GENERATIVE ADVERSARIAL NETWORK FOR INTRUSION DETECTION SYSTEM

Master Thesis

Completed Date: 20-OCT-2021

DECLARATION OF ORIGINALITY

"I certify that this work does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university and that to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where due reference is made in the text."

20-OCT-2021

ABSTRACT

Generative Adversarial Network (GAN) has been extensively studied in image processing and computer vision (CV). However, their influence on the cyber security field remains an area of open research. Cyber security attacks and incidents on the Internet of Things (IoT) have been increased. Many deep learning algorithms are being proposed to detect and mitigate such intrusion attacks. In modern days where new methods to prevent such attacks are introduced, attackers also improve their techniques to attack. Adversarial attacks became the most emerging technique to fool the network security system by giving deceptive input. This research project investigates GAN generated data to deceive the machine learning-based IDS system. We used the BoT-IoT dataset in our analysis. First, we trained our ML models and developed a benchmark; second, we generated fake data with GAN to deceive the IDS system. In the first approach, three machine learning/ deep learning-based IDS models, Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN), are trained, which gave an accuracy of up to 99% on binary class training data. After generating the fake data through GAN, these three trained models were tested again for the fake data and compared. Results showed that GAN successfully fooled RF and ANN, the accuracy of both the models dropped to 50% and 38.63%, respectively.

On the contrary, SVM outperformed RF and ANN and proved more robust against them. In the second approach, Wasserstein GAN is used to develop a trained model on the training dataset then tested with test data, which predicted the malicious data as normal (malicious data as non-malicious data). RF and SVM models were used for the IDS. Likewise, in the first approach, RF and SVM also demonstrate accuracy of 99.9% in binary and multiclass classifications on the actual dataset. The evaluation of this research concluded that ML-based IDS could be compromised by using GAN.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to all who assisted me to complete this research project.

I would first like to extend my sincere thanks to Dr SR from the College of Science and Engineering, whose knowledge was invaluable to nurture me to this project's design. Your insightful remarks and feedback enhanced my thinking and levelled up my skills to complete this research project.

I would also like to extend my gratitude to Dr BW from the College of Science and Engineering for his assistance, consultation and support during the whole venture.

In addition, I am deeply indebted to my parents for their sympathetic ear and wise counselling. You always stand firm with me in any hardship. Finally, this research project could not have been completed without the provision of my friends MSS, SAH and MFM, who provided happy disruptions and simulating discussion to rest my mind outside of this project.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

The internet rapidly becomes prevalent through novel gadgets and technologies that increase attack risk, allowing cybercriminals to control insecure devices. Since the sudden growth of the interconnected digital world [1], such as the Software Defined Networks (SDNs), Internet of Things (IoT), cyber-attack and their related risks have increased substantially. The internet is evolving towards growing connectivity between devices. It is often referred to under the IoT, in which physical devices are vigorous players in business activities [2]. Sensors, computing chips, and other technologies are integrated, allowing them to gather and share data through the internet. The goal of IoT networks is to enhance the productivity of cloud platforms, such as industrial systems and intelligent buildings. IoT devices are expected to exceed fifty billion by the end of 2020 [2]. As a result of this expansion, the number of cyber-attack events and the related risk has increased. As a result, corporations and organisations are exploring innovative ways to secure personal and corporate data held on network nodes. Unfortunately, current IoT system security mechanisms have been revealed unreliable in the face of unprecedented threats [3]. In 2017, for example, attackers used an IoT fish tank thermometer to infiltrate a casino's sensitive information. In 2018, more than 2.4 million new malicious types were developed, referring to the Symantec Internet Security Threat Report [4]. As a result, there has been a surge in interest in enhancing the ability of network Intrusion Detection Systems (IDS) to identify novel attacks. As a result, advanced ways are necessary to proliferation the efficiency of IDS in detecting attacks.

An IDS monitors traffic flows in a network to detect potential threats and secure digital assets [5]. It aims to preserve the three security principles of information systems, namely confidentiality, availability, and integrity [5]. It is developed to give cyber solid security protection in operating infrastructures. For a long time, the principal purpose of IDS has been to detect cyber-attacks and threats. IDS are divided into two categories signature and anomaly-based Detection: The goal of signature-based attacks detection is to match and compare signatures from incoming communications to a database of predefined signatures from previously known attacks [6]. Signature-based detection, the best variation and identification relying on anomalies, can be referred to as detecting deviations from a model of worthy traffic. This method relies on deep learning methodologies, and various machine learning algorithms need to be used to get the desired result. They usually give good detection accuracy for previously detected attacks, but they fail to identify the latest or modified threats, not in the database. IDS must adapt to new detection tactics, as attackers constantly vary their concepts and methods for executing attacks to avoid current security measures. The existing mechanism for tweaking signatures to keep up with changing attack vectors is unstable. Anomaly-based IDS strive to overcome the drawbacks of signature IDS by employing advanced statistical approaches that have allowed researchers to discover behavioural trends of network traffic. Intrusion detection is accomplished using a variety of ways, including statistical knowledge and Machine Learning (ML) based algorithms [6]. They can achieve significant accuracy and Detection Rate (DR) for zero-day attacks because they match attack behaviours rather than signatures [7]. On the other hand, Anomaly IDS has high False

Alarm Rates (FARs) because they can label any innocuous traffic that differs from particular behaviour as an anomaly.

Existing signature IDS have shown to be ineffective at identifying zero-day attack signatures as they move across IoT networks [8]. This is due to the system's registry lacking known attack signatures. Many strategies, including ML, have been developed and implemented with some effectiveness to prevent similar occurrences from reoccurring. ML is a modern technology that can learn and extract hazardous patterns from network data, which can aid in the detection of security problems [9]. Deep Learning (DL) is a new field of ML that has shown to be particularly effective in detecting complex data patterns [10]. Its algorithms are based on biological brain systems, which convey data signals through a network of connected layers. A computational activation function in each unit translates input to output. Hidden layers in all of these algorithms can extract even more complicated patterns in network activity. Network attack vectors, which can be derived from numerous features communicated by network traffic, such as packet services, protocols, count/size, and signals, are used to understand these patterns. Each attack type has a distinct identification pattern, which is defined as a series of actions that, if left undetected, can weaken network security standards.

Researchers have created and tested various ML models, frequently paired with Feature Reduction (FR) techniques to increase efficiency. Although ML's detecting skills have been encouraging results using a set of assessment parameters, these models are not yet reliable for real-world IoT networks. Rather than acquiring insights into an ML-based IDS application, the trend in this discipline has been to outperform state-of-the-art outcomes for a given dataset [11]. As a result, the vast volume of academic research undertaken far outnumbers the number of real deployments in real life. Although this could be attributed to the high cost of errors in this sector compared to others [11], it is also possible that these strategies are unreliable in practice. This is since they are frequently evaluated using a single dataset, including a list of features that may not be viable to collect or store in a live IoT network stream. Furthermore, because of the architecture of ML, there is frequently room to improve in its hyper-parameters when applied to specific data. As a result, this work aims to assess the generalizability of Feature Extraction (FE) techniques and ML model combinations on various IDS datasets.

Generative Adversarial Network (GAN) is an example of a deep generative model that determines a density function across the data distribution by using different training methods. The fundamental idea behind GAN is that a GAN is to create two adversarial networks: a discriminator and a generator. The generator network aims to create real-looking images that can confuse the discriminator. On the other hand, the discriminator attempts to label generated pictures (generated from the generated network) as fake and then identify the real images of that original image as authentic [12]. Figure 1 demonstrates the GAN functionality; the generator generates fake images on actual data and random noise. Then, the discriminator trained on the actual data identifies the fake images as real or fake. GAN is used in the most advanced tasks of realistic generation like the prediction of video frames, increasing the resolution of images, image-to-text translation and generative image manipulation. GAN has proven to be at the top of the technology for creating accurate and precise images [12].

Image removed due to copyright restriction you can see the image from below link

https://www.xenonstack.com/insights/generative-adversarial-networks

Figure 1. Generative Adversarial Network [13]

GAN is extensively used in computer vision (CV), but its impact on cyber security is still a subject of ongoing research. Cyber security threats and attacks related to attacks on the IoT have increased. Various (DL) algorithms are currently being employed to detect and prevent such intrusion attacks. Today, new techniques are being developed to stop these attacks; attackers are also improving their strategies to target. This research project investigates the GAN models that could alter the signature of IoT data and create fake data that can deceive the Intrusion Detection System (IDS) based on classical ML algorithms. It also focuses on another GAN model, which predicts the malicious data as normal (malicious data as non-malicious data).

2. PROBLEM AND SOLUTION STATEMENT

The critical problem within the IoT network is the cyber-attacks through which sensitive data leaks. Although modern techniques are used to detect and prevent these attacks, the attacker also developed their approaches to hacking the security system. So, the use of IDS with the GAN is a necessary approach individual must take. Two research questions discovered through the examination of network security has been outlined as follows.

- How can the IDS network be compromised by the interference of adversarial models in the network?
- How can the performance of the IDS algorithms be increased by training them alongside adversarial examples?

As for the solution, this project suggests two GAN models that use a structured and representative BoT-IoT data set. Based on the different functions of the two different GAN models, this project is divided into two approaches.

First Approach:

- Train and test the IDS model on the actual data set.
- GAN model creates bogus IoT data that could fool the ML/ DL- based Intrusion detection system (IDS).
- On the newly generated bogus data, apply the trained IDS model and evaluate the results.

Second Approach:

- Use Wasserstein GAN (WGAN) to develop a model that predicts fraudulent IoT data as normal.
- Use some IDS models on the same data used for the WGAN
- Compare the results of WGAN and IDS models.

3. LITERATURE REVIEW

The advancement of technology and scientific progress has enabled the industry to become more efficient and effective. The adoption of the new technology has enabled the industries to generate more productive results, enabling the industry to be a significant contributor to the country's overall growth. The new technology such as IoT, ML, AI and DL, and Blockchain technology has increased the efficiency of the various existing industry and the innovative solution has enabled various industries such as Health care, Education, Manufacturing, Finance and Food industry and various other industry to be more impactful in the market and enable this industry to become a significant contributor to a nation's GDP.

These technologies have opened several opportunities for the various industries, enabling the industry to be more robust, flexible and prosperous. However, some significant adverse implications have become a significant roadblock to implementing this technological solution into the existing business module with all these opportunities. Cyber-attack is one of the major obstacles in adopting these types of technologies. Due to the data revolution, data consumption has increased, leading businesses to adopt data-driven decisions to become more effective in their existing market. The rise of data consumption has created a potential bottleneck for technological adoption in the business, as the businesses are continuously adopting the technological solution; every communication is recorded at the server. Hence, if the server does not have the highest security practices, it leads to severe data breaching incidents. In recent times, various events have demonstrated the importance of practising the highest quality of cyber security to create a safe and secure environment for the users.

