INTERNATIONAL JOURNAL ENGINEERING RESEARCH

International Journal for Innovative Engineering and Management Research

A Peer Revieved Open Access International Journal

www.ijiemr.org

IJIEMR Transactions, online available on 24th Dec 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=Issue 12

DOI: 10.48047/IJIEMR/V11/ISSUE 12/39

Title **NETWORK INTRUSION DETECTION USING DCGAN: SEMI-SUPERVISED APPROACH**

Volume 11, ISSUE 12, Pages: 283-291

Paper Authors

Koppolu Naga Shivaji, Dr.M.Arathi

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

NETWORK INTRUSION DETECTION USING DCGAN: SEMI-SUPERVISED APPROACH

1 Koppolu Naga Shivaji, M.TECH in Computer Networks and Information Security (CNIS), SIT, JNTUH
2 Dr.M.Arathi, M.Tech(CS),Ph.d,Professor, SCHOOL OF IT, JNTUH

ABSTRACT: Network intrusion detection is a crucial task since malicious traffic occurs every second these days. Various research has been studied in this field and shows high performance. However, most of them are conducted in a supervised manner that needs a range of labeled data but it is hard to obtain. This paper proposes a semi-supervised Generative Adversarial Networks (GAN) model for network intrusion detection that requires only 10 labeled data per each flow type. Our model is evaluated using the publicly available CICIDS-2017 dataset and outperforms other malware traffic classification models.

Keywords- *Network Intrusion Detection, Semi-supervised learning, Generative Adversarial Network.*

1. INTRODUCTION

As network traffic among systems and devices has grown up, the number of network attacks has increased accordingly. Various studies have been conducted on network intrusion detection to prevent unexpected cyber attacks. Starting with the less accurate port-based method, payload-based approach called deep packet inspection (DPI) emerged. However, payload-based technique is only applicable to unencrypted traffic and has high computational overhead. A machine learning approach can solve these problems but requires the manual feature extraction process. To overcome these drawbacks, deep learning methods [1] [2] such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) automatically extract the feature from raw data and get better results than the machine learning approach in network intrusion detection. However, they are concentrated on supervised learning, a training method that uses only labeled data. The supervised learning approach needs a large amount of labeled data. It is challenging to obtain labeled datasets due to the constraints of capturing the specific malicious flow. On the other hand, unsupervised learning approaches use only unlabeled data to perform classification tasks. However, classifying unlabeled data is a difficult task because it requires numerous datasets and significant effort, and has a lower performance. Therefore, we choose a semi-supervised approach that utilizes a few labeled data and a large amount of unlabeled data simultaneously during training. he supervised learning approach needs a large amount of labeled data. It is challenging to obtain labeled datasets due to the constraints of capturing the specific malicious flow. On the other hand, unsupervised learning approaches use only unlabeled data to perform classification tasks. However, classifying unlabeled data is a difficult task because it requires numerous datasets and significant effort, and has a lower performance.

