A New Approach for Detecting Network Intrusion Based on Anomalies Using a Deep Clustering Variational Auto-Encoder

Aravapalli Sai Venkata Subba Rao[1], Kommineni Maheshbabu[2], Srikanth Yadav.M[3]
[1, 2, 3] Department of IT, VFSTR Deemed to be University, Guntur, A.P., India

ABSTRACT - Semi-supervised network intrusion detection systems are becoming more vital in today's fast-evolving digital ecosystem. While increasing interest in commercial and academic contexts is rising, specific accuracy difficulties still need to be resolved. Two significant challenges contributing to this fear are accurately learning the probability distribution of standard network data and identifying the boundary between normal and abnormal data locations in the latent space. Several methods have been proposed for semi-supervised learning of the latent representation of standard data, including clustering-based Autoencoders (CAEs) and hybridized approaches combining Principal Component Analysis (PCA) and CAEs. Inadequate handling of high-dimensional data and excessive dependence on feature engineering remain limitations of current methods. To combat these problems and boost the efficiency of network intrusion detection, we introduce a novel deep learning model called Cluster Variational Autoencoder (CVAE). This approach allows for a more condensed and dominant representation of the latent space. Thanks mainly to the VAE's ability to comprehend the fundamental probability distribution of specific network data, we have broken through these barriers. The proposed model is tested on eight different network intrusion benchmark datasets. These datasets include NSL-KDD, UNSW-NB15, and CICIDS2017. Experimental findings demonstrate that our method outperforms state-of-the-art semi-supervised methods.

Keywords: Intrusion Detection, Variational Autoencoder, NSLKDD, PCA, Autoencoder

I. INTRODUCTION

As the use of the internet and other forms of communication has multiplied in recent years, so has the volume of data produced by its many associated programs and services. Numerous fields, including medicine, academia, and e-commerce, have seen a boon thanks to the advent of big data. However, as the number of IoT devices proliferates, so does the volume of network data and the number of connections between those devices. Intruders aim for this data since it often includes personal details. Detecting, preventing, and responding to cyber assaults have become more challenging as sophisticated attack methods have evolved, especially zero-day attacks. As a protective measure, a Network Intrusion Detection System (NIDS) is a great way to face these threats.

The two most common forms of NIDS detection are the Signature-based Intrusion Detection System (SIDS) and the Anomaly-based Intrusion Detection System (ADIS) (AIDS). Unlike SIDS, which can only identify previously seen assaults, AIDS may identify previously unseen attacks by establishing normative profiles of network activity. However, AIDS's effectiveness is limited by the challenge of distinguishing between typical and pathological network activity. Researchers have been using machine learning techniques to make NIDS more effective in recent years. However, these methods have a significant false alarm rate because they rely too much on feature engineering and cannot handle substantial dimensional data.

Successes with deep learning models have been shown recently in several areas, including those dealing with images, texts, voices, and autonomous vehicles. However, the use of such models in the development of anomaly intrusion detection systems (AIDS) remains in its infancy. It has not yet fully exploited the potential of deep learning in network data analysis. Researchers have employed several deep-learning models to identify network abnormalities. These include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Deep Belief Networks (DBNs), and Autoencoders (AEs). Compared to the other models, AE is the most effective. Hence it is now often employed in anomaly detection systems for networks. There are, however, two significant caveats to working with such models. First, they aren't very good at detecting abnormalities since they can't create robust profiles from regular data. Second, the latent representation space of AE makes it challenging to pinpoint the transition between normal and aberrant data areas. This is because the trained model cannot still discover the correct sample placement in the latent space. This research presents a unique deep-learning strategy that overcomes these restrictions by introducing a deep generative model on a single class of standard data. Together with a clustering method, the model learns a condensed standard feature space that helps distinguish between typical and out-of-the-ordinary data. Clustering Variational Autoencoder is the name of the suggested method, which employs the K-means clustering algorithm at the latent layer of a variational AE model (CVAE).