This research will address the cyber security gap and provide an effective solution to meet these gaps. In recent times the IoT has gathered attention from various industries. Due to its compatibility and robustness, the technology helps various other industries to be transformed and enable the other industries to use the intelligent solution to their existing problem, The intrusion detection system is one such facility of the IoT technology and the recent event of the cyber-attack has increased the need of the IDS system for the various other industries. There are various ways to detect intrusion in the network. However, this research project will use the GAN to provide effective results. This DL algorithm will enable the IDS system to detect the anomaly within the network and enable the system to be safe and secure. This chapter will discuss the various aspects of the GAN network for IDS systems and enable readers to understand the impact and the significant implications of these combinations. Various works of literature will be studied by the different authors to provide authentic information. This approach will enable the readers to form foundational information about the subject and enable the stakeholders to decide.

3.1 Empirical Study

This section is the most significant section of the Literature review. This section will enable the readers to gain introductory awareness about the topic and increase their knowledge. The following section will be based on reviewing numerous literature and empower the researchers to develop intuitive abilities about the subject, which will enable them to form a crucial understanding of the subject and facilitate them to evaluate the existing gap in the research and further help them to create more effective solutions to mitigate the existing problem.

In the recent study of Papadopoulos et al., IDS based on ML have proved their worth in detecting unknown threats with high accuracy. Nevertheless, these models are also susceptible to attack. Examples of adversaries can be used to assess the strength of a model before it is implemented. Additionally, using examples from adversaries is essential to create an efficient model to work in this environment. Their research evaluates the robustness of traditional ML models and models based on DL's reliability with the Bot-IoT dataset. Our approach was based on two primary strategies. They used the Fast Gradient Sign Method (FGSM) against Artificial Neural Network (ANN) and investigated the adversarial example using the Bot-IoT dataset [14]. However, they did not produce any new or fake data with the help of GAN that can trick the IDS. According to Wang, Deep Neural Network (DNN) has proven efficacy in all ML applications, including intrusion detection. In the past, researchers have discovered the DNN is susceptible to adversarial attack within the image classification; they provide a few prospects for a hacker to trick the networks into misclassification through making subtle variations to the pixels of the image. This susceptibility has raised particular questions regarding the use of DNN to secure areas such as intrusion detection. The author analysed the performance of the most advanced attacks alongside DL-based IDS using the NSL-KDD data set and examined the viability and effectiveness of attack strategies. The weaknesses of neural networks (NN) used by IDS have been practically confirmed. The role of each feature in the creation of adversarial scenarios is investigated [15].

Conversely, the NSL-KDD dataset had lost importance in the current IDS algorithms due to recent attacks' nonexistence and new protocols' introduction. This dataset has been identified as inaccurate and distorted [14]. Koroniotis et al. assess the accuracy of the BoT-IoT dataset by implementing various ML-based and statistical analysis approaches for forensic analysis of Bot-IoT datasets [16]. However, they did not implement the GAN.

According to Wang, IoT is the recent buzzword in the industry. One of the reasons for getting this massive attention is that this technology is one of the most compatible technologies, and it has several benefits over conventional hardware. Due to this reason, it enables the users to have various opportunities to grow in their business [17]. However, there is one problem: businesses are rapidly adopting the technology due to its massive popularity, and the different hardware characteristics have become potential vulnerabilities. Due to this reason, an additional security component needs to be added to the system to enhance the security module of the system and enable the system to be more secure and safe. According to Yilmaz, IDS is the initiative

that will allow IoT devices to be more safe and secure for their users. The author further stated that few industries had piloted this concept. Those industries have been able to generate promising results; due to this reason, the author has stated that the adoption of IDS methodology can increase the security element of the IoT devices [18]. Currently, the Wireless industry is using this type of concept, and the wireless industry has seen promising results and the industry has been able to detect various attacks and abnormalities in the system.

According to Nie, a few things must be considered before implementing this technology in IoT devices [19]. The fundamental characteristics are different from the wireless industry; the wireless industry has a one-dimensional functionality; hence, it is a static function. However, this industry is dynamic for the IoT industry, and their multiple objectives need to be met by an IoT device. Due to this reason designing an effective IDs system is complicated for IoT devices [19]. Nevertheless, a few points will enable the business to develop an effective IDS system for the IoT network.

- **Authorisation:** The authorisation is the privileges or the features that a user has access to over the network. To create an effective IDS system, the user access over the network has to be predefined. The entire process needs to be based on information confidentiality, and the designer must consider all the appliances when designing an effective authorisation process for the users [20].
- **Authentication:** This process is referred to as the verification process. The IoT devices have unique hardware characteristics. Due to these unique characteristics, the classification of the objects in the early stage is crucial. These types of devices have a vital confirmation process to provide authorisation to the users. Due to this reason, the exchange of information between two variables such as IoT environment and data within the IoT devices will increase and provide adequate service to the users.
- **Data Confidentiality:** Data confidentiality is one of the critical aspects of IoT devices. Suppose the business does not practice effective data confidentiality practices. Then there can be severe consequences. For example, in the health sector, if the company does not practice an efficient approach, then the incident of the data alteration, and data modification should not be allowed to the users, if there is any change occurred by the users, then it leads to severe consequences [20]. Hence, confidentiality is the critical factor for getting effective results of IDS system
- **Data integrity:** Data integrity is one key deterministic factor of a reliable IoT system. The IoT system is considered to be a heterogeneous system. The devices have several components such as internet devices, sensor devices and computational devices [20]. Due to this various system inclusion, it is essential to practice and define data integrity. Data integrity within the IoT device can be achieved by verifying the source of the data. Along with that, the IoT team needs to report the malicious attack within the system.
- **Data Availability:** This is a vital process of IoT devices. The availability can be defined as the gathered information from the IoT framework which is accessible to users. If the device fails to

provide access, the device will fail, and the users no longer want to use such devices [20]. It is an essential aspect of the IoT devices that will allow the users to have access 24*7, and along with that, the devices must be aware of potential vulnerabilities in the network and provide users with a safe and secure network to access at any given time.

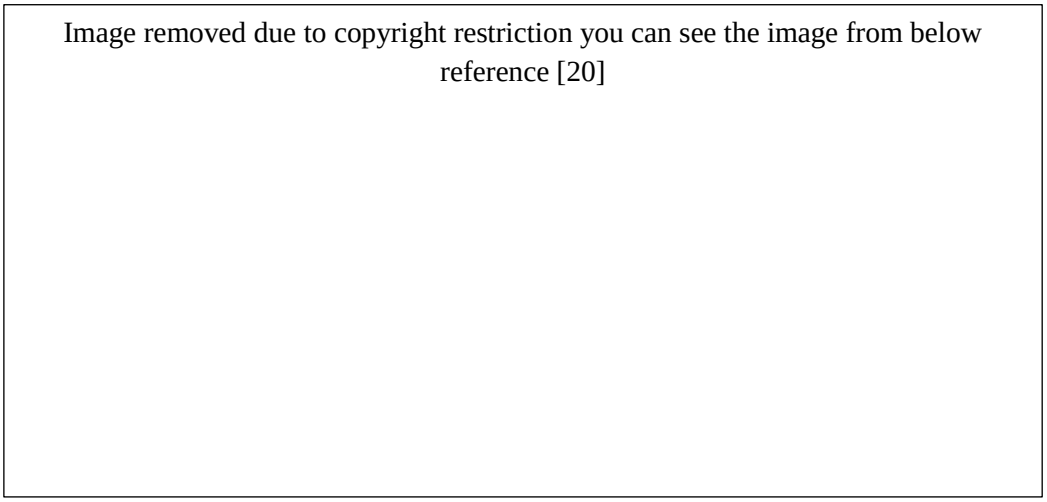

Figure 2. Different Intrusion Detection [20]

Above **Error! Reference source not found.**demonstrates the different types of intrusion detections. Figure 2 hows the multiple approaches to detect intrusion within the network. These types of systems can be classified as a device, or the software responsible for detecting the malicious activity or any violation within the network is known as the intrusion detection system [20]. Some sophisticated systems are further responsible for recovering the attacks and enabling the system to state agile and prevent such malicious attacks; those systems are known as intrusion prevention systems. There are multiple approaches to detect intrusion within the system. The first approach is known as Network intrusion detection systems" or shortly known as NIDS. Another approach is known as Host-based intrusion detection systems" or shortly known as HIDS [21]. The first approaches analyse the entire network and enable the stakeholders to get the crucial data regarding the intrusion. The other approach is based on monitoring operating system files to identify the intrusion.

Along with these two methods, two additional subsets are signature-based and anomaly-based. The first subset allows the IDS to use the pattern recognition system to identify the similar kinds of the attack and enable the users to take the safety measures. However, the pattern recognition system cannot detect the new types of attacks; it only prevents similar attacks encountered in the past. To further improve the feature of detecting the attack in the network. Recently, anomaly detection methodology has been used in the IDS system. This methodology allows the IDS to be more proactive in the new kinds of attack and enable the IDS to make the proper measures that allow the IDS to be more effective in intrusion detection and provide better results for detecting the intrusion and enable the users to have access in the safe and secure network. The new approaches practice the ML algorithm to detect the new kinds of attacks in the network. The sudden surge in the malware attack and the anomaly detection enables the users to have a safer and more secure environment against malware attackers and enables the company to practice the best security protocol. Figure 2 demonstrates the exciting insights of the IDS system. It depicts both the approach of the IDS

system; from Figure 2, it can be observed that the key objective of this IDS system is to monitor the data traffic in the network. There are two approaches to monitor the network's data traffic: the host-based sensor approach and the network-based sensor. All these approaches are instrumental in finding the intrusion within the network. The host-based approach uses anomaly detection to detect the intrusion within the network. These are highly effective in detecting new attacks. This approach uses DL methods such as GAN, Convolutional Neural Network (CNN), and many more to provide effective results. In other approaches, the network-based approaches monitor data traffic using the pattern recognition system to detect the misuse of any devices within the network.

