International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Enhancing DDoS Attack Detection in SDNs with GAN-Based Imbalanced Data Augmentation

Konda Srikar Goud[1*, 3] , Srinivasa Rao Gidutur[2]
[1*]Department of CSE,
GITAM School of Technology, GITAM University,
Andhra Pradesh, INDIA.
kondasrikargoud@gmail.com
[2]Department of CSE,
GITAM School of Technology, GITAM University,
Andhra Pradesh, INDIA.
sgidutur@gitam.edu
[3]Department of Information Technology,
BVRIT HYDERABAD College of Engineering for Women,
Hyderabad, INDIA.
kondasrikargoud@gmail.com

Abstract—Securing computer networks has become crucial due to the ongoing emergence of diverse network attacks. The popularity of Software Defined Networks (SDN) has risen because of its ability to enhance network agility, efficiency, and adaptability to recent networking challenges. However, it is essential to note that SDNs, which depend on centralized controllers, can be severely affected by Distributed Denial of Service (DDoS) attacks. The threat of DDoS attacks has grown exponentially, resulting in the evolution of robust Machine Learning-based DDoS attack detection systems within SDN. DDoS attack detection systems may deliver poor performance when trained on imbalanced datasets. Traditional techniques for handling imbalanced datasets need to be revised. Recent advances in generative adversarial networks (GANs) have revealed significant potential in generating synthetic data while preserving the probability distribution of the original data. This innovative procedure offers a promising solution to mitigate the challenges of imbalanced data in DDoS attack detection. To address challenges originating from imbalanced training datasets, we employed Generative Adversarial models to generate adversarial attacks from one viewpoint and evaluate their quality from another perspective. We chose Generative Adversarial Networks (GANs), Bidirectional GANs (Bi-GANs), and Wasserstein GANs (WGANs) based on extensive usage and reliability criteria in various domains. We conducted a comprehensive assessment to evaluate their effectiveness and resilience in generating high-quality attacks. It helps to develop, train, and fine-tune machine and deep learning models to estimate their impacts. We utilized NSL-KDD and CICIDS-2017 datasets to ensure generalization, implementing both ML and DL approaches. The outcomes demonstrate that the WGAN model outperformed GAN, Bi-GAN, and the models trained on the original imbalanced dataset and traditional sampling techniques in binary and multiclass classifications for both datasets.

Keywords-SDN; DDoS; GAN; Imbalanced dataset; Wasserstein GAN

I. INTRODUCTION

In the ever-evolving landscape of network security, Software-Defined Networking (SDN) has brought tremendous advantages and unique challenges. SDN's centralized control and dynamic adaptability have significantly enhanced network management and efficiency. However, these benefits come with an increased vulnerability to malicious threats, particularly DDoS attacks. DDoS attacks pose a grave risk to the availability and performance of network services. They involve massive malicious traffic intended to overwhelm network resources and disrupt normal operations. In SDN, where centralized controllers play a pivotal role in network decision-making, DDoS attacks can directly impact the controller, potentially leading to severe network-wide congestion and service degradation.

In response to this critical security concern, developing effective DDoS attack detection systems tailored for SDN environments has become imperative. These systems leverage Machine Learning (ML) and Deep Learning (DL) techniques to monitor network traffic patterns, identify anomalies, and swiftly mitigate DDoS threats. This integration of advanced technologies promises to fortify SDN networks against the escalating menace of DDoS attacks, ensuring modern network infrastructures' continued reliability and security.

However, the practical implementation of machine learning-based models in the real world poses unique challenges. A key challenge lies in acquiring an extensive dataset encompassing attack and normal data samples, an essential prerequisite for training robust detection models. However, the collection of attack data, characterized by its scarcity and associated costs, starkly contrasts the relative abundance of normal data. Consequently, this imbalance in data distribution poses a formidable hurdle, impacting the accuracy and effectiveness of machine learning-based intrusion detection.

This research paper introduces an innovative solution to tackle the pervasive issue of imbalanced datasets in IDS. This approach pioneers the generation of synthetic attack data by employing the proficiencies of Generative Adversarial Networks (GANs). These artificially crafted attack instances are seamlessly integrated with the original dataset, forming an augmented training dataset. We train various Machine and Deep Learning classifiers on this augmented dataset. The research presents comprehensive experimental findings spanning datasets, including NSL-KDD, and CICIDS2017 dataset. The

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

results unequivocally demonstrate that the MDL algorithms exhibit superior performance when trained on the augmented dataset generated by GANs, surpassing the performance of models trained on the datasets generated through SMOTE technique and the original dataset.

This research addresses the pervasive challenge of imbalanced datasets in intrusion detection with an innovative solution grounded in Generative Adversarial Networks. By augmenting the machine learning-based IDS with synthetically generated attack data, this approach seeks to elevate the precision and efficacy of intrusion detection, thus contributing significantly to the enhancement of network security in a rapidly evolving digital landscape.

The primary objective of this work lies in demonstrating the effectiveness of Generative Adversarial Networks (GANs) in dealing with the pervasive issue of imbalanced datasets in Intrusion Detection System (IDS) datasets. In this paper, we present an enhanced version of using GANs to generate synthetic data and thoroughly examine their performance across a diverse range of benchmark datasets. The article is organized as follows to provide a complete understanding of our procedure. Section II briefly examines earlier research on ML applications detecting DDoS attacks and examines existing techniques for dealing with imbalanced datasets, which form the background for our novel approach. Section III delivers the materials and methods for developing the proposed method. Section IV furnishes a detailed illustration of our approach, explaining how GANs are employed to generate synthetic attack data and seamlessly incorporate it with the original dataset. Section V provides the results of our comprehensive experimentation. We examine the performance of our approach on various network attack datasets and thoroughly investigate the results. This section offers valuable insights into the effectiveness of GANs in enhancing the model's performance. Lastly, Section VI concludes our results and examines potential avenues for future research. Overall, this paper offers a comprehensive exploration of the use of GANs to address imbalanced datasets in detecting DDoS arracks, focusing on enhancing network security in an ever-evolving digital landscape.

II. RELATED WORKS

The significance of a DDoS attack detection system depends on several aspects. They are a) the number of features utilized. b) the characteristics of data within each feature, c) the number of samples used for attack training and testing, and d) the selection of a classifier (machine learning or deep learning). Generative adversarial models can enhance the performance of DDoS attack detection by providing improved training data.

