KIPS Trans. Softw. and Data Eng.
Vol.10, No.7 pp.271~278
ISSN: 2287-5905 (Print), ISSN: 2734-0503 (Online)

https://doi.org/10.3745/KTSDE.2021.10.7.271

Host-Based Intrusion Detection Model Using Few-Shot Learning

Park DaeKyeong† · Shin DongIl†† · Shin DongKyoo†† · Kim Sangsoo†††

ABSTRACT

As the current cyber attacks become more intelligent, the existing Intrusion Detection System is difficult for detecting intelligent attacks that deviate from the existing stored patterns. In an attempt to solve this, a model of a deep learning-based intrusion detection system that analyzes the pattern of intelligent attacks through data learning has emerged. Intrusion detection systems are divided into host-based and network-based depending on the installation location. Unlike network-based intrusion detection systems, host-based intrusion detection systems have the disadvantage of having to observe the inside and outside of the system as a whole. However, it has the advantage of being able to detect intrusions that cannot be detected by a network-based intrusion detection system. Therefore, in this study, we conducted a study on a host-based intrusion detection system. In order to evaluate and improve the performance of the host-based intrusion detection system model, we used the host-based Leipzig Intrusion Detection-Data Set (LID-DS) published in 2018. In the performance evaluation of the model using that data set, in order to confirm the similarity of each data and reconstructed to identify whether it is normal data or abnormal data, 1D vector data is converted to 3D image data. Also, the deep learning model has the drawback of having to re-learn every time a new cyber attack method is seen. In other words, it is not efficient because it takes a long time to learn a large amount of data. To solve this problem, this paper proposes the Siamese Convolutional Neural Network (Siamese-CNN) to use the Few-Shot Learning method that shows excellent performance by learning the little amount of data. Siamese-CNN determines whether the attacks are of the same type by the similarity score of each sample of cyber attacks converted into images. The accuracy was calculated using Few-Shot Learning technique, and the performance of Vanilla Convolutional Neural Network (Vanilla-CNN) and Siamese-CNN was compared to confirm the performance of Siamese-CNN. As a result of measuring Accuracy, Precision, Recall and F1-Score index, it was confirmed that the recall of the Siamese-CNN model proposed in this study was increased by about 6% from the Vanilla-CNN model.

Keywords : Machine Learning, LID-DS, Few-Shot Learning, Siamese Network, HIDS

Few-Shot Learning을 사용한 호스트 기반 침입 탐지 모델

박 대 경† · 신 동 일†† · 신 동 규†† · 김 상 수†††

요 약

현재 사이버 공격이 더욱 지능화됨에 따라 기존의 침입 탐지 시스템(Intrusion Detection System)은 저장된 패턴에서 벗어난 지능형 공격을 탐지하기 어렵다. 이를 해결하려는 방법으로, 데이터 학습을 통해 지능형 공격의 패턴을 분석하는 딥러닝(Deep Learning) 기반의 침입 탐지 시스템 모델이 등장했다. 침입 탐지 시스템은 설치 위치에 따라 호스트 기반과 네트워크 기반으로 구분된다. 호스트 기반 침입 탐지 시스템은 네트워크 기반 침입 탐지 시스템과 달리 시스템 내부와 외부를 전체적으로 관찰해야 하는 단점이 있다. 하지만 네트워크 기반 침입 탐지 시스템에서 탐지할 수 없는 침입을 탐지할 수 있는 장점이 있다. 따라서, 본 연구에서는 호스트 기반의 침입 탐지 시스템에 관한 연구를 수행했다. 호스트 기반의 침입 탐지 시스템 모델의 성능을 평가하고 개선하기 위해서 2018년에 공개된 호스트 기반 LID-DS(Leipzig Intrusion Detection-Data Set)를 사용했다. 해당 데이터 세트를 통한 모델의 성능 평가에 있어서 각 데이터에 대한 유사성을 확인하여 정상 데이터인지 비정상 데이터인지 식별하기 위해 1차원 벡터 데이터를 3차원 이미지 데이터로 변환하여 재구성했다. 또한, 딥러닝 모델은 새로운 사이버 공격 방법이 발견될 때마다 학습을 다시 해야 한다는 단점이 있다. 즉, 데이터의 양이 많을수록 학습하는 시간이 오래 걸리기 때문에 효율적이지 못하다. 이를 해결하기 위해 본 논문에서는 적은 양의 데이터를 학습하여 우수한 성능을 보이는 Few-Shot Learning 기법을 사용하기 위해 Siamese-CNN(Siamese Convolutional Neural Network)을 제안한다. Siamese-CNN은 이미지로 변환한 각 사이버 공격의 샘플에 대한 유사성 점수에 의해 같은 유형의 공격인지 아닌지 판단한다. 정확성은 Few-Shot Learning 기법을 사용하여 정확성을 계산했으며, Siamese-CNN의 성능을 확인하기 위해 Vanilla-CNN(Vanilla Convolutional Neural Network)과 Siamese-CNN의 성능을 비교했다. Accuracy, Precision, Recall 및 F1-Score 지표를 측정한 결과, Vanilla-CNN 모델보다 본 연구에서 제안한 Siamese-CNN 모델의 Recall이 약 6% 증가한 것을 확인했다.