According to Zixu, the operation of the entire IDS system can be divided into three phases; these three phases are universal for both the different approaches. These three phases for detection are monitoring, analysis and detection [22]. The monitoring is the first phase of the IDS system, which allows the IDS systems to monitor data traffic in the network. The second phase is the analysis. This phase is based on analysing the data traffic in the server per second. During this second phase, the different approaches use different processes; for example, the host-based uses the feature extraction, and the network-based approaches use the pattern identification process [23]. The last stage is the detection stage which is liable for detecting the intrusion within the network. IDS helps to improve data confidentiality and data integrity in the heterogeneous computing system.

According to Ferdowsi, The IDS system is a widely studied area. It enables the business to get effective results in detecting the anomalies within the network, empowering the industry to practice a secure and safe framework to protect the user's data [24]. However, there are a few challenges that have been observed in using this IDS system. The typical IDS system helps detect intrusion for moderate datasets and enables businesses to detect real-time threats in the network. However, when the dataset is large, the IDS system is ineffective, and the healthcare and financial industries contain sensitive information of the users [25]. The dataset of these two industries is private. Hence, the conventional methodology is not appropriate to get to the solution.

According to the Shah, using the GAN network will help mitigate this existing problem, increase the performance of the IDS system, and enable the IDS system to detect anomalies within the network in an efficient manner [26]. According to the author, the accuracy of the combination for GAN using IDS is much higher than the conventional IDS methodology. The author stated accuracy of using the GAN network in the IDS system has increased by 20%, and precision has increased by 25%. GAN network is based on the unsupervised machine learning module. Currently, this technique is one of the emerging techniques for unsupervised anomaly detection [27]. Currently, the GAN architecture is used in CV and image processing. This technology uses the centralised GAN module, creating communication overhead within distributed computing, especially in heterogeneous computational infrastructure. Using the centralised GAN can increase complexity and get effective and efficient results from GAN architecture. An association rule mining algorithm needs to be implemented, which will increase the computational cost. Along with that, it will also increase complexity in implementing the algorithm. The primary objective of the GAN network is

to learn from the unknown probability distribution of the population [28]. The entire sample data is collected from that unknown probability distribution. Once the model is trained, then the network will generate new observations from the existing sample dataset. The GAN network consists of two crucial functional parts that enable the network to create more effective observation and enable the researchers to get more insight into the existing dataset. The functional parts of GAN networks are: The first part is Generator, and the second is Discriminator.

Image removed due to copyright restriction you can see the image from below link

https://developers.google.com/machine-learning/gan/discriminator

Figure 3. Backpropagation in Discriminator [29]

Figure 3. Backpropagation in Discriminator [29] demonstrates the GAN network within the system. It can be identified that there are essential components in the entire network structure. The model generates the new sequences, and this type of network is created to detect the new attack within the system. There are various forms of GAN architecture present in the deep learning algorithm; this is just one of the general structures of the GAN network, which allow the system to create a more accurate image. The two elements of the GAN architecture. The first one is the generator and the second one is the discriminator. The generator job is to generate the data, whereas the other discriminator job is to differentiate the fake data and the actual data; if the resultant data is fake, then the generator improves the data accuracy and further enables the higher accuracy in the generated data. Both these elements of the GAN architecture is based on NN. The output of the generator is connected to the input of the discriminator. The discriminator uses the backpropagation algorithm. This is one of the fundamental algorithms on NN. It enables the system to perform gradient descent. To provide the practical result, first, the output of each node within the network is calculated. The resultant value is stored in a forward pass [30]. In the next step, the graph indicates the partial derivative of the error with each parameter in the backward pass. The above Figure 3 demonstrates that the generator creates the sample. The sample is connected to the discriminator module; the discriminator module is connected to the two major components: discriminator loss and generator loss. The discriminator discriminates between real and fake data and classifies both kinds of data within the system to differentiate between the data. The loss function reproaches the discriminator; the backpropagation algorithm updates the entire loss function through the discriminator network whenever the discriminator misclassifies the data.

According to ying, the distributed GAN based IDS for IoT system effectively detects the intrusion within the network. The author further stated that standalone GAN based approaches would enable the system to detect the internal intrusion. however, in external attacks. This methodology will not be helpful; hence using the distributed GAN based approach is critical. Every IoT device consists of distributed computed devices that enable the system to generate data from the various segments and enable users to real-time data, to create a more practical approach to reduce the vulnerabilities and enable the system to be incorporated with the practical traffic monitoring approaches. The Distributed GAN is an excellent approach in detecting external intrusion in the various other computing devices and enables the system to be more secure and safe. According to the author, the goal of decentralised GAN is to find a discriminator through each IoT Device (IoTD) without exchanging information, so each IoTD's discriminator could tell if a new piece of evidence matches the overall data distribution. The critical distinction between an isolated IDS and a decentralised IDS would be that a standalone IDS trains to match a new sample to its own data distribution [31]. The suggested decentralised IDS, but on the other hand, allows each IoTD to evaluate a new instance to the distributions of the existing data. As a result, because each IoTD's discriminator recognises the distribution of the entire dataset inside the decentralised IDS, each IoTD may recognise infringement on all other IoTDs" as well, even though a single GAN could be able to recognise intrinsic invasions. For example, if an adversarial has tampered with the IoTD's information, an attacker could also exploit the IDS to remain undetected at the targeted IoTD. Implementing a centralised IDS that analyses all the IoTDs is one way of stopping the adversary from being inconspicuous in intrinsic incursions. Due to the enormous characteristics of an IoT system. However, the communication overhead in this technique could be exceedingly significant. Furthermore, the centralised IDS must have accessibility to all the IoTDs' information in this circumstance, which might not be realistic in a confidential IoT network.

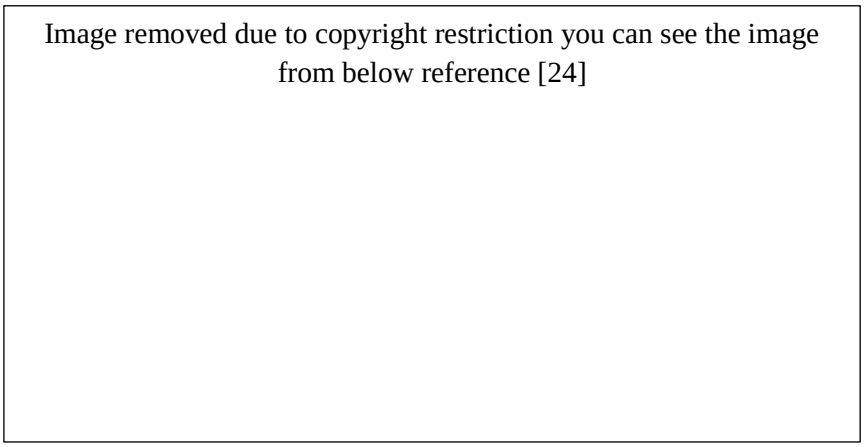

Figure 4. Different phases of the GAN network [24]

Figure 4 demonstrates the different phases of the GAN network within the IDS system. The entire architecture consists of two phases. The first phase is the training phase, and the second phase is the intrusion detection phase. During the preprocessing step of decentralised GAN, the author's proposed a centralised system with a generative Model G, where the strengths of the generator's ANN [32]. Additionally, each IoTD only has one discriminator, which is indicated by Di, wherein ‘I’ is the strengths of each discriminator's

ANN [33]. Each IoTD in a wireless system is connected to at least another IoTD in this configuration, requiring the IoTDs' connectivity network to form a process. Likewise, each IoTD is an interconnected unit throughout the training process. T is the number of epochs throughout which IoTDs interact with the centre, whereas E is the number of epochs during which IoTDs connect. Then there is no need for such a central entity just after decentralised GAN converges because all of the discriminators at IoTDs would be able to perceive the invasion to the network. As a result, each IoTD will pass its recorded real-time data itself via discriminator as well as the discriminator one of its neighbours. For a standard state piece of information, the best discriminator would produce 0.5. Accordingly, the result of the discriminator could be evaluated to 0.5 to determine a network intrusion, and if the result is near 0.5, the IoTD would be in a resting condition. If the result is near zero or 1, then the network is considered under attack. Because every IoTD can examine its nearby discriminator data, this strategy allows the IoT system to recognise an intruder without relying on a central entity [34]. The recommended networked GAN-based IDS is an excellent tool for detecting intrusion detection phases in the entire network.

According to Cheng, there are three main characteristics of the attack. These characteristics define the classification of the attack. These three characteristics are Influence, Security Violation, and Specificity [35]. Causative attacks fall under the influence of learning. These attacks alter the entire process. In the same category, the next attack is Exploratory attacks, this attack is responsible for causing the DoS, and further, it rejects good input. The second characteristic is a security violation responsible for an integrity attack that compromises assets via false negatives and only accepts malicious input. The next one is Availability attacks which cause the DoS via false positive. The third characteristics specificity, the attackers focus on the particular scenario to exploit, and the last one is indiscriminate attacks lets several variables without inspecting those variables. These are the fundamental combination that the attackers can use. Typically, a hacker chooses one characteristic from each category as choosing from the same category will not be effective for hackers. IDS system is more susceptible to exploratory attack, which emphasises the entire system's flooding by the mass input request. All these requests are false negative, which confuses the existing authentication system and result in four different attacks. These attacks are DoS, Probe, R2L, and U2R [36]. These attacks have different outcomes.

According to Huang, with rising potential attacks, ensuring data security, particularly inflexible and unstructured ad-hoc channels, is becoming more and more essential. Intrusion detection, essentially determines abnormal activity according to road characteristics, is a fundamental element of cyber protection. However, class-imbalanced data posed a challenge in the case of significantly lower anomalous values than normal samples [37]. This difficulty with class imbalances restricts the effectiveness of intrusion processors to unknown abnormalities. The author suggested a new GAN network for the adversarial unbalancing of the class disruption challenge. The authors model's main innovation has used standard GAN with an unequalled data filter and fully convolutional layer, creating new illustrative examples for different classifiers. In order to deal with subclass unbalanced intrusion detection, a GAN oriented intrusion detection system called the

GAN-IDS is also built in the GAN instances. Specifically, IGAN-IDS comprises three components: extraction of functions, IGAN and the profound neural network.

According to Song, the author mainly focused on reducing the research limitation and proposed a more effective network for detection. The model's decision limits are influenced by adverse defensive techniques against opposing instances, as such numerical simulations are unaltered over a narrow area of the inputs. This goal is nonetheless optimised about training examples [38]. Two novel techniques of assessment that leverage resilient representations to the structure and composition of adverse altered data. The empirical assessment shows that adversary defensive strategies can enhance the risk of the target image against inference-related assaults in comparison with (unsupported) spontaneous training strategy.