The authors in [1] introduced the Deep Convolutional GAN (DCGAN), which utilizes a Long Short-Term Memory (LSTM) algorithm to generate adversarial attacks from original data. The DCGAN technique is employed to extract accurate features, reducing false detections. For feature selection, a Simple Recurrent Unit algorithm is applied for feature extraction from original data, followed by LSTM execution for real-time attack detection. In the same year, the authors [2] employed Auto Encoders (AE) with GANs to address data imbalance issues and improve anomaly detection using the Random Forest (RF) algorithm. Three models were considered: RF, hybrid of Auto Encoders with RF, and a hybrid of Auto Encoders with Conditional GAN, the AE-RF model utilizes AE to reduce data dimensionality for learning and RF for detection. AE-CGAN incorporates CGAN after AE feature extraction, achieving better performance than Single-RF and AE-RF. However, this work was applied to only 20% of the NSL-KDD training set.

In 2020, a new model was introduced based on AE and statistical analysis techniques for AIDS [3]. Pre-processing includes outlier analysis using the Median Absolute Deviation Estimator method, min-max normalization, and one-hot encoding for non-numeric features. The authors in [4] proposed a model that combines Conditional Wasserstein GANs (CWGAN) and Cost-Sensitive Stacked Auto-Encoders (CSSAE). The CWGAN component was used to obtain more samples representing rare attacks, while the CSSAE component extracted relevant features. This model effectively addressed imbalanced data, especially for rare attacks. It introduced a cost function that heavily penalized minor classes, resulting in high accuracy for detecting minority attacks. In the same year, the authors in [5] developed various classifiers for an IDS using a machine learning approach. They applied AE and PCA for dimensionality reduction and utilized the Uniform Distribution Based Balancing approach to address data imbalance.

In 2020, the authors in [6] proposed reducing loss function in GANs through supervised learning, improving AIDS performance for various classifiers, including RF, SVM, KNN, and ANN. During the same year, the authors in [7] introduced the Monte Carlo tree search algorithm to obtain more samples for cross-site scripting (XSS) attacks using GANs. GANs were also applied for detecting adversarial attacks, significantly improving AIDS performance. In 2023, the authors in [8] suggested using GANs in IDS to enhance attack detection on the NSL-KDD dataset. GANs were applied to synthesize instances, improving attack detection with ML classifiers, including KNN, Decision Tree (DT), RF, SVM, and ANN. However, the study did not apply the official NSL-KDD split and used different evaluation metrics, emphasizing Recall and F1 score. The author in [9] employed Weighted Support Vector machine to handle class imbalance by assigning highest weights to the minority class which allows the model to give more importance to those samples during training and reduce bias. Much of the research has focused on detecting DDoS attacks with various architectures employing machine learning and deep learning [10]. Deep learning has consistently demonstrated improved performance in terms of detection rates and F1 scores. However, no study has evaluated the effectiveness of different GAN-based AIDS for binary and multi-class classification on NSL-KDD and CICIDS-2017 datasets. Additionally, many studies have focused on subsets of the dataset, leading to bias and potentially misleading results.

The contributions of our research are as follows:

- Identifying significant features related to label classes for building a robust model using a hybrid of Information Gain and Random Forest Feature Importance feature selection techniques, a novel approach.
- Developing and utilizing generative adversarial models (GAN, BiGAN, and WGAN) for generating realistic adversarial attacks, considering various parameter values.
- Providing a comparative analysis of Original, SMOTE, GAN, BiGAN, and WGAN models and their impact on DDoS attack detection model performance across classification types and DML approaches (ML and DL).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

- Identifying the most beneficial generative adversarial models for diverse attack classes using cross-datasets and domain adaptation to enhance generalization and robustness.

III. MATERIALS AND METHODS

A. *Generative Adversarial Network*

The Generative Adversarial Network (GAN) is a modern and potent DL technique for creating artificial data. GAN comprises two neural networks i.e., a Generator (G) and a Discriminator (D). In this adversarial framework, the generator takes random noise as input and generates synthetic data samples, while the discriminator receives two kinds of input: genuine data samples from the training dataset and synthetic samples created by the generator [11]. The primary goal of the discriminator is to differentiate between real and synthetic data. These two networks engage in an ongoing training process, with the generator learning to produce increasingly realistic samples, and the discriminator striving to enhance its ability to discriminate among genuine and generated data. This competitive learning dynamic often results in generated samples that are hard to distinguish from real ones.

From a mathematical standpoint, if we denote the random noise as 'z' and the generator's output as 'X_fake = G(z),' the discriminator (D) accepts either a real data sample 'x' or a generated sample 'G(z)' as input and generates a value representing the likelihood of the input being real data. In other words, the discriminator is trained to maximize the probability of correctly identifying real data and minimize the probability of incorrectly classifying generated data. This training objective is captured by the loss function outlined in Equation (1):

$$L = E[\log D(x)] + E\left[\log\left(1 - D\left(G(z)\right)\right)\right] \quad (1)$$

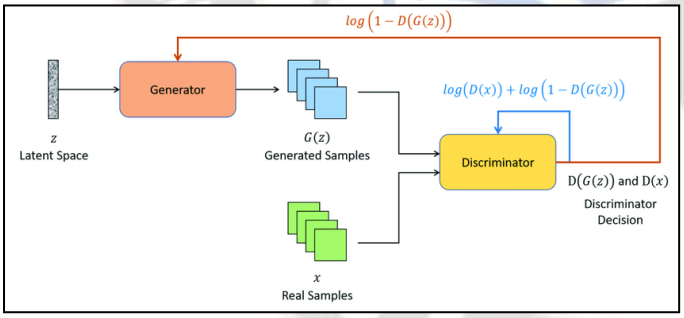

Figure 1. Architecture of Generative Adversarial Networks

One significant advantage of GANs is their completely unsupervised nature, enabling them to be trained without requiring labeled data. This unsupervised training can then be utilized to use the GAN's generator and discriminator as feature extractors for various supervised responsibilities. Nevertheless, it's crucial to recognize that GANs can be challenging to train due to their inherent instability. The adversarial training process entails the training of two networks opposing each other, which is a departure from the typical single back-propagation employed in neural networks. As a result, GANs may generate outputs that may seem illogical or unrealistic. Figure 1 illustrates the framework of the GAN.

B. *Wasserstein Generative Adversarial Network*

Wasserstein Generative Adversarial Network (WGAN) is a significant advancement in the field of generative models, aimed at addressing the training stability challenges that have plagued traditional Generative Adversarial Networks (GANs). Unlike the standard GANs, WGAN introduces the Wasserstein distance, also known as the Earth Mover's distance, as the primary metric for quantifying the dissimilarity between probability distributions. This shift from metrics like the JSD or KL divergence results in remarkable improvements in training stability, a critical concern in vanilla GANs.

