Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.33, NO.6, Dec. 2023 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2023.33.6.1033

스텍앙상블과 인접 넷플로우를 활용한 침입 탐지 시스템*

성 지 현,[1†] 이 권 용,[1] 이 상 원,[1] 석 민 재,[1] 김 세 린,[1] 조 학 수[2‡]
[1,2]호서대학교 (학생, 교수)

Intrusion Detection System Utilizing Stack Ensemble and Adjacent Netflow*

Ji-Hyun Sung,[1†] Kwon-Yong Lee,[1] Sang-Won Lee,[1] Min-Jae Seok,[1]
Se-Rin Kim,[1] Harksu Cho[2‡]
[1,2]Hoseo University (Undergraduate student, Professor)

요 약

본 논문은 네트워크에서 침입 행위를 하는 플로우를 탐지하는 네트워크 침입 탐지 시스템을 제안한다. 대다수 연구에 활용되는 데이터세트는 시계열 정보를 포함하고 있지 않으며, 공격 사례가 적은 공격은 샘플 데이터 수가 부족해 탐지율 향상이 어렵다. 하지만 탐지 방안에 대해 연구 결과가 부족한 상황이다. 본 연구에서는 ANN(Artificial Neural Network) 모델과 스택 앙상블 기법을 활용한 선행 연구를 토대로 하였다. 앞서 언급한 문제점을 해결하기 위해 인접 플로우를 활용하여 시계열 정보를 추가하고 희소 공격의 샘플을 강화하여 학습하여 탐지율을 보강하였다.

ABSTRACT

This paper proposes a network intrusion detection system that identifies abnormal flows within the network. The majority of datasets commonly used in research lack time-series information, making it challenging to improve detection rates for attacks with fewer instances due to a scarcity of sample data. However, there is insufficient research regarding detection approaches. In this study, we build upon previous research by using the Artificial neural network(ANN) model and a stack ensemble technique in our approach. To address the aforementioned issues, we incorporate temporal information by leveraging adjacent flows and enhance the learning of samples from sparse attacks, thereby improving both the overall detection rate and the detection rate for sparse attacks.

Keywords: NIDS, Network IDS, Stack Ensemble, ANN

I. 서 론

현대 사회에서 인터넷은 일상생활에서 떼어 놓을 수 없는 중요한 요소가 되었다. 디지털 기술의 급속한 발전 속의 누구나 어디서든 인터넷을 사용하면서 네트워크에 문제가 생기면 피해를 보는 개인, 기업 또한 점차 늘어나고 있다. 이에 따른 네트워크 보안의 중요성이 더욱 부각되고 있다. 디지털 기술의 급속한 발전 속에 공격 기술 또한 점차 고도화하고 발전하며 새로운 위협 형태를 만들어 내고 있다. 과학기술정보통신부의 '2023년 상반기 주요 사이버 위협 동향 분석'에 따르면 공격 대상 개인과 기업들의 시스템, 모바일 기기 등을 면밀히 분석하여 취약점을 노리는 고도화된 공격이 증가하고 있다. 또한 침해사고 신고가 21년 640건에서 22년 1,142건으로 전년 대비 약 2배가 증가하였으며, 2023년 상반기 침해 사고 신고 건수는 664건으로 전년 동기 대비 약

Received(10. 20. 2023), Accepted(11. 21. 2023)
* 본 연구는 2023년도 한국데이터산업진흥원의 데이터 청년 캠퍼스 운영기관 사업(2023-01140001) 지원을 받아 수행되었습니다.
* 본 연구는 2023년도 호서대학교 AI+X 연구센터의 지원을 받아 수행되었습니다(2023-97087)
† 주저자, tjdwlgus0204@naver.com
‡ 교신저자, marius71@hoseo.edu(Corresponding author)

40%가 증가하였다 [1]. 빈번하게 통신사의 개인정보 유출, 은행의 디도스(DDoS) 공격에 의한 업무지연 등 일상생활 가까이에서도 위협을 많이 느낄 수 있다. 네트워크 보안은 데이터 보안, 서비스 가용성, 국가 안보 등의 측면에서 매우 중요하며 개인, 기업, 기관은 안전한 네트워크 환경을 유지하며 보안 위협으로부터 보호받을 수 있다. 네트워크 보안 강화와 공격을 탐지·대응하기 위해 NIDS(Network Intrusion Detection System)의 중요성이 커지고 있다. 기존의 NIDS는 주요 탐지 방법에 따라 새로운 공격이나 변형된 공격 탐지에 제한적이거나 이상 행위 탐지를 위한 임곗값 정의, 거짓 양성이 발생하는 한계가 있다. 기존의 NIDS의 한계를 해결하기 위해 여러 방면에서 딥러닝을 활용한 NIDS가 연구되고 있다.

본 연구는 선행 연구에서 참고했던 논문인 Lamia Parven의 논문을 토대로 진행된 연구이다 [2][3].

