RaPP 기반 네트워크 이상탐지 시스템

김미르, 계효선, 권민혜
숭실대학교

{48rlaalfm, ksh8301000}@soongsil.ac.kr, minhae@ssu.ac.kr

RaPP-based Network Anomaly Detection Systems

Miru Kim, Hyoseon Kye, Minhae Kwon
Soongsil University

요 약

네트워크 상의 트래픽이 증가함에 따라 네트워크 침입 탐지 시스템(Network Intrusion Detection System; NIDS)의 필요성이 대두되고 있다. 본 논문에서는 오토인코더(Autoencoder; AE)의 은닉층의 정보를 활용하여 기존의 오토인코더 방식보다 더 높은 성능으로 이상을 탐지할 수 있는 RaPP(Reconstruction along Projection Pathway)기반의 이상탐지 시스템을 제안한다. 오토인코더의 은닉층의 정보를 효과적으로 표현할 수 있도록 SAP (Simple Aggregation along Pathway)와 NAP(Normalize Aggregation along Pathway)를 활용하였고 기존의 오토인코더 방식과 성능을 비교하였다. 그 결과 RaPP 기반의 NAP 방식이 기존의 오토인코더 방식보다도 우수한 성능을 보임을 확인하였다.

Ⅰ. 서 론

네트워크 통신 기술의 발달로 네트워크 통신을 이용하는 기기의 종류와 양이 기하급수적으로 증가하면서 네트워크 트래픽 또한 급증하고 있다. 이에 따라 네트워크상에서의 악의적인 침입 시도의 방법과 빈도가 계속해서 늘어나고 있다. 따라서 공격을 사전에 탐지하는 네트워크 침입 탐지 시스템의 필요성이 대두되고 있다.

네트워크 침입 탐지 시스템은 기존의 알려진 공격 데이터의 패턴을 이용해서 침입을 탐지하는 오용탐지와 보편적인 정상 데이터의 패턴을 이용해서 침입을 탐지하는 이상탐지로 나누어진다[1,2,3]. 오용탐지는 사전에 알려진 공격 유형만을 탐지하는 방식으로 새로운 공격 유형에 즉각적인 대응을 하기 어려운 방식이다. 그러나 네트워크 트래픽 상에서 네트워크 공격 유형이 점차 지능화되고 다양화되고 있어 이에 대한 대비가 필수적이다. 이상탐지는 기존의 공격 유형뿐만 아니라 새로운 공격 유형을 탐지할 수 있어 현재의 문제점에 적합한 탐지 방식이기에 이에 대한 연구가 활발히 진행되고 있다.

머신러닝 기반의 대표적인 이상탐지 모델인 오토인코더[2,4]는 인코더(encoder)에서 입력데이터의 차원을 효과적으로 축소하고, 디코더(decoder)에서 축소된 데이터의 차원을 복원한다. 이렇게 복원된 데이터와 입력데이터 사이의 오차를 복원오차(reconstruction error)라고 정의 한다. 정상 데이터의 복원오차는 학습을 통해 작은 값을 가지게 되므로 오토인코더는 설정한 임계점보다 큰 복원오차를 가지는 데이터를 비정상 데이터로 탐지하여 이상탐지를 진행한다.

본 논문에서는 학습된 오토인코더의 성능을 효과적으로 향상시킬 수 있도록 은닉층의 정보를 활용하는 RaPP[5]를 사용함으로써 기존의 네트워크 침입 탐지 시스템의 성능을 향상시키는 새로운 시스템을 제안한다.

Ⅱ. RaPP 기반 이상탐지 시스템

오토인코더는 차원을 축소하는 인코더 부분과 축소된 차원을 다시 복원하는 디코더 부분으로 구성된다. 인코더는 g로 정의하며 l개의 층(layer)으로 구성되어 있다.