The Wasserstein distance plays a crucial role in WGAN by seeking the optimal mapping between samples from two distributions: the original data distribution and the synthetic data distribution. This mapping minimizes the overall cost and effectively measures the discrepancy between the distributions. As a result, the loss function for WGAN is redefined in Equation (2), reflecting this innovative approach [12, 13].

$$L(p_r, p_g) = W(p_r, p_g) = max_{w \in W} E_{x \sim p_r}[f_w(x)] - E_{z \sim p_{r(z)}}[f_w(g_\theta(z))] \quad (2)$$

The choice of the Wasserstein distance over JSD and KL-divergence is driven by its ability to handle overlapping distributions seamlessly. Unlike its counterparts, the Wasserstein distance maintains its properties even when the gradient descent value (θ) equals zero, resulting in a zero distance. In contrast, KL and JSD divergences become unbounded (infinity) and non-differentiable under similar conditions. This intrinsic stability and favorable gradient behavior make WGAN a powerful tool for various generative tasks, particularly in the context of training generative models with improved ease and reliability.

C. *Bi-Directional Generative Adversarial Network*

A BiGAN is a generative model that evolves from the standard GAN [14]. The structure of a BiGAN is akin to that of a GAN, with an additional Autoencoder (AE) component. This AE comprises an Encoder (E) and a Decoder (D). The Encoder takes high-dimensional input and transforms it into a lower-dimensional (latent) space, represented as E(input). Subsequently, the Decoder reconstructs the encoded data to produce data resembling the original input.

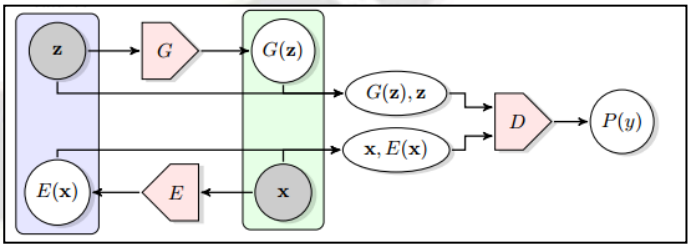

Figure 2. Architecture of Bi-Directional Generative Adversarial Networks

Within a BiGAN, the Generator (G) and Encoder (E) collaborate closely. The Encoder compresses real data into a latent space represented as 'z,' and the Generator then reconstructs this encoded data, generating data that closely resembles the original input, denoted as 'x.' The inclusion of an Autoencoder in the GAN serves to mitigate the issue of mode collapse, where the model might focus on learning from one

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

class while ignoring the others. As a result, the Autoencoder enhances the stability of a GAN by linking real data to the latent space 'z.' Moreover, this leads to a more abstract representation of the data, making it less susceptible to minor modifications, and the GAN's generator can produce high-quality, novel samples.

In contrast to the GAN's generator, which learns from the original data distribution, a BiGAN learns from both the original data and the latent space (Z). However, the role of the discriminator in a BiGAN differs somewhat from that in a vanilla GAN. The discriminator in a BiGAN distinguishes between fake samples and genuine data by assessing a joint probability distribution. Consequently, it discriminates between (G(z); z) and (x; E(x)) and aims to maximize the MiniMax objective function across three components: the discriminator, encoding, and generator. The optimization process for this model is akin to that of the standard GAN but includes the joint distribution for both the latent space (z) and the original data (x). Figure 2 demonstrates the architecture of the BiGAN.

IV. PROPOSED METHODOLOGY

This section outlines the comprehensive framework for developing an efficient DDoS Detection System, comprising seven essential steps. The process commences in step one by acquiring raw input records from the NSL - KDD and CICIDS - 2017 datasets. Subsequently, in step two, the dataset undergoes necessary preparation to facilitate further processing and eventual input to generative models. In step three, feature selection is conducted to identify and retain the most relevant attributes. Step four involves incorporating additional rare attack data through standard generative adversarial models to enhance the dataset's ability to detect rare attacks. Step five encompasses establishing the proposed model architecture, combining both ML and DL techniques, followed by training and optimization to determine the most effective hyper parameter settings. In step six, the model's performance is rigorously assessed based on a training dataset, evaluating its ability to detect known attacks effectively. In step 7, the model is tested against formerly unrecognized (0 - day) attacks, allowing for evaluating its performance under real-world circumstances. Figure 3 visually depicts these sequential steps, illustrating the order followed to build a robust DDoS detection system.

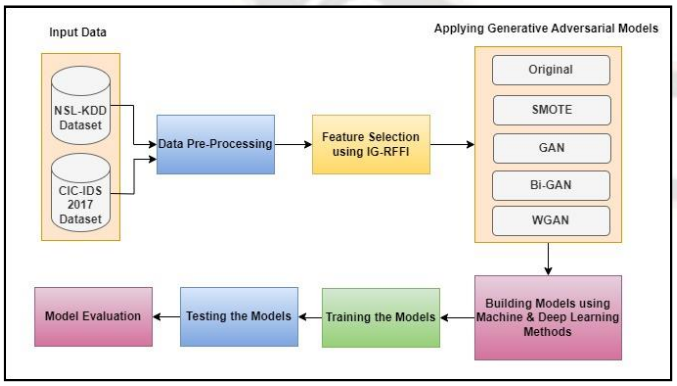

Figure 3. Architecture of the Proposed Framework

A. *Datasets*

In this study, selecting datasets is vital in assessing the effectiveness of the DDoS attack detection system against generative adversarial attacks. We employed the NSL - KDD [15] and CICIDS - 2017 [16] datasets due to their importance in DDoS attack detection and applicability to generate generative adversarial attack traffic. These datasets are widely determined for providing realistic network traffic data, encompassing various attack classes and types. The NSL - KDD and CICIDS - 2017 datasets are known for their high diversity and complexity, guaranteeing robust, reliable, and generalized results. They also function as valuable benchmarks for evaluating the ability of generative models like GANs, BiGAN, and WGAN to produce adversarial attacks by training on diverse and rare attacks, yielding high-quality outcomes to detect DDoS attacks in SDN.

The NSL - KDD dataset is derived from the KDD CUP 99 dataset [17], which contains 15,000 records, ignoring redundancy and reducing training time. This dataset presents a comprehensive evaluation of a model compared to the original KDD CUP 99 dataset. Figure 4 outlines the distribution of attacks across classifications within the NSL - KDD dataset. The NSL - KDD dataset delivers an exciting characteristic where the training set includes attacks not present in the testing set and vice versa. Additionally, the User to Root (U2R) and Root to Local (R2L) attack classes contain minimal samples, posing challenges to the model performance in these types.