본 논문의 주요 목표는 참고 논문을 토대로 구현한 선행 연구의 모델의 문제점을 찾고, 보다 높은 성능을 얻고자 한다[2]. 선행 연구에서 구현한 모델은 각기 다른 하이퍼파라미터 설정을 한 4개의 ANN모델을 스택 앙상블 기법을 활용하여 입력받는 최종 모델이다[3]. 데이터 세트는 NF-UQ-NIDS-V2를 사용하였고 테스트 결과 약 97%의 정확도를 보였다. NF-UQ-NIDS-V2 데이터 세트는 NetFlow 기반 특성 집합의 데이터 세트로 네트워크 트래픽 세부 정보를 담고 있다. 하지만 모델 구현 결과 공격 트래픽의 수가 균일하지 않은 분포의 불균형 문제가 있다. 딥러닝 모델 구축에서 데이터 세트는 매우 중요하다. 소수 클래스의 패턴을 더 정확하게 학습하고, 데이터 불균형으로 인한 과적합 문제를 방지해야 한다. 또한 기존의 모델은 특정 Netflow의 한 패킷으로만 공격 여부를 판단하는 경우가 있다. 이는 일반화 문제, 거짓양성(False Positive) 문제 등이 발생할 수 있다. 이들을 해결하고 좋은 성능을 위해 One-hot Encoding, Smote를 이용한 데이터 세트 증폭, 인접 데이터 활용 학습, Binary Classification 등을 활용해 더욱 성능 좋은 모델을 만들고 성능 평가 지표들을 통해 검증한다.

II. 관련 연구

기존의 딥러닝 및 머신러닝 기반 NIDS 연구에 대해 알아보고 모델을 학습 및 평가하기 위한 데이터 세트에 대해서 소개한다.

2.1 기존 NIDS연구 분석

자기 교사 학습(Self-taught learning)기법, 희소 오토인코더(Sparse autoencoder)와 소프트맥스 회귀(Soft-max regression)를 기반으로 한 NIDS를 구현했을 때, 이전에 구현된 NIDS들과 비교해 우수한 성능을 보였다. 성능은 Stacked Autoencoder와 같은 기술을 적용하여 향상시킬 수 있다[4].

심층 신경망(Deep Neural Network:DNN)을 사용하여 모델을 구성하였으며, 공격탐지는 이진분류, 다중 클래스 분류로 수행되었으며 모두 높은 정확도를 보여줬다. 하지만 KDD Cup 99데이터 세트를 사용하였기 때문에 데이터의 수가 작은 U2R 공격의 경우 정밀도, 재현율, F1-score에서 매우 낮은 성능을 보였다[5]. 비슷한 예시로, NDNN(New Deep Neural Network)모델을 통해 정확도를 더 향상시키고 잘 알려지지 않은 침입 데이터를 식별·분류한다. 침입 공격을 분류할 때 SVM보다 성능이 우수했으며 다른 머신러닝 알고리즘보다 좋은 성능을 보이지만 KDD99와 NSL-KDD 데이터 세트를 사용했기 때문에 데이터의 증폭의 필요성을 말하고 있다.[6]

인공신경망(Artificial Neural Network ;ANN) 기반의 NIDS는 다양한 유형의 DOS 공격을 빠른 시간 내에 감지하는 데에 사용될 수 있다고 설명한다. 제안된 NIDS는 낮은 오류율, 높은 학습률 및 빠른 응답률을 보였다. 또한 인공 신경망 구성 시 은닉층의 수를 증가시킴으로써 오진과 미탐지율을 최소화할 수 있다는 점을 제안했다.[7]

입자 군집 최적화(PSO) 및 트리기반 분류기를 연구에서는 입자의 수가 커질수록 피쳐를 적게 선택하여 정확도가 떨어져 50의 입자를 사용한 PSO-50+(C4.5+RF+CART) 앙상블 조합이 99.80로 높은 정확도가 나왔다.[8]

NSL을 통한 배깅, 부스팅 및 스태킹을 사용하여 C4.5 분류기와 그 조합의 성능을 평가한 연구에서는 분류기 조합을 사용하여 분산으로 인한 오류가 감소하였다. 또한 NSL-KDD 데이터 세트를 사용하여 분류 성능을 향상 시켰다.[9]

새로운 분류함수를 위해 진행된 연구에서는 많이

쓰는 SoftMax 함수와 ReLU 함수를 분류함수로 넣었을 때의 성능을 비교하고 있다. 모델은 CNN과 FFNN이 쓰였으며 결과는 Softmax 함수가 우세하였다[10].

Spark를 이용하여 다중 계층 앙상블(Multi-layer ensemble), SVM, 심층 특징 추출(Deep Feature Extraction)을 분산 방식으로 조합하여 이상행동을 탐지하였을 때, 제안된 접근 방식은 다른 분류기(Naive Bayers, Random Forest, SYM, J48)에 비해 네트워크 이상 행위을 탐지하는데 더 높은 성능을 보였다[11]. 앙상블을 이용한 또 다른 연구에서는, 선형회귀(LR), 의사결정트리, 인공신경망 및 K-최근접이웃(KNN)과 같은 4가지 알고리즘의 조합을 사용하여 스태킹하였을 때 다른 알고리즘과 비교하여 최고의 정확도 결과를 보여주었다.[12] 여러 모델을 결합할 때 서로 다른 아키텍처를 효율적으로 결합한 연구에서는 데이터 세트가 많을수록 드롭아웃 효과가 증가했다. 또한 데이터 정규화, 높은 학습율(learning rate)과 모멘텀을 사용하여 유의미한 학습 속도 향상을 이루었다[13].

