# Hiatus: Unsupervised Generative Approach for Detection of DoS and DDoS Attacks

Sivaanandh Muneeswaran(✉), Vinay Sachidananda, Rajendra Patil, Hongyi Peng, Mingchang Liu, and Mohan Gurusamy

National University of Singapore, Singapore, Singapore
e0503509@u.nus.edu, {comvs,rspatil,dcslium,gmohan}@nus.edu.sg, dcshongp@nus.edu

**Abstract.** Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks pose a serious threat to the internet community by disrupting the availability of services. The current methods for detecting DoS and DDoS attacks have several drawbacks including a high false-positive rate and are mostly supervised techniques. The datasets used lack recent attack types. To overcome these limitations, we propose Hiatus: two independent generative models as anomaly detectors: (1) Variational Auto Encoder (VAE), and (2) Generative Adversarial Network (GAN) to classify the traffic flow as either benign or DoS or DDoS. We make the following contributions: (1) two learning algorithms (VAE and GAN) are trained in an unsupervised fashion to detect DoS and DDoS traffic without the involvement of labeled data, (2) avoid external feature engineering, (3) both the learning algorithms are trained and tested on CICDDoS2019 dataset which consists of latest exploitation and reflection based attacks, and the models are benchmarked by testing them with CICIDS2017 and UNSW-NB15 dataset. With the evaluated results, the proposed approaches outperform existing state-of-the-art techniques and could be used for effective DoS and DDoS detection.

**Keywords:** Denial of Service · Distributed Denial of Service · Unsupervised learning · VAE · GAN · UNSW-NB15 · CICDDoS2019

## 1 Introduction

Network attacks pose a serious threat to the growing internet traffic. With high-speed internet access and the rapid expansion of computer networks, we see an increase in the number of cyberattacks. One of the most common network attacks is the Denial of Service or Distributed Denial of Service attacks. Such attacks tend to disrupt the availability of the resources in the network by overwhelming traffic from various sources. These DoS and DDoS attacks are focused on obtaining financial or economical benefits, besides being used as a tool for revenge and also for political and military advantages. According to the quarterly reports on


F. Li et al. (Eds.): SecureComm 2022, LNICST 462, pp. 529–546, 2023.
https://doi.org/10.1007/978-3-031-25538-0_28

DDoS attacks by Kaspersky[1], for the second quarter of 2021, the top three most attacked countries are the United States (36.00%), China (10.28%), and Poland (6.34%). Almost 60% of the attacks were UDP flood attacks, and the next being SYN flood attacks accounting for 23.67%.

Network Intrusion Detection Systems (NIDS) are employed to detect such DoS and DDoS attacks. NIDS can be classified into signature-based and anomaly-based. Signature-based NIDS attempts to match with the known attack signatures to detect intrusions. The main drawback of such systems is that they cannot detect unknown attacks. On the other hand, anomaly-based NIDS tries to capture the normal behavior of the network, and any deviation from such behavior is flagged as an intrusion. The main advantage of anomaly-based NIDS over signature-based NIDS is that it could even detect unknown attacks and hence the former is preferred over the latter.

Such anomaly-based detection systems are built through statistical methods besides using machine learning techniques. Deep learning has shown to be effective in various tasks due to the availability of data. Many such learning techniques have been proposed to detect the normal and DoS and DDoS records from the network flow. There are some drawbacks to the existing techniques. One such drawback is the limitation in the availability of labeled data which is required for the supervised learning techniques. To overcome this, unsupervised detection of DoS and DDoS traffic is required.

### 1.1 Motivation and Problem Statement

Most of the state-of-the-art approaches are based on supervised learning techniques. The major challenge with supervised learning techniques is the collection of large-scale labeled data which is quite tedious. Moreover, supervised learning techniques do not generalize well to unknown attacks. Even though some unsupervised learning techniques exist for detecting DDoS attacks, they suffer from a high false-positive rate. Moreover, the datasets used in existing works do not address the detection of up-to-date DDoS attacks.

The existing works rely on using the network traffic to detect the attacks. But all these network traffic features do not contribute to the detection of DoS or DDoS attacks. Therefore, the important features from the network traffic have to be extracted. Existing works apply different feature engineering techniques to extract the necessary data from network traffic. But these feature engineering techniques require expertise and vary with different data and algorithms. In addition, as the attacks are evolving on a daily basis, it is necessary for the system to be able to detect recent attacks.

In summary, existing DDoS detection systems suffer from heavy dependence on labeled data for training, unsatisfactory performance, complex feature engineering techniques. To solve the limitations of existing approaches, there is a need for a DoS and DDoS detection system which is capable of detecting up-to-date DoS and DDoS attacks from network traffic without relying much on labeled

[1] https://securelist.com/ddos-attacks-in-q2-2021/103424/.

data (i.e. in an unsupervised fashion). The proposed model should be capable of generating abstract features from high dimensional data without involving domain expertise and without deteriorating performance.

**Shortcomings.** Most of the above-mentioned techniques implement feature extraction through various techniques in order to develop the model which impedes the development of a model suitable across data from different origins and has to adapt to different algorithms. Moreover, highly accurate models require labeled data for development. Thus, we need a methodology to detect DoS, DDoS attacks without heavy dependence on labeled data and complex feature engineering techniques.

**Research Gaps.** Even though numerous solutions exist for the detection of DDoS attacks, the setbacks in the existing works lead to the following research questions:

- RQ-1: how to develop systems capable of detecting DoS, DDoS attacks without much reliance on labeled data?
- RQ-2: how to generate or extract abstract features from high dimensional data without any domain expertise?
- RQ-3: how to develop a detection system with a low false-positive rate along with less inference time capable of detecting recent attacks?

### 1.2 Approach Overview

To build a model for detecting DoS, DDoS attacks, we train two generative models in an unsupervised fashion. This will eliminate the reliance on labeled data to a larger extent. Therefore we train Variational Autoencoder (VAE) and Generative Adversarial Network (GAN) on a single category of data (either benign or malicious). Through this, we limit the usage of labeled data. We are aware that collecting one particular category of data without contamination is not feasible. Therefore, one of our models i.e. VAE is trained with a particular category of data along with a few records of another category as outliers. Through this, our model is robust to outliers in the training data.