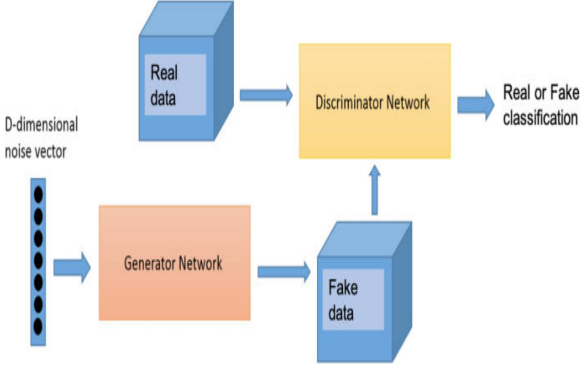


Fig.1: GAN model

2. LITERATURE REVIEW

2.1 Autonomous unknown-application filtering and labeling for dl-based traffic classifier update:

Network traffic classification has been widely studied to fundamentally advance network measurement and management. Machine Learning is one of the effective approaches for network traffic classification. Specifically, Deep Learning (DL) has attracted much attention from the researchers due to its effectiveness even in encrypted network traffic without compromising neither user privacy nor network security. However, most of the existing models are created from closed-world datasets, thus they can only classify those existing classes previously sampled and labeled. In this case, unknown classes cannot be correctly classified. To tackle this issue, an autonomous learning framework is proposed to effectively update DL-based traffic classification models during active operations. The core of the proposed framework consists of a DL-based classifier, a self-learned discriminator, and an autonomous self-labeling model. The discriminator and self-labeling process can generate new dataset during active operations to support classifier update. Evaluation of the proposed framework is performed on an open dataset, i.e., ISCX VPN-nonVPN, and independently collected data packets. The results demonstrate that the proposed autonomous learning framework can filter packets from unknown classes and provide accurate labels. Thus, corresponding DL-based classification models can be updated successfully with the autonomously generated dataset.

2.2 HAST-IDS: Learning Hierarchical Spatial-Temporal Features Using Deep Neural Networks to Improve Intrusion Detection:

The development of an anomaly-based intrusion detection system (IDS) is a primary research direction in the field of intrusion detection. An IDS learns normal and anomalous behavior by analyzing network traffic and can detect unknown and new attacks. However, the performance of an IDS is highly dependent on feature design, and designing a feature set that can accurately characterize network traffic is still an ongoing research issue. Anomaly-based IDSs also have the problem of a high false alarm rate (FAR), which seriously restricts their practical applications. In this paper, we propose a novel IDS called the hierarchical spatial-temporal features-based intrusion detection system (HAST-IDS), which first learns the low-level spatial features of network traffic using deep convolutional neural networks (CNNs) and then learns high-level temporal features using long short-term memory networks. The entire process of feature learning is completed by the deep neural networks automatically; no feature engineering techniques are required. The automatically learned traffic features effectively reduce the FAR. The standard DARPA1998 and ISCX2012 data sets are used to evaluate the performance of the proposed system. The experimental results show that the HAST-IDS outperforms other published approaches in terms of accuracy, detection rate, and FAR, which successfully demonstrates its effectiveness in both feature learning and FAR reduction.

2.3 Anomaly-based intrusion detection from network flow features using variational autoencoder:

The rapid increase in network traffic has recently led to the importance of flow-based intrusion detection systems processing a small amount of traffic data. Furthermore, anomaly-based methods, which can identify unknown attacks are also integrated into these systems. In this study, the focus is concentrated on the detection of anomalous network traffic (or intrusions) from flow-based data using unsupervised deep learning methods with semi-supervised learning approach. More specifically, Autoencoder and Variational Autoencoder methods were employed to identify unknown attacks using flow features. In the experiments carried out, the flow-based features extracted out of network traffic data, including typical and different types of attacks, were used. The Receiver Operating Characteristics (ROC) and the area under ROC curve, resulting from these methods were calculated and compared with One-Class Support Vector Machine. The ROC curves were examined in detail to analyze the performance of the methods in various threshold values. The experimental results show that Variational Autoencoder performs, for the most part, better than Autoencoder and One-Class Support Vector Machine.

2.4 Improved Techniques for Training GANs:

We present a variety of new architectural features and training procedures that we apply to the generative adversarial networks (GANs) framework. We focus on two applications of GANs: semi-supervised learning, and the generation of images that humans find visually realistic. Unlike most work on generative models, our primary goal is not to train a model that assigns high likelihood to test data, nor do we require the model to be able to learn well without using any labels. Using our new techniques, we achieve state-of-the-art results in semi-supervised classification on MNIST, CIFAR-10 and SVHN. The generated images are of high quality as confirmed by a visual Turing test: our model generates MNIST samples that humans cannot distinguish from real data, and CIFAR-10 samples that yield a human error rate of 21.3%. We also present ImageNet samples with unprecedented resolution and show that our methods enable the model to learn recognizable features of ImageNet classes.