2.2 데이터 세트 분석

과거부터 현재까지 많이 사용되는 데이터세트에는 KDD Cup 99, NSL-KDD, UNSW-NB15, Bot-IOT, Ton-IoT, CIC-CSE-IDS2018 등이 있다. 하지만 KDD Cup 99의 경우 많은 양의 중복 레코드가 포함되어 있어 성능이 제대로 측정되지 않을 가능성이 높다. NSL-KDD의 경우 공격 데이터 세트이 고르게 분포되어 있지 않으며, 현대의 적은 흔적 공격 시나리오(modern low footprint attack scenarios)를 담고 있지 않다 [14]. UNSW-NB15는 현대의 적은 흔적 공격 시나리오를 담고 있지만[14] 다른 클래스에 비해 상대적으로 적은 인스턴스를 가지는 희소클래스 불균형 문제가 발생한다 [15]. 기존에 연구되었던 ANN과 스택 앙상블 기법을 이용한 모델은 NF-UQ-NIDS-V2 데이터세트를 이용하고 있다. NF-UQ-NIDS-V2 데이터세트는 여러 네트워크 설정과 공격 시나리오를 담고 있다. 기존 데이터세트의 제한 사항들을 개선하고 일관된 평가를 위해 표준 기능 세트로 제안된 데이터세트이다[16]. 하지만 NF-UQ-NIDS-V2 데이터세트 또한 각 공격들의 불균형 문제를 포함하고 있고 시계열 정보를 포함하고 있지 않다는 문제점이 있다.

따라서 본 논문은 ANN모델을 이용하여 스택 앙상블 기법을 이용하여 네트워크 침입 탐지 모델을 구현하고, NF-UQ-NIDS-V2 데이터세트의 문제점을 개선하기 위하여 인접 데이터 활용하여 데이터를 병합한다.

III. 방법론

제안된 모델의 기반은 ANN이며 1개의 정상 트래픽과 20개의 공격트래픽을 분류하는 ANN 모델 4개를 Stacking하여 구성하였다.

3.1 데이터 전처리

3.1.1 데이터 정제

Null값은 0으로 대체했다. IP 주소와 같은 object 유형은 long int 데이터 유형으로 변환했다. 그 후 모든 데이터형을 float으로 변환했다.

3.1.2 스케일링(Data Scaling)

DST_TO_SRC_SECOND_BYTES와 SRC_TO_DST_SECOND_BYTES, SRC_TO_DST_AVG_THROUGHPUT 그리고 DST_TO_SRC_AVG_THROUGHPUT 네 개의 피처에서 데이터값이 1.0e+9을 초과하는 값이 수록되어 있음을 확인하여 최대값을 1.0e+9으로 대체하였다.

3.1.3 이상값 처리

모든 피처에 대해 Z-score를 사용하여 정규화 처리했다. 하지만 피처의 값이 ±3을 초과하는 이상치에 대해서는 ±3으로 대체하였다.

3.2 실험 설계

3.2.1 희소클래스 데이터 강화(SMOTE)

NF-UQ-NIDS-V2 데이터셋은 공격 트래픽 분포의 불균형 문제가 있다. 이 데이터셋은 20개의 공격 유형 중 10개의 공격이 발생 빈도가 0.1% 미만

으로 나타난다. 본 논문에서는 이러한 데이터 불균형 문제를 해결하기 위해 SMOTE(Synthetic Minority Over-sampling Technique)를 활용하였다. SMOTE를 이용하여 희소 클래스 데이터를 합성함으로써 데이터셋을 균형 있게 강화하였다.

3.2.2 원핫인코딩 데이터(ENCODE)

TCP_FLAGS 데이터를 One-Hot Encoding을 하였다.

3.2.3 정상/공격 이진분류 강화(BINARY)

희소하게 발생하는 공격을 탐지하기 위해, 기존 4개의 앙상블 모델에 이진 분류 모델을 추가했다. 이진 분류 모델은 정상과 비정상 트래픽을 분류함으로써 비정상인 네트워크 활동을 탐지한다.

3.2.4 인접플로우 데이터 병합(MERGE)

NF-UQ-NIDS-V2 데이터셋에는 시계열 정보가 없다. 하지만 'Scanning', 'DoS', 'DDoS', 'Brute Force'와 같이 반복적인 트래픽 흐름을 통해 공격의 분류가 가능한 공격기법들은 하나의 플로우만 가지고 분류하는 것이 어려운 측면이 있다. 따라서 동일 근원지 IP에 대해 과거 또는 미래의 인접 플로우의 유사한 행동 정보가 필요하다. 이 문제를 해결하기 위해서 본 연구는 Fig 1.과 같은 데이터 구조를 제안한다. 이 데이터 구조는 IP 번호를 기반으로 데이터를 그룹화하고, 시간적 순서에 따라 나열하여 데이터의 시간적 관계를 명확하게 유지한다. 이러한 구조를 통해 인접 패킷의 유사한 행동을 파악하여 공격을 감지하는데 도움을 준다.