Although many works exist for detecting such attacks through shallow and deep ML models, we propose the use of generative models to design the detection system. The main advantage of these generative models is the modeling and learning of the distribution of the training data. Since the model learns the distribution of the data, additional feature selection techniques are not required. Hence generative models are best suited to avoid domain expertise and complex feature engineering procedures.

### 1.3 Results Overview

We have conducted multiple sets of performance evaluation and benchmarking experiments. We evaluated `Hiatus`(VAE and GAN models) on multiple datasets

like CICDDoS2019, CICIDS2017, and UNSW-NB15 datasets containing recent DoS, and DDoS attacks. The proposed GAN model achieves around 99% recall with a false positive rate of 4.16% on the CICDDoS2019 dataset. The proposed VAE model achieves 96.27% recall with a false positive rate of 0.08% on the CICIDS2017 dataset. Moreover, both the proposed models i.e. GAN and VAE achieve 99.93% and 98.87% accuracy along with the false-positive rates of 0.35% and 2.7% respectively.

**Our Contributions.** We make the following contributions to address the existing research gaps.

- We propose `Hiatus` - a DoS, DDoS detection system with two independent generative models VAE, and GAN trained in an unsupervised fashion to detect DoS and DDoS attacks.
- We train the generative models without any external feature engineering process. These generative models are robust to noise in training data.
- We train the models on CICDDoS2019, consisting of modern reflection-based attacks and exploitation-based attacks. Also, the models are trained and tested on datasets like CICIDS2017, UNSW-NB15 and benchmarked.

The remainder of this paper is organized as follows: Section 2 investigates the existing works, Sect. 3 provides a detailed discussion on the proposed models. In Sect. 4 and Sect. 5, the performance of the model along with its results in addition to comparison with existing state-of-the-art is provided. The evaluation results are discussed in Sect. 6, conclusion is provided in Sect. 7 with references at the end.

## 2 Related Work

Various methods for detecting DDoS attacks have been proposed in the literature. Some techniques use statistical methods to detect such attacks which resulted in increased computational complexity. Other techniques include data mining and machine learning methods to detect DDoS attacks. This section summarizes various machine learning techniques to detect DoS and DDoS attacks along with their missing gaps.

**Supervised Learning Techniques:** In [15], an ensemble of deep learning techniques (CNN, LSTM, RNN) is trained on the CICIDS2017 dataset to detect DDoS attacks. Two binary classifiers are trained individually and ensembled. In [29], DeepDefense - several variations of recurrent deep neural networks like LSTM, GRU, CNN-LSTM are trained on the ISCX2012 dataset to differentiate normal traffic from DDoS attacks. Generally, deep learning models involve a higher number of parameters than shallow machine learning models and hence involve higher inference time. However, there is only a 0.1% difference in F1 score between a Random Forest model and the proposed LSTM model. [14] proposed a deep Autoencoder for feature extraction followed by classification of

normal and DDoS traffic through the k-Nearest Neighbour algorithm. The proposed work includes both binary and multi-class classification with the hyperparameters of k-NN and Autoencoder optimized through Bayesian optimization. General ML techniques like Random Forest, outperform the proposed AE + k-NN technique. [16] analyzes statistical features of four different DDoS attacks (SSH Brute-force, DNS Reflection, ICMP flood, TCP SYN) obtained from a simulated dataset. The technique aims to detect DDoS attacks from the source side. The pre-trained model in addition to several machine learning models like Decision Tree, Naive Bayes, K-means acts as an online learning mechanism i.e. using these predictions the pre-trained model can be updated. A combination of different algorithms could have affected the performance, especially in the case of unsupervised learning algorithms. [9] uses a RNN-autoencoder (autoencoder with RNN layers) as a feature extractor in the pre-training stage and a softmax classifier is used in fine-tuning stage to classify the normal and malicious traffic in CICDDoS2019 dataset. Though the dataset consists of around 11 types of DDoS attacks, the model is trained as a binary classifier. RNN has its benefits for sequence data. But the advantage of using it over feed-forward neural networks is not illustrated. One of the major issues of the supervised learning approach is the lack of labeled data. Obtaining labeled data on a large scale is costly in terms of computation. Moreover, these approaches employ additional feature engineering techniques which would vary over data and algorithms.

**Unsupervised Learning Techniques:** In the case of unsupervised learning, labeled data is not required for building the classifier. In [10], multivariate correlation analysis is performed on network features to show the degree of dependency. Clustering through DBSCAN algorithm is used to cluster normal and DDoS traffic with experiments on CAIDA DDoS dataset. Accuracy of 99.99% with only 3000 testing records could not account for the validity of the model. But DBSCAN falls behind with datasets where the density of normal and DDoS records are similar. In [4], K-means clustering is used for determining the cluster with experiments on CAIDA DDoS dataset. But K-means clustering does not work well with non-spherical cluster shapes. [22] proposed an autoencoder trained on CICIDS2017 dataset to obtain low-dimensional data. The low-dimensional normal traffic is used to train the One Class-SVM model to classify the DDoS traffic. Although the model has good accuracy of 99.35% accuracy, the false positive rate is quite high and also it requires noiseless normal traffic for training which is difficult to obtain in the real world setup. [19] proposed two Self-Organizing Maps (SOM) to label the unlabeled data. One SOM is used to mark the record as normal or suspicious and the other to mark the suspicious record as normal or DDoS attack. The labeled data is used to train an ensemble of Random Forest, Decision Tree, and Gradient Boosted Tree through max voting. If the input data has several hot spots, SOM might generate several smaller groups instead of one larger group. If there are discrepancies with the initial SOM model, it would adversely affect the classifier. The main drawback of unsupervised learning algorithms is the high false-positive rate.