According to Seo, DNN is a practical approach to assisting in detecting the attack. Possible ways of adverse attack are essential to comprehend how strong and exceptionally well tested DNN can be built. In this work [39] author presented an algorithm of black-box attack that can beat both Vanilla DNN and the protection methods which were previously introduced. The technique identifies a probable population distribution across a limited area centre on the input because a sample taken from this distribution is probably an adverse example, knowing the underlying layers or parameters of both the DNN.

In the research of Hasan et al., GAN is a powerful tool for deep learning. IDS is another essential method in Cyber-Physical Systems (CPS). Both GAN and IDS models were combined and implemented on the NSL KDD'99 dataset to get the results that significantly perform well than impartial IDS. Experimental analysis shows that new developed GAN-IDS model predicts with greater accuracy than standalone IDS [40]. Figure 5 describes the full implementation of the GAN-IDS framework, which is included in the four different modules. However, the GAN-IDS model proved as centralised, time-consuming and computationally complicated. In the modern IDS solution, the NSL KDD'99 dataset loses its relevance. So, using the Bot-IoT dataset, which is more reliable for the current IDS, there is essential to improve the algorithm, which should be decentralised, efficient and dynamic.

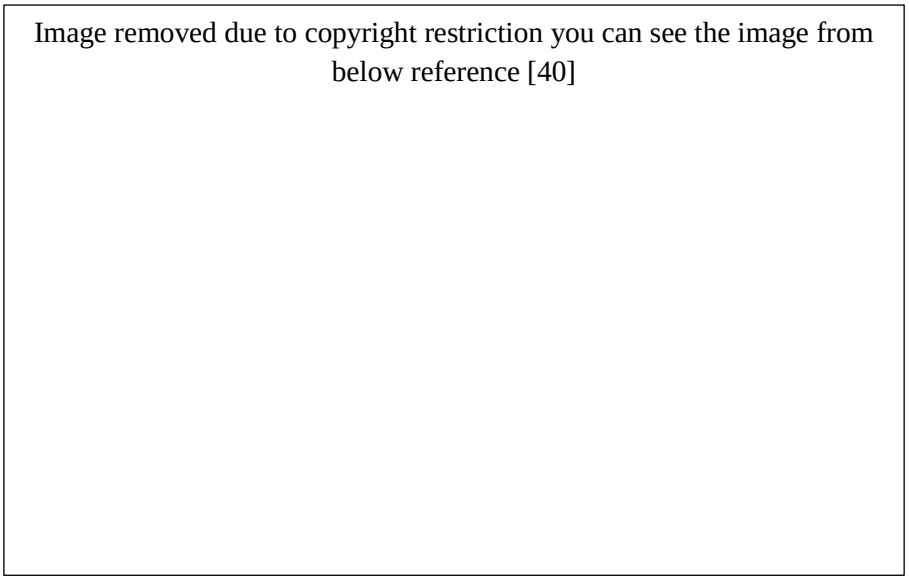

Figure 5. GAN helped IDS [40]

3.2 Theories and Models

3.2.1 Internet of Things (IoT)

The IoT relates to specific objects or groups of these kinds of objects equipped with sensors, computing power, firmware, and other technologies, but that communicate and interchange information between devices and organisations over the Online platform or even other network infrastructure. The subject has progressed because of diverse technologies, such as interconnected devices, inexpensive sensors, immensely influential embedded devices, and deep learning. Conventional domains like embedded applications, wireless sensor networks, control systems, and automation such as home and building automation enable the IoT individually and collectively [41]. Inside the mainstream market, IoT Infrastructure is most closely involved in the design that supports the concept of a smart home, such as light fittings, heating systems, security systems and camera systems, as well as other household appliances which can be governed by diverse mechanisms, such as smartphones and smart speakers. The IoT can also be used in medicine. Together with that, various industries are using IoT technology. Many efforts have been made about the risks associated with the emergence of IoT technologies and services, specifically in the realm of privacy protection. As a result, private and public sector attempts to address such concerns have started to establish global and local benchmarks, instructions, and legal requirements. This research project has used the advanced AI technique to improve the privacy mechanism in IoT devices by deploying the IDS system, and the entire IDS system will be based on GAN architecture.

3.2.2 Deep Learning (DL)

Deep learning is one of the advanced mechanisms of Artificial Intelligence (AI). The entire function of deep learning is based on the neural network. The neural network mirrors the human brain functionality and enables the machines to perform a particular task. The neural network consists of three layers, allowing the layers to perform the crucial computational logic to get the desired data. Several incidents have been observed. The accuracy of the deep learning methodologies has surpassed human efficiency and enables the researchers to get insightful information about the various aspects that strengthen the current understanding of certain aspects. Deep learning has versatile applications in various industries and enables the researchers to get more insightful information about the existing problem and get more innovative ideas to solve the existing problem. All the (DL) algorithm variations use three different approaches to learn from the data that enable the researchers to get meaningful information. These three approaches are supervised, semi-supervised and unsupervised [42]. These learning approaches are applied to the hidden layer of the neural network, which enables the system and the network to learn from the entire dataset and enable the researchers to get valuable insight into the output layer.

3.2.3 Intrusion Detection System (IDS)

An intrusion detection system is a hardware and software program that analyses networks or devices for fraudulent attacks or regulation. Any intrusive activity or infringement is often reported to an administration or gathered centrally to use a Security Information and Event Management (SIEM) system. SIEM system

intelligent to investigate security alerts from network hardware and applications in real-time. A SIEM system captures data from various streams and employs alarm thresholding techniques to differentiate between intentional and false alarms [43].

Usually, three primary methods are popular for network security. In the first method, a few triggers are set up to detect the attack on a network as the threshold value is exceeded. This detection approach informs the administration upon the threshold level but does not prevent the network from attack. The second option is to avoid an attack by implementing defence policies that prevent the attack, but this poses an issue if a valid operation is judged illegitimate, resulting in a denial service. The last method to block an attack is setting up the protection system that reviews the attack and prevents the attack when it happens in the future. In the last two methods, the intrusion detection system (IDS) is setting up an intrusion prevention system (IPS) [44]. The IDS can be configured on two locations based on the source of information. Firstly, the sensor can be fixed on a Host-based Intrusion Detection System (HIDS) or, finally, can be set up on the Network Intrusion Detection System (NIDS) [45][46]. The HIDS is a system to monitor an essential part of an organisation's computer system, whereas a NIDS is a method that analyses traffic on the network. IDS can also be classified using a detection strategy. The Network Intrusion Detection Systems are mainly responsible for observers of the network traffic by inspecting the different parameters like protocol usage, packet inspection, and the checking of IP addresses [45-47]. An IDS is a crucial tool for the network's security, and its success is mainly calculated by the accuracy of predicting legitimate and illegitimate events.

3.2.4 Literature Gap

The literature gap can be defined as the particular research segment where the researchers have not done the research yet. In other words, it can be stated as the uncharted territory of the research that researchers are entirely unaware of. Empirical studies are one of the effective ways to understand the current gap in the research study and enable the researchers to look for new opportunities to explore the uncharted territories that allow the researchers to effectively understand the existing research problem. This approach also enables researchers to get more innovative solutions for the existing cyber security problem.

While conducting this research, a few things have been observed; most of the DL studies are based on the application of the medical industry. The deep learning algorithms efficiently create high-resolution images, allowing the doctors to get more insight into the patients' health conditions and make more practical steps to help the patients recover. However, there are various other fields where deep learning has promising potential and can increase the efficiency of other industries. One example is the cyber security domain. Cyber security has become more relevant today due to the rise of malicious attackers. The existing methodology of detection intrusion is no longer effective. Hence new methodologies need to be developed to counter the malware attack. Finally, the GAN architecture is relatively new; the architecture was proposed in 2014. The architecture is already famous in digital image processing. However, the technology can be used to improve various other technologies. GAN based IDS architecture is still at the preliminary stage. However, some study conducted on this architecture, but the dataset they used is too old and redundant.

Hence various researches need to be conducted to understand the overall aspect of the technology. The architecture shows promising results in detecting exploratory integrating attacks; nevertheless, hackers might use some different kinds of attacks. Hence the evaluation of the different characteristics needs to consider.

3.2.5 Conceptual Framework

This section will provide information about the conceptual framework of the research project and enable the readers to understand the concept in a better way which will further help them to create more robust and innovative solutions to improve the research problem and along with this approach will enable the researchers to get a better understanding of the problem. There are two crucial variables in this conceptual framework that will help the researchers and the readers to understand the dependencies and relationships with the other variables. These two variables are independent variables and dependent variables. For this research, the independent variable is the IDS system, and the dependent variable is GAN architecture; within the architecture, there are mainly three elements that are responsible for improving the performance of the algorithm and enabling the entire system to function effectively and efficiently and enable the system to conduct an effective search which allows the systems to detect the intrusion within the network.

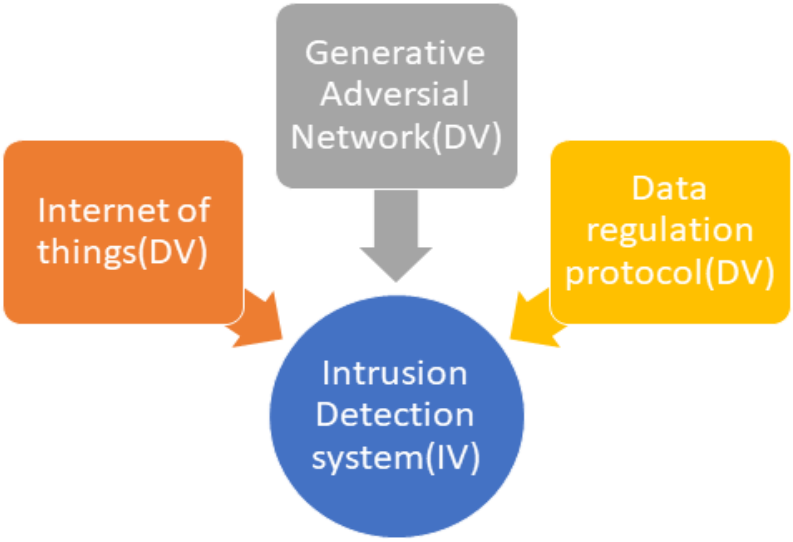

Figure 6. Conceptual Framework

These critical parameters of dependent variables are discriminator, generator and backpropagation algorithm [48]. These are three main components of the algorithm that determine the accuracy and efficiency of the entire network and enable effective invasion detection, promoting more safe and secure networks for the users. The generator here generates the sample. The output section of the generator is connected to the discriminator section [49]. The discriminator section evaluates the fake and real data and enables the entire network to produce more accurate data. Suppose the discriminator detects fake data then, the generator will create more accurate data to improve the algorithm's accuracy; the entire system uses the backpropagation algorithm to increase the accuracy of the given dataset.