II. RELATED WORK

Diverse models, including AE, VAE, and CAE, have been utilized in other research on network anomaly detection, and their results are given here. The VAE model is trained using the reconstruction distribution rather than the reconstruction error, as suggested by one study's proposal for anomaly detection utilizing the VAE model's reconstruction probability. Additionally, research combined dictionary learning with VAE modeling inside a sparse representation framework to enhance the efficiency of anomaly detectors. To shed light on traffic irregularities, another research used a VAE model to extract features using a gradient-based fingerprinting approach. In another investigation, the authors tried out different AE topologies to see how they affected the efficiency with which networks detected anomalies. Last, recent research used PCA and CAE techniques to enhance network anomaly identification.

These studies aim to inform the development of a new approach that uses the clustering algorithm in tandem with a hidden layer of the VAE model. By fusing AE's latent representation learning power with the generative model features of VAE, this technique aims to develop a more compact representation of the latent probability space of the standard network data. The proposed strategy is expected to improve network anomaly detection.

An increasing number of organizations are turning to autoencoders (AEs) to spot outliers. A bottleneck layer of AEs may be utilized as a feature representation block for subsequent classifiers [3, 4, 5, 7, 10, 15, 16], making them useful both as a solo classifier and in conjunction with other classification techniques. It is possible to generate the latent feature representation using supervised, semi-supervised, or unsupervised learning techniques [3, 4, 5, 7, 10, 15, 16]. The encoder may then be employed as a feature representation block to enhance anomaly detection after the AE training. The dimensionality of the input data may be decreased, and stronger characteristics representative of typical behavior can be extracted, thanks to the bottleneck layer. Specific tasks associated with the above-mentioned latent representation strategies are discussed in this section.

To represent typical and out-of-the-ordinary data in separate parts of VAE's middle-hidden layer, developed a multi-distribution for the supervised method. The input data in traditional VAEs may be transformed into a Gaussian distribution N (0,1) at the bottleneck level. The class labels are included in the VAE loss function to split the two data classes into areas with the same Gaussian distribution shape but different mean values. Both publicly accessible network security datasets were used to assess the proposed model, which showed encouraging results.

III. PROPOSED ARCHITECTURE

In this part, we introduce the Clustering Variational Autoencoder, a novel model developed to solve the problems found in earlier research. The objective is to create an anomaly detector for networks that can operate with regular network input and only use semi-supervised learning.

Learning the underlying probability distribution of normal network data was a primary motivation for the model's creation since it would improve its capacity to spot anomalies. Based on the idea that typical network data consists of several sub-clusters, the model combines the latent representation learning of Autoencoders (AEs), the efficient clustering abilities of K-means, and the probability distribution understanding of deep generative model VAEs.

The VAE model employs the K-means clustering algorithm inside a hidden layer of the encoder to help effectively group data points. We refer to this underlying structure as the clustering layer. After the data's most concise, dominant, and fundamental aspects have been learned using an AE-based model, the VAE model is used to nudge the points back toward the original distribution. Figure 1 depicts the overall design of the proposed paradigm.

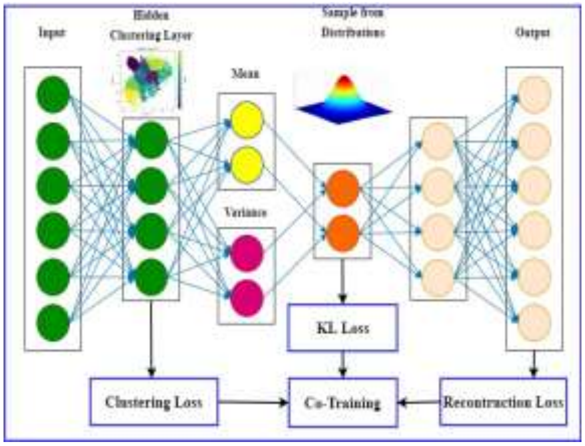

Fig. 1. Proposed Model

IV. RESULTS AND OBSERVATION

Our suggested model has been tested on eight benchmark datasets, and the results are shown here. The experimental findings reveal that across all eight benchmark datasets, the model consistently outperforms the DAE and VAE baseline models and the CAE and PCA+CAE state-of-the-art models. Table II summarizes the results of the AUC scores used to estimate, assess, evaluate, and compare the trained model to the baseline and state-of-the-art models.