**Semi-Supervised Learning Techniques:** To tackle the problem of high false rate, and the need for a large amount of labeled data, semi-supervised learning approaches have been proposed that work on labeled and unlabeled datasets. In [1], agglomerative clustering and K-means cluster the unlabeled data into two clusters, and initial labelling is done based on entropy. Labeling is done based on the voting of the two clustering techniques followed by supervised training and testing through SVM, Random Forest, k-Nearest Neighbours. A simulated dataset is used for training, and the model is evaluated using CICIDS2017 dataset. Even very slight differences in entropy values may mislead the initial labeling process and lead to bad clusters. In [17], Co-clustering is employed for dimensionality reduction, and the dataset (NSL-KDD and ISCX2012) is split into three clusters: 1. the cluster with DDoS traffic, 2. the cluster with normal traffic, 3. the cluster with DDoS-normal traffic. Cluster 2 is considered to be noisy normal traffic and thus it is eliminated and the remaining clusters are combined and trained with Extra-tree classifier.

In [21], the centroid of 14 clusters (13 attack types and normal) obtained through Fuzzy C-means clustering is predefined based on the available labeled data. The cluster for unlabeled data is determined using the membership value. Botnet 2014 dataset containing 13 DDoS attack types was used for experimentation. Noise in the available labeled data used for determining the initial clusters may lead to instability. [13] proposed constrained K-means clustering for distinguishing normal traffic from DDoS traffic. The centroid for the clusters is initialized through the labeled seed set with Lincoln Laboratory Scenarios(DDoS) 1.0 dataset. Though the time to converge is less than K-means, it works on the assumption that the initial labeling of the seed set is noise-free. In [12], the candidate feature set is obtained based on entropy difference and ranked through K-means based on the ratio of average sum of squares error to cluster distance (RSD) followed by Sequential Forward Selection. After feature selection, K-means clustering is performed with an allocation of initial cluster centers based on labeled data density. Experiments were performed on DARPA, CAIDA, CICIDS2017, and a real-world dataset independently. Feature selection leads to a very less number of features used for training and testing. With such less number of features and noise in the data points used for allocating initial cluster centers, this technique becomes unstable. [3] surveyed the machine learning approaches used to detect DDoS attacks.

## 3 Our Proposed Approach

We propose `Hiatus` - an unsupervised DoS and DDoS detection method based on generative models like Variational AutoEncoder (VAE) and Generative Adversarial Network(GAN) trained on datasets with recent DDoS attacks. VAE is preferred over AutoEncoder because the classification of normal and DDoS traffic is based on the reconstruction score which considers the variability of the distribution of variables rather than the reconstruction error of AutoEncoder which is deterministic. GAN is trained to fit the distribution of normal samples and based on its

ability to reconstruct a sample from certain latent representations, the classification between normal and DDoS traffic is done. Both VAE and GAN can handle complex high dimensional data, thus eliminating the curse of dimensionality. Moreover, both techniques are trained in an unsupervised fashion.

### 3.1 Variational Autoencoder

Variational Autoencoder (VAE) [18] is an unsupervised directed probabilistic model whose structure is similar to that of an Autoencoder. It consists of an encoder, a latent distribution, and a decoder. The difference between Autoencoder and VAE is that Autoencoder is a deterministic model and could not produce new samples. But VAE is a probabilistic model and can produce new samples. The probabilistic encoder $e_\theta$ and the probabilistic decoder $d_\phi$ together form the Variational Autoencoder. The objective of VAE is the variational lower bound of the marginal likelihood of the data. The marginal likelihood of individual data points can be written as

$$logp_\theta(x^{(i)}) = -D_{KL}(q_\phi(z|x^{(i)}||p_\theta(z)) + E_{q_\phi(z|x^{(i)})}[logp_\theta(x|z)] \tag{1}$$

where $q_\phi(z|x)$ is the approximate posterior to be modeled. This posterior can be denoted as $\mathcal{N}(\mu_\phi(x), \sigma_\phi(x))$ where $\mu_\phi(x)$ and $\sigma_\phi(x)$ are the mean and standard deviation of the posterior distribution derived through the VAE, $p_\theta(z)$ is the prior distribution of the latent variable $z$. $p_\theta(x|z)$ is the likelihood of $x$ given the latent variable $z$. The first term of equation(1) is the KL divergence between the approximate posterior and the prior. The second term of equation(1) is the reconstruction of $x$. VAE models the parameters (mean and standard deviation) of the distribution. VAE applies reparameterization by using a random variable from a standard normal distribution. The latent variable $z$ is reparameterized through a transformation $h_\phi(\epsilon, x)$ where $\epsilon$ is the random variable from a standard normal distribution.

$$z = h_\phi(\epsilon, x), \epsilon \sim \mathcal{N}(0, 1) \tag{2}$$

This will ensure that the latent variable $z$ follows the distribution of the approximate posterior.

VAE is trained with one class of data (e.g. normal or with DDoS records) and with little noise from the other class. During testing, a number of samples are drawn from the encoder. For every sample, the decoder outputs the mean and standard deviation. Based on this, the probability of generating the input data from this distribution is calculated as the reconstruction score. The average reconstruction score is used as the score for detecting DDoS records.

The main advantage of VAE over Autoencoder is that VAE takes variability of the data into account which is not in the case of Autoencoder. It is possible that both normal and DDoS records share the same mean value but their variance can differ. Most of the techniques for anomaly detection including GAN require only one particular category of data without noise. Noise in such data will deteriorate the performance of the model. Since VAE takes variability into account, it could even work well with noise in the input data. The working of the proposed VAE is depicted in algorithm 1.

**Algorithm 1:** VAE for DoS and DDoS detection

**INPUT:** Training dataset $X$, Validation dataset $X_{val}$, Testing dataset $X_{test}^{(i)}$ $i = 1, ..., M$
**OUTPUT:** benign or anomalous
$\phi, \theta \leftarrow$ train a VAE using $X$ ;
$\alpha \leftarrow$ obtain through Validation dataset $X_{val}$;
**for** $i \leftarrow 0$ **to** $M$ **do**
  $\mu_{z^{(i)}}, \sigma_{z^{(i)}} = e_\theta(z|x^{(i)})$;
  draw N samples from $z \sim \mathcal{N}(\mu_z(i), \sigma_z(i))$;
  **for** $j \leftarrow 0$ **to** $N$ **do**
    $\mu_{\hat{x}^{(i,j)}}, \sigma_{\hat{x}^{(i,j)}} = d_\phi(x|z^{(i,j)})$;
  **end**
  reconstruction score$(i) = \frac{1}{N} \sum_{n=1}^{N} p_\theta(x^{(i)}|\mu_{\hat{x}^{(i,n)}}, \sigma_{\hat{x}^{(i,n)}})$;
  **if** *reconstruction score* $(i) < \alpha$ **then**
    $x^{(i)}$ is not an anomaly
  **end**
  **else**
    $x^{(i)}$ is an anomaly
  **end**
**end**