On the other hand, the CICIDS - 2017 dataset is a valuable resource for IDS research as it reflects real-world attacks and encompasses various protocols and attack categories. It includes Brute Force Attacks, Heartbleed Attacks, Botnets, Denial of Service (DoS) Attacks, Distributed Denial of Service (DDoS) Attacks, Web Attacks, and Infiltration Attacks. This dataset was captured over five days, featuring network traffic with diverse protocols and attack scenarios.

One noteworthy aspect of the CICIDS - 2017 dataset is the class imbalance, with approximately 80% representing benign traffic and only 0.1% comprising rare attacks like Infiltration, Heartbleed, and specific Web Attacks. This imbalance presents a significant challenge for IDSs, particularly in accurately detecting and classifying rare and adversarial attacks.

Overall, the choice of these datasets ensures the rigorous evaluation of IDSs and the effectiveness of generative adversarial models in producing high-quality adversarial attacks for enhancing the capabilities of IDSs in handling real-world network security threats. Figure 4 shows the class distribution of the CICIDS 2017 dataset.

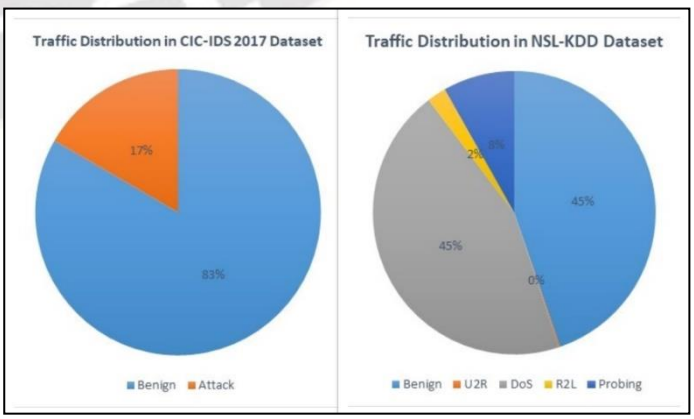

Figure 4. Traffic distribution in NSL-KDD and CIC-IDS 2017 Datasets.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

B. Data pre-processing and feature extraction

Data pre-processing and feature extraction are essential in preparing the NSL-KDD and CICIDS-2017 datasets for DDoS attack detection model and generative adversarial attack generation. This process involves several key steps to ensure data quality and relevance.

In the initial step, benign records are labelled as "one," while attack records are labelled as "zero," establishing a binary classification scheme that distinguishes normal from adversarial network traffic. The second step addresses missing values within both datasets by employing median imputation. This strategy ensures that incomplete data does not impede the subsequent analysis and modelling efforts. The third step focuses on enhancing computational efficiency by identifying and removing quasi-constant features. These are features with values that remain the same across most data records, and eliminating them streamlines the model learning process. In this case, twelve quasi-constant features were dropped. Numeric data undergoes scaling in the fourth step using the Standard Scaler. Scaling ensures that data values are on a consistent scale, a prerequisite for many machine learning algorithms to perform optimally. Non-numeric or categorical features are handled in the fifth step using One-Hot Encoding. This conversion transforms categorical attributes into a numerical format that can be effectively utilized by machine learning models.

TABLE I. TOP FEATURES WITH HIGHEST IMPORTANCE SCORES FROM NSL-KDD AND CIC-IDS 2017 DATASETS.

NSL-KDD Dataset		**CIC-IDS 2017 Dataset**
Duration	num file creations	DestinationPort
source bytes	num shells	FlowDuration
destination bytes	num access files	BwdPacketLengthMin
wrong fragment	count	FlowIATMean
urgent	srv count	FlowIATMax
hot	dst host count	FwdIATMean
number failed logins	dst host srv count	FwdPackets/s
num compromised	level	BwdPackets/s
su attempted		MinPacketLength
num root,		min_seg_size_forward

Feature selection, the final step, is a critical phase to identify the most relevant attributes for both DDoS attack detection and generative adversarial models. Two distinct techniques, Information Gain (IG) [18] and Random Forest Feature Importance (RFFI) [19], are employed for this purpose. IG is a filter-based approach that quantifies attribute relevance by calculating entropy. Features with higher IG values are deemed more relevant, and these top-ranking features are selected based on their ability to distinguish between normal and DDoS traffic. On the other hand, RFFI is an embedded feature selection technique known for its accuracy. It involves training a Random Forest classifier with the selected features and evaluating each feature's contribution to predictive accuracy. The top features with the highest importance scores are then chosen for further modelling and analysis. The table 1 presents the top features with highest feature importance scores from both the datasets.

By meticulously following these data pre-processing and feature extraction steps, the NSL-KDD and CICIDS-2017 datasets are optimized for effective use in DDoS attack detection model evaluation and generating high-quality adversarial attacks, ultimately bolstering network security systems' capabilities.

C. Apply generative adversarial models

Assessing the generative adversarial models' performance involves a crucial step of testing our model against unseen attacks. To ensure the robustness of the model, synthetic adversarial samples for all attack categories are combined with the training data. Importantly, the entire dataset is used for constructing the model instead of relying on a subset. For the NSL-KDD dataset, the official split is applied, while no official split exists for CICIDS-2017. Nonetheless, an equal number of samples, 70,343, is utilized for each attack category in both datasets, promoting balanced evaluation.

The process of obtaining adversarial attacks using GAN, BiGAN, and WGAN models consists of several key steps:

- Select the training sample set (either NSL-KDD or CICIDS-2017) as input for the generative model.
- Apply the generator (G) to the selected samples from Step 1 using a latent space (Z) in 2D format.
- Calculate the loss error value for the original data (loss of real data) and the generated data (loss of fake data).
- Obtain the discriminator results by adding the loss error values of the real and original data, then multiplying the sum by 0.5.
- Update the weights of the discriminator and generator using gradient descent in the back propagation process.
- Monitor the loss error values between the generator and discriminator across multiple epochs, with specific conditions established through extensive experiments. These conditions ensure that the training process is optimized and that the generated adversarial attacks meet predefined quality criteria, such as low discriminator loss and stable generator loss.

This systematic approach to obtaining adversarial attacks through GANs, BiGANs, and WGANs is critical for training robust DDoS attacks detection model capable of effectively detecting a wide range of network attacks, including rare and previously unseen ones.