키워드 : 기계학습, LID-DS, 퓨샷 러닝, 샴 네트워크, 호스트 기반 침입 탐지 시스템

1. 서 론

현재 사이버 공격이 더욱 지능화됨에 따라 공격자들은 알려지지 않은 취약점을 악용하고 지능적으로 다양해지고 있다. 지능적으로 다양해지는 공격을 방어하는 것은 매우 중요한 문제이다. 문제를 해결하는 방법으로 가장 많이 사용하는

※ 본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었음 (UD2000014ED).
† 준 회 원 : 세종대학교 컴퓨터공학과 지능형드론 융합전공 석사과정
†† 종신회원 : 세종대학교 컴퓨터공학과 지능형드론 융합전공 교수
††† 비 회 원 : 국방과학연구소 사이버/네트워크 기술센터 책임연구원
Manuscript Received : December 30, 2020
First Revision : February 2, 2021
Accepted : February 25, 2021
* Corresponding Author : Shin DongKyoo(shindk@sejong.ac.kr)

솔루션 중 하나는 침입 탐지 시스템(Intrusion Detection System)이다. 침입 탐지 시스템은 네트워크 기반인 NIDS (Network-based Intrusion Detection System), 호스트 기반인 HIDS(Host-based Intrusion Detection System) 두 가지 방식으로 나눌 수 있다. 호스트 기반 침입 탐지 시스템은 네트워크 기반 침입 탐지 시스템과 달리 시스템 내부와 외부를 전체적으로 관찰해야 하는 단점이 있다. 하지만 네트워크 기반 침입 탐지 시스템에서 탐지할 수 없는 침입을 탐지할 수 있는 장점이 있어서 많은 연구가 필요하다. 또한, 침입 탐지 시스템에는 두 가지 유형이 있다[1]. 오용탐지와 이상탐지로 나눌 수 있다. 오용탐지는 알려진 시그니처를 기반으로 사용자나 시스템 또는 프로그램의 행동이 공격 패턴과 일치하는지 검사하는 방법이다. 이상 탐지는 오용탐지 방법과 달리 정상적인 패턴을 기반으로 비정상 행위를 탐지하는 방법이다. 오용탐지는 알려지지 않은 새로운 공격을 탐지하기 어렵다는 단점이 있지만, 이상 탐지는 알려지지 않은 새로운 공격을 탐지 할 수 있다는 장점이 있다. 그러나 이상 탐지는 다양한 정상적인 사용 패턴을 정의하기가 어렵고 학습되지 않은 정상적인 패턴은 이상 행위로 간주하기 때문에 오경보율이 증가한다는 문제점이 있다[2]. 최근 딥러닝 기술의 발전으로 인해 ICT(Information & Communication Technology) 분야 및 IoT(Internet of Things) 분야에 많은 연구가 이루어져 다양한 지능형 서비스들이 제공되고 있다. 이러한 기술 발전에 따라 보안 분야에서는 침입 탐지 시스템에 딥러닝 기술이 적용되는 연구가 진행되는 사례가 있다. 딥러닝은 심층 신경망을 통해 자체 기능을 학습하여 앞서 말한 약점을 보완할 수 있는 기술이다. 즉, 기계 학습(Machine Learning)과 딥러닝은 자체적으로 이상 행위를 학습하고 정상적인 패턴을 구분하여 오경보를 줄일 수 있다. 현재 다양한 연구들이 비정상 행위를 탐지하기 위해 딥러닝을 침입 탐지 시스템 연구에 사용한다[3].

실험에 사용된 데이터 세트는 2018년에 공개된 호스트 기반 LID-DS 데이터 세트이다. LID-DS 데이터 세트는 기존 공개되었던 데이터들과 다르게 구성되어 있다. 기존에 공개된 데이터 세트들보다 최신 컴퓨터 시스템의 다양한 특징들과 공격 방법 및 시나리오로 구성되어 있다[4].

본 논문에서는 벡터 데이터를 이미지로 변환하여 비정상 행위에 대한 딥러닝 기반 탐지 모델을 생성하여 연구를 진행한다. 딥러닝 모델은 새로운 공격이 발견될 때마다 학습해야 하는 문제점 때문에 많은 데이터를 학습할 때 효율적이지 못하다. 따라서, 적은 양의 데이터를 학습하여 우수한 성능을 보이는 Few-Shot Learning 기법을 사용하기 위해 Siamese-CNN을 제안한다. 정확성은 Few-Shot Learning 기법을 사용하여 정확성을 계산했으며, Siamese-CNN의 성능을 확인하기 위해 Vanilla-CNN과 성능을 비교했다. 그 후, Vanilla-CNN과 Siamese-CNN 중 어떠한 모델이 공격 유형을 가장 잘 찾아낼 수 있는지에 대한 최종 결과를 서술하였다.

2. 관련 연구

2.1 침입 탐지 데이터 세트

KDD99 데이터 세트는 침입 탐지 시스템을 평가하기 위해 DARPA(Defense Advanced Research Projects Agency) 및 AFRL(Air Force Research Laboratory)의 후원으로 MIT에서 최초로 침입 탐지 시스템에 대한 표준 데이터 공개했다. 데이터 종류는 서비스 거부 (DoS), 사용자 대 루트 (U2R), 원격 대 로컬 공격 (R2L) 및 Probe 공격의 네 가지 공격 범주로 구성되어 있다. KDD99 데이터 세트는 침입 탐지 시스템 평가를 위해 많은 연구가 진행되고 있다[5].

UNM 데이터 세트는 KDD99 데이터 세트보다 최신에 공개된 데이터 세트이지만 데이터들이 일련의 시스템 콜 형태로 구성되어 있다[6].

침입 탐지 시스템 분야에서 KDD99 및 UNM과 같은 데이터 세트는 오래된 데이터 세트로 현재 사용 중인 컴퓨터 시스템의 특징을 포함하지 않는다. 이를 해결하기 위해 2013년 Australian Defense Force Academy(ADFA)에서 호스트 기반 침입 탐지 시스템을 평가하기 위해 ADFA 데이터 세트를 공개했다. ADFA 데이터 세트에는 정상 및 공격 데이터를 일련의 시스템 콜로 구성되어 있다[5,6].

2.2 침입 탐지 관련 연구

침입 탐지 시스템은 사이버 공격 패턴에 대한 정상 패턴과 비정상 패턴을 비교하여 비정상 행위를 탐지하고 차단하는 시스템이다[7,8].