### 3.2 Generative Adversarial Network

Our model is based on [11]. A generative module and discriminative module are trained for the detection of DoS and DDoS attacks. Both the generator network $G$ and the discriminator network $D$ are trained in an adversarial fashion. The generator is involved in mapping uniformly distributed noise sampled from the latent variable $z$ to the input space $\mathcal{X}$ through the mapping $G(z)$. The objective of the generator is to improve the generation of realistic data. Since the proposed work is based on traffic flows, instead of CNN, ANN is used in both generator and discriminator. The discriminator is aimed at mapping the input data to the probability that the given input to D is real or generated by the generator. The objective of the discriminator is to improve the identification of real and generator data. During training, both the generator and discriminator are optimized through a two-player min-max game.

$$\begin{aligned} min_G max_D V(D,G) = E_{x \sim p_{data}(x)}[logD(x)] \\ + E_{z \sim p_z(z)}[log(1 - D(G(z))] \end{aligned} \tag{3}$$

After training, the generator is capable of mapping the latent variable $z$ to realistic data. To detect the anomaly, two components are used: the *residualloss* and *discriminatorloss*. The residual loss deals with the similarity between the generated data from the generator $G$ and the query data and it can be defined as

$$\mathcal{L}_R(z) = \sum |x - G(z)| \tag{4}$$

If the generator is perfectly trained, the residual loss will be zero and the discrimination loss is defined as:

$$\mathcal{L}_D(z) = \sum |f(x) - f(G(z))| \tag{5}$$

where f(.) represents the output of an intermediate layer of the discriminator. The idea is to obtain a better feature representation through the intermediate layer rather than relying on the scalar output of the discriminator. The GAN model is trained only with one particular category of data (either benign or anomalous data). In this way, the GAN model learns the representation of the input data. As the model is being trained only with one particular category of data, the GAN model could reconstruct that particular category of data. This ability could be used to find the anomalous data since both the loss components remain high for any other category of data. Therefore, the overall anomaly score can be defined as the sum of both residual loss and discriminator loss. The overall algorithm of the proposed GAN technique is given in Algorithm 2.

$$\mathcal{L}(z) = \mathcal{L}_R(z) + \mathcal{L}_D(z) \tag{6}$$

This anomaly score is used to detect DDoS attacks based on a threshold $\gamma$.

**Algorithm 2:** GAN for DoS and DDoS detection

**INPUT:** Training dataset $X$, Validation dataset $X_{val}$, Testing dataset $X_{test}^{(i)}$ $i = 1, ..., M$
**OUTPUT:** benign or anomalous
$G, D \leftarrow$ train a GAN using $X$ ;
$\alpha \leftarrow$ obtain through Validation dataset $X_{val}$;
**for** $i \leftarrow 0$ **to** $M$ **do**
  draw $z \sim \mathcal{N}(0, 1)$;
  $\mathcal{L}_R(z) = |x^{(i)} - G(z)|$;
  $\mathcal{L}_D(z) = |f(x) - f(G(z))|$;
  anomaly score$(i) = \mathcal{L}_R(z) + \mathcal{L}_D(z)$;
  **if** *anomaly score*$(i) < \alpha$ **then**
    $x^{(i)}$ is not an anomaly
  **end**
  **else**
    $x^{(i)}$ is an anomaly
  **end**
**end**

## 4 Performance Evaluation

### 4.1 Datasets

In this section, we list the publicly available datasets consisting of DDoS attacks.

The CAIDA "DDoS Attack 2007" dataset [5] contains one hour of anonymous traffic traces from a DDoS attack. DARPA dataset [20] consists of LLDOS 1.0, which includes a DDoS attack by a novice attacker against a naive defender, LLDOS 2.0.2 which includes a DDoS attack by a stealthy attacker yet novice against a naive defender, and Windows NT Attack Dataset. NSL-KDD [27] contains four categories of attacks: Probe, DoS, R2L, and U2R. It contains 10 types of DoS attacks like Neptune, back, Teardrop, Pod, etc.

The main drawback of the above-mentioned datasets is that almost all of them are outdated. They do not contain recent types of DDoS attacks. In order to solve this issue, we use the CICDDoS2019 dataset [26] which remains as one of the largest public dataset and addresses the gaps in the existing datasets. It contains the most up-to-date DDoS attacks like SSDP, NTP, NETBIOS, etc. It consists of both reflection-based and exploitation-based attacks. 12 DDoS attacks were included during the training day and 7 DDoS attacks were included during the testing day.

Also, to validate the performance of our proposed approach and to compare it with existing literature, we use two more datasets. UNSW-NB15 dataset [23] is labeled and contains nine categories of attacks including DoS. It consists of 49 features with around 16,353 DoS attack records. CICIDS2017 dataset [25] is a labeled dataset and contains seven categories of attacks. It consists of both DoS and DDoS attack types like Hulk, GoldenEye, Slowloris, Slowhttptest, Heartleech, and Low Orbit Ion Canon attacks. We train our proposed model with these datasets and benchmark them. Table 1 contains the distribution of the above-mentioned datasets for both the proposed VAE and GAN models in case of training, validation, and testing sets.

**Table 1.** Distribution of datasets for VAE and GAN models.

| Dataset | Approach | Type of record | Training | Validation | Testing |
|---|---|---|---|---|---|
| CICDDoS2019 | VAE | DDoS | 207880 | 36685 | 61142 |
| | | Benign | 1000 | 4327 | 38948 |
| | GAN | DDoS | 181894 | 32100 | 91713 |
| | | Benign | – | 4427 | 39848 |
| CICIDS2017 | VAE | DDoS | 22326 | 3941 | 11258 |
| | | Benign | 100 | 5990 | 539108 |
| | GAN | DDoS | 57127 | 10082 | 16803 |
| | | Benign | – | 6000 | 54000 |
| UNSW-NB15 | VAE | DDoS | 100 | 2819 | 11276 |
| | | Benign | 89250 | 15750 | 45000 |
| | GAN | DDoS | – | 2839 | 11356 |
| | | Benign | 76500 | 13500 | 60000 |

**Pre-Processing.** The first step in pre-processing is to obtain the DDoS records from various datasets. Benign and DDoS records are extracted from datasets like UNSW-NB15 and CICIDS2017 since these datasets contain other types of attacks also. Since CICDDoS2019 contains only DDoS records, the records are equally sampled from all the different DDoS attacks.