D. Proposed Framework

This research has developed a comprehensive model leveraging both ML and DL algorithms to enhance the effectiveness of DDoS attack detection. In the realm of machine learning, the Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) algorithms were employed. For fine-tuning purposes in Decision Tree and Random Forest, parameters such as tree depth and the number of jobs were adjusted. The tree depth parameter was varied within the range of 1 to 13, while the number of jobs was set to one, enabling parallel computation of trees. In contrast, the deep learning approach involved the utilization of Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) algorithms for investigation.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

In the case of the NSL-KDD dataset, the model's framework comprises of two key components: Convolutional and Dense layers. The convolutional layer employs a (3*3) kernel size with standard kernel initialization parameters. A flattened layer seeks the convolutional layer, introducing a dropout strategy with a value of 0.1. After that, we employ a dense layer with a standard kernel initializer and ReLU activation function. Then, we employ a Kernel regularization to optimize the model, with an L1 parameter value of 0.02. In the CICIDS-2017 dataset, a similar architecture is employed, with variations in dropout, kernel regularization, and activation function.

We implement a three-layer architecture for the GRU algorithm in the NSL-KDD dataset. The first layer comprises the GRU layer, followed by three dense layers, all utilizing the ReLU activation function. A dropout strategy is involved immediately after the GRU layer, with a value of 0.1. In contrast, the model for CICIDS-2017 consists of four layers, with the first three being GRU layers and the final being a dense layer with an L1 value of 0.005. Finally, we employ the tanh activation function across all layers. The LSTM algorithm comprises two layers, i.e., an LSTM and a dense layer. The LSTM layer possesses a dropout layer with a value of 0.5 and employs ReLU activation functions on both layers. These algorithmic preferences and architectural configurations contribute to the model's robustness and efficacy in detecting DDoS attacks across diverse datasets.

V. Experimental Setup and Results Evaluation

A. *Evaluation Metrics*

Several metrics are employed to assess the performance of the proposed technique quantitatively. These metrics include Accuracy, Precision, Recall, and F1-Score.

1) *Accuracy:* This metric computes the ratio of correctly classified samples to the total samples.

$$Accuracy = \frac{T_P+T_N}{T_P+T_N+F_N+F_P} \quad (3)$$

2) *Precision:* It specifies the quality of an optimistic prediction made by the model by calculating the ratio of true positive predictions to the overall instances that are predicted correctly.

$$Precision = \frac{T_P}{T_P+F_P} \quad (4)$$

3) *Recall:* Recall denotes the ratio of correctly identified positive class samples to the total number of positive instances. It is also known as True Positive Rate.

$$Recall = \frac{T_P}{T_P+F_N} \quad (5)$$

4) *F-Score (F1 Score):* The F1-Score is an evaluation metric, functioning as the weighted harmonic mean of Precision and Recall. It evaluates a model's performance by integrating Precision and Recall into a single score.

$$F\ Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (6)$$

B. *Experimental Results*

Here, we perform a broad comparison of various ML and DL models when trained on diverse datasets, including augmented datasets generated by GAN, Bi-GAN, and WGAN, as well as the original dataset and datasets generated using the Synthetic Minority Over-sampling Technique (SMOTE) [20]. The ML algorithms considered for evaluation include DT, RF, and SVM. On the DL side, the methods include CNN, LSTM, and GRU. The aim is to assess and compare these algorithms' accuracy when encountering various data sources, including synthetic data generated by generative adversarial networks and oversampled data from SMOTE. Table 3 exemplifies the performance of various DML Learning models in binary classification for both the datasets.

TABLE II. Results of various DML methods on Multi classification for NSL-KDD datasets.

Dataset Type	Classifier (Accuracy in %)	Attack Type			
		DoS	Probe	R2L	U2R
Original	DT	90	58	30	16
	RF	89	65	24	20
	SVM	89	65	24	0
	CNN	76	52	0	0
	LSTM	84	72	18	0
	GRU	80	79	0	0
SMOTE	DT	90	60	35	16
	RF	90	76	24	10
	SVM	89	65	24	0
	CNN	80	56	0	0
	LSTM	86	75	15	0
	GRU	82	82	0	0
GAN	DT	89	77	42	20
	RF	92	82	24	21
	SVM	94	89	63	28
	CNN	86	86	45	20
	LSTM	92	89	52	26
	GRU	94	92	64	18
Bi-GAN	DT	84	76	44	30
	RF	89	76	28	0
	SVM	88	72	38	20
	CNN	76	58	0	24
	LSTM	82	76	24	24
	GRU	82	74	36	28
WGAN	DT	86	76	49	40
	RF	89	8	43	70
	SVM	88	82	44	18
	CNN	82	65	28	10
	LSTM	86	82	48	14

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

	GRU	88	78	38	21

TABLE III. OUTCOMES OF VARIOUS DML METHODS ON BINARY CLASSIFICATION FOR NSL-KDD AND CICIDS2017 DATASETS.

ataset Type	**Classifier**	**NSL-KDD**				**CIC-IDS 2017**			
		Accuracy in (%)	**Precision in (%)**	**Recall in (%)**	**F-score in (%)**	**Accuracy in (%)**	**Precision in (%)**	**Recall in (%)**	**F-score in (%)**
	DT	84	86	86	91	92	93	92	95
	RF	86	87	89	90	94	96	96	97
	SVM	89	87	86	89	93	92	96	95
	CNN	80	87	85	87	96	94	99	97
	LSTM	89	87	85	89	97	97	99	98
original	**GRU**	90	87	87	92	98	97	98	97
	DT	86	86	89	89	92	93	92	95
	RF	89	87	90	90	95	96	98	98
	SVM	91	90	90	92	97	98	98	98
	CNN	92	90	91	92	98	98	99	98
	LSTM	90	92	89	92	96	97	99	98
SMOTE	**GRU**	92	94	92	94	98	97	98	98
	DT	94	96	93	94	94	93	96	94
	RF	94	96	89	91	97	96	99	98
	SVM	92	96	94	94	97	96	98	98
	CNN	96	95	95	96	98	99	98	99
	LSTM	95	95	94	95	98	97	99	98
GAN	**GRU**	94	96	94	95	99	98	99	98
	DT	82	86	89	93	94	92	96	92
	RF	89	87	86	90	96	95	99	99
	SVM	86	83	84	83	97	97	98	97
	CNN	89	86	84	89	98	97	99	98
	LSTM	89	87	82	83	98	97	98	98
Bi-GAN	**GRU**	88	86	83	89	98	97	100	98
	DT	93	95	92	93	91	89	89	92
	RF	89	87	84	89	97	97	99	98
	SVM	89	93	88	90	97	96	100	98
	CNN	90	92	92	92	98	97	99	98
	LSTM	88	87	82	89	98	97	98	98
WGAN	**GRU**	89	87	84	90	98	98	98	98

1) Experiment with GAN

This investigation assesses the Generative Adversarial Network (GAN) model's ability to generate high-quality attack traffic and its impact on binary and multi-class classification tasks using the NSL-KDD and CICIDS-2017 datasets. First, we fine-tune the GAN model to produce high-quality synthetic attack traffic. We use cosine similarity as a metric to measure the quality of generated attack traffic. In the NSL-KDD dataset, the GAN achieved a cosine similarity of 18.1343, while in the CICIDS-2017 dataset, it achieved 4.6018. The difference in cosine similarity between the datasets is attributed to variations in the type and number of attacks in each variety. Observing the generation of synthetic attacks by the GAN, the model's performance is evaluated using GAN-generated, original, and SMOTE technique-generated data. We develop the DDoS attack detection system using ML approaches such as DT, RF, and SVM, as well as DL approaches including CNN, LSTM, and GRU.