2.5 Semi-supervised learning with generative adversarial networks:

We extend Generative Adversarial Networks (GANs) to the semi-supervised context by forcing the discriminator network to output class labels. We train a generative model G and a discriminator D on a dataset with inputs belonging to one of N classes. At training time, D is made to predict which of N+1 classes the input belongs to, where an extra class is added to correspond to the outputs of G. We show that this method can be used to create a more data-efficient classifier and that it allows for generating higher quality samples than a regular GAN.

2.6 Unsupervised representation learning with deep convolutional generative adversarial networks:

In recent years, supervised learning with convolutional networks (CNNs) has seen huge adoption in computer vision applications. Comparatively, unsupervised learning with CNNs has received less attention. In this work we hope to help bridge the gap between the success of CNNs for supervised learning and unsupervised learning. We introduce a class of CNNs called deep convolutional generative adversarial networks (DCGANs), that have certain architectural constraints, and demonstrate that they are a strong candidate for unsupervised learning. Training on various image datasets, we show convincing evidence that our deep convolutional adversarial pair learns a hierarchy of representations from object parts to scenes in both the generator and discriminator. Additionally, we use the learned features for novel tasks - demonstrating their applicability as general image representations..

2.7: Toward generating a new intrusion detection dataset and intrusion traffic characterization:

With exponential growth in the size of computer networks and developed applications, the significant increasing of the potential damage that can be caused by launching attacks is becoming obvious. Meanwhile, Intrusion Detection Systems (IDSs) and Intrusion Prevention Systems (IPSs) are one of the most important defense tools against the sophisticated and ever-growing network attacks. Due to the lack of adequate dataset, anomaly-based approaches in intrusion detection systems are suffering from accurate deployment, analysis and evaluation. There exist a number of such datasets such as DARPA98, KDD99, ISC2012, and ADFA13 that have been used by the researchers to evaluate the performance of their proposed intrusion detection and intrusion prevention approaches. Based on our study over eleven available datasets since 1998, many such datasets are out of date and unreliable to use. Some of these datasets suffer from lack of traffic diversity and volumes, some of them do not cover the variety of attacks, while others anonymized packet information and payload which cannot reflect the current trends, or they lack feature set and metadata. This paper produces a reliable dataset that contains benign and seven common attack network flows, which meets real world criteria and is publicly avaliable. Consequently, the paper evaluates the performance of a comprehensive set of network traffic features and machine learning algorithms to indicate the best set of features for detecting the certain attack categories.

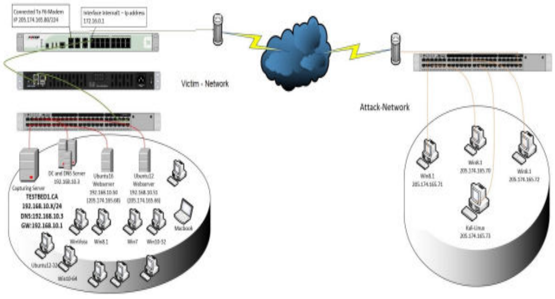


Fig.2: Testbed architecture

3. IMPLEMENTATION

Existing work [3] using semisupervised learning approach in network intrusion detection mostly utilize autoencoder. This paper suggests the use of a semi-supervised GAN (SGAN) [4] [5] in malware trafficclassification to achieve higher performance than autoencoder. With the SGAN model shown in Fig. 3, it can be trained by making the best use of numerous unlabeled data with just a few labeled data. High-performance Deep Convolutional Generative Adversarial Network (DCGAN) [6] and CNN can be applied to each operator to train data. Extensive simulations have been performed using the latest benchmark CICIDS- 2017 [7], and our model outperforms other malware traffic classification models. To the best of our knowledge, it is the first attempt to use a SGAN for malware traffic classification.