먼저, 데이터 전처리를 통해 동일 IP를 기준으로, 가장 가까이 있는 플로우를 과거1, 두 번째로 가까이 있는 플로우를 과거2로 묶어 컬럼을 확장하였다.

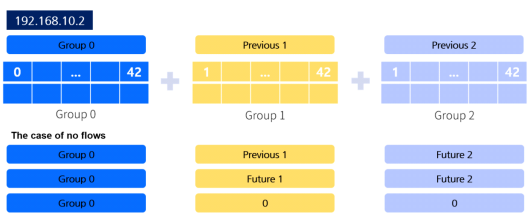

Fig. 1. Data structure of the adjacent netflow

단 10,000개 플로우 이내에 존재하는 경우만 인접 플로우로 처리하였다. 이 그룹화는 IP를 기반으로 플로우를 그룹화하여 관련 플로우를 함께 묶는 역할을 한다. 그룹화된 플로우는 시간 순서에 따라 인접한 플로우로 나열된다. 이렇게 함으로써 플로우의 시간적 순서를 보존하며 패턴을 분석할 수 있다.

만약 두 번째로 가까운 과거 플로우가 없는 경우, 가장 가까운 미래에 있는 플로우를 가져온다. 과거 플로우가 전혀 없는 경우, 미래에 가장 가까운 플로우를 2개의 그룹으로 가져온다. 이렇게 함으로써 플로우 부재 문제를 처리하고 시간적 연속성을 유지한다. 과거와 미래 데이터가 모두 없는 경우 해당 위치의 값을 0으로 대체한다. 이는 플로우의 누락을 처리하고 모델 학습에 안정성을 제공한다.

3.2.5 이진분류와 병합 동시 사용(BM)

MERGE의 방법을 통해 생성된 인접 NetFlow 데이터셋을 기반으로 모델 1,2,3은 기존과 동일하게 진행하였으며 다른 하나의 모델은 희소 공격 여부에 대해 이진분류를 수행하였다.

3.2.6 SMOTE와 병합 동시 사용(SM)

MERGE의 방법을 통해 생성된 인접 NetFlow 데이터셋을 기반으로 모델 1,2,3은 기존과 동일하게 학습하고 모델S의 경우 SMOTE를 이용하여 추가로 희소 클래스 데이터를 강화하였다.

3.3 ANN 학습 모델

본 논문에 모델1~모델4의 ANN 모델은 선행연구에서 참고한 논문[2]의 Fig 2.의 ANN 모델 구조를 참고하고 노드의 수를 일부 수정하여 적용했다. ANN에 SMOTE를 적용하는 경우 Fig 3.과 같이 Modle-4를 Model-S(SMOTE)로 대체하였다. Fig 4.는 학습에 사용한 ANN모델의 노드수와 드롭아웃 비율을 정리한 표이다. 학습 데이터를 인접 플로우 병합 데이터를 사용하는 Merge 모델의 경우 인접 플로우 병합으로 입력 피쳐가 127개로 기존 대비 3배 증가하였다 따라서 은닉 계층의 노드수를 다소 증가시킨 512개에서 64개 사이로 구성하였다.

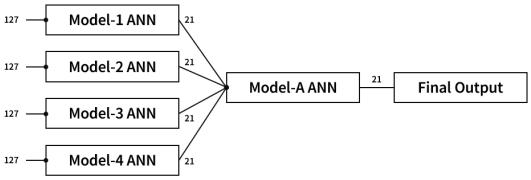

Fig. 2. Stack ensemble ANN models

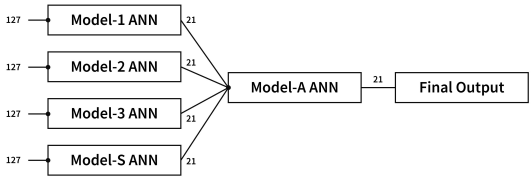

Fig. 3. 3 ANN and SMOTE models generate input for the model-a

Model	Input	Dropout	Hid1	Dropout	Hid2	Dropout	Out
Model-1	64	-	128	0.3	128	0.3	64
Model-2	64	0.3	128	0.2	256	0.1	128
Model-3	512	0.1	256	0.3	128	0.2	128
Model-4	128	0.3	64	0.1	64	0.1	64
Model-B	256	-	128	0.3	128	0.3	64
Model-S	256	0.3	128	0.2	128	0.1	64

Fig. 4. ANN structure of base Model

3.4 앙상블 - 스태킹(Stacking)

모델의 분류를 강화하기 위해 여러 모델을 함께 스택 앙상블 기법을 사용했다.