4. METHODOLOGY

4.1 Dataset Overview

There are many trained ML models available using the dataset of the literature. There are options for the dataset such as UNSW-NB15, KDD-CUP99 and Bot-IoT. The UNSW-NB15 and KDD-CUP99 lost their relevance according to the modern IDS and they also have some permission issue. However, the Bot-IoT dataset is the new dataset and in its early days. The Bot-IoT dataset was developed by the Cyber Range Lab of UNSW Canberra from the attack and victim machines in the stimulated network environment [16]. In this environment, the network traffic is captured in argus and pcap files, then exported into comma-separated value (CSV) files after proper processing and analysis. The Bot-IoT dataset consisted of more than seventy-three million instances with forty-six feature values and three classification types. The dataset was categorised into five types of classes (Normal, Reconnaissance, DDoS, DoS, Information Theft) and randomly extracted the 5% sample with 19 features from each class [16]. The total counts for each category are presented in Table 1[14].

Table 1: The total count of each category in the Bot-IoT dataset.

Category	Total Samples	5% Samples	Training Samples	Testing Samples
DoS	33,005,194	1,650,260	1,320,148	330,112
DDoS	38,532,480	1,926,624	1,541,315	385,309
Reconnaissance	1,821,639	91,082	72,919	18,163
Theft	1587	79	370	14
Normal	9543	477	370	107
Total	73,370,443	3,668,522	2,934,817	733,705

This research only uses one million rows from the 5% version of data. Table 2 demonstrate the data used in this project. Eight hundred thousand rows are used for training purposes and two hundred thousand for testing purposes.

Table 2. Dataset used in this research

Category	Testing Data	Training Data	Total Data
Theft	3	14	17
Reconnaissance	4,911	19,694	24,605
Normal	107	370	477
DoS	89,797	359,899	449,696
DDoS	105,182	420,023	525,205
Total	200,000	800,000	1,000,000

For the feature selection, the processing steps of the study of Koroniotis et al. [16] was followed. The joint entropy and correlation coefficient methods were implemented for feature selection from the Bot-IoT dataset. For the first approach of our research, the top ten score features were extracted using the score metrics of both algorithms (see Table 3). The selected features depended on the dataset's three classification

features (attack, category, subcategory). The attack is the first classification feature proposed for binary classification. The label of attack feature was either true (1) or false (0) directly mapped with malicious and normal traffic, respectively. All the classes of malicious traffic are labelled as True. The following classification feature of the Bot-IoT dataset is five class (DoS, DDoS, Normal, Information Theft, Reconnaissance) multi-classification feature labelled as a category. The category classification feature is consist of five-string values (class) shown in Table 4. The last classification attribute is a subcategory that is the most described form of attack. The subcategory is made up of ten classification values by dividing each category into a subcategory. Like the DoS category is further divided based on protocol (TCP, UDP, HTTP), reconnaissance is categorised into OS fingerprinting and service scanning, information theft is split into data theft and key logging. For the second approach, 14 features are used to achieve the desired results (see the Table 4).

Table 3: Top 10 features for the first approach of Research

Top-10 Selected Features	**Description**
min	Minimum duration of records
max	The average duration of aggregated records
stddev	The standard deviation of records
seq	sequence number
mean	The average duration of records
srate	Source-to-destination packets per second
drate	Destination to source packets per second
state_number	Numeric representation of transaction state
N_IN_Conn_P_SrcIP	The number of inbound connections per source IP.
N_IN_Conn_P_DstIP	Number of inbound connections against destination IP

Table 4. 14 Features used in the second approach of Research

14 Features along with 2 classes	**Description**
min	Minimum duration of records
max	The average duration of aggregated records
stddev	The standard deviation of records
proto	Transaction protocol
mean	The average duration of records
srate	Source-to-destination packets per second
drate	Destination to source packets per second
state_number	Numeric representation of transaction state
N_IN_Conn_P_SrcIP	The number of inbound connections per source IP.
N_IN_Conn_P_DstIP	Number of inbound connections against destination IP

Saddr	IP address of the source
Daddr	IP address of the destination
Sport	port number of source
Dport	port number destination
Attack	Binary Class 1 for malicious data and 0 for Normal data
Category	Category of traffic

4.2 Machine Learning (ML) and Deep Learning (DL) Models Overview

ML is an area of AI that enables machines to learn and progress independently without being statically programmed. ML is concerned with forming automated models that can retrieve data and self learn from this data. The training procedure instigates data observations, finds the patterns and outlines in data and save them in memory for future prediction. The ultimate goal of the algorithm is to learn self-sufficiently, without human involvement, and to change their behaviour consequently. The first stage of this study is to replicate the random forest model proposed by the study [16]. Replication aims to equate the machine ML and DL model using evaluation measures. The structure and activation function for normal and adversarial data was identical for a fair evaluation. The processes focused on generating adversarial data for the random forest, support vector machine, and ANN model. DL is an area of ML that deals with ANN, which are algorithms inspired by the human brain's biological nervous system and function. AI algorithms are usually divided into the following categories.

4.2.1 Supervised ML Model

A supervised ML model is a type of ML in which the algorithm is trained with the tagged or labelled data that help the model predict the output. This labelled data guide the machine to learn and predict the result more accurately. The machine algorithm compares the results with the labelled data’s output to improve its performance [50].

4.2.2 Unsupervised ML Model

On the other hand, Unsupervised ML techniques are utilised when the trained data is not categorised and the classes are not classified. Unsupervised learning inspects how computers might infer a function from unlabelled data to explore an underlying pattern. The system does not determine the appropriate result but investigates the input data and can indicate hidden patterns from unlabelled data [50].

As the proposed study is based on the BoT-IoT dataset in which all the samples are classified, our problem is to fall in the supervised machine learning category. The detail and methodology of different ML and DL algorithms for executing the proposed problem are as follows.

4.2.3 Random Forest Model

Random forest (RF) is a convenient and flexible ML technique that, in most situations, provides tremendous results even without hyper-parameter optimisation. Because of its flexibility, simplicity and adaptability, it is also one of the most extensively used ML algorithm (it can be used equally for regression and classification problems). RF is a supervised ML model and builds a 'forest' ensemble of decision trees usually trained on the 'bagging' method [51]. The first stage-trained the random forest model on attack and category classification features for binary and multiclass classification. The default hyperparameters of random forest are n-estimator=20 and random-state=0 in the scikit-learn package. The cross-validation technique was used to train the model rather than the train test split approach. The four cross-validations were used that split the 0.75% data for training and 0.25% data for testing in each fold. The cross-fold validation technique iterated the 4^{th}, 3^{rd}, 2^{nd}, and 1^{st}, quarter as testing set in each fold, respectively. Confusion matrix and evaluation measures were also generated by combining training and testing data as proposed in [16]. The confusion matrix also calculated the accuracy, f1-score, precision and recall. For the multiclass classification, RF was also trained with category feature as the label. The cross fold technique was also used rather than the train test split approach as the category feature was based on the five-string value (classes). So, before the model's training, the feature was encoded with the OneHotEncoder function of scikit-learn. The hyper parameters were tuned with the same values of the random forest classification model. The evaluation of the model was done by calculating evolution measures using a confusion matrix. The ANN was also trained for binary and multiclass classification on attack and category features,

4.2.4 Artificial Neural Network (ANN) Model

ANN is used to simulate complex systems and predict target values associated with the input parameters based on training experiences. ANN is based on the biological nervous system of the brain, but it uses the reduced form of biological neural network. ANN mainly simulate the electrical activity of the nervous system and brain. An ANN is composed of many cores that work parallel and are arranged in tiers (layer). The first layer obtains the raw input samples corresponding to the optic neurons in human visual perception. Similarly, as neurons auxiliary move from the optic nerve receives signals from those nearer to it, each subsequent layer takes the information from the previous layer rather than the raw input. The output of the system is created by the last layer [52].

The ANN model had a five-layer including one input layer and one output layer. The input layer was the map to the total sum of features and had ten nodes. The 3 hidden layers among the input and output layers are composed of 20, 60, 80 and 90 nodes [14]. All the hidden layers were fully connected convolutional layers. The input layer was directly mapped with the length of input features. The nodes on the output layer were either two or five for binary and multiclass classification. The activation function of the hidden layers is 'tanh', and Sigmoid on output layer for binary classification as sigmoid base models are more robust [53]. Softmax is the activation function used on the output layer for multi-class classification, and the architecture of both models is shown in Figure 7 and Figure 8. Both models were evaluated based on accuracy and loss function graphs.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 20)                220
_________________________________________________________________
dense_2 (Dense)              (None, 60)                1260
_________________________________________________________________
dense_3 (Dense)              (None, 80)                4880
_________________________________________________________________
dense_4 (Dense)              (None, 90)                7290
_________________________________________________________________
dense_5 (Dense)              (None, 1)                 91
=================================================================
Total params: 13,741
Trainable params: 13,741
Non-trainable params: 0
```

Figure 7:Structure of ANN model for binary classification.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 20)                220
_________________________________________________________________
dense_1 (Dense)              (None, 60)                1260
_________________________________________________________________
dense_2 (Dense)              (None, 80)                4880
_________________________________________________________________
dense_3 (Dense)              (None, 90)                7290
_________________________________________________________________
dense_4 (Dense)              (None, 5)                 455
=================================================================
Total params: 14,105
Trainable params: 14,105
Non-trainable params: 0
```

Figure 8: Structure of ANN model for multiclass classification.

4.2.5 Support Vector Machine (SVM)

SVM, a supervised ML algorithm, can learn data classification patterns with excellent accuracy. SVM is a popular tool for classifying data. It has high versatility and can be used in various AI and data science scenarios, including IDS research. The primary purpose of SVM seeks to locate a hyperplane in multi-dimensional space that classifies the data points. SVM is highly favoured in many ML problems because it provides high accuracy and uses less computing power. Although it could be used for classification and regression tasks, it is extensively used for classification purposes. The decision function of SVM is more accurately an ideal hyperplane that assists in distinguishing observations fitting to unlike classes based upon different features. The hyperplane can then assist in deciding the most prospective label of concealed

information. The variety of hyperparameters and kernel options, i-e linear, polynomial, sigmoid tanh, and user-defined kernel, enhance performance [54].