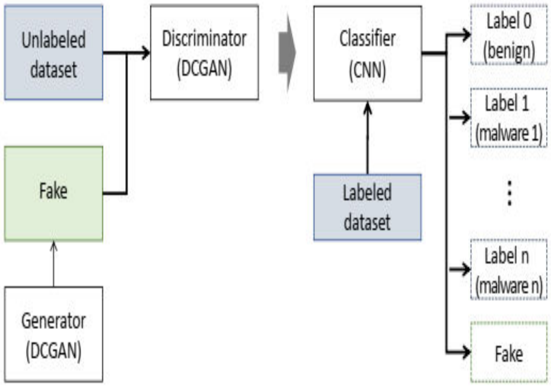


Fig.3: System architecture

EXISTING SYSTEM:

The supervised learning approach needs a large amount of labeled data. It is challenging to obtain labeled datasets due to the constraints of capturing the specific malicious flow. On the other hand, unsupervised learning approaches use only unlabeled data to perform classification tasks. However, classifying unlabeled data is a difficult task because it requires numerous datasets and significant effort, and has a lower performance. Therefore, we choose a semi-supervised approach that utilizes a few labeled data and a large amount of unlabeled data simultaneously during training.

PROPOSED SYSTEM:

This paper suggests the use of a semi-supervised GAN (SGAN) in malware traffic classification to achieve higher performance than autoencoder. With the SGAN model, it can be trained by making the best use of numerous unlabeled data with just a few labeled data. High-performance Deep Convolutional Generative Adversarial Network (DCGAN) and CNN can be applied to each operator to train data.

GAN model is composed of the generator and the discriminator. The generator is trained to deceive the discriminator by making the fake data with random noise so that it looks like real data. The discriminator is then reviewed using the input data to identify fake and real data correctly.

The original GAN model is composed of the generator and the discriminator. The generator is trained to deceive the discriminator by making the fake data with random noise so that it looks like real data. The discriminator is then reviewed using the input data to identify fake and real data correctly. SGAN [4] [5] is an advanced model that utilizes a semisupervised approach. The difference between GAN and SGAN comes from the discriminator. The discriminator in SGAN receives fake data and two kinds of real data, consisting of labeled and unlabeled data. The discriminator also produces a multi-class output to identify the correct labels in real data as well as fake data. Unlike the original GAN, whose goal is to get the generator with high performance, the target operator of the SGAN model is the trained discriminator. For the generator and discriminator of SGAN, we adopt DCGAN [6] which replaces fully connected layers of original GAN with CNN to derive more stable learning than GAN. Our semi-supervised GAN model shown in Fig. 3 has separated the discriminator and classifier. By dividing them, we can treat he discriminator and classifier as modules and each of them can be replaced or developed if necessary.

4. ALGORITHMS

GAN:

Generative Adversarial Networks, or GANs for short, are an approach to generative modeling using deep learning methods, such as convolutional neural networks. Generative modeling is an unsupervised learning task in machine learning that involves

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original dataset. GANs are a clever way of training a generative model by framing the problem as a supervised learning problem with two sub-models: the generator model that we train to generate new examples, and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated). The two models are trained together in a zero-sum game, adversarial, until the discriminator model is fooled about half the time, meaning the generator model is generating plausible examples. GANs are an exciting and rapidly changing field, delivering on the promise of generative models in their ability to generate realistic examples across a range of problem domains, most notably in image-to-image translation tasks such as translating photos of summer to winter or day to night, and in generating photorealistic photos of objects, scenes, and people that even humans cannot tell are fake.

Generative Adversarial Networks (GANs) can be broken down into three parts: Generative: To learn a generative model, which describes how data is generated in terms of a probabilistic model.

Adversarial: The training of a model is done in an adversarial setting.

Networks: Use deep neural networks as the artificial intelligence (AI) algorithms for training purpose. In GANs, there is a generator and a discriminator. The Generator generates fake samples of data(be it an image, audio, etc.) and tries to fool the Discriminator. The Discriminator, on the other hand, tries to distinguish between the real and fake samples. The Generator and the Discriminator are both Neural Networks and they both run in competition with each other in the training phase. The steps are repeated several times and in this, the Generator and Discriminator get better and better in their respective jobs after each repetition. The working can be visualized by the diagram given below:

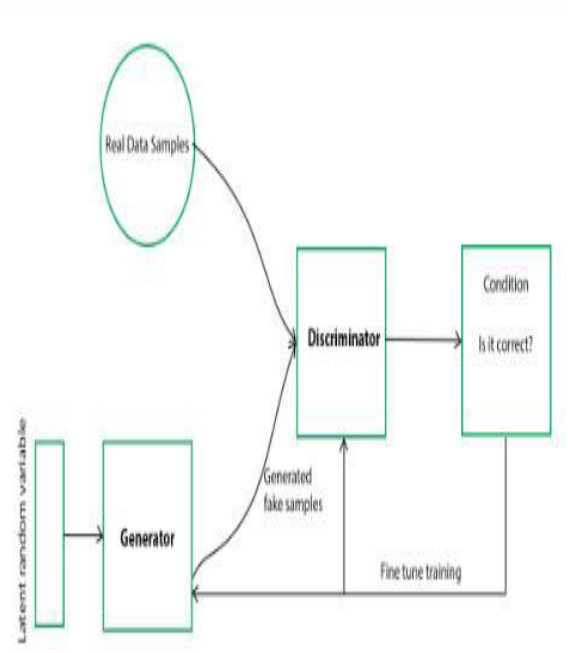
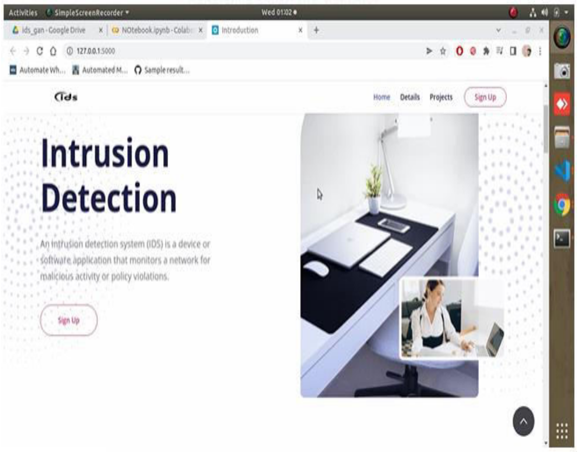


Fig.4: Working of GAN model

The generative model captures the distribution of data and is trained in such a manner that it tries to maximize the probability of the Discriminator in making a mistake. The Discriminator, on the other hand, is based on a model that estimates the probability that the sample that it got is received from the training data and not from the Generator. The GANs are formulated as a minimax game, where the Discriminator is trying to minimize its reward V(D, G) and the Generator is trying to minimize the Discriminator’s reward or in other words, maximize its loss.

5. EXPERIMENTAL RESULTS



Intrusion Detection

An intrusion detection system (IDS) is a device or software application that monitors a network for malicious activity or policy violations.

Fig.5: Output screen

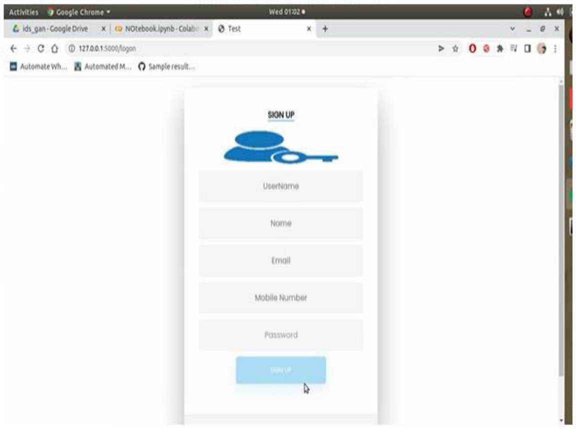
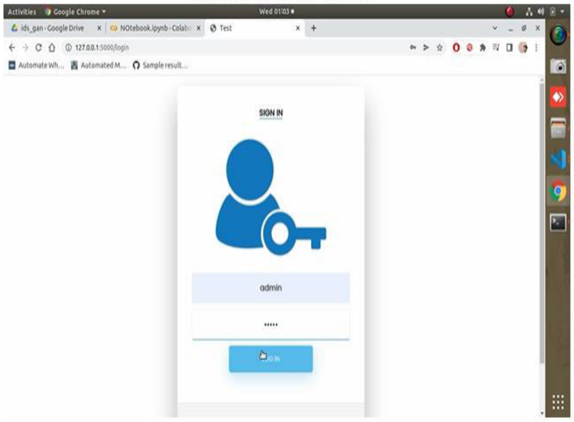