The features containing flow details like timestamp, source IP, destination IP, source port, destination port, etc. are removed. Also, the categorical features are one-hot encoded. Outliers in the dataset are removed using z-score. Normalization is important to convert all the features into a common scale. Hence the data is normalized using min-max normalization.

### 4.2 Experiments

In our experiments, VAE and GAN are trained and tested for different distributions of datasets. For both, VAE and GAN methods, a threshold $\gamma$ is required to classify the data as benign or malicious. This threshold $\gamma$ is calculated through the validation set.

**Model Setup for VAE.** The encoder consists of two hidden layers with 16 and 8 dimensions. The decoder consists of two hidden layers with 8 and 16 dimensions. The latent space consists of 2 dimensions. In addition to this, batch size, the number of epochs for every dataset is determined through hyper-parameter tuning. VAE is trained with only one class of data (either benign or DDoS data) with little noise from the other. During training, the data is preprocessed and passed into the probabilistic encoder where the latent vector learns the distribution of the training data. The average reconstruction loss of the testing data is calculated and the threshold is determined using the validation data.

**Model Setup for GAN.** The generator of GAN consists of two fully connected layers with 64 and 128 units respectively. The units in the output layer of the generator are the number of features. Therefore during training, the generator tunes the latent space accordingly to generate data similar to training data. The discriminator of GAN consists of three fully connected layers with 256, 128, 128 units respectively. The input to the discriminator is either the data generated by the generator or the original data and the discriminator classifies the input as original or fake. All these fully connected layers are activated with LeakyRelu and 20% dropout is applied. By calculating the sum of residual loss and discriminator loss from the generator and discriminator respectively, and comparing it against a threshold calculated through the validation set, the testing data is classified.

## 5 Results and Comparison

This section comprises the results of VAE and GAN implemented for the three datasets: CICDDoS2019, CICIDS2017, and UNSW-NB15. To evaluate the performance of the model, the metrics like Recall, Precision, F1-score, False Positive

Rate are calculated. Since the testing data is imbalanced, robust scoring metrics like F1-score, Area under ROC curve, and Area under PRC curve are calculated in order to avoid bias from the imbalanced data. Moreover to benchmark against the existing works, other metrics like Accuracy, Precision and Recall are also calculated. Results are also obtained by varying the size of datasets. The training size of datasets is varied from 50% to 90% and Receiver operating characteristic curve, Precision-Recall curve are plotted.

### 5.1 CICDDoS2019 Dataset

Table 2 shows the performance of our proposed GAN and VAE models for CICDDoS2019 dataset in comparison with the existing models. It can be seen that GAN performs a little better than VAE because the training data for VAE contains noise (both categories of data) but GAN simply contains one category of data. It is a trade-off between the inclusion of noise in training data and the performance of the models. Although [6,9] are supervised approaches, our approach being unsupervised performs on par and better respectively in terms of higher accuracy and low false positive rate. Figure 1 shows the ROC curve along with Area Under the ROC value and Precision-Recall Curve (PRC) along with Area Under PRC value respectively for different sizes of the dataset for GAN and VAE models. Figure 1e and Fig. 1f show the comparison between the proposed VAE and GAN models in terms of ROC curve and Precision-Recall Curve.

**Table 2.** Performance of `Hiatus` on CICDDoS2019 dataset.

| Method | Precision(%) | Recall(%) | FPR(%) | F1(%) |
|---|---|---|---|---|
| DDoSNet [9] | 99% | 99% | – | 99% |
| Automatic Feature Selection [6] | 91.16% | 79.41% | – | 79.39% |
| `Hiatus-GAN (Our approach)` | 97.33 | 99.05 | 4.16 | 98.18 |
| `Hiatus-VAE (Our approach)` | 95.76 | 97.52 | 10.15 | 96.63 |

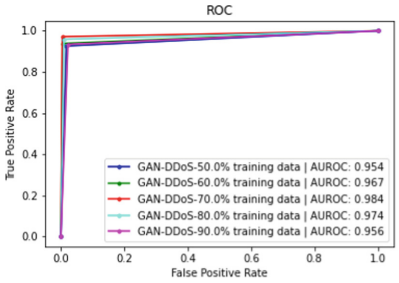
### 5.2 CICIDS2017 Dataset

Table 3 shows the performance of GAN and VAE for CICIDS2017 dataset. [28] generalizes higher-order features from attributed network flow graph and detects the network attack. [8] utilized convolutional neural networks to detect benign or malicious traffic flows. [24] performs unsupervised feature selection and computes initial cluster centers using a set of semi-identical instances and performs clustering. [7] classifies DDoS records from normal records through the Kernel Online Anomaly Detection algorithm which is unsupervised. [8] and [28] are supervised techniques. Although our proposed work is unsupervised, and it is not a head-to-head comparison, our proposed VAE model could perform equivalent to the state of the art and in fact could achieve a better false-positive rate than the state of the art. Our proposed GAN model could perform better than the

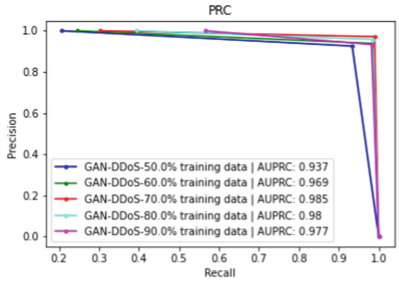
other two existing works. Figure 2 shows the ROC curve along with Area Under ROC value and Precision-Recall Curve (PRC) along with Area Under PRC value respectively for different sizes of the dataset for GAN and VAE models.