On the NSL-KDD dataset, the suggested model outperformed the DAE and VAE baseline models (accuracy of 0.723 and 0.646, respectively) and the CAE and PCA+CAE models (accuracy of 0.963 and 0.966, respectively). The studies on the UNSWNB15 dataset also show that the model outperforms the others, with a remarkable result of 0.897 compared to the accuracy of 0.583, 0.528, 0.804, and 0.855 produced by the DAE, VAE, CAE, and PCA+CAE models, respectively.

The experimental findings show that the proposed model outperforms the VAE and CAE models across all tested datasets. Similarly, the model scored 0.846 on the CICIDS2017 dataset, whereas the best score among the current models was just 0.830. Experiments on five situations from the CTU13 dataset further validated the model's capabilities.

This is because the objective function's three parts worked together more effectively than the aim of the VAE function's two pieces (reconstruction loss and KL-divergence) and the CAE objective function's combination (reconstruction loss and clustering loss). Because of this, it can now identify abnormal samples with greater accuracy thanks to extracting more potent and noteworthy latent characteristics of standard data.

Table 1. Performance comparison of the proposed model

Model	**Datasets**		
	NSL-KDD	**UNSW-NB15**	**CICIDS 2017**
DAE	0.72	0.58	0.71
VAE	0.65	0.53	0.66
CAE	0.96	0.80	0.81
PCA+CAE	0.97	0.86	0.83
CVAE	0.97	0.90	0.85
CVAE	0.97	0.90	0.85

Three benchmark datasets, NSL-KDD, UNSW-NB15, and CICIDS2017, are summarized in the table below. Different types of DAES, VAE, CAE, PCA+CAE, and models are considered. The AUC score is used as a statistic for assessment.

The AUC values for the DAE and VAE models on the NSL-KDD dataset are 0.72 and 0.65, while those for the CAE and PCA+CAE models are 0.96 and 0.97. The score of 0.97 for the suggested model is considerably higher.

Both the DAE and VAE models perform poorly on the UNSW-NB15 dataset, scoring 0.58 and 0.53, respectively. With scores of 0.80 and 0.86, respectively, the CAE and PCA+CAE models fall short of the suggested model's 0.90. The CICIDS2017 dataset is best modeled by CAE, which scores 0.81, followed by PCA+CAE, which scores 0.83. With an overall score of 0.85, the suggested model outperforms the best-competing models.

Experimental findings reveal that the suggested model consistently achieves the highest AUC values across all three datasets.

V. CONCLUSION

This research proposes a novel deep clustering variational Autoencoder model to construct a semi-supervised anomaly-based NIDS. By understanding the underlying probability distribution of average network data and re-arranging the usual data points to identify the border between normal and anomalous data areas more precisely, this model hopes to overcome the limitations of earlier approaches. The proposed model combines the strength of deep AE for learning the most important aspects of the data with the capability of the VAE for learning the most latent and descriptive elements of the normal data. To further facilitate proper data point organization, the K-means clustering algorithm is embedded in a hidden layer of the encoder. The proposed technique outperforms prior methods on all chosen datasets, as shown by experimental results on eight standard benchmark datasets, including NSL-KDD, UNSW-NB15, CICIDS2017, and five scenarios in CTU13. The research will be expanded to investigate more deep generative models and do trials on additional state-of-the-art data sets.

VI. REFERENCES

[1]. K. Xie et al., "Fast tensor factorization for accurate Internet anomaly detection," IEEE/ACM Trans. Netw., vol. 25, no. 6, pp. 3794–3807, Dec. 2017.

[2]. Aishah Abdullah, Reem Hamad, Mada Abdulrahman, Hanan Moala, and Salim Elkhediri. Cybersecurity: A review of internet of things (iot) security issues, challenges, and techniques. In 2019 2nd International Conference on Computer Applications Information Security (ICCAIS), pages 1–6, 2019.

[3]. Ly Vu, Van Loi Cao, Quang Uy Nguyen, Diep N. Nguyen, Dinh Thai Hoang, and Eryk Dutkiewicz. Learning latent representation for iot anomaly detection. IEEE Transactions on Cybernetics, 52(5):3769– 3782, 2022.

[4]. Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. ACM Computing Surveys (CSUR), 54(2):1–38, 2021.