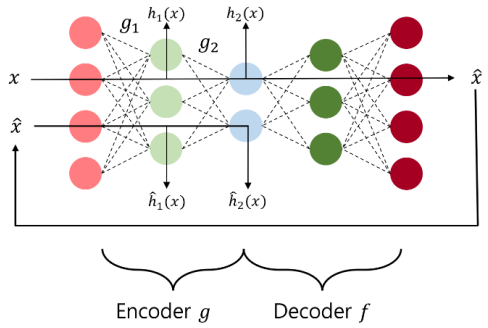

그림 1. RaPP 구조

또한 인코더의 $i\ (1 \le i \le l)$ 번째 은닉층의 축소과정은 g_i, 출력은 $h_i(x)$로 정의하고 디코더는 인코더와 동일하게 l개의 층으로 구성되어 있으며 f로 정의한다. 이렇게 정의된 오토인코더에 대해서 입력 데이터 x의 복원 데이터는 $\hat{x}$으로 표현한다.

$$h_i(x) = g_i \circ \cdots \circ g_1(x) \quad (1)$$

$$\hat{x} = f \circ h_l(x) \quad (2)$$

RaPP 기반 이상탐지 시스템은 복원 데이터 $\hat{x}$을 다시 인코더에 입력시켜 인코더의 i 번째 층의 출력 $\hat{h}_i(x)$ 을 확인한다.

$$\hat{h}_i(x) = h_i(\hat{x}) = h_i \circ f \circ h_l(x) \quad (3)$$

따라서, 입력 데이터 x와 복원 데이터 $\hat{x}$ 의 복원 오차만을 기반으로 이상탐지를 수행하는 기존의 오토인코더 방식과는 달리, RaPP 기반의 방식은 $h_i(x)$와 $\hat{h}_i(x)$의 차이를 비교한다. RaPP 구조는 그림 1 에 설명하였다. RaPP 기반 이상탐지 시스템에서는 은닉층의 복원오차를 확인하기 위해 SAP 와 NAP 적용한다. 입력데이터 x를 $h_0(x)$, 복원 데이터 $\hat{x}$를 $\hat{h}_0(x)$로 표현하면 기존의 오터인코더 방식(AE)과 SAP, NAP 는 (4), (5), (6)와 같은 수식으로 나타낼 수 있다.

표 1. Test set 데이터의 성능 측정 결과

	Pre-cision	Re-call	Fall out	Speci-ficity	Accu-arcy	F1-score	MCC
AE	0.91	0.99	0.09	0.90	0.94	0.95	0.90
SAP	0.83	0.99	0.20	0.79	0.89	0.90	0.82
NAP	0.97	0.99	0.02	0.97	0.98	0.99	0.97

$$s_{AE} = \left\| h_0(x) - \hat{h}_0(x) \right\|_2^2 \quad (4)$$

$$s_{SAP} = \sum_{i=0}^{l} \left\| h_i(x) - \hat{h}_i(x) \right\|_2^2 \quad (5)$$

$$s_{NAP} = \sum_{i=0}^{l} \left\| (h_i(x) - \hat{h}_i(x))^T V \Sigma^{-1} \right\|_2^2 \quad (6)$$

$$M = U \Sigma V^T \quad (7)$$

AE 와 SAP 방식은 복원오차 값에 L2 norm 을 적용하였다. SAP 방식에서 각 은닉층의 복원 오차 분포의 스케일이 상이할 경우, 분포의 스케일을 고려하지 않고 단순히 원점으로부터의 거리를 구하는 것은 모델의 정확도를 떨어뜨리게 되므로 이상탐지의 성능을 저하시킬 수 있다. 이에 SAP 방식의 문제점을 보완하기 위해 NAP 는 특이값 분해(Singular Value Decomposition; SVD)를 사용하여 은닉층의 복원오차 값을 정규화한 후에 L2 norm 을 적용하였다. SVD 는 $m \times n$ 행렬 M을 U, Σ, V의 3 개의 행렬로 분해할 수 있는 방식이며 (7)과 같은 수식으로 나타낼 수 있다. U는 좌특이벡터로 정의되며 V는 우특이벡터로 정의된다. U와 V는 기하학적으로 선형변환의 의미를 가지는 직교행렬(orthogonal matrix) 이고, Σ는 기하학적으로 스케일 변환을 의미하는 대각행렬(diagonal matrix)이다. NAP 는 은닉층의 복원오차 $(h_i(x) - \hat{h}_i(x))^T$를 SVD 하여 얻은 V로 선형변환 시킨 뒤에 Σ로 스케일을 변환하여 은닉층의 복원오차 분포 스케일을 하나의 스케일로 통일한다. 이에 따라 SAP 보다 높은 성능으로 이상탐지를 수행할 수 있다.