In the NSL-KDD dataset, the model performance improved for all algorithms with CNN achieving the best results with a 16% increase in the accuracy, followed by decision tree with 10% on original data. Whereas, DT improved 8% and RF, CNN

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

by 5% when compared with SMOTE data. In contrast, in the CICIDS-2017 dataset LSTM, GRU, and DT algorithms had limited impact. Whereas, SVM showed an improvement of 4% on original dataset.

TABLE IV. TABLE 4: RESULTS OF VARIOUS DML METHODS ON MULTI CLASSIFICATION FOR CIC-IDS 2017 DATASETS.

Attack Type	Original			SMOTE			GAN			Bi-GAN			WGAN		
	SVM (%)	LSTM (%)	GRU (%)	SVM (%)	LSTM (%)	GRU (%)	SVM (%)	LSTM (%)	GRU (%)	SVM (%)	LSTM (%)	GRU (%)	SVM (%)	LSTM (%)	GRU (%)
Bot	76	92	84	80	91	85	95	74	82	73	82	74	82	20	87
Brute force	93	85	65	94	85	72	75	66	100	95	95	96	95	96	96
DDoS	94	95	95	93	92	95	68	92	95	95	98	92	96	99	96
DoS GoldenEye	0	0	96	0	0	0	98	92	95	0	0	0	0	45	0
DoS hulk	99	99	90	97	99	99	96	78	88	99	99	98	99	100	100
DoS slowhttptest	90	72	94	95	76	94	95	92	94	92	92	98	86	94	89
DoS slowloris	0	0	90	0	0	0	94	86	92	0	0	0	100	89	0
FTP Patator	88	77	98	90	95	96	92	99	99	96	96	78	96	98	89
Heartbleed	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0	0
Infiltration	98	98	25	98	95	26	26	0.9	0	99	99	99	99	99	99
PortScan	54	62	98	66	72	96	99	99	98	68	68	62	68	68	68
SSH Patator	95	87	98	96	91	98	99	99	99	93	94	76	95	96	94
Sql injection	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XSS	0	0	0	0	0	0	0	0	0	98	99	\97	99	99	100

CNN, GRU, SVM showed limited improvement, DT, RF, and LSTM algorithms showed an improvement of 2% on SMOTE data, suggesting that GANs could learn from specific attack distributions.

The study also investigates the model performance in multi-class classification, as shown in Tables 2 and 4. In the NSL-KDD dataset (Table 2), ML algorithms demonstrated similar behavior, with exceptional performance for Probe, R2L, and U2R categories. In the CICIDS-2017 dataset (Table 4), the model performance in multi-class classification outperformed binary classification, particularly for Bot, SSH Patator, and FTP Patator attacks, which had a larger number of samples. However, certain attacks with fewer than 25 samples, such as SQL Injection, Infiltration, and Heartbleed, remained undetected.

To address stability issues and mode collapse observed in GANs, a second experiment using the BiGAN model was conducted in the research. This demonstrates the exploration of alternative generative models to improve attack generation and subsequent IDS performance.

2) Expermiment with BiGAN

This experiment was conducted to assess the impact of the Bidirectional Generative Adversarial Network (BiGAN) model on the performance of DDoS attack Detection Systems. The goal was to evaluate whether the addition of an encoder component in the BiGAN architecture enhances the detection of attacks across different datasets and classification types. The BiGAN model was constructed and fine-tuned for generating high-quality attacks. When measuring cosine similarity, the model achieved a similarity of 16.5455 for the NSL-KDD dataset and 3.2155 for the CICIDS-2017 dataset.

Comparing the BiGAN model to the GANs model, the cosine similarity values were generally higher for BiGAN, indicating better quality in the generated attacks. However, this metric alone may not reflect the result of the model based on the BiGAN model at the attack level. Therefore, the model based on BiGAN was evaluated for both datasets in binary classification, as shown in Table 3.

The results exposed that the model performance in the NSL-KDD dataset saw a slight enhancement in the accuracy compared with the GANs model. However, the performance of the model based on BiGAN for DT and RF algorithms was poorer, leading to negative impacts. For the CICIDS-2017 dataset, the model using DL algorithms generally showed positive impacts, except for the LSTM algorithm. Further investigation was conducted to assess BiGAN's impact on the model through multi-class classification, presented in Tables 2 and 4 for NSL-KDD and CICIDS-2017 datasets, respectively.

In the NSL-KDD dataset, the model performance based on GANs was slightly better than that of the BiGAN model. In the CICIDS-2017 dataset, BiGAN's main contribution was in detecting Brute Force and DDoS attacks, which were not detected by GANs. Moreover, negative impacts on Infiltration, DoS Slowhttptes, and DoS Hulk attacks were reduced compared to AIDS based on the GANs model. However, the detection of DoS Slowhttptest and GoldenEye attacks was negatively affected by BiGAN, likely due to the limited number of samples for these attacks. In summary, the model performance in detecting attacks varied between GANs and BiGAN models, with some attacks being detected by GANs and others exclusively by the BiGAN model. Accordingly, we perform

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

additional investigations using the Wasserstein Generative Adversarial Network (WGAN) model, which employs a different loss function, and we expect that it will detect a broader range of attacks compared to other generative models.