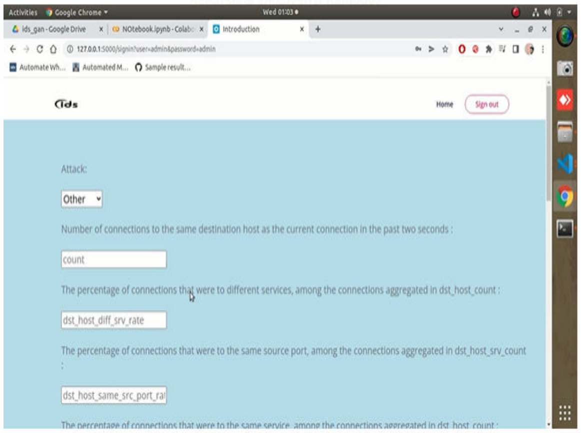
Fig.6: Signup screen



admin

•••••

Fig.7: Login screen



Attack:

Other

Number of connections to the same destination host as the current connection in the past two seconds :

count

The percentage of connections that were to different services, among the connections aggregated in dst_host_count :

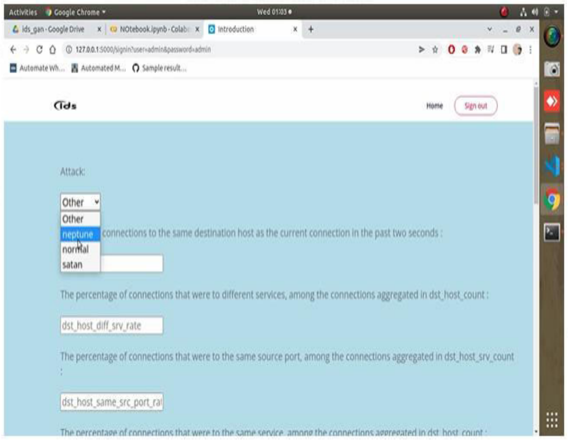
dst_host_diff_srv_rate

The percentage of connections that were to the same source port, among the connections aggregated in dst_host_srv_count :

dst_host_same_src_port_rat

The percentage of connections that were to the same service, among the connections aggregated in dst_host_count :

Fig.8: Main screen



Attack:

Other
Other
neptune
normal
satan

connections to the same destination host as the current connection in the past two seconds :

The percentage of connections that were to different services, among the connections aggregated in dst_host_count :

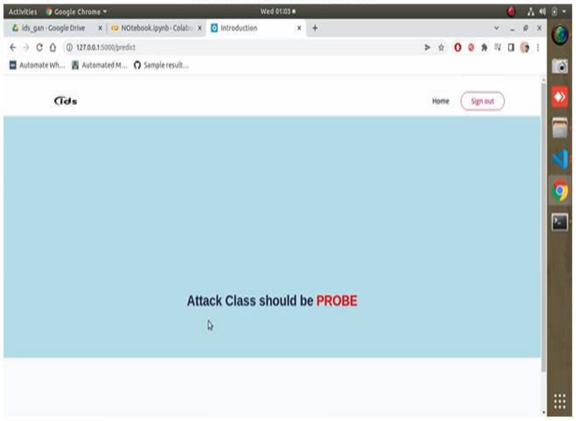
dst_host_diff_srv_rate

The percentage of connections that were to the same source port, among the connections aggregated in dst_host_srv_count :

dst_host_same_src_port_rat

The percentage of connections that were to the same service, among the connections aggregated in dst_host_count :

Fig.9: Input giving screen



Attack Class should be PROBE

Fig.10: Prediction

6. CONCLUSION

In this paper, we applied semi-supervised GAN in network intrusion detection. Our model utilizes numerous unlabeled data with the discriminator using DCGAN and only 10 labeled data per the flow type with the classifier using CNN architecture to distinguish each flow label. Without a feature extraction process, the model surpassed other models in malware traffic classification.