### 5.3 UNSW-NB15 Dataset

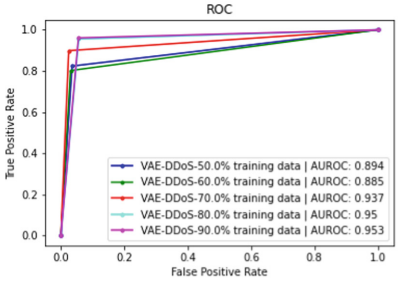
Table 4 shows the performance of GAN and VAE for UNSW-NB15 dataset. In [2], Feature Correlation Map is extracted to detect malicious traffic from normal traffic. [17] utilizes network entropy estimation, co-clustering, and extra-tree



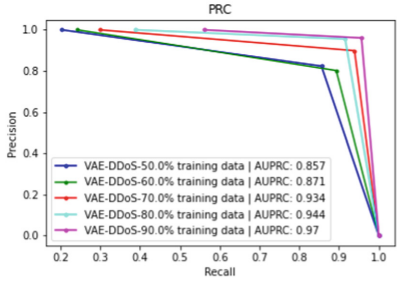

(a) ROC curve for different sizes of dataset for GAN.



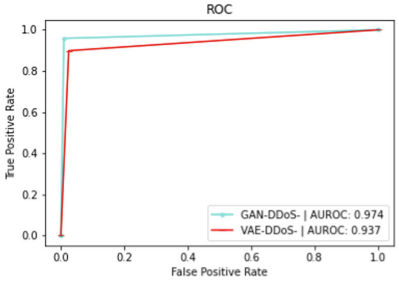

(b) Precision-Recall curve for different sizes of dataset for GAN.



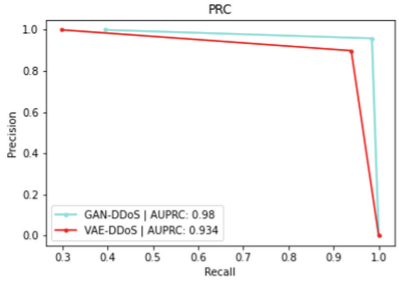

(c) ROC curve for different sizes of dataset for VAE.




(d) Precision-Recall curve for different sizes of dataset for VAE.

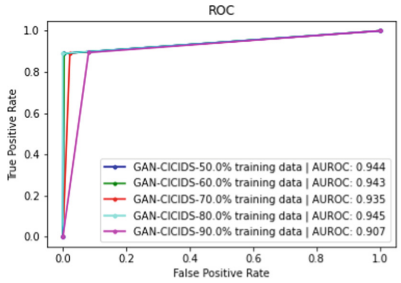



(e) Receiver Operating Characteristic curve for VAE and GAN.

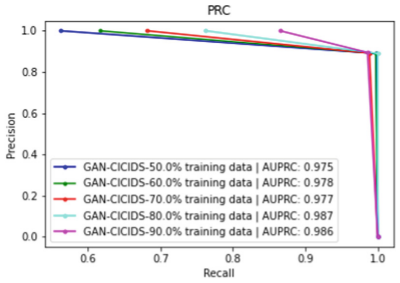



(f) Precision-Recall curve for VAE and GAN.

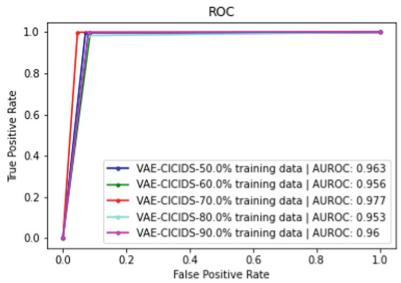
**Fig. 1.** Results of VAE and GAN with CICDDoS2019 dataset.



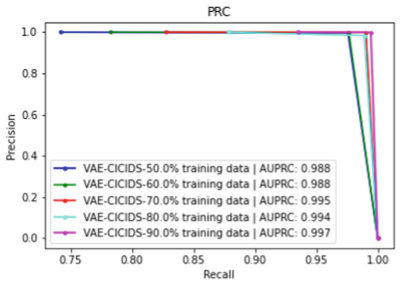

(a) ROC curve for different sizes of dataset for GAN.



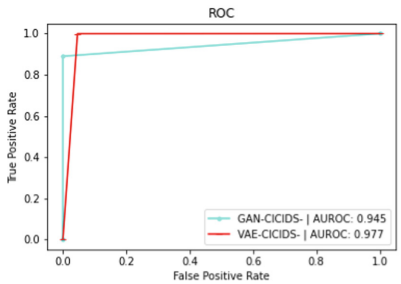

(b) Precision-Recall curve for different sizes of dataset for GAN.



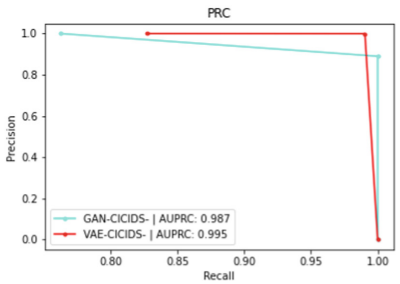

(c) ROC curve for different sizes of dataset for VAE.




(d) Precision-Recall curve for different sizes of dataset for VAE.




(e) Receiver Operating Characteristic curve for VAE and GAN.




(f) Precision-Recall curve for VAE and GAN.

**Fig. 2.** Results of VAE and GAN with CICIDS2017 dataset.

algorithm to detect DDoS attacks. Table 4 shows that the proposed VAE and GAN methods outperform the existing methods in the literature. The main reason for low performance in existing works is that the normal and DDoS records in UNSW-NB15 are similar. Hence most of the models fail to perform better or result in increased False Positive Rates. Since VAE takes the variability of the data into account, it could differentiate between normal and DDoS records effectively.