Laskov 등[9]은 Decision Tree, K-NN(K-Nearest Neighbor), MLP(Multi-Layer Perceptron), K-means, SVM(Support Vector Machine) 등 다양한 기계학습 알고리즘을 침입 탐지에 적용했고, 각 알고리즘을 ROC(Receiver Operator Characteristic) 곡선을 사용해 비교했다.

Kim 등[10]은 침입 탐지 시스템에서 SVM과 K-NN 같은 기계학습 알고리즘을 사용하여 높은 오경보율을 보이는 문제점을 해결하기 위한 연구를 진행했다.

Kim 등[11]은 비정상 행위 기반의 호스트 침입 탐지 시스템을 설계하는 데 있어, LSTM-Based System-Call Language Modeling 방법을 제안하였다. 기존 방법들에서 자주 발생하는 높은 오탐율(False-Alarm Rate) 문제를 해결하기 위해서, 저자는 새로운 앙상블(Ensemble) 방법을 사용하여 문제점을 해결했다.

Ravipati 등[12]은 LID-DS 데이터 세트의 특징과 가장 비슷한 KDD99 데이터 세트를 이용하여 8가지의 기계학습 알고리즘을 실험한 결과로 성능 평가 및 오탐율 수치를 보여주었다.

최근 딥러닝 기술의 발전으로 인해 CNN을 기반으로 바이너리 및 다양한 범주의 공격을 탐지하는 수많은 연구도 진행되고 있다[13,14].

Khan 등[15]은 새로운 침입 탐지 모델을 제안하기 위해 기계학습 알고리즘을 사용할 때의 단점을 지적했다. CNN 기반 네트워크 침입 탐지 모델과 소프트 맥스 알고리즘을 결합

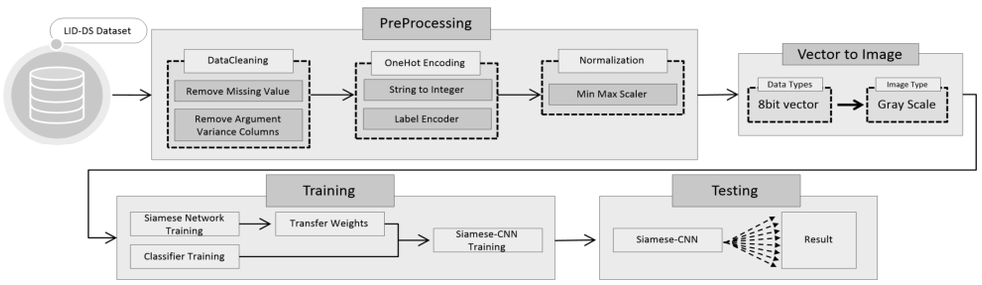

Fig. 1. Proposed Host-based Intrusion Detection Model Structure

하는 방법을 제안했다. 또한, KDD 데이터 세트를 사용하여 제안된 모델을 평가하고 실험 결과는 모델이 SVM 및 DBN (Deep Belief Network) 알고리즘보다 침입 탐지에 더 효율적임을 보여주었다.

Upadhyay 등[16]은 41개의 KDD 데이터 세트의 컬럼 중에서 무작위로 선택된 36개의 컬럼을 사용했다. 그 후 데이터 세트를 6 × 6 크기의 이미지로 변환한 다음 나머지 기능을 다른 변수에 저장하여 CNN 모델을 학습했다. 실험 결과 제안된 모델의 침입 탐지 오류가 2% 미만인 것으로 나타났으며, 데이터를 이미지로 변환하여 분석하는 것이 더 효율적임을 보여주었다.

2.3 Siamese Networks

Siamese Network는 같은 형태의 네트워크를 사용하여 서로 다른 2개의 입력 데이터를 처리하는 네트워크이다. 네트워크들은 가중치를 공유하며 입력 이미지에 대한 특징 벡터를 생성한다. 같은 클래스의 이미지들은 벡터 공간에서 가깝게, 서로 다른 클래스의 이미지들은 멀게 표현되도록 학습한다. 즉, 거리 함수를 사용하여 생성된 특징 벡터 간의 거리를 계산하여 두 이미지가 같은 클래스인지의 아닌지 판단한다. 거리 함수는 일반적으로 유클리드 거리(Euclidean distance)나 코사인 거리(cosine distance)와 같은 일반적인 유사도 함수를 사용한다.

Hsiao 등[17]은 Siamese Network를 사용하여 샘플 간의 유사성 순위를 매기도록 훈련했다. 또한, N-way one-shot 작업을 통해 정확성을 계산했다. 그 결과 일반적인 딥러닝 모델보다 더 효율적임을 보여주었다.

Moustakidis 등[18]은 특징 벡터를 이미지로 변환하는 Vec2im 방법과 새로운 특징을 추출하는 파이프라인을 제안했다. 또한, 입력 데이터 차원을 1차원으로 줄이기 위해 Siamese convolutional neural network를 사용하여 NSL-KDD 침입 탐지 데이터 세트에 적용했다.

Taigman 등[19]은 표준 교차 엔트로피 손실과 오류 역전파를 사용하여 Siamese Network를 훈련했다. 각 샘플 간의 L1 거리에 유사성을 예측하고 각 샘플의 얼굴이 같은 얼굴인지 예측한다.

2.4 Few-Shot Learning

Few-Shot Learning은 데이터가 충분한 데이터 세트를 사용하여 Meta Learning을 진행하고, 각 클래스에 포함된 데이터가 적은 데이터 세트를 분류하기 위한 학습 방법이다[20].

본 논문에서는 Siamese Network에서 두 이미지에 대한 특징 벡터 학습과 벡터 사이의 거리를 비교한다. 또한, 각 사이버 공격 방법에 대한 유사성 점수를 비교하여 같은 공격인지 아닌지 탐지하는 모델을 제안한다.