[5]. Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, and Farhan Ahmad. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Transactions on Emerging Telecommunications Technologies, 32(1):e4150, 2021.

[6]. Dylan Chou and Meng Jiang. A survey on data-driven network intrusion detection. ACM Computing Surveys (CSUR), 54(9):1–36, 2021.

[7]. Sharmila Kishor Wagh, Vinod K Pachghare, and Satish R Kolhe. Survey on intrusion detection system using machine learning techniques. International Journal of Computer Applications, 78(16), 2013.

[8]. Geeta Singh and Neelu Khare. A survey of intrusion detection from the perspective of intrusion datasets and machine learning techniques. International Journal of Computers and Applications, pages 1–11, 2021.

[9]. Shi Dong, PingWang, and Khushnood Abbas. A survey on deep learning and its applications. Computer Science Review, 40:100379, 2021.

[10]. Arwa Aldweesh, Abdelouahid Derhab, and Ahmed Z Emam. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. Knowledge-Based Systems, 189:105124, 2020.

[11]. Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407, 2019.

[12]. Ly Vu, Quang Uy Nguyen, Diep N. Nguyen, Dinh Thai Hoang, and Eryk Dutkiewicz. Deep generative learning models for cloud intrusion detection systems. IEEE Transactions on Cybernetics, pages 1–13, 2022.

[13]. Van Quan Nguyen, Viet Hung Nguyen, Nhien-An Le-Khac, and Van Loi Cao. Clustering-based deep autoencoders for network anomaly detection. In International Conference on Future Data and Security Engineering, pages 290–303. Springer, 2020.

[14]. Van Quan Nguyen, Viet Hung Nguyen, Nhien-An Le Khac, Nathan Shone, et al. A robust pca feature selection to assist deep clustering autoencoder-based network anomaly detection. In 2021 8th NAFOSTED Conference on Information and Computer Science (NICS), pages 335–341. IEEE, 2021.

[15]. Diederik P Kingma and Max Welling. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013.

[16]. Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In International conference on machine learning, pages 1278–1286. PMLR, 2014.

[17]. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[18]. Stuart Lloyd. Least squares quantization in pcm. IEEE transactions on information theory, 28(2):129–137, 1982.

[19]. Christopher M Bishop and Nasser M Nasrabadi. Pattern recognition and machine learning, volume 4. Springer, 2006.

[20]. Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. Special Lecture on IE, 2(1):1– 18, 2015.

[21]. Jiayu Sun, Xinzhou Wang, Naixue Xiong, and Jie Shao. Learning sparse representation with variational auto-encoder for anomaly detection. IEEE Access, 6:33353–33361, 2018.

[22]. Quoc Phong Nguyen, Kar Wai Lim, Dinil Mon Divakaran, Kian Hsiang Low, and Mun Choon Chan. Gee: A gradient-based explainable variational autoencoder for network anomaly detection. In 2019 IEEE Conference on Communications and Network Security (CNS), pages 91–99. IEEE, 2019.

[23]. Wen Xu, Julian Jang-Jaccard, Amardeep Singh, Yuanyuan Wei, and Fariza Sabrina. Improving performance of autoencoder-based network anomaly detection on nsl-kdd dataset. IEEE Access, 9:140136–140146, 2021.

[24]. Van Quan Nguyen, Viet Hung Nguyen, Nhien-An Le Khac, et al. Automatically estimate clusters in autoencoder-based clustering model for anomaly detection. In 2021 RIVF International Conference on Computing and Communication Technologies (RIVF), pages 1–6. IEEE, 2021.

[25]. Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A Ghorbani. A detailed analysis of the kdd cup 99 data set. In 2009 IEEE symposium on computational intelligence for security and defense applications, pages 1–6. Ieee, 2009.

[26]. Nour Moustafa and Jill Slay. Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). In 2015 military communications and information systems conference (MilCIS), pages 1–6. IEEE, 2015.

[27]. Iman Sharafaldin, Arash Habibi Lashkari, and Ali A Ghorbani. Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSp, 1:108–116, 2018.

[28]. Sebastian Garcia, Martin Grill, Jan Stiborek, and Alejandro Zunino. An empirical comparison of botnet detection methods. computers & security, 45:100–123, 2014.