행동 단위로 집계하여 해석하는 방법을 제안하였다. 정상 분포 기반의 편차 계산과 행동 그룹 집계를 통해, 개별 특징 나열 대신 이상 트래픽의 행동적 특성을 중심으로 설명할 수 있음을 보였다. NSL-KDD와 NetML 실험 결과, 행동 수준 설명은 특징 수준 설명보다 더 낮은 엔트로피를 보여 이상 설명이 보다 집중적인 구조를 갖는다는 점을 확인하였다. 또한 개별 샘플 예시를 통해 다수의 저수준 특징 편차가 소수의 행동 축으로 요약되어 보다 직관적인 해석이 가능함을 보였다.

향후에는 행동 그룹의 데이터 기반 자동 구성, 범주형 특징 및 시간적 맥락을 포함하는 확장, 실제 보안 분석가를 대상으로 한 사용자 평가가 필요하다. 또한 실무 적용을 위해서는 분석 목적에 맞는 그룹 설계와 함께 그룹 구성 민감도 결과를 함께 제시할 필요가 있다.

이 연구는 2025년도 산업통상자원부 및 한국산업기술기획평가원(KEIT) 연구비 지원에 의한 연구임(과제번호 RS-2025-02653102)

참고문헌

[1] V. Yepmo et al., “Anomaly explanation: A review,” Data & Knowledge Engineering, 137, 101946, 2022.

[2] A. Nascita et al., “A survey on explainable artificial intelligence for internet traffic classification and prediction, and intrusion detection,” IEEE Communications Surveys & Tutorials, 27(5), 3165–3198, 2024.

[3] Z. Li et al., “A survey on explainable anomaly detection,” ACM Transactions on Knowledge Discovery from Data, 18(1), 1–54, 2023.