3. 데이터 세트 소개 및 모델 구현

본 논문에서 제안하는 구조는 Fig. 1과 같다. LID-DS Dataset, PreProcessing, 이미지 생성, Siamese Network, Siamese-CNN, N-way K-Shot Learning으로 구성되어 있으며, 본 절에서 각 파트에 대해서 나뉘어 설명한다. 3.1절은 LID-DS Dataset 에 대한 설명을 한다. 3.2절은 PreProcessing 파트이며, 데이터 형식에 따른 데이터 정규화 과정을 서술한다. 3.3절은 1차원 벡터 데이터를 3차원 이미지 데이터로 변환하는 과정을 서술한다. 3.4 절은 Siamese Network 파트로 같은 형태를 가지는 두 개의 Convolutional Neural Network에 대한 구조를 살펴본다. 3.5절은 N-way K-Shot Learning 파트로, N과 K에 대한 설명 및 특징에 관해서 설명한 뒤 3.6절을 통해 본 논문에서 제안하는 Siamese-CNN의 구조를 살펴본다. 3.7절은 Train Test Split 파트이며, 실험에 사용한 Train Test 데이터의 비율을 서술한다.

3.1 LID-DS Dataset

본 논문에서 사용한 LID-DS 데이터 세트는 2018년 Leipzig University에서 호스트 기반 침입 탐지 시스템의 이상 탐지 연구를 위한 LID-DS 데이터 세트를 공개했다.

LID-DS 데이터 세트는 기존에 공개되었던 데이터 세트보다 최신 컴퓨터 시스템의 특징들과 사이버 공격 방법 및 사이버 공격 시나리오로 구성되어 있다. Table 1은 기존에 공개되었던 침입 탐지 데이터 세트와 LID-DS 데이터 세트를 비교한 표이다.

Table 1과 같이 기존에 공개되었던 데이터 세트들은 너무 오래되어 현재 시스템 특징을 반영하고 있지 못하거나, 일련의 시스템 콜 형태로 구성되어 있어 침입 탐지 시스템 연구에 사용하기에 적합하지 않다. 따라서 본 논문에서는 현재 시스템 특징들과 다양한 사이버 공격 방법으로 구성된 LID-DS 데이터 세트를 사용하여 실험을 진행했다.

3.2 PreProcessing

LID-DS 데이터 세트는 Fig. 1의 PreProcessing 파트와 같이 모든 데이터에 대해 Argument Feature와 결측값은 삭제하였고 event_time Feature는 콜론(:)을 제거했다. event_ direction과 event_type은 LabelEncoder를 사용하여 숫자로 변환했다.

Process 카테고리는 총 16개로 구성되어 있다. 또한, Process의 개수는 공격 방법마다 다르다. 그에 따라, 각 사이버 공격 방법에 사용된 Process들은 하나의 Process로 통합했다. 그 결과 총 10개의 Process로 구성된 라벨들은 LabelEncoder를 사용하여 각 Process에 라벨을 붙여 사용했다. 그리고 MinMaxScaler는 0에서 255값을 사용하여 Normalization을 진행했다.

3.3 이미지 생성

각 샘플들은 0에서 255 사이의 값을 가지고 있다. 본 논문에서는 Fig. 1의 Vector to Image 파트와 같이 샘플들을 8bit vector로 변환하고, grayscale 유형의 이미지 데이터를 생성했다. Fig. 2는 데이터 전처리하는 과정과 이미지로 생성된 과정을 나타낸 그림이다. grayscale 이미지는 한 가지 유형의 컬러로 구성된 구조를 가지며 최종적으로, MxNx1픽셀 배열로 변환한다.

M과 N은 각각의 열과 행의 수를 나타낸다. 변환한 샘플을 64x64픽셀의 이미지로 변환했다. Table 2는 LID-DS 데이터 세트를 변환한 이미지 개수이다.

Table 1. Feature Comparison LID-DS with other Datasets

Feature	LID-DS	ADFA-LD	UNM	KDD99
Arguments	o	x	x	o
Returnvalues	o	x	x	o
Timestamps	o	x	x	o
Process ID	o	x	o	o
Data buffers	o	o	x	x
Meta data	o	x	x	o
Thread	o	x	o	o

3.4 Siamese Network

본 논문에서 사용한 Siamese Networks는 Fig. 3의 Siamese Network 파트와 같이 같은 형태를 가지는 두 개의 Convolutional Neural Network로 구성되어 있다. 두 개의 입력 이미지는 Convolution Layer를 통해 각각의 특징 벡터를 생성한다. 생성된 두 개의 특징 벡터 간의 거리를 유클리드 거리 방법으로 계산하여 유사성 점수를 통해 두 이미지가 같은 클래스인지 아닌지 판단한다.

Table 3을 통해 본 논문에서 제안된 Siamese Network의 구조를 확인할 수 있다. 마지막 층을 제외한 Layer의 활성화 함수로 LeakyReLU를 사용하였다.

Table 2. Number of Images Per Each Cyber Attack Method

Attack type	Number of Images
Bruteforce	825
CVE-2012	1019
CVE-2014	786
CVE-2017	1157
CVE-2018	1022
CVE-2019	1073
EPS	1060
PHP	1112
SQL	1078
Zip	1062