모델 A의 효율성을 높이기 위해 두 개의 은닉 계층만 사용했다. 입력 계층의 경우 4개 모델의 출력으로 한다. 은닉 계층의 노드 수는 128개이며 ReLU를 사용했다. 출력 계층에는 softmax 활성화 함수를 사용했다.

IV. 실험 환경 및 평가지표

4.1 데이터 세트 및 세부 사항

NetFlow는 네트워크 트래픽 수집과 저장을 위한 방법으로 네트워크 트래픽 흐름을 통계적으로 집계하여 처리한다. NetFlow 데이터는 출발지/목적지 IP 주소, 포트 번호, 프로토콜 등의 네트워크 트래픽 세부정보를 담고 있으며, 이를 활용하여 보안 이벤트를 분석할 수 있다.

본 연구에서 네트워크 공격 탐지용으로 사용한 넷플로우 기반의 NF-UQ-NIDS-v2에는 46개의 컬럼이 있으며 1개의 정상 트래픽과 20개의 공격 트래픽으로 레이블이 되어있다[16].

본 연구는 NF-UQ-NIDS-V2 데이터세트의 300만 개 행을 이용한다. 992,765(33.09%)는 정상적인 네트워크 트래픽이며 2,007,235(66.91%)은 악성 네트워크 트래픽이다. 데이터 세트에서 가장 빈번한 공격은 DDoS이며 드문 공격은 Worms이다. 하지만 표 1에서 확인할 수 있듯 NF-UQ-NIDS-V2 데이터세트는 공격 트래픽 분포의 불균형 문제가 있다. Table 1.은 공격별 분포를 보여주고 있다. 20개의 공격 유형 중 10개의 공격은 0.1% 미만이다. 이로 인해 특정 공격에 대한 감지 능력의 저하, 과적합 문제 등이 발생한다. 하위 분포 10개의 공격을 희소 공격으로 정의하였다.

또한 해당 데이터세트는 시계열 정보가 없으며 특정 Netflow의 한 패킷으로 공격을 판단하고 있다. 네트워크 공격에서 시계열 정보는 중요한 역할을 한다. 시계열 정보를 통해 공격 패턴을 식별할 수 있으며, 특정 시간대나 특정 주기에 공격 행위가 늘어나는 등의 패턴을 판단할 수 있다. 해당 시간대의 정보를 연결하여 공격을 판단할 수도 있다.

Table 1. The distribution ratie for each attack category except benign

Attack	ratio	Attack	ratio
DDoS	42.79%	Exploits	0.062%
DoS	35.17%	Fuzzers	0.044%
scanning	7.459%	Backdoor	0.037%
Reconnaissance	5.167%	Generic	0.034%
XSS	4.851%	MITM	0.015%
Password	2.258%	Ransomware	0.007%
Injection	1.348%	Analysis	0.005%
Bot	0.284%	Theft	0.004%
Bruteforce	0.248%	Shellcode	0.003%
Infilteration	0.235%	Worms	0.0003%

4.2 실험 환경

4.2.1 ANN 및 스택앙상블 설계

본 실험에서는 스택앙상블하기 위한 ANN을 총 여섯 종류를 설계하여 학습하였다. Model-1에서 Model-4는 각 레이어의 노드수와 드롭아웃 수치만

변화를 주었으며, 희소 공격에 대한 이진분류 학습기인 Model-B는 출력 데이터를 희소 공격 여부에 대해 이진분류 값으로 설정하였다. 희소 공격에 대해 오버샘플링을 한 SMOTE 모델의 경우 점유율이 10%가 넘는 공격에대해서는 언더샘플링으로 10%로 샘플의 수를 축소하고 1%가 안되는 공격에 대해서는 1%로 증폭하여 학습을 진행하였다. 실험에 사용한 각 모델의 레이어 설계는 Fig. 4와 같다.

실험은 파이썬 언어를 사용하였으며 데이터 전처리, 모델의 구축과 학습, 성능 평가, 데이터 처리, 시각화 등에 PyTorch, tensorflow, keras, scikit-learn, imblearn, pandas, matplotlib 등과 같은 파이썬 라이브러리를 사용하였다. 실험은 주피터(Jupyter) 및 구글 코랩에서 진행하였다. 코랩은 제공되는 GPU를 사용하였다.

4.3 평가

평가 지표들은 선행 연구에서 참고했던 논문[2]의 ANN 스택-앙상블 논문에서 사용한 정밀도(Precision), 재현율(Recall), F1값(F1-Score), 지원 개수(support)를 동일한 방법을 사용해 각 공격 범주별 평가한다. 정확도(Accuracy)를 통해 모델의 정확도를 말한다. 다음 평가 지표들은 기계학습에서 모델의 성능을 평가하고 비교하는 데 많이 사용된다. 각각의 특성에 따라 적절하게 선택하여 모델의 성능을 평가할 수 있다.