4.2.6 Generative Adversarial Network (GAN) Model

A GAN is a type of modelling in which samples are generated using DL techniques such as CNN (convolutional neural network) and ANN (artificial neural networks. In ML, GAN modelling is an unsupervised learning job that automatically detects and learns constancies or patterns in given data so that the system may be used to produce or output training examples that could have been taken from the given data. GANs are an innovative way of training a generative model by describing the problem as a supervised learning problem with two different models: the generator model, which we prepare to generate new instances, and the discriminator model, which tries to categorise examples as real (from the domain) or fake (not from the domain) [55].

The GAN is used in the first approach of the research project to generate adversarial data by adding noise in real data. GAN model was trained using generator and discriminator functions to generate fake malicious data. The above described CNN structure is also used in generator and discriminator functions of GAN. The generator model generates the new instances for malicious attacks, and the generator function predicts them as these are malicious instances or benign instances. The combination of training and testing data was used in GAN, and then one million fake malicious samples were generated for binary classification. The purpose of generated samples is to test the robustness of RF, SVM and ANN binary classification models. The already trained models are then implemented on newly generated data and evaluate the performance of models by seeing how much these malicious samples deceive the models.

4.2.7 Wasserstein Generative Adversarial Network (WGAN)

The WGAN is a type of GAN practised to measure the dissimilarity between model and target distribution using Wasserstein distance instead of JS-Divergence. This minor modification has significant implications; not only does WGAN improve its training efficiency and produces remarkable results [56]. WGAN seeks a different method to train the generator model to better predict the data distribution from the training dataset. In place of employing a discriminator to predict or identify images as actual or fake, the WGA create new or modify the existing discriminator model by an objective critic, which evaluates the image as real or fake. The advantage of using the WGAN is that the learning procedure is much more reliable and less susceptible to model structure and the choice for hyperparameters settings. The vital thing is that the quality of images produced by the generator seems to relate to the discriminator's loss [57]. WGAN is used to achieve the second research approach, which helps develop a GAN-based model capable of predicting malicious data as normal.

4.3 Software and Hardware

Due to the plenty of the packages and libraries, the python programming language is used to investigate proposed research. Mainly Scikit-learn package and TensorFlow (TF), along with the Keras library, has been

used. TF is an open-source platform developed by Google and Keras library is mainly allows for analysing data and use to develop the neural network. Google Colab and Jupyter notebook is used for the implementation environment. Google Colab is a project of Google research that offers free GPU and requires zero configuration. However, the free version of Google Colab only allows executing two notebooks at a time, and the code execution time of a notebook is a maximum of 12 hours [58]; after that, it disconnects. Jupyter notebook is used to overcome this problem where the algorithm takes excessively execution time.

4.4 Implementation Flow

For the first approach of the project RF, SVM and ANN are trained on the training data and attained the results on the testing data. After that GAN model is trained with the same training data used for the IDS models; at that time, this trained GAN model generated the new fake data. The newly generated data is tested on the already trained IDS models, which will produce different results. These results were compared with the IDS results acquired from actual data. Figure 9. Implementation of First Approach by using GAN demonstrate the whole implementation method.

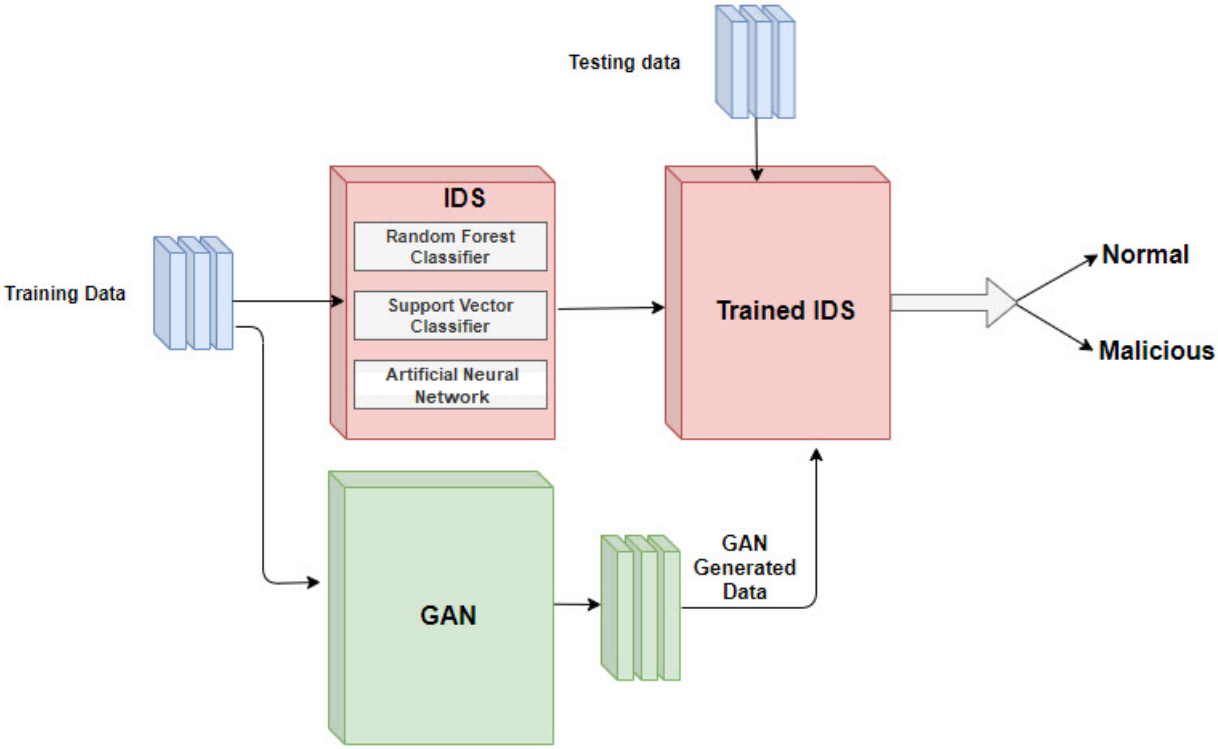

Figure 9. Implementation of First Approach by using GAN

In the second research approach, RF and SVM are trained and tested. Subsequently, using WGAN develops a GAN model on the actual data that is intelligent enough to predict all the malicious data as benign and compare the results with the IDS models (RF, SVM). Figure 10. Implementation Flow of Second Approach by using WGAN describe the implementation flow of this approach.

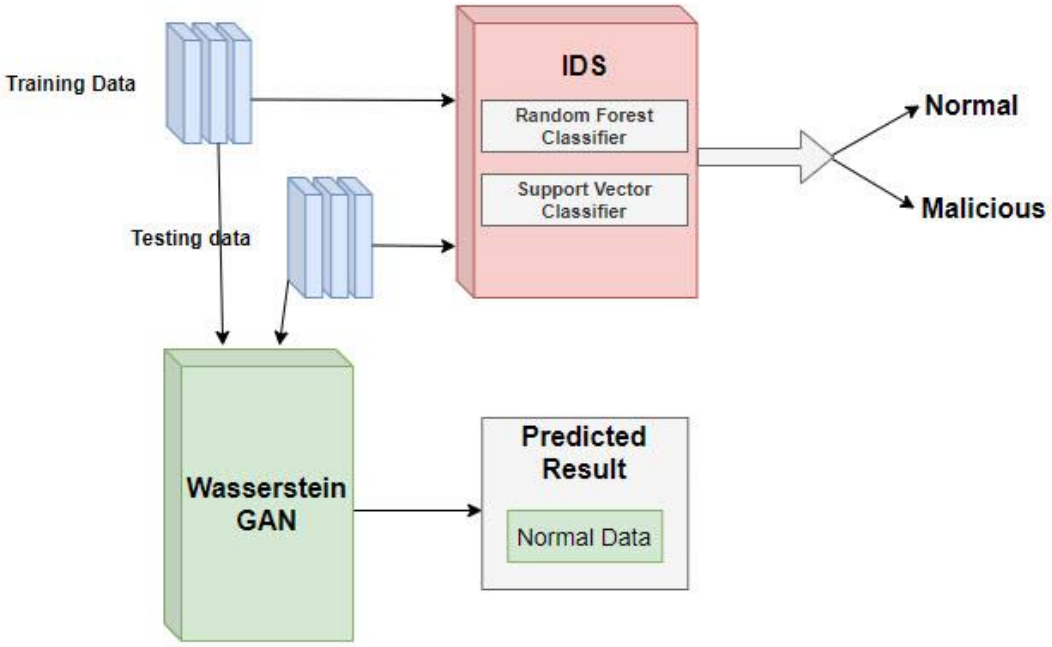

Figure 10. Implementation Flow of Second Approach by using WGAN

4.5 Evaluation Measures

The last stage of this research is to evaluate the results of the trained model using by comparing actual data with generated samples. Numerous evaluation measures including precision, recall, F1-score and accuracy were used (Eq. 1-4) to evaluate all models. These evaluation measures were also calculated by generating a graphical representation of confusion metrics rather than numeric values. The recall score is considered a key evaluation measure. The decrease in values showed that the model increases false negatives and increases false negatives allow hackers to launch cyber security attacks that the trained model would not detect. For the ANN, binary and multiclass classification accuracy and loss function graphs were also plotted to evaluate both models. For the generated sample same evaluation criteria were used for a fair comparison.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (Eq. 1)$$

$$Precision = \frac{TP}{TP + FP} \qquad (Eq. 2)$$

$$Recall = \frac{TP}{TP + FN} \qquad (Eq. 3)$$

$$F1\ Score = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \qquad (Eq. 4)$$

5. RESULTS AND DISCUSSION

The Bot-IoT dataset was downloaded and prepared for ML and DL models. The Bot-IoT dataset total has 19 features that can be used for statistical analysis. The features saddr, daddr, proto, dport, and sport are used to uniquely identify the data points and therefore removed from the training set for model deployment. The

pkSeqID feature is a sample identity key that is also removed from the training set. Then the feature selection was made to reduce the number of features to avoid overfitting and get significant results. The feature selection methods aim to find those features that play a significant role in training models. Different types of statistical features are implemented on the features to rank the features according to their significance. Lastly, the top significant features are selected as training data. Correlation and joint Entropy feature selection algorithm were used and extracted top ten features of Bot-IoT dataset for training Table 3: Top 10 features for the first approach of Research. As the dataset is too large and imbalanced in terms of classes, selecting the whole dataset for training is not suitable. For the manageable training of the models, the 5% samples of the dataset were extracted, resulting in 3.6 million samples [16]. Then further 1 million samples are selected for this research and split into training and testing set with the ratio of 80 and 20%, respectively. After splitting the dataset, the training set has 800,000 instances, and the testing set has 200,000 instances.