7. FUTURE SCOPE

The following three dimensions requires future works. Firstly, the recent work applies semi-supervised learning for the IDS using Variational Auto-Encoder (VAE) [9]. VAE assumes that the underlying distribution for the latent variables is a Gaussian while the AAE assumes that an arbitrary distribution is used the underlying distribution for the latent variables. Even though we used a mixture of Gaussian and categorical distribution in this paper, study on applying other types of distribution is left for future work, which we will work for the next step. We aim to realize the higher detection rate and lower misdetection. Optimization of the distribution imposed on the latent variable as well as the dimensionality of the latent variable, which is an important parameter in using AAE need to be investigated. Secondly, we succeeded in reducing False Negative Rate (FNR) compared with DNN, but it is not sufficient when we take practical use into account. It is necessary to capture the features of each attack and improve our method by focusing on each attack. Thirdly, we used NSL-KDD dataset. Although it is still used as a reference in recent study, it is old and not a perfect representative of existing real networks. We plant to use more recent dataset [1] to consider the latest attacks.

REFERENCES

[1] J. Zhang, F. Li, F. Ye, and H. Wu, “Autonomous unknown-application filtering and labeling for dl-based traffic classifier update,” in Proc. IEEE INFOCOM, 2020.

[2] W. Wang, Y. Sheng, J. Wang, X. Zeng, X. Ye, Y. Huang, and M. Zhu, “Hast-ids: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection,” IEEE Access, vol. 6, pp. 1792–1806, Dec. 2017.

[3] S. Zavrak and M. ˙ Iskefiyeli, “Anomaly-based intrusion detection from network flow features using variational autoencoder,” IEEE Access, vol. 8, pp. 108 346–108 358, Jun. 2020.

[4] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” in Proc. NIPS, 2016.

[5] A. Odena, “Semi-supervised learning with generative adversarial networks,” 2016. [Online]. Available: http://arxiv.org/abs/1606.01583

[6] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” in Proc. ICLR, 2016.

[7] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization,” in Proc. ICISSP, 2018.

[8] Saurabh Mukherjee and Neelam Sharma. Intrusion detection using naive bayes classifier with feature reduction. Procedia Technology, 4:119–128, 2012.

[9] Genki Osada, Kazumasa Omote, and Takashi Nishide. Network intrusion detection based on semi-supervised variational auto-encoder. In Simon N. Foley, Dieter Gollmann, and Einar Snekkenes, editors, Computer Security – ESORICS 2017, pages 344–361, Cham, 2017. Springer International Publishing.

[10] M. S. Pervez and D. M. Farid. Feature selection and intrusion classification in nsl-kdd cup 99 dataset employing svms. In The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014), pages 1–6, Dec 2014.

[11] Ngoc Tu Pham, Ernest Foo, Suriadi Suriadi, Helen Jeffrey, and Hassan Fareed M Lahza. Improving performance of intrusion detection system using ensemble methods and feature selection. In Proceedings of the Australasian Computer Science Week Multiconference, ACSW '18, pages 2:1–2:6, New York, NY, USA, 2018. ACM.

[12] Ali Shiravi, Hadi Shiravi, Mahbod Tavallaee, and Ali A Ghorbani. Toward developing a systematic approach to generate benchmark datasets for intrusion detection. computers & security, 31(3):357–374, 2012.

[13] R. Sommer and V. Paxson. Outside the closed world: On using machine learning for network intrusion detection. In 2010 IEEE Symposium on Security and Privacy, pages 305–316, May 2010.

[14] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho. Deep recurrent neural network for intrusion detection in sdn-based networks. In 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft), pages 202–206, June 2018.

[15] Tuan A. Tang, Lotfi Mhamdi, Des McLernon, Syed Ali Raza Zaidi, and Mounir Ghogho. Deep learning approach for Network Intrusion Detection in Software Defined Networking. Proceedings - 2016 International Conference on Wireless Networks and Mobile Communications, WINCOM 2016: Green Communications and Networking, pages 258– 263, 2016.