**Table 3.** Performance of Hiatus on CICIDS2017 dataset.

| Method | Accuracy(%) | FPR(%) | Precision(%) | Recall(%) | F1(%) |
|---|---|---|---|---|---|
| LUCID [8] | 98.88 | 1.79 | 98.27 | 99.52 | **98.89** |
| DeepGFL [28] | – | – | 75.67 | 30.24 | 43.21 |
| Cluster center initialization [24] | 81.98 | 59.68 | 79.16 | 81.98 | 80.54 |
| E-KOAD [7] | **99.55** | 0.23 | 95.24 | 95.24 | 95.24 |
| Hiatus-GAN (Our approach) | 91.59 | 11.01 | 73.85 | **100** | 84.96 |
| Hiatus-VAE (Our approach) | 99.28 | **0.08** | **99.55** | 96.27 | 97.86 |

Figure 3 shows the ROC curve along with Area Under ROC value and Precision-Recall Curve (PRC) along with Area Under PRC value respectively for different sizes of the dataset for GAN and VAE models. Figure 3e and 3f show the comparison between the proposed VAE and GAN models in terms of ROC curve and Precision-Recall Curve.

**Table 4.** Performance of Hiatus on UNSW-NB15 dataset.

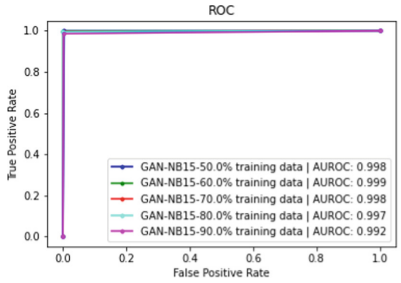
| Method | Accuracy(%) | TPR(%) | FPR(%) | F1(%) |
|---|---|---|---|---|
| Feature Correlation Map [2] | 91.82 | 60.92 | 0.46 | 72.65 |
| Semi supervised machine learning [17] | 93.71 | – | 1.41 | – |
| **Hiatus-GAN (Our approach)** | **99.93** | **99.99** | **0.35** | **99.95** |
| **Hiatus-VAE (Our approach)** | 98.87 | 99.29 | 2.7 | 99.29 |

## 6 Discussion

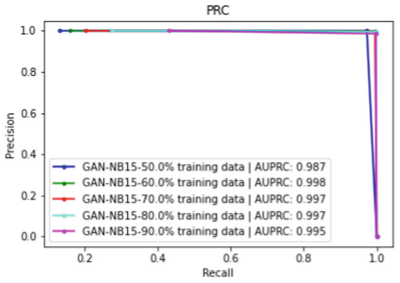
The proposed generative models (VAE and GAN) have the advantage of less reliance on labeled data during training which makes the feasibility of collection of large-scale data easier. Considering the noise within a single category of data in the real-world environment, the proposed VAE model proves to be robust to noise and has on-par performance with the GAN model.

For different datasets and attacks, the existing works have relied on multiple rounds of feature selection in order to achieve good performance. However, the proposed approaches have eliminated the need for such expensive feature engineering techniques. The proposed approaches can handle all the features of the data and model it completely without any loss in potential information. Therefore, *Hiatus* can ingest network traffic data in real time and could detect the DoS and DDoS attacks without much latency as the system does not involve complex feature engineering practices and also considers all the information for detection.

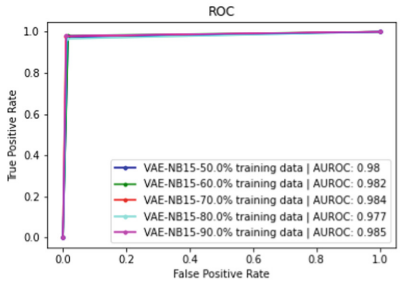
In order to prove the robustness of our approaches, we have conducted multiple experiments with varying proportion of training and testing data in addition to using robust scoring metrics like Receiver Operating Characteristic (ROC)



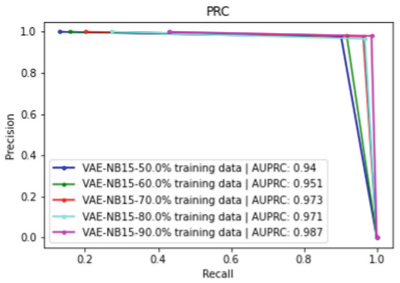

(a) ROC curve for different sizes of dataset for GAN.



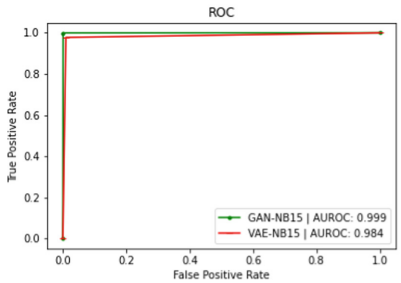

(b) Precision-Recall curve for different sizes of dataset for GAN.



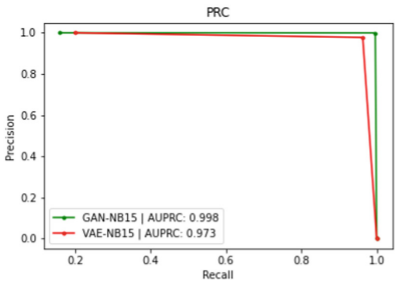

(c) ROC curve for different sizes of dataset for VAE.




(d) Precision-Recall curve for different sizes of dataset for VAE.




(e) Receiver Operating Characteristic curve for VAE and GAN.




(f) Precision-Recall curve for VAE and GAN.

**Fig. 3.** Results of VAE and GAN with UNSW-NB15 dataset.

curve, and Precision Recall Curve (PRC) and plotted their graphs for each of the dataset to show that our models achieve good performance while handling the data imbalance problem. Moreover, datasets containing recent attack types are used and the model is capable of classifying new attack patterns due to its unsupervised training nature.

## 7 Conclusion

In this work, we have presented two generative models for the detection of DDoS attacks which are capable of outperforming the performance of state-of-the-art models. The benefit of both models is that they do not require additional domain expertise for the feature selection and are unsupervised without any dependency on labels. Despite being an unsupervised technique, our models could achieve a low false-positive rate. To show the reliability of our approach, we have tested the models on benchmark datasets and produced the results.