Secondly, the training features data and attack classification feature were already numeric and did not require label encoding. However, the category classification feature used five classes in string format that required the label encoding. The five classes of category features were encoded using the oneHotEncoder function of scikit-learn. The oneHotEncoder function of scikit-learn replaces each class category feature with one corresponding numeric value. The DoS, DDoS, Normal, Reconnaissance, and Information Theft are replaced by the 0 to 4 numeric values. By replicating the Scaling, the data had normalised without training and testing split to equally scale the data. The MinMaxScaling Function of scikit learn was used to normalise data between the range of -1 and 1 according to Eq. 5 and 6. After preparing the dataset, the machine learning and deep learning models were trained on trusted (normal) data and then tested on trusted and manipulated data. Lastly, the result of trusted data compares with manipulated data using evaluation measures.

$$X - Std = \frac{X - X_{min}(axis = 0)}{X_{max}(axis = 0) - X_{min}(axis = 0)} \quad (Eq. 5)$$

$$X - Scaled = X - Std * (\max - \min) + \min \quad (Eq. 6)$$

5.1 Results of First Approach

The random forest is a ML model that was trained on trusted data for binary classification. The hyperparameters of the RF was the same as the default parameters used in scikit-learn module except for the n-estimator=20 and random-state=0. For the RF, the four cross-validation technique was used rather than the train test split approach described in the methodology section. In four cross-validate, the data is split 75% for training and 25% in testing for each iteration. The evaluation measures were also calculated with cross-fold validation. Random forest showed 99% accuracy for trusted data, and the graphical confusion metrics with cross-validation results are shown in Figure 11. By viewing the confusion metrics, the model inaccurately predicts the 10 malicious attacks as benign. Appendix A shows the implementation of the RF model.

The random forest model was also trained on category feature columns for multiclass classification. The hyper parameters were the same as for binary classification. Random forest used the top ten features in Table

3: Top 10 features for the first approach of Research as input and the category feature as output. The label encoding scheme was used on the category feature to encode the binary value into numeric values. The oneHotEncoder encodes the DOS, DDos, Normal, Information Theft and Reconnaissance into 0, 1, 2, 3, and 4, respectively. After the model's training, RF multiclass classification model showed 99.99% accuracy for trusted data. Figure 12 showed that the model inaccurately predicted the 5 malicious samples as benign instances. All the defined evolution measures for this study were calculated using a confusion matrix. Appendix A indicates the implementation of RF for multiclass classification.

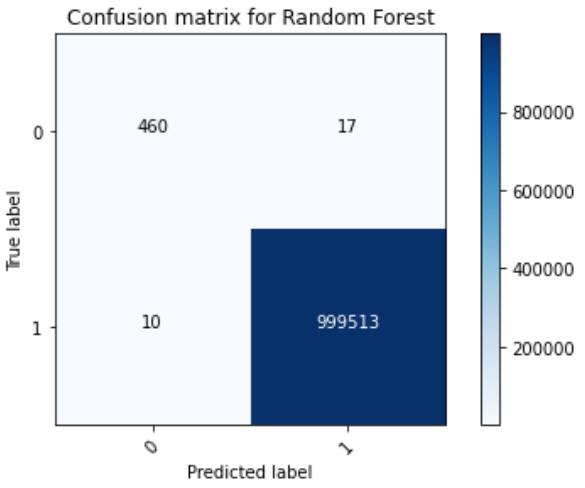

Figure 11. Confusion Metrix of RF for Binary Class

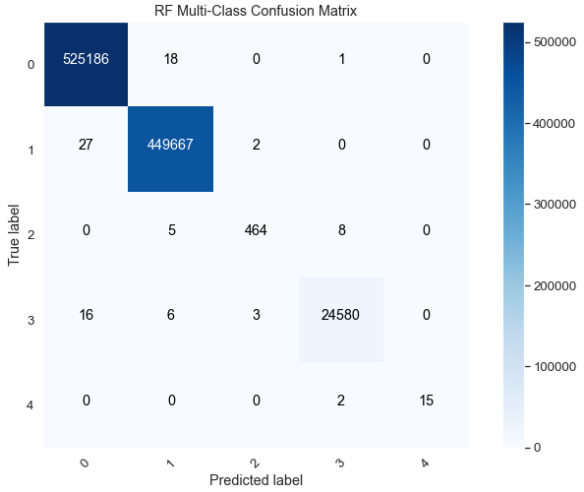

Figure 12. Confusion Metrix of RF for Multi-Class

The ANN model was also trained for binary and multi-classification on attack and category features, respectively. Both models were made with five layers, including input and output layers. The train test split on data was used for the training of both ANN models. For Binary classification, the sigmoid activation function was used with the Adam optimiser. The 'sparse_categorical_crossentropy' loss function and accuracy evaluation matrix was used with 20 epochs. The ANN showed 99% accuracy for binary classification. The loss score for the ANN binary classification model was minimal. The accuracy and loss of the model are presented in Figure 13 and Figure 14, respectively. The graphical confusion matrix showed that the model false-negative rate is nearly zero. Hence, the model can detect malicious traffic over normal traffic in the real environment. The Softmax was used as an activation function on the output layer with Adam optimiser for the multiclass classification. The rest of the parameters were the same set as for the binary classification model. The multiclass classification models showed 97% accuracy with near to the ground loss score. The accuracy and loss score of the multiclass classification model is presented in Figure 15 and Figure 16. Further, the confusion matrix in Figure 17 and Figure 18 demonstrate the exact accuracy measure of ANN for the binary and multiclass classification. Appendix C shows the detailed implementation of the ANN model for binary and multiple class classification, respectively.

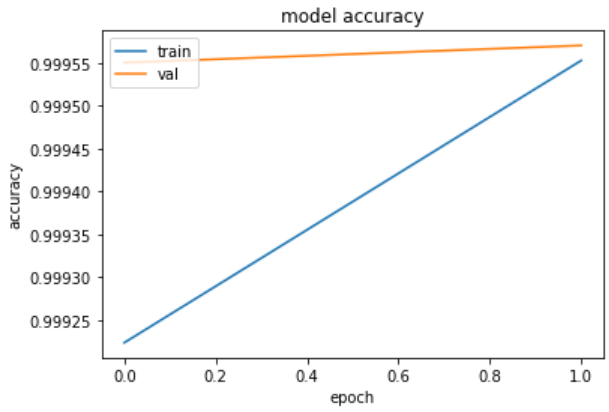

Figure 13. Binary class accuracy plot

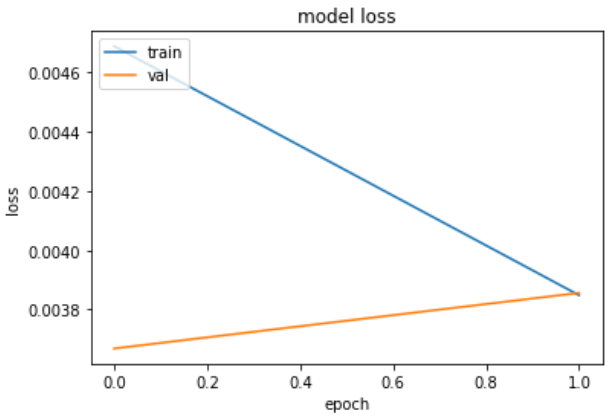

Figure 14. Binary class loss plot

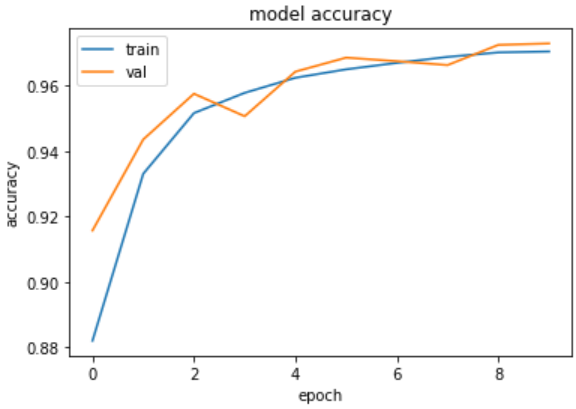

Figure 15. Multiclass accuracy plot

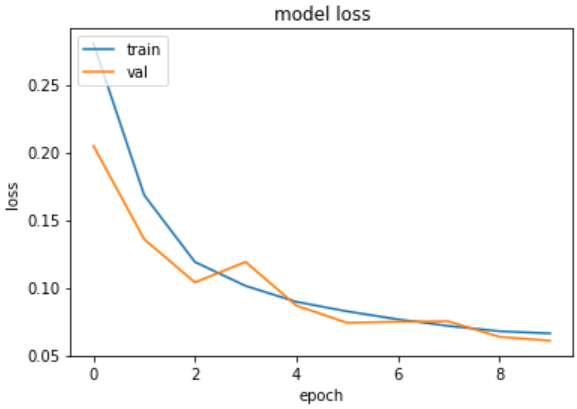

Figure 16. Multiclass loss plot

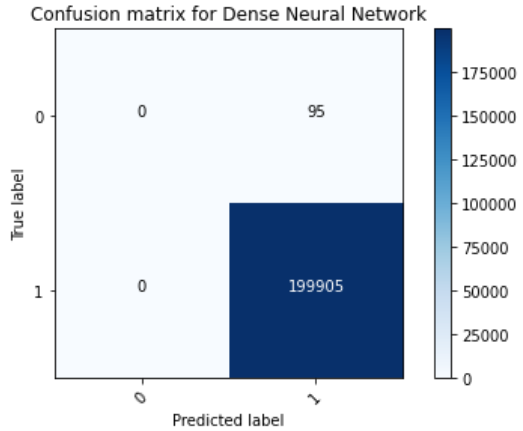

Figure 17. Confusion matrix of ANN for binary class

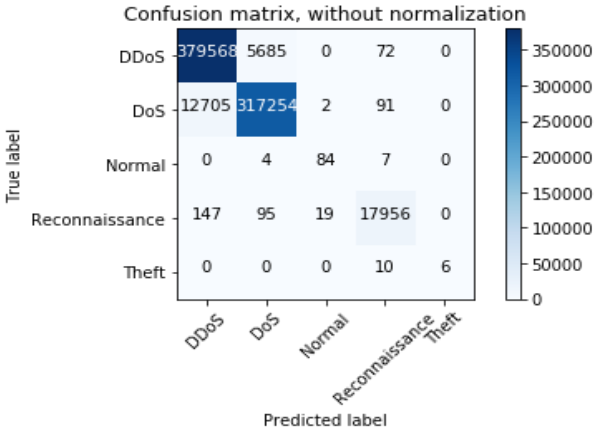

Figure 18. Confusion matrix of ANN for multiclass class

In the same way, as RF and ANN model was trained for the binary and multiclass, the SVM classifier is also trained. The hyperparameters for SVM was verbose = 1, random_state = 42 and used default kernel with 4 cross-validation. The accuracy with these parameters of SVM was 99.9%. The confusion matrix in Figure 19 graphically explains the classification report of SVM. These results are produced by the SVM classifier see Appendix B.