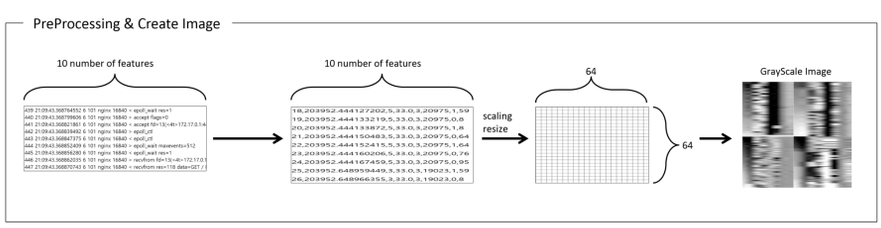

Fig 2. The Detailed Dataset Preprocessing Steps and Create Image Steps

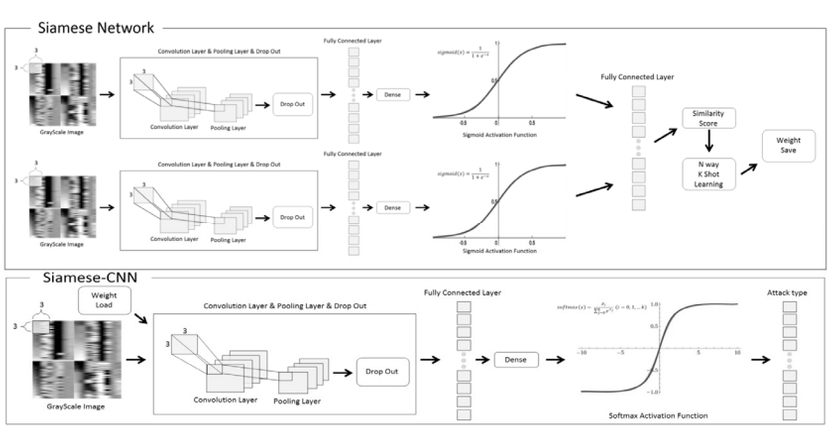

Fig. 3. The Structure of Siamese Convolutional Neural Networks and Convolutional Neural Networks

Table 3. The Siamese Network Configuration used in the Experiment

Parameter	Value
Layer	64-128-128-256
MaxPooling	2, 2
Dropout	0.25
Activation	LeakyReLU, sigmoid
Optimizer/ Learning rate	Adam / 0.0004
Loss	binary_crossentropy

Table 4. The Neural Network Configuration used in the Experiment

Parameter	Value
Layer	64-128-128-256
MaxPooling	2, 2
Dropout	0.25
Activation	relu, softmax
Optimizer/ Learning rate	Adam / 0.0004
Loss	categorical crossentropy

3.5 N-way K-Shot Learning

Siamese Network 모델의 학습이 제대로 이루어졌는지 확인하기 위해 Few-Shot Learning의 방법의 하나인 N-way K-shot Learning 방법을 사용했다. N-way K-shot Learning은 데이터 세트를 훈련에 사용하는 서포트 데이터(Support Data)와 테스트에 사용하는 쿼리 데이터(Query Data)로 구성한다. N은 범주의 수, K는 범주별 서포트 데이터의 수를 의미한다. N-way K-shot Learning은 N의 값이 작아질수록 더 정확한 예측이 가능하고 N의 값이 클수록 정확성이 낮아진다. 일반적으로 실험에서는 N은 2~10개 이하, K를 1개 또는 5개로 설정한다.

3.6 Siamese-CNN and Vanilla-CNN

본 논문에서는 침입 탐지 분류 모델로 합성곱 신경망을 사용한다.Table 4와 Fig. 3의 Siamese-CNN 파트와 같이 제안된 합성곱 신경망의 구조를 확인할 수 있으며, Siamese Network의 Weight를 사용하여 학습을 진행한다.

마지막 층을 제외한 Layer의 활성화 함수로 relu를 사용하였다. input_shape는 64x64 크기와 1개의 컬러 채널을 가지고 있으므로 (64, 64, 1)의 튜플 값을 가진다. 그리고 사소한 변화를 무시하기 위해서 Maxpooling2D를 통해 주요 값만 추출하여 작은 출력값을 만들어 사용했다. 또한, Conv2D와 Maxpooling은 2차원을 주로 다루기 때문에 전 결합 층에 전달하기 위해서는 1차원으로 전달해야 하므로 Flatten 함수를 사용하여 1차원으로 변환하였다.

Vanilla-CNN의 제안된 합성곱 신경망의 구조는 Table 4와 같다. 즉, Fig. 3의 Siamese-CNN 학습 과정 중 Siamese Network의 Weight를 사용하는 과정을 제외한 모든 구조는 같다.

3.7 Train Test Split

훈련 데이터와 테스트 데이터는 train_test_split 모듈을 사용하여 일반적으로 많이 사용하는 8:2 비율로 나누어 실험을 진행했다. 훈련 데이터를 학습시킨 후 테스트 데이터를 사용하여 모델을 평가하는 과정에서 과적합(Overfitting) 현상을 발견했다. 과적합이란 모델이 너무 과적합 되도록 학습한 나머지, 올바른 예측을 하지 못하는 현상을 말한다. 과적합을 방지하기 위하여 다른 비율로 나누어진 데이터를 포함하여 Cross-validation으로 모델평가를 진행한 결과 7:3 비율로 나누어진 데이터 세트의 모델 성능이 가장 높았다. 따라서 본 논문에서는 훈련 데이터와 테스트 데이터를 7:3 비율로 나누어 사용했다.

4. 평가 지표 및 실험 결과

4.1 평가 지표

학습된 모델의 성능 평가는 Precision, Recall, F1 Score를 사용했으며 성능 평가 및 정확도의 Equation은 다음과 같다.

Equation (1)은 Precision 공식의 한 예이다.

$$Precision(\text{정밀도}) = \frac{TP}{FP+TP} \tag{1}$$

Equation (2)은 Recall 공식의 한 예이다.

$$Recall(\text{재현율}) = \frac{TP}{FN+TP} \tag{2}$$

Equation (3)은 F1 Score 공식의 한 예이다.

$$F1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = 2 * \frac{precision * recall}{precision + recall} \tag{3}$$

Equation (4)은 Accuracy 공식의 한 예이다.

$$Accuracy(\text{정확도}) = \frac{TN+TP}{TN+FP+FN+TP} \tag{4}$$

Equation (1), (2), (3), (4) 수식에 대한 속성은 Fig. 4, Table 5와 같다.

Precision(정밀도)은 True라고 분류한 것 중에서 실제 True인 것의 비율이며, Recall(재현율)은 실제 True라고 분류한 것 중에서 실제 True인 것의 비율이다. F1-Score(조화평균)는 Precision과 Recall의 조화평균이다. 즉, 모델의 성능을 측정하는 데 있어서 Precision과 Recall은 유용하게 사용되지만 실제로 모델이 얼마나 효과적인지 설명할 방법이 없으므로 F1-Score라는 방법을 사용하여 실제로 모델이 효과적인지 아닌지 판단하는 데 사용한다. Accuracy(정확성)는 Precision, Recall과 달리 False를 False라고 예측한 예도 옳은 경우로 계산하기 때문에 Equation (4)와 같이 계산한다.

4.2 실험 결과

64x64의 크기로 이미지를 변환한 LID-DS 데이터 세트를 사용하여 3.5 파트에서 생성한 Siamese Network를 사용하여 실험을 진행하였다. Fig. 5는 생성한 Siamese Network의 학습 과정을 시각화한 그림이다.

학습된 Siamese Network의 모델의 성능을 확인하기 위해서 3.6 파트에서 제안한 N-way One-shot Learning을 사용하였다. N은 1, 2, 3, 4에 대해 N-way 테스트를 진행했다. 성능 평가를 하기 위해 사용한 데이터는 3.7 파트의 테스트 데이터를 사용했다.

		Real correct answer	
		True	False
Classification result	True	True Positive(TP)	False Positive(FP)
	False	False Negative(FN)	True Negative(TN)

Fig. 4. Confusion Matrix

Table 5. Properties and Descriptions used in Equations

Property	True Value	Prediction	Result
TN(TrueNegative)	False	False	Correct
FP(FalsePositive)	False	True	Wrong
FN(FalseNegative)	True	False	Wrong
TP(TruePositive)	True	True	Correct

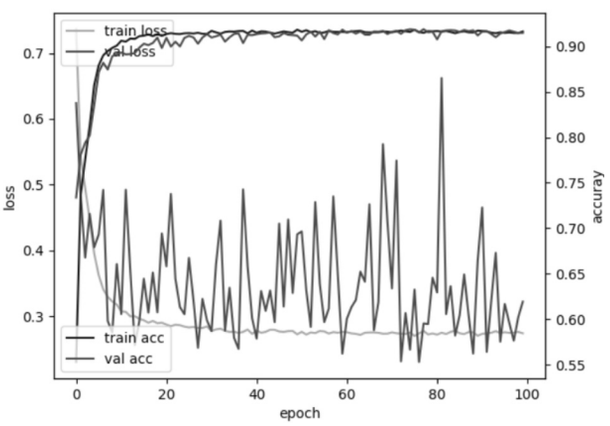

Fig. 5. Visualize the Learning Process of Siamese Network

테스트는 N값의 테스트에 대해 step_epoch를 2000번 수행하였고 2000번을 수행한 평균 정확성을 계산했다. 그 결과 Table 6과 같다.

LID-DS 데이터 세트의 각 사이버 공격 방법을 분류하기 위해 Fig. 3과 같이 저장한 Siamese Network의 Weight를 Siamese-CNN에 불러와 실험을 진행했다.

Table 7은 Siamese Network의 Weight를 사용하여 생성된 Siamese-CNN과 Weight를 사용하지 않은 Vanilla-CNN에 대한 성능 비교 결과이다.

Table 7의 결과를 보면 Siamese-CNN이 Vanilla-CNN보다 정확성이 약 3%, Recall이 약 4% 높은 결과를 보여주었다. 각 사이버 공격 방법이 제대로 분류가 되었는지 확인하기 위해 Siamese-CNN의 Confusion Matrix를 확인했으며 그 결과 Fig. 6과 같다. 또한, Fig. 6에 분류된 사이버 공격 방법과 이름은 Table 8과 같다.

Fig. 6을 보면 (0, 2)와 (4, 5), (7, 8) 사이버 공격 방법이 제대로 분류되지 않고 있는 점을 확인했다. Bruteforce와 CVE-2014 사이버 공격 방법을 제외한 (CVE-2018, CVE-2019)는 정보 유출에 대한 사이버 공격 방법으로 CVE-2019는 CVE-2018의 취약점을 보완한 사이버 공격 방법이다. (PHP, SQL) 사이버 공격 방법은 OWASP(Open Web Application Security Project)에서 식별한 취약점으로 PHP와 SQL은 공격자가 인터프리터에 적대적인 데이터를 보낼 수 있는 공격으로 분류했다. 따라서 (CVE-2018, CVE-2019), (PHP, SQL)

Table 6. Siamese Network Performance with N-way One-Shot Learning

N-way	1	2	3	4
Accuracy	100%	98%	94%	92%

Table 7. Classifier Model Performance Results using LID-DS Data

	Precision	Recall	F1-Score	Accuracy
Vanilla-CNN	87%	86%	87%	88%
Siamese-CNN	89%	90%	90%	91%

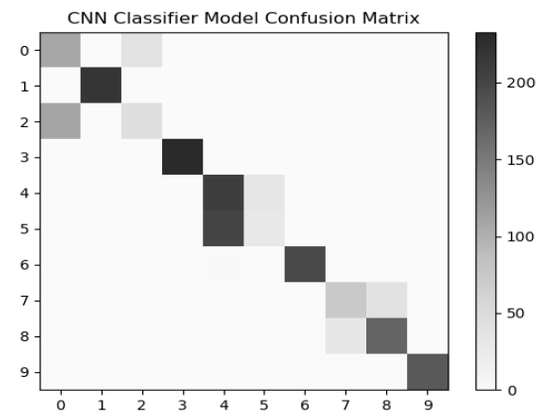

Fig. 6. Confusion Matrix for CNN Model Classification Performance

사이버 공격 방법은 서로 같은 유형의 사이버 공격 방법이기 때문에 각각 하나의 공격 방법으로 볼 수 있다.

따라서, 10개의 사이버 공격 방법을 8개의 사이버 공격 방법으로 변환했다. Table 9는 10개의 사이버 공격 방법에서 8개의 사이버 공격 방법으로 변환한 표이다.

변환한 8개의 사이버 공격을 사용하여 Siamese-CNN의 성능을 다시 확인했다.

Fig. 7은 8개의 사이버 공격을 사용한 Siamese-CNN의 학습 과정을 시각화한 그림이다. 학습한 Siamese-CNN의 성능을 확인한 결과는 Table 10과 같으며, 성능에 대한 Confusion Matrix는 Fig. 8과 같다.

위의 결과와 같이 8개의 사이버 공격 방법으로 변환한 뒤 실험을 진행한 결과 10개의 사이버 공격 방법으로 분류한 Siamese-CNN보다 정확성이 약 2%, Recall이 약 2% 향상된 것을 확인했다.

5. 결론 및 추후 연구

본 논문에서 소개하는 LID-DS 데이터 세트는 시스템의 보안 취약점을 최신 상태로 구성되어 있다. 기본 스레드 정보가 사라지지 않고 다양한 유형의 HIDS를 평가하는 데 사용할 수 있다. 각 데이터에 대한 유사성을 확인하기 위해 1차원 벡터 데이터를 3차원 이미지 데이터로 변환하여 재구성했다.

딥러닝 기반의 침입 탐지 시스템은 새로운 공격이 발견될 때마다 다시 학습을 진행해야 한다는 단점이 있다. 이를 해결하기 위해, 적은 양의 데이터를 학습하여 우수한 성능을 보이는 Few-

Table 8. LID-DS Data Cyber Attack Method

Num	Cyber attack type
0	Bruteforce
1	CVE-2012
2	CVE-2014
3	CVE-2017
4	CVE-2018
5	CVE-2019
6	EPS
7	PHP
8	SQL
9	Zip

Table 9. Modified LID-DS Data Cyber Attack Method

Num	Cyber attack type	Changed cyber attack type
0	Bruteforce	Bruteforce
1	CVE-2012	CVE-2012
2	CVE-2014	CVE-2014
3	CVE-2017	CVE-2017
4	CVE-2018	XSS
	CVE-2019	
5	EPS	EPS
6	PHP	Script
	SQL	
7	Zip	Zip

Table 10. Model Performance Results for Classifying 8 Cyber Attack Methods

	Precision	Recall	F1-Score	Accuracy
Siamese-CNN	93%	92%	91%	93%

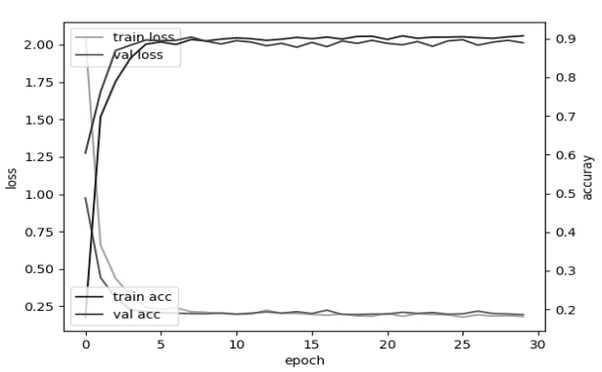

Fig. 7. Visualize the Learning Process of Classifier Model

Shot Learning 기법을 사용하기 위해 Siamese Network를 생성하여 모델의 성능을 확인했다. 그 후, 각 사이버 공격 방법을 분류하기 위해 본 논문에서 제안하는 Siamese Network의 Weight를 사용하는 Siamese-CNN과 Vanilla-CNN의 성능 평가를 진행했다. 그 결과 일반적인 Vanilla-CNN보다 Siamese-CNN의 정확성이 약 3%, Recall이 약 4% 높은 것을 확인했다.

각 사이버 공격 방법이 제대로 분류되었는지 확인하기 위해서 Confusion Matrix를 확인해본 결과 4가지의 사이버 공격 방법을 2가지의 사이버 공격 방법으로 변환할 수 있는

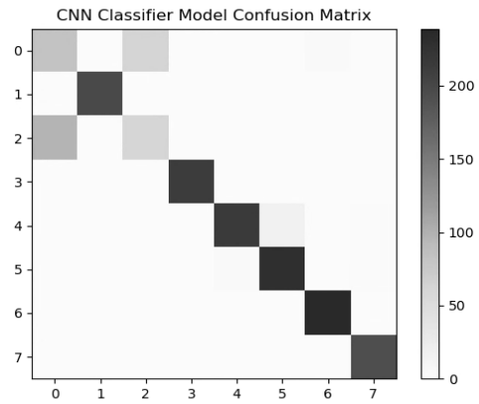

Fig. 8. Confusion Matrix for Model Performance Classifying 8 Cyber Attack Methods

것을 확인했다. 따라서, 10개의 사이버 공격 방법을 8개의 사이버 공격 방법으로 줄여서 성능을 다시 평가한 결과 10개의 사이버 공격 방법을 분류한 Siamese-CNN보다 정확성은 약 2% 증가하였고 Precision 4%, Recall 2%, F1-Score 1% 증가한 것을 확인했다.

향후 연구로 이미지로 변환한 LID-DS 데이터 세트를 활용하여 다양한 사이버 공격에 대한 침입 탐지하는 연구를 진행할 것이다. 제안한 모델의 Hyper Parameter 값을 최적화하여 새로운 사이버 공격과 내부 사이버 공격에 대한 침입 탐지 정확성을 더욱 높이는 연구를 진행할 것이다. 또한, 최근 생성되고 있는 침입 탐지 데이터 세트에 대해서 실험을 확장 시킬 수 있다.