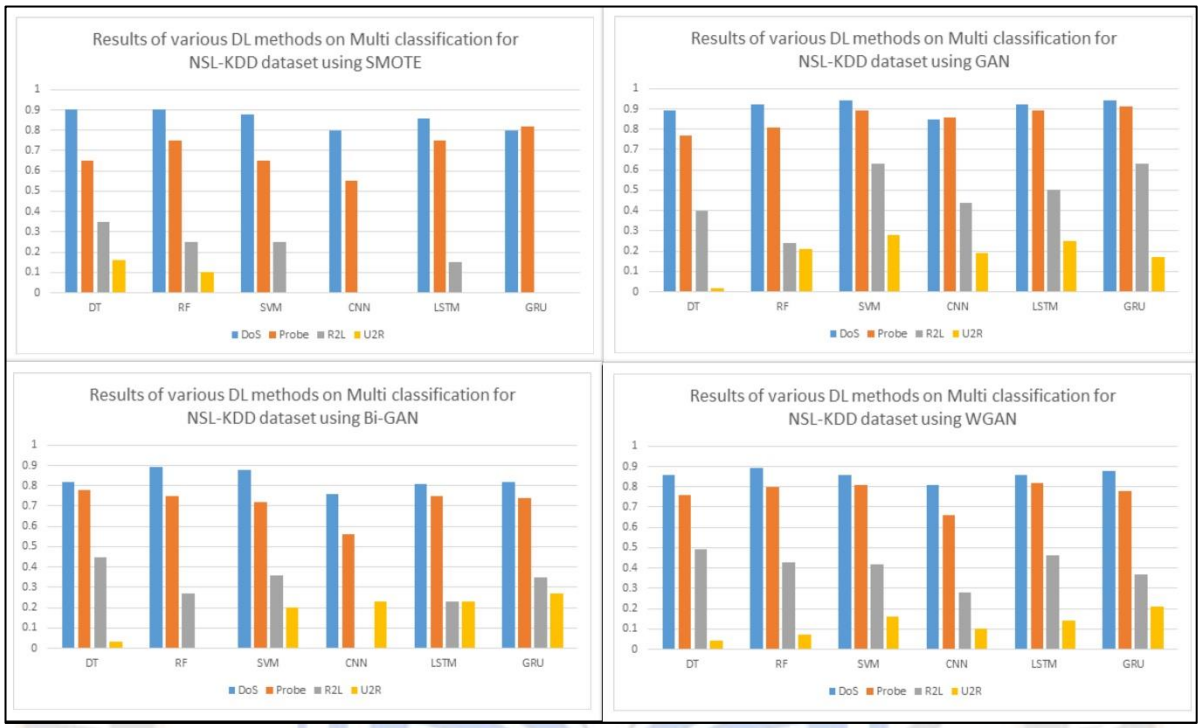

Figure 5: Results of various DML methods on Multi classification for NSL-KDD datasets

.

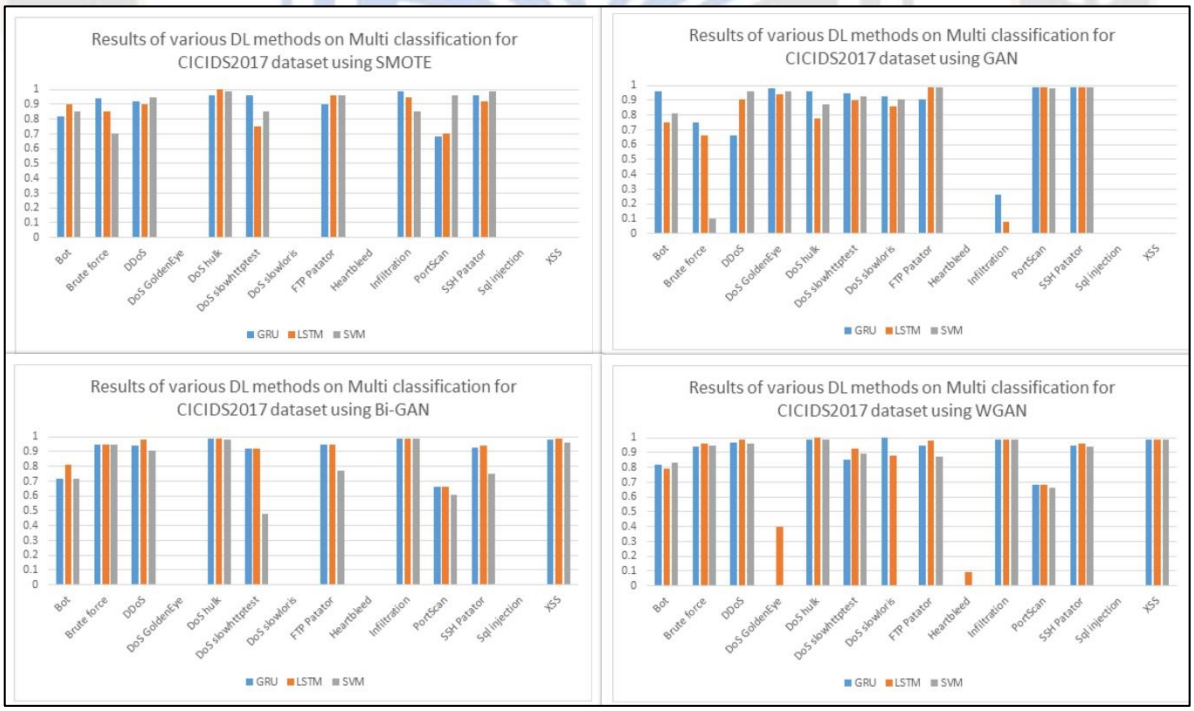

Figure 6: Results of various DML methods on Multi classification for CIC-IDS 2017 datasets.

3) Experiment with WGAN

This section assesses the model's effectiveness on WGAN and other generative models and investigates the deviations between these models in detecting attacks. To assess the quality of the generated attacks produced by WGAN, cosine similarity metrics were calculated to compare the distributions of original

and generated attack traffic across all three generative models (GANs, BiGAN, and WGAN). For the NSL-KDD dataset, the absolute cosine distance was 19.2529 for WGAN, 20.3715 for GANs, and 21.9501 for BiGAN. In the CICIDS-2017 dataset, the individual values were 3.2244 for WGAN, 4.6028 for GANs, and 4.7007 for BiGAN. These results indicate that WGAN produces attack traffic closer to the original attacks than GANs and BiGAN in both datasets.

Table 3 presents the performance of the models in binary classification for NSL-KDD and CICIDS-2017 datasets using WGAN. Notably, WGAN consistently improves the models performance for most approaches (ML and DL) in both datasets, except for some exceptions like GRU in NSL-KDD and RF in CICIDS-2017. The CNN algorithm achieved the highest accuracy in NSL-KDD, while the DT and LSTM algorithms were less effective for CICIDS-2017.