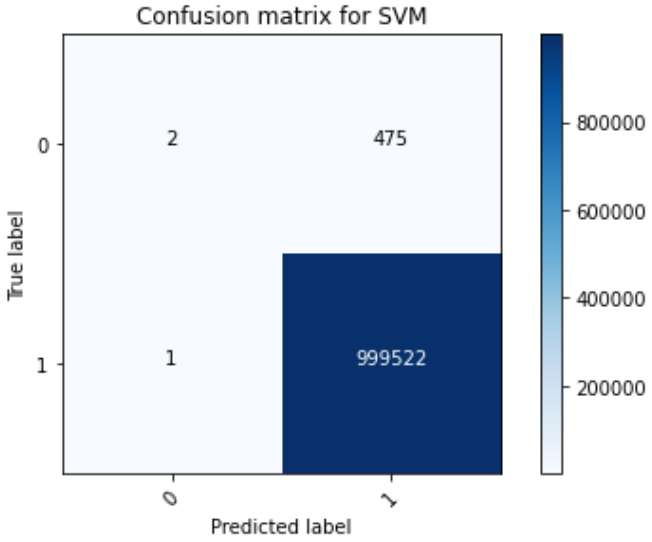

Figure 19. SVM binary class confusion matrix

Lastly, the GAN was trained on a combined dataset to generate manipulated data for binary classification. The GAN network was composed of a generator function followed by the discriminator function. The purpose of the generator function is to generate the samples, and the discriminator aims to categorise the real and fake samples. The 100 iterations were performed on the combination of training and testing data by filtering attack records only. It firstly induced the noise in trusted samples and then started training on them. The generator generates the sample after inducing noise in each iteration, and the discriminator categorises them into normal and malicious classes. After the 100 iterations, the generator model was trained enough to deceive the discriminator for malicious traffic. Then on the actual data, one million fake samples were generated by the trained GAN model. These generated samples were passed through the preprocessing pipeline line described in the methods section. After that, these manipulated samples were predicted by the already trained RF classifier and ANN model for binary classification to test the robustness of the model. To see the implementation code look into Appendix D and Appendix E. RF model was loaded and predicted with the one million generated samples. The accuracy of RF and ANN dropped to 50% and 38.63%, respectively, for the malicious samples generated by the GAN model. The model is again evaluated with predefined evaluation measures. The confusion matrix of RF for generated samples was plotted, and the graphical confusion matrix calculated all the evaluation measures. Figure 20 of the confusion matrix of RF showed that the model predicted the 495,839 generated malicious samples as benign. The False Negative (FN) rate was bearable for the manipulated samples to deploy in a real environment. The ANN accuracy was 38.63% on the fake data, and Figure 21 indicate ANN predicted 613,681 samples as normal data.

Appendix F explains the algorithm, and Figure 23 demonstrate the final results of the SVM that shows all the one million data is predicted as malign. The rest of the evaluation measures were also calculated the confusion matrix of the ANN binary classification model. The FN rate for manipulated data showed the robustness of the trained model. Figure 22 shows the comparison of all the three classifiers on trusted and GAN generated data. Graphical results explain that the accuracy of all the classifiers was nearly 100% on the trusted data. However, on the GAN generated fake data, the accuracy of RF and ANN dropped to 50% and

38.63%, respectively. On the contrary, SVM proved as more robust than the other model, the accuracy of SVM is 100% as all the data was maligned and SVM correctly predicted all the data malicious. Conversely, SVM was substantiated as a vigorous model against the fake data.

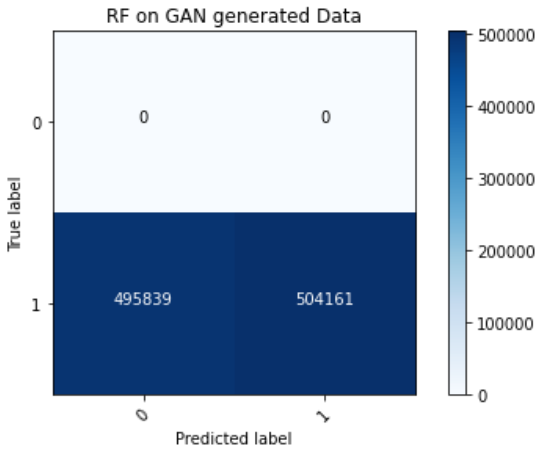

Figure 20. Random Forest on GAN generated fake data

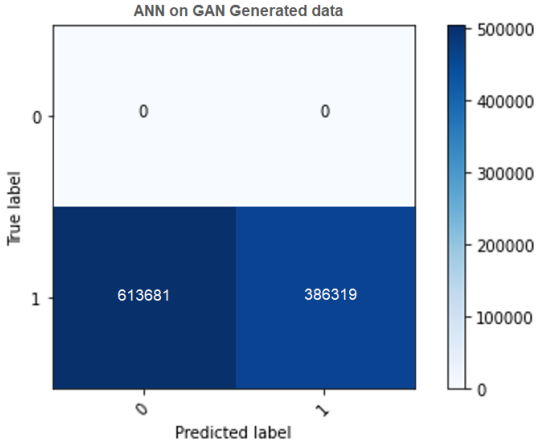

Figure 21. ANN on GAN generated fake data

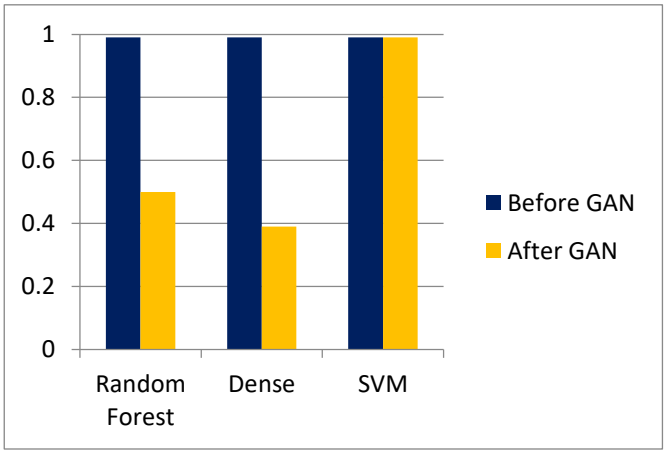

Figure 22. Before and After GAN comparison of 3 models

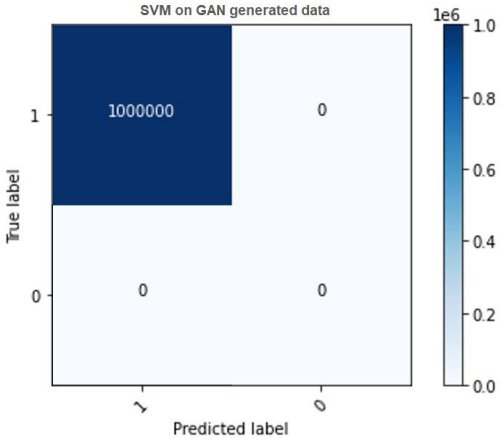

Figure 23. SVM on GAN generated fake data.

5.2 Results of Second Approach

In this approach, RF and SVM models are used for the IDS with some minor changes. Both models used 80% for training and 20% for testing data. In this approach, we used 14 features, including saddr, daddr, proto, dport and sport. The hyperparameters used for the RF model are verbose=1 and random_state=42 with kernel set as default see the Appendix G. SVM use the same hyperparameters which are used in the previous approach, look into the Appendix H. RF and SVM models trained and tested on the 80% and 20% data respectively for both binary and multiclass classification. In this approach, both classifiers give approximately 100% accuracy in both classifications. The precision, recall and f1-score results of the RF model's multiclass classification are exhibited in Figure 24 and Figure 25. The classification report and graph illustrate the 100% accuracy by using 14 features. The binary and multiclass confusion matrix reports in Figure 26 and Figure 27 indicate that with nearly 100% accuracy, only a few samples were mispredicted.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	105182
1	1.00	1.00	1.00	89797
2	0.99	0.99	0.99	107
3	1.00	1.00	1.00	4911
4	1.00	0.67	0.80	3
accuracy			1.00	200000
macro avg	1.00	0.93	0.96	200000
weighted avg	1.00	1.00	1.00	200000

Figure 24. RF Classification report

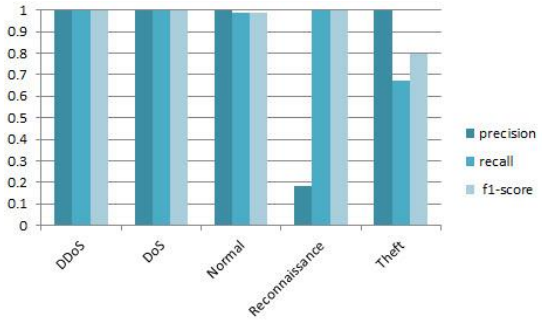

Figure 25. Graph of RF Classification report

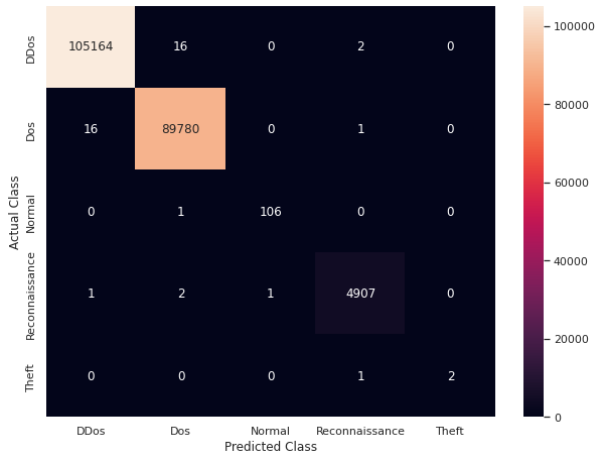

Figure 26. Confusion matrix of RF for the multiclