AutoEncoder Based Feature Extraction for Multi-Malicious Traffic Classification

Sungwoong Yeom
Chonnam National University
Gwangju City, South Korea
yeomsw0421@gmail.com

Chulwoong Choi
Chonnam National University
Gwangju, South Korea
sentilemon02@gmail.com

Kyungbaek Kim
Chonnam National University
Gwangju, South Korea
kyungbaekkim@jun.ac.kr

ABSTRACT

In recent years, research is being activated to classify deep learning-based malicious network traffic. Malicious network traffic classification has a problem of wasting time by learning meaningless features due to a large number of traffic and high-dimensional features. In this paper, we propose a technique for feature extraction based on AutoEncoder and classifying malicious network traffic through a random forest classifier. This technique reduces the time and spatial complexity required in the intrusion detection system by extracting features from high-dimensional data. To evaluate this technique, the performance of AE-RF and Single-RF classifiers is measured for Accuracy, Precision, Recall and F-Score using the CICIDS 2017 data set. The evaluation showed that AE-RF has an accuracy of 98% or more, which shows excellent performance and detection speed.

CCS CONCEPTS

• **Networks** → **Firewalls**.

KEYWORDS

IDS, Feature extraction, Classification

ACM Reference Format:
Sungwoong Yeom, Chulwoong Choi, and Kyungbaek Kim. 2020. AutoEncoder Based Feature Extraction for Multi-Malicious Traffic Classification. In *The 9th International Conference on Smart Media and Applications (SMA 2020), September 17–19, 2020, Jeju, Republic of Korea.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3426020.3426093

1 INTRODUCTION

Recently, COVID-19 made it necessary to adopt telecommuting to continue running services and businesses globally. At this point, the mobile office and telecommuting concept allows employees to access the corporate network and who need to share sensitive information via mobile or computer networks to access the corporate network. Because a large amount of confidential information must be shared through a network by the activation of a remote operation mechanism, network security that is vulnerable to cyber attacks needs to be further strengthened [5].

SMA 2020, September 17–19, 2020, Jeju, Republic of Korea

ACM ISBN 978-1-4503-8925-9/20/09…$15.00
https://doi.org/10.1145/3426020.3426093

Recently, as machine learning technology matures, research on classifying malicious traffic based on machine learning is being activated. Malicious traffic is classified by learning traffic characteristics with classifiers such as CNN [9] and Random Forest [8]. However, these techniques have a problem that time consuming due to data imbalance and unnecessary learning due to a large number of traffic and high-dimensional features collected in the network intrusion detection system. In order to reduce such data imbalance and waste of time, feature extraction is necessary [1].

In this paper, we propose AutoEncoder based feature extraction technique for classifying malicious network traffic through a random forest classifier. To evaluate this technique, we use CICIDS2017, a benchmarking dataset built out of 11 evaluation criteria [7]. The accuracy of the proposed AutoEncoder based random forest technique was 98%, showing excellent performance.

2 RELATED WORK

Recently, as machine learning matures, machine learning-based research is being activated in anomaly detection-based intrusion detection systems [4]. The paper[8] efficiently detects application layer DoS attacks by using the random forest algorithm for DoS attack detection. In this paper, the 500 trees were used to improve accuracy while lowering the false positive rate. However, as the number of trees increases, time and resources are more wasted. Feature extraction or dimensionality reduction is needed to reduce the time used to test the random forest. The paper[1] compares feature extraction techniques for high-dimensional data that are difficult to analyze. Among the introduced techniques, AutoEncoder showed that it can reduce the time and spatial complexity of the intrusion detection system by extracting meaningful features. The paper[2, 3, 6] extracts features based on Auto-encoder and classifies malicious traffic through machine learning, but does not measure performance based on Random Forest, which has very good performance. In this paper, we propose a method of extracting features based on AutoEncoder and classifying malicious traffic through a random forest classifier.

3 AUTOENCODER BASED FEATURE EXTRACTION FOR MALICIOUS TRAFFIC DETECTION

3.1 AutoEncoder based Feature Extraction

In this paper, we use AutoEncoder, a deep learning algorithm to reduce dimensions of high-dimensional features to low-dimensional features. As shown in Figure 1, AutoEncoder is composed of Input Layer, Hidden Layer, and Output Layer, and is a self-supervised

Table 1: DNS Request Dataset

Model	Encoder Structure	Accuracy	Precision	Recall	F1-Score	Train time(s)	Test time(s)
Single RF	.	98.29	99	98.49	98.49	338.37	11.37
AE_RF	[77, 64]	98.84	99	98.49	98.49	912.13	16.86
AE_RF	[77, 64, 49]	98.84	99	98.49	98.49	994.19	18.46
AE_RF	[77, 64, 49, 36]	98.74	99	98.49	98.49	863.78	19.9
AE_RF	[77, 64, 49, 36, 25]	98.84	99	98.49	98.49	609.72	20.1
AE_RF	[77, 64, 49, 36, 25, 16]	98.85	99	99	99	447.24	21.05

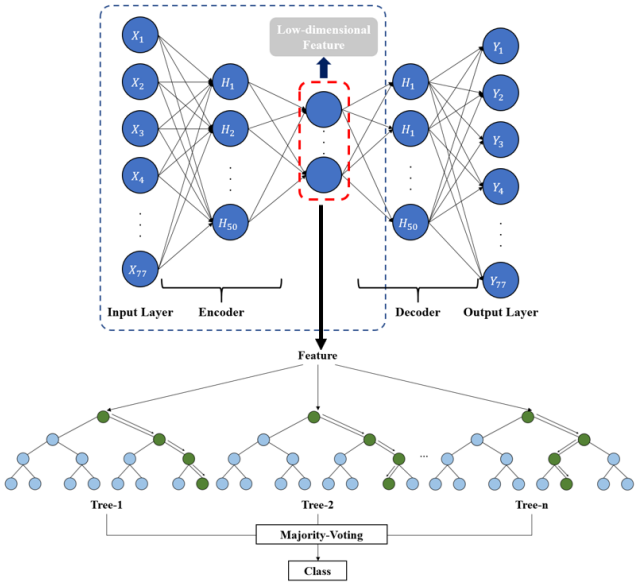

Figure 1: Architecture of AutoEncoder based Random Forest

neural network. The self-supervised neural network is a methodology to improve understanding data itself by solving a predefined problem using data without labels. AutoEncoder is composed of encoder and decoder which has symmetric hidden layers. In case of the encoder, it is structured to reduce the dimension gradually. In case of decoder, the dimension reduced by the encoder is gradually expanded. AutoEncoder creates an encoder model through self-supervised learning that makes that input and output are the same without a label. This encoder is responsible for extracting features. In this study, features are extracted using the only encoder among AutoEncoders. The extracted features classify malicious traffic through a machine learning classifier.

3.2 Random Forest-based Malicious Traffic Classification

In this paper, we use a random forest, a representative machine learning algorithm for classification based on the extracted low-dimensional features. Random forest is an algorithm that uses an ensemble learning technique that generates several decision trees and predicts the class that has received the most selection among the predicted values in each tree. The random forest uses the bagging method to generate each decision tree. Bagging (bootstrap aggregation) generates a classifier by randomly and repeatedly extracting samples of the same size from a training dataset and applying an algorithm for each. Also, each branch of the tree uses a different subset of features, because each node uses only a subset of the features. These 2 mechanisms combine to make all the trees in a random forest different. Therefore, the random forest can be said to be an algorithm suitable for generalization by preventing overfitting.

4 EVALUATION

4.1 DATASETS AND EVALUATION OVERVIEW

To evaluate this technique, we use the CICIDS2017 dataset, which contains the latest attack and sparse classes, and the results of network traffic analysis [7]. This dataset comprises over 80% of normal traffic and 12 lastest attack traffic. These dataset include the most modern and common attacks such as Brute Force FTP, Brute Force SSH, DoS, DDoS, Heartbleed, Web Attack, Infiltration, and Botnet. At this time, Infiltration and Heartbleed are rare classes that occupy less than 0.01% of the total data set. The number of features in the CICIDS2017 dataset is 77.

In the proposed technique, the number of hidden layers and neurons are adjusted to find optimal deep learning and machine learning models. In case of AutoEncoder, the batch size is set to 154, the learning rate is 0.001, and Epoch is set to 100 times. In case of a random forest, n_estimators, which means the number of decision trees, was set to 500. In Deep Neural Network, we can see the difference performance according to the number of hidden layers and the number of neurons in each layer. If the number of hidden layers is small and the number of neurons in each layer is insufficient, high−dimensional features cannot be extracted effectively. Table 1 compares the performance and AutoEncoder learning time and compression time when different numbers of hidden layers are applied. As a result of the experiment, the deeper the network we apply, the longer the learning time of the random forest and the final classification time we take. In this experiment, the performance of the AE_RF[77, 64, 49, 36, 25, 16] model was the best, and the training time and test time were not slow compared to single RF.

5 CONCLUSION

In this paper, we propose a technique that extracts features based on AutoEncoder and classifies malicious network traffic through a

random forest classifier. The evaluation results show that the accuracy of the proposed AE_RF is 0.55% higher than that of Single RF. This technique was able to reduce the time and spatial complexity required in the intrusion detection system while maintaining high performance by extracting features from high−dimensional data. However, since the performance of the random forest technique was very high, it was difficult to make a clear comparison. In the future AutoEncoder−based feature extraction will be performed for the CNN classifier.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2017R1A2B4012559). This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2016-0-00314) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

REFERENCES

[1] Bahareh Abolhasanzadeh. 2015. Nonlinear dimensionality reduction for intrusion detection using auto-encoder bottleneck features. In *2015 7th Conference on Information and Knowledge Technology (IKT)*. IEEE, 1–5.

[2] Md Zahangir Alom and Tarek M Taha. 2017. Network intrusion detection for cyber security using unsupervised deep learning approaches. In *2017 IEEE National Aerospace and Electronics Conference (NAECON)*. IEEE, 63–69.

[3] Kazuki Hara and Kohei Shiomoto. 2020. Intrusion Detection System using Semi-Supervised Learning with Adversarial Auto-encoder. In *NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 1–8.

[4] Vinod Kumar, Vinay Choudhary, Vivek Sahrawat, and Vinay Kumar. 2020. Detecting Intrusions and Attacks in the Network Traffic using Anomaly based Techniques. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 554–560.

[5] Yair Meidan, Michael Bohadana, Yael Mathov, Yisroel Mirsky, Asaf Shabtai, Dominik Breitenbacher, and Yuval Elovici. 2018. N-baiot—network-based detection of iot botnet attacks using deep autoencoders. *IEEE Pervasive Computing* 17, 3 (2018), 12–22.

[6] Soosan Naderi Mighan and Mohsen Kahani. 2018. Deep learning based latent feature extraction for intrusion detection. In *Electrical Engineering (ICEE), Iranian Conference on*. IEEE, 1511–1516.

[7] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A Ghorbani. 2018. Toward generating a new intrusion detection dataset and intrusion traffic characterization.. In *ICISSP*. 108–116.

[8] Shreekh Wankhede and Deepak Kshirsagar. 2018. DoS attack detection using machine learning and neural network. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. IEEE, 1–5.

[9] Sungwoong Yeom and Kyungbaek Kim. 2019. Detail Analysis on Machine Learning based Malicious Network Traffic Classification. In *the Eighth International Conference on Smart Media and Applications (SMA 2019)*. KISM, 49–53.