정확도(Accuracy)는 간단하고 직관적인 평가 지표이다. 전체 샘플 중에서 얼마나 정확하게 예측하는지를 나타내는 지표이다. 정밀도(Precision)는 양성(Positive)으로 예측한 데이터 중 실제로 양성인 데이터의 비율을 나타낸다. 클래스 간의 데이터가 불균형할 때 정밀도는 양성 클래스를 올바르게 예측하는 데 중요한 역할을 한다. 재현율(Recall)은 양성인 데이터 중에서 모델이 양성으로 예측한 데이터의 비율을 나타낸다. 실제 양성 데이터를 얼마나 잘 찾아내는지를 나타내므로 높은 재현율 값은 모델이 양성 클래스를 잘 감지하고 있음을 의미한다. F1값(F1-Score)는 정밀도와 재현율의 조화평균이다. 데이터 클래스가 균일하지 못할 때 모델의 성능을 더 정확하게 평가하는 데 도움이 된다. 지원 개수(support)는 각 클래스에 대한 샘플 수를 나타내며 지원 개수를 고려하여 다른 지표들을 해석하면, 각 클래스의 예측 성능을 더욱 정확하게 평가할 수 있다. 본 실험에서는 정확도와 F1 값이 차이가 크지 않아 이후로는 정확도 기준으로 언급하였다.

V. 연구 결과

5.1 수정 모델들의 연구 결과

데이터의 문제점을 해결하고 기존 모델의 성능을 향상시키기 위해서 총 여섯가지의 방법을 실험하였다. Table 2.는 SMOTE, ENCODE, BINARY, MERGE는 Fig.3의 모델에 입력만 넷플로우 병합 전의 43개의 피쳐를 입력으로 하여 총 100만 건의 레코드를 학습한 결과이다.

Table 2. result of SMOTE, ENCODE, BINARY models

Model	Acc.	Prec.	Recall	F1
SMOTE	0.9719	0.97	0.97	0.97
ENCODE	0.9731	0.97	0.97	0.97
BINARY	0.9733	0.97	0.97	0.97

5.1.1 SMOTE

공격유형들의 불균형한 데이터를 맞추기 위하여 데이터 증폭을 진행하여 비율이 10%가 넘는 과점 공격 샘플은 언더샘플링을 통해 데이터 수를 10%로 감소시켰고 비율이 1%가 안되는 희소 공격 샘플은 오버샘플링을 통해 데이터 수를 1%로 늘려 학습을 진행하였다. 모델의 정확도는 0.9719로 나왔다.

5.1.2 ENCODE

중요한 피쳐라고 생각한 TCP_FLAGS 데이터를 one-hot 인코딩을 이용하여 총 83개로 늘어난 피쳐로 학습하였다. 최종 모델의 정확도는 0.9731로 기존 모델보다 성능이 향상되었다.

5.1.3 BINARY

기존 앙상블한 4개의 모델에 추가로 이진 분류 모델을 추가하였다. 공격인지 아닌지 판단하는 모델을 추가함으로써 성능을 향상하려고 하였다. 최종 모델의 정확도는 0.9733으로 기존 모델보다 성능이 향상되었다.

5.1.4 MERGE

Table 3.은 인접 데이터를 병합하여 각 ANN별로 학습한 결과이다. 넷플로우 인접 데이터를 병합하여 학습한 모델 평가 결과로 최종 모델의 정확도는 Table 4.에서와 같이 MERGE의 경우 0.9852가 나왔다.

기존 모델은 데이터 세트의 시계열 정보를 고려하지 않았다. Scanning 공격 같은 시계열 정보가 중요한 공격의 피쳐를 기존 데이터 세트를 인접 데이터와 함께 확장 시킴으로써 시계열 정보를 추가해 학습시켜 기존 모델보다 성능이 좋아진 것을 확인할 수 있다.

Table 3. Result of ANN each base models

Model	loss	Acc.	val loss	val.acc
Modle-1	0.0600	0.9839	0.0568	0.0948
Modle-2	0.0617	0.9836	0.0570	0.9848
Modle-3	0.0611	0.9838	0.0592	0.9846
Modle-4	0.0679	0.9817	0.0593	0.9847
Modle-B	0.0012	0.9998	0.1074	0.9994
Modle-S	0.1004	0.9689	0.0792	0.9794

Table 4. result of Merge, BM, SM ensemble models (weighted average)

Model	Acc.	Prec.	Recall	F1
Merge	0.9852	0.99	0.99	0.98
BM	0.9853	0.99	0.99	0.98
SM	0.9854	0.99	0.99	0.98

5.1.5 BM(BINARY+MERGE)

인접 데이터를 활용한 모델에서 BINARY모델을 추가하였고 모델의 정확도는 0.9853, F1값은 0.98가 나왔다. 정확도가 0.985를 약간 넘긴 것으로 보아 F1값 또한 0.985를 약간 하회한 것으로 추정된다. 성능이 대폭 향상된 MERGE모델에 이진 분류 모델을 추가하여 희소 공격의 탐지율을 향상하고자 한 것으로서 전체에서 희소공격이 차지하는 비율이 작으므로 정확도와 F1값이 크게 변화하지 않은 것으로 이해된다.