References

[1] Y. G. Choi and S. S. Park, "Reinforcement Mining Method for Anomaly Detection and Misuse Detection using Post-processing and Training Method," *Proceedings of the Korean Information Science Society Conference*, pp.238-240, 2006.

[2] S. O. Choi and W. N. Kim, "Control system intrusion detection system technology research trend," *Review* of *Korea Institute of Information Security & Cryptology*, Vol.24, No.5, pp.7-14, 2014.

[3] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," arXiv preprint arXiv: 2007.02500 (2020).

[4] M. M. Röhling, M. Grimmer, D. Kreubel, J. Hoffmann, and B. Franczyk, "Standardized container virtualization approach for collecting host intrusion detection data," *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*, IEEE, 2019.

[5] O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," *2017 IEEE International Conference on Big Data (Big Data)*, IEEE, 2017.

[6] M. Pendleton and S. Xu, "A dataset generator for next generation system call host intrusion detection systems," *MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM)*, IEEE, 2017.

[7] L.N. Tidjon, M. Frappier, and A. Mammar, "Intrusion detection systems: A cross-domain overview," *IEEE Communications Surveys & Tutorials*, Vol.21, No.4, pp.3639-3681, 2019.
[8] H. Kwon, Y. Kim, H. Yoon, and D. Choi, "Optimal cluster expansion-based intrusion tolerant system to prevent denial of service attacks," Applied Sciences, Vol.7, No.11, pp.1186, 2017.
[9] P. Laskov, P. Düssel, C. Schäfer, and K. Rieck, "Learning intrusion detection: supervised or unsupervised?," *International Conference on Image Analysis and Processing*, Springer, Berlin, Heidelberg, 2005.
[10] J. H. Kim and H. W. Kim, "An effective intrusion detection classifier using long short-term memory with gradient descent optimization," *2017 International Conference on Platform Technology and Service (PlatCon)*, IEEE, 2017.
[11] G. Kim, H. Yi, J. Lee, Y. Paek, and S. Yoon, "LSTM-based system-call language modeling and robust ensemble method for designing host-based intrusion detection systems," arXiv preprint arXiv:1611.01726, 2016.
[12] R. D. Ravipati and M. Abualkibash, "Intrusion Detection System Classification Using Different Machine Learning Algorithms on KDD-99 and NSL-KDD Datasets-A Review Paper," *International Journal of Computer Science & Information Technology*, Vol.11, 2019.
[13] A. K Verma, P. Kaushik, and G. Shrivastava, "A Network Intrusion Detection Approach Using Variant of Convolution Neural Network," *2019 International Conference on Communication and Electronics Systems (ICCES)*, IEEE, 2019.
[14] J. Kim, J. Kim, H. Kim, M. Shim, and E. Choi, "CNN-Based Network Intrusion Detection against Denial-of-Service Attacks," *Electronics*, Vol.9, No.6, pp.916, 2020.
[15] R. U. Khan, X. Zhang, M. Alazab, and R. Kumar, "An improved convolutional neural network model for intrusion detection in networks," *2019 Cybersecurity and Cyberforensics Conference (CCC)*, IEEE, 2019.
[16] R. Upadhyay and D. Pantiukhin, "Application of convolutional neural network to intrusion type recognition," *Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics, Udupi*, India, pp.13-16, 2017.
[17] S. C. Hsiao, D. Y. Kao, Z. Y. Liu, and R. Tso, "Malware image classification using one-shot learning with Siamese networks," *Procedia Computer Science*, Vol.159, pp.1863-1871, 2019.
[18] S. Moustakidis and P. Karlsson, "A novel feature extraction methodology using Siamese convolutional neural networks for intrusion detection," *Cybersecurity*, Vol.3, No.1, pp.1-13, 2020.
[19] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
[20] S. E. Jang and J. T. Kim, "Few-shot classification of Histopathology image using Batch Hard Loss-based Siamese Networks," *The Korean Institute of Information Scientists and Engineers*, pp.634-636, 2019.

박 대 경
https://orcid.org/0000-0003-1195-9017
e-mail : dkpark@sju.ac.kr
2020년 ~ 현 재 세종대학교 컴퓨터공학과 지능형드론 융합전공 석사과정
관심분야 : 디지털 포렌식, 기계학습, 정보보안, 데이터 마이닝

신 동 일
https://orcid.org/0000-0002-8621-715X
e-mail : dshin@sejong.ac.kr
1988년 연세대학교 컴퓨터과학과(학사)
1993년 Washington State University 컴퓨터과학과(석사)
1997년 North Texas University 컴퓨터과학과(박사)
1998년 ~ 현 재 세종대학교 컴퓨터공학과 지능형드론 융합전공 교수
관심분야 : 정보보안, 기계학습, 데이터 마이닝, 생체신호 데이터처리

신 동 규
https://orcid.org/0000-0002-2665-3339
e-mail : shindk@sejong.ac.kr
1986년 서울대학교 계산통계학과(학사)
1992년 Illinois Institute of Technology 컴퓨터과학과(석사)
1997년 Texas A&M University 컴퓨터과학과(박사)
1998년 ~ 현 재 세종대학교 컴퓨터공학과 지능형드론 융합전공 교수
관심분야 : 정보보안, 기계학습, 유비쿼터스 컴퓨팅, 생체신호 데이터처리

김 상 수
https://orcid.org/0000-0001-7975-673X
e-mail : wisdory@naver.com
1997년 경북대학교 전자공학과(학사)
2003년 경북대학교 컴퓨터공학과(석사)
2003년~현 재 국방과학연구소 사이버/네트워크 기술센터 책임연구원
관심분야 : 사이버전 기술, 위협 헌팅, 사이버 상황인식