Ⅲ. 이상탐지 성능 평가 결과 비교

본 논문에서 정의한 SAP, NAP 의 성능평가를 위하여 시뮬레이션을 통해 AE 와 성능비교를 진행하였다. 네트워크 이상탐지 시스템의 성능을 평가하기 위해 Precision(정밀도), Recall(재현율), Fall-out, Specificity (특이도), Accuracy(정확도), F1-score, MCC(Matthews Correlation Coefficient)의 7 가지 성능지표를 사용하였다. 본 실험에서 사용한 데이터셋은 네트워크 이상데이터를 포함하고 있는 NSL-KDD 데이터셋[6]을 사용했으며, 10 개의 랜덤시드에 대한 성능결과의 평균값으로 성능 값을 구하였다. 훈련과 테스트에 모두 사용되는 정상데이터는 6:2:2 의 비율로 train set, validation set, test set 으로 구분하였다. Test set 은 학습과정에서는 사용하지 않고 성능평가를 위해서만 사용하였으며, test set 에 포함되는 비정상 데이터의 양은 정상 데이터와 동일하도록 설정하였다. 성능평가 결과는 표 1 에 정리하였다.

SAP 를 사용하였을 때는 기존 AE 방식에 비해 저하된 성능을 보였다. 이는 RaPP 의 은닉층의 복원오차가 다양한 스케일을 가지게 되기 때문이다. 반면, 은닉층의 복원오차에 정규화를 수행한 NAP 방식은 모든 성능 지표에서 다른 두 방식 대비 우수한 것을 확인하였다. 특히, F1-score 는 0.99 로 매우 높은 이상탐지 성능을 가짐을 확인하였다. 또한, Specificity 와 MCC 는 AE 대비약 8%의성능향상이 이루어짐을 확인하였다. 그림 2 는 test set 데이터의 NAP 기반 복원 오차 값 분포를 확인한 것이다. 이를 통해 정상과

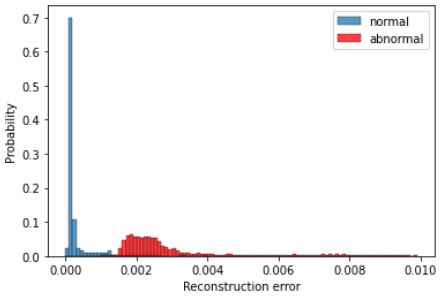

그림 2. NAP 방식의 복원오차 분석

비정상 데이터가 서로 다른 분포를 가지고 명확하게 구분되는 것을 확인할 수 있다. 정상데이터의 경우 성공적인 복원으로 복원오차 값이 0 에 가깝게 분포하였다. 반면, 비정상 데이터의 경우 복원오차 값이 정상데이터에 비해 큰 값인 0.002 를 중심으로 분포한다.

Ⅳ. 결론

본 논문에서는 은닉층의 정보를 활용하여 수행하는 RaPP 기반의 네트워크 이상탐지 시스템을 제안하였다. 은닉층의 정보를 활용하지 않는 기존 오토인코더 방식과, 복원오차의 SAP, NAP 의 성능 검증 결과, 기존 방식 대비 은닉층의 복원오차를 정규화 하지 않은 SAP 방식은 오히려 성능이 저하되는 반면, NAP 방식은 나머지 두 성능에 비해 우수한 성능임을 입증하였다. 따라서 NAP 방식을 적용하여 우수한 성능의 네트워크 이상탐지 시스템 구축을 기대할 수 있다.

ACKNOWLEDGMENT

이 논문은 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업(IITP- 2021-2020-0-01602)과 방송통신산업기술개발사업(IITP-2021-0-00739)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] O. Depren, M. Topallar, E. Anarim, and M. K. Ciliz, “An Intelligent Intrusion Detection System for Anomaly and Misuse Detection in Computer Networks,” Expert Systems with Applications, vol. 29, no. 4, pp. 713-722, Nov. 2005.

[2] 김승주, 계효선, 권민혜, “오토인코더 기반 네트워크 침입탐지 시스템,” 한국통신학회 추계종합학술발표회, Nov. 2020.

[3] 계효선, 권민혜, “주성분 분석기반 저복잡 이상탐지 기술 연구,” 한국통신학회 논문지, vol. 46, no. 6, pp. 941-955, 2021 년 6 월.

[4] C. Zhou, and R.C. Paffenroth, “Anomaly Detection with Robust Deep Autoencoders,” ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2017.

[5] K. H. Kim, S. Shim, Y. Lim, J. Jeon, J. Choi, B. Kim, and A. S. Yoon, "RaPP: Novelty Detection with Reconstruction along Projection Pathway," International Conference on Learning Representations, Sep. 2019.

[6] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” IEEE Symposium on Computational Intelligence for Security and Defense Applications, Jul. 2009.