## References

1. Aamir, M., Zaidi, S.M.A.: Clustering based semi-supervised machine learning for ddos attack classification. Journal of King Saud University - Computer and Information Sciences (2019). https://doi.org/10.1016/j.jksuci.2019.02.003, http://www.sciencedirect.com/science/article/pii/S131915781831067X
2. Amma, N.G.B., Subramanian, S.: Feature correlation map based statistical approach for denial of service attacks detection. In: 2019 5th International Conference on Computing Engineering and Design (ICCED), pp. 1–6 (2019)
3. Bhardwaj, A., Mangat, V., Vig, R., Halder, S., Conti, M.: Distributed denial of service attacks in cloud: State-of-the-art of scientific and commercial solutions. Comput. Sci. Rev. **39** 100332 (2021). https://doi.org/10.1016/j.cosrev.2020.100332, https://www.sciencedirect.com/science/article/pii/S1574013720304329
4. Bhaya, W., Manna, M.: A proactive ddos attack detection approach using data mining cluster analysis. J. Next Gener. Inform. Technol. **5** 36– 47 (D2014)
5. CAIDA: The caida ucsd "ddos attack 2007" dataset (2007)
6. Can, D.-C., Le, H.-Q., Ha, Q.-T.: Detection of distributed denial of service attacks using automatic feature selection with enhancement for imbalance dataset. In: Nguyen, N.T., Chittayasothorn, S., Niyato, D., Trawiński, B. (eds.) ACIIDS 2021. LNCS (LNAI), vol. 12672, pp. 386–398. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-73280-6_31
7. Daneshgadeh Çakmakçı, S., Kemmerich, T., Ahmed, T., Baykal, N.: Online ddos attack detection using mahalanobis distance and kernel-based learning algorithm. J. Netw. Comput. Appl. **168** 102756 (2020). https://doi.org/10.1016/j.jnca.2020.102756, http://www.sciencedirect.com/science/article/pii/S1084804520302307
8. Doriguzzi-Corin, R., Millar, S., Scott-Hayward, S., Martinez-del Rincon, J., Siracusa, D.: Lucid: A practical, lightweight deep learning solution for ddos attack detection. IEEE Trans. Netw. Serv. Manage. **17**(2), 876–889 (2020). https://doi.org/10.1109/tnsm.2020.2971776, http://dx.doi.org/10.1109/TNSM.2020.2971776
9. Elsayed, M.S., Le-Khac, N.A., Dev, S., Jurcut, A.D.: Ddosnet: A deep-learning model for detecting network attacks. In: 2020 IEEE 21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), pp. 391–396 (2020). https://doi.org/10.1109/WoWMoM49955.2020.00072
10. Girma, A., Wang, P.: An efficient hybrid model for detecting distributed denial of service (ddos) attacks in cloud computing using multivariate correlation and data mining clustering techniques. Issues Inform. Syst. **19**(2), 12 (2018)
11. Goodfellow, I., et al.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) Advances in Neural Information Processing Systems. vol. 27. Curran Associates, Inc. (2014). https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

12. Gu, Y., Li, K., Guo, Z., Wang, Y.: Semi-supervised k-means ddos detection method using hybrid feature selection algorithm. IEEE Access **7**, 64351–64365 (2019)
13. Gu, Y., Wang, Y., Yang, Z., Xiong, F., Gao, Y.: Multiple-features-based semisupervised clustering ddos detection method. Math. Prob. Eng. **2017** (2017)
14. Görmez, Y., Aydin, Z., Karademir, R., Gungor, V.: A deep learning approach with bayesian optimization and ensemble classifiers for detecting denial of service attacks. Int. J. Commun. Syst. **33**(6), e4401 (2020). https://doi.org/10.1002/dac.4401
15. Haider, S., et al.: A deep cnn ensemble framework for efficient ddos attack detection in software defined networks. IEEE Access **8**, 53972–53983 (2020)
16. He, Z., Zhang, T., Lee, R.B.: Machine learning based ddos attack detection from source side in cloud. In: 2017 IEEE CSCloud, pp. 114–120 (2017)
17. Idhammad, M., Afdel, K., Belouch, M.: Semi-supervised machine learning approach for ddos detection. Appl. Intell. **48**(10), 3193–3208 (2018)
18. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. CoRR abs/1312.6114 (2014)
19. Ko, I., Chambers, D., Barrett, E.: Self-supervised network traffic management for ddos mitigation within the isp domain. Future Gener. Comput. Syst. **112**, 524–533 (2020). http://www.sciencedirect.com/science/article/pii/S0167739X20302193
20. Laboratory, M.L.: 2000 darpa intrusion detection scenario specific datasets (2000)
21. Lysenko, S., Savenko, O., Bobrovnikova, K.: Ddos botnet detection technique based on the use of the semi-supervised fuzzy c-means clustering. In: ICTERI Workshops, pp. 688–695 (2018)
22. Mhamdi, L., McLernon, D., El-moussa, F., Raza Zaidi, S.A., Ghogho, M., Tang, T.: A deep learning approach combining autoencoder with one-class svm for ddos attack detection in sdns. In: 2020 IEEE Eighth International Conference on Communications and Networking (ComNet), pp. 1–6 (2020). https://doi.org/10.1109/ComNet47917.2020.9306073
23. Moustafa, N., Slay, J.: 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), pp. 1–6. IEEE (2015)
24. Prasad, M., Tripathi, S., Dahal, K.: Unsupervised feature selection and cluster center initialization based arbitrary shaped clusters for intrusion detection. Comput. Secur. **99**, 102062 (2020). https://doi.org/10.1016/j.cose.2020.102062, http://www.sciencedirect.com/science/article/pii/S0167404820303357
25. Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A.: Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: ICISSP, pp. 108–116 (2018)
26. Sharafaldin, I., Lashkari, A.H., Hakak, S., Ghorbani, A.A.: Developing realistic distributed denial of service (ddos) attack dataset and taxonomy. In: 2019 International Carnahan Conference on Security Technology (2019)
27. Tavallaee, M., Bagheri, E., Lu, W., Ghorbani, A.A.: 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)
28. Yao, Y., Su, L., Lu, Z.: Deepgfl: Deep feature learning via graph for attack detection on flow-based network traffic. In: MILCOM 2018–2018 IEEE Military Communications Conference (MILCOM), pp. 579–584 (2018)
29. Yuan, X., Li, C., Li, X.: Deepdefense: Identifying ddos attack via deep learning. In: 2017 IEEE International Conference on Smart Computing, pp. 1–8 (2017)