5.1.6 SM(SMOTE+MERGE)

인접데이터 모델에서 3개의 모델에 언더샘플링과 오버샘플링을 한 SMOTE모델을 추가하였다. 데이터 불균형을 맞추기 위해 SMOTE를 사용함으로써 성능향상을 하고자 하였다. 최종 모델의 정확도는 0.9854가 나왔다.

5.2 최종 모델 연구 결과

Table. 2와 Table 4.의 실험결과에 의하면 원본 데이터로 Smote, Encode, Binary 학습한 결과 대비 인접데이터를 병합한 Merge, BM, SM 모델의 결과가 0.01이상 개선 됨을 알 수 있다. Table 4.의 결과에서 Merge와 BM, SM의 결과를 비교하면 정확도, 정밀도, 재현율, F1값 측면에서는 큰 차이가 없음을 알수 있다. 하지만 SM의 도입 목적이 희소공격에 대한 탐지율을 개선하기 위함이었으므로 희소공격을 전체에서 점유율 0.1% 미만인 공격을 희소공격으로 정의하고 이에 대해 정밀도, 재현율, F1 값의 가중치를 배제한 산술 평균을 비교했다. Table 5.에서 보듯이 F1값 기준으로 SM모델이 Merge와 BM 대비 희소 공격에 대한 F1 값 또한 각 0.02, 0.04가 개선됨을 확인하였다.

Table 6.에서 Benign과 각 공격 별로 F1값을 살펴보면 Benign과 DDoS, Dos, Bot공격이 0.99 이상으로 가장 높게 나왔고 Worms는 두 모델 모두 0이 나왔고 Shellcode, Theft 또한 F1값이 낮게 나왔다. 하지만 희소공격의 경우 희소공격을 강화학습한 SM 모델의 F1 값이 Generic과 Shellcode 공격을 제외하고는 Merge 모델보다 높게 나왔음을 알 수 있다.

점유율 상위 10개의 공격에 대한 탐지율은 전반적으로 높게 나왔는데 Bruteforce 만이 탐지율이 낮음을 알 수 있다. 이는 DoS 공격과 Bruteforce 공격의 차이점이 적어 DoS 공격으로 오탐이 일어난 횟수가 많아 탐지율과 F1값이 낮은 것으로 판단된다.

Table 5. Total and the sparse attacks detection results (non-weighted averagei)

Model		Acc.	Prec.	Recall	F1
Merge	Tot.	0.9852	0.87	0.70	0.74
	Sparse		0.65	0.57	0.58
BM	Tot.	0.9853	0.82	0.73	0.75
	Sparse		0.77	0.50	0.56
SM	Tot.	0.9854	0.81	0.75	0.76
	Sparse		0.63	0.61	0.60

Table 6. The F1-value and sample counts of SM and Merge model

Attack	SM-F1	MergeF1	Samples
Benign	0.99	0.99	198,477
DDoS	0.99	0.99	171,544
DoS	0.99	0.99	141,449
Scanning	0.98	0.97	29,808
Reconnaissance	0.96	0.96	20,974
XSS	0.96	0.96	19,343
password	0.93	0.93	9,108
injection	0.83	0.83	5,374
Bot	1.00	1.00	1,126
Bruteforce	0.39	0.38	966
Infilteration	0.99	0.99	954
Exploits	0.72	0.65	271
Fuzzers	0.59	0.60	184
Backdoor	0.93	0.93	162
Generic	0.77	0.84	135
MITM	0.48	0.45	53
Ransomware	0.92	0.93	31
Analysis	0.92	0.79	17
Theft	0.23	0.00	12
Shellcode	0.47	0.62	10
Worms	0.00	0.00	2
Total			600,000

VI. 결 론

네트워크의 사용이 보급화되면서 다양한 네트워크 공격이 생겨나고 있다. 네트워크 공격은 개인 뿐만아니라 기업, 국가까지 대상으로 사람들의 편의와 경제적 손실에도 영향을 미치고 있다. 이를 해결하기 위한 수단인 네트워크 침입 탐지 시스템은 앞으로 더 중요한 해결책으로 여겨질 것이다. 본 논문에서는 NF-UQ-NIDS-V2 데이터 세트를 사용하여 네트워크 공격을 탐지하는 방법론을 제시 하였다. NF-UQ-NIDS-V2의 경우 시계열 정보를 포함하고 있지 않기 때문에 이를 해결하기 위해 인접 플로우 데이터를 병합하여 모델 학습에 적용시켰다. 제시하고 있는 모델은 선행 연구로 진행되었던 레이어별 노드의 설계에 변화를 준 스택형 앙상블 ANN모델에 대해 플로우 데이타의 시계열 성격을 보완하고, 희소공격에 대한 탐지율을 보강한 점에 의미가 있다.

모델의 성능을 향상 시키기 위해서 One-hot encoding, 이진분류, Smote와 같은 방법 등을 활용해 보았지만 성능 향상에 가장 큰 영향을 미쳤던 방법은 인접 데이터를 활용한 MERGE였다. 또한 희소공격의 탐지율 개선을 위해 희소공격 데이터를 보강하여 학습한 SMOTE 방법으로 희소공격의 탐지율을 개선하여 SM모델 기준 최종 정확도는 0.9854로 선행 연구보다 약 0.01의 성능 향상되었다.