The model based on WGAN demonstrated enhanced performance in both binary and multi-class classification and across both datasets. The results show that WGAN outperformed BiGAN and performed equally to GANs for rare attack categories (R2L and U2R) in the NSL-KDD dataset. Tables 2 and 4 provide detailed results for the model based on the WGAN model using NSL-KDD and CICIDS-2017, respectively.

In the CICIDS-2017 dataset (Table 4), WGAN detected nine attacks, compared to five attacks by GANs and six by BiGAN using the LSTM algorithm. This indicates that WGAN detected more attacks and achieved enhancements in the detection of specific attacks, such as Brute Force, DDoS, DoS Slowhttptest, DoS slowloris, FTP Patator, and Heartbleed, compared to GANs and BiGAN models. Figure 5 and 6 illustrates the Results of various DML methods on Multi classification for NSL-KDD and CIC-IDS 2017 datasets.

When comparing the findings of this study to our prior research efforts [9] and [10], we observe notable differences in approach and outcomes. In [10], we utilized the SMOTE technique to address the challenges associated with imbalanced datasets, while in [9], we employed a Weighted Support Vector Machine to mitigate class imbalances by assigning greater weights to the minority class, thereby emphasizing its significance during training and reducing bias. This research reveals that the application of Generative Adversarial Networks has resulted in a more comprehensive detection of attacks and noteworthy improvements in identifying specific attack types.

VI. Conclusion

This research underscores the significant impact of imbalanced data on DDoS attack detection model performance, particularly for rare attacks. Generative adversarial models have been employed to address this issue by generating additional samples for these rare attacks. However, the influence of such generative models on DDoS attack detection model performance can vary, with some attacks benefiting while others are negatively affected. For example, DoS slowloris and PortScan exhibited improved performance, while DoS Hulk and SSH Patator suffered from decreased performance. Furthermore, the type of classification, binary or multi-class, displayed differing performances across models. Therefore, both classification types were applied to both datasets to evaluate their impacts on model performance. Feature selection, when coupled with adequate training data, can significantly enhance model performance, especially when the selected features are relevant to the target class. After pre-processing both datasets, the Mutual Information and Random Forest Feature Importance techniques were used for feature selection. However, the limited sample size for some rare attacks may pose challenges for generative models in generating similar adversarial samples for these attacks. In summary, this research conducted an extensive evaluation of SMOTE, GAN, BiGAN, and WGAN on various datasets, classification types, and classifiers. The results highlight that WGAN outperforms GAN, BiGAN, and SMOTE models in detecting a broad spectrum of attacks, following optimization of several classifiers and feature selection.

Future research directions could involve the development of new generative models based on GANs, BiGAN, and WGAN in the context of DDoS attack detection, particularly for combating complex and adversarial attacks. Additionally, these models could be optimized for malware analysis by gaining insights into malware behavior at an early stage. The WGAN, with its significant performance in detecting a wide range of attacks across popular datasets, holds promise for applications in the cybersecurity field.

References

[1] Y Yang, Jin, Tao Li, Gang Liang, Wenbo He, and Yue Zhao. "A simple recurrent unit model based intrusion detection system with DCGAN." IEEE Access 7 (2019): 83286-83296.

[2] Lee, JooHwa, and KeeHyun Park. "AE-CGAN model based high performance network intrusion detection system." Applied Sciences 9, no. 20 (2019): 4221.

[3] Ieracitano, Cosimo, Ahsan Adeel, Francesco Carlo Morabito, and Amir Hussain. "A novel statistical analysis and autoencoder driven intelligent intrusion detection approach." Neurocomputing 387 (2020): 51-62.

[4] Zhang, Guoling, Xiaodan Wang, Rui Li, Yafei Song, Jiaxing He, and Jie Lai. "Network intrusion detection based on conditional Wasserstein generative adversarial network and cost-sensitive stacked autoencoder." IEEE access 8 (2020): 190431-190447.

[5] Abdulhammed, Razan, Hassan Musafer, Ali Alessa, Miad Faezipour, and Abdelshakour Abuzneid. "Features dimensionality reduction approaches for machine learning based network intrusion detection." Electronics 8, no. 3 (2019): 322.

[6] Yin, Chuanlong, Yuefei Zhu, Shengli Liu, Jinlong Fei, and Hetong Zhang. "Enhancing network intrusion detection classifiers using supervised adversarial training." The Journal of Supercomputing 76, no. 9 (2020): 6690-6719.

[7] Zhang, Xueqin, Yue Zhou, Songwen Pei, Jingjing Zhuge, and Jiahao Chen. "Adversarial examples detection for XSS attacks based on generative adversarial networks." IEEE Access 8 (2020): 10989-10996.

[8] Mari, Andrei-Grigore, Daniel Zinca, and Virgil Dobrota. "Development of a Machine-Learning Intrusion Detection System and Testing of Its Performance Using a Generative Adversarial Network." Sensors 23, no. 3 (2023): 1315.

[9] Goud, Konda Srikar, and Srinivasa Rao Giduturi. "OML-SDN: Detection of DDoS attacks in SDN using Optimized Machine Learning Methods." International Journal of Intelligent Systems and Applications in Engineering 11, no. 4 (2023): 197-208.

[10] Konda Srikar Goud, Srinivasa Rao Gidituri, "Security Challenges and Related Solutions in Software Defined Networks: A Survey", International Journal of Computer Networks and Applications (IJCNA), 9(1), PP: 22-37, 2022, DOI: 10.22247/ijcna/2022/211595.

[11] Goodfellow, Ian. "Nips 2016 tutorial: Generative adversarial networks." arXiv preprint arXiv:1701.00160 (2016).

[12] Gulrajani, Ishaan, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C. Courville. "Improved training of wasserstein gans." Advances in neural information processing systems 30 (2017).

[13] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks." In International conference on machine learning, pp. 214-223. PMLR, 2017.

[14] Donahue, Jeff, Philipp Krähenbühl, and Trevor Darrell. "Adversarial feature learning." arXiv preprint arXiv:1605.09782 (2016).

[15] Tavallaee, Mahbod, 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, pp. 1-6. Ieee, 2009.

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

[17] Hu, Yinan, and Quanyan Zhu. "Game of Travesty: Decoy-based Psychological Cyber Deception for Proactive Human Agents." arXiv preprint arXiv:2309.13403 (2023).

[18] Lee, Changki, and Gary Geunbae Lee. "Information gain and divergence-based feature selection for machine learning-based text categorization." Information processing & management 42, no. 1 (2006): 155-165.

[19] Rogers, Jeremy, and Steve Gunn. "Identifying feature relevance using a random forest." In International Statistical and Optimization Perspectives Workshop" Subspace, Latent Structure and Feature Selection", pp. 173-184. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[20] Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.