References

[1] MSIT, “Analysis of major cyber threat trends in the first half of 2023”, https://www.msit.go.kr/bbs/view.do?sCode=user&mId=239&mPid=113&bbsSeqNo=94&nttSeqNo=3183355, Jul. 2023

[2] Khan, Lamia Parven, Tasfia Tahsin Anika, Suraka Iban Hanif, and Rashedur M. Rahman. "Network Intrusion Detection Using Stack-Ensemble ANN", 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 1104-1109, Jun. 2022

[3] Kwon-yong Lee, Se-rin Kim, Min-Jae Seok, Sang-Won Lee, Ji-Hyun Sung and Harksu Cho, “A network intrusion detection system model using stack ensemble technique and netflow traffic”. 2023 Chungcheong Cyber Security Conference, pp. 1-9, Sep, 2023

[4] Javaid A, Niyaz Q, Sun W, Alam M. “A deep learning approach for network intrusion detection system”, InProceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS). pp. 21-26, May. 2016

[5] Maithem M, Al-Sultany GA. "Network intrusion detection system using deep neural networks",InJournal of Physics: Conference Series. Vol. 1804, no.1, p.012138. Feb. 2021

[6] Jia Y, Wang M, Wang Y. "Network intrusion detection algorithm based on deep neural network", IET Information Security 13(1), pp. 48-53. Jan. 2019

[7] Tangi SD, Ingale MD. "A Survey: Importance of ANN basedNIDS in Detection of DoS Attacks", International Journal of Computer Applications, 83(11), pp. 1-83, Jan. 2013

[8] Tama BA, Rhee KH. "A combination of PSO-based feature selection and tree-based classifiers ensemble for intrusion detection systems", In Advances in Computer Science and Ubiquitous Computing. Springer, pp. 489-495, 2015

[9] Gyanchandani M, Yadav RN, Rana JL. "Intrusion detection using C4. 5: performance enhancement by classifier combination", ACEEE Int. J. on Signal & Image Processing, 1(03), pp. 46-49. Jun. 2010

[10] Agarap AF. "Deep learning using rectified linear units (relu)", arXiv preprint arXiv:1803.08375. Mar. 2018

[11] Marir N, Wang H, Feng G, Li B, Jia M. "Distributed abnormal behavior detection approach based on deep belief network and ensemble SVM using spark", IEEE Access 6, pp. 59657-59671. Oct. 2018

[12] Murat, U. Ç. A. R., Emine, U. Ç. A. R., İncetaş, M. O.. "A Stacking Ensemble Learning Approach for Intrusion Detection System". Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 9(4), 1329-1341, July 2021

[13] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R., "Dropout:A Simple Way to Prevent Neural Networks from Overfitting", The journal of machine learning research 15(1), pp. 1929-1958. Jan. 2014

[14] Moustafa N, Slay J., "Unsw-nb15: a comprehensive data set for network intrusion detection systems (unswnb15 network data set)", 2In2015 military communications and information systems conference (MilCIS), IEEE. pp. 1-6. Nov. 2015

[15] Dae-Bum Lee and Jae-Hyun Seo, "Classification Performance Improvement of UNSW-NB15 Dataset Based on Feature Selection", Journal of the Korea Convergence Society, vol.10, no.5, pp. 35-42. May. 2019

[16] Sarhan M, Layeghy S, Portmann M. , "Towards a standard feature set for network intrusion detection systemdatasets", Mobile networks and applications. pp. 1-14, Feb. 2022

〈저 자 소 개〉

성 지 현 (Ji-Hyun Sung) 학생회원
2020년 3월: 호서대학교 컴퓨터공학부 재학중
〈관심분야〉 컴퓨터공학, 정보보안

이 권 용 (Kwon-Yong Lee) 학생회원
2018년 3월: 호서대학교 컴퓨터공학부 재학중
〈관심분야〉 컴퓨터공학, 정보보안

이 상 원 (Sang-Won Lee) 학생회원
2019년 3월: 호서대학교 컴퓨터공학부 재학중
〈관심분야〉 컴퓨터공학, 정보보안

석 민 재 (Min-Jae Seok) 학생회원
2018년 3월: 호서대학교 컴퓨터공학부 재학중
〈관심분야〉 컴퓨터공학, 정보보안

김 세 린 (Se-Rin Kim) 학생회원
2020년 3월: 호서대학교 컴퓨터공학부 재학중
〈관심분야〉 컴퓨터공학, 정보보안

조 학 수 (Harksu Cho) 정회원
1997년: 서울대학교 계산통계학과 전산과학전공 학사
1999년: 서울대학교 전산과학과 석사
2001년~2022년: ㈜윈스 부사장
2023년~현재: 호서대학교 컴퓨터공학부 특임교수
〈관심분야〉 네트워크보안, 클라우드보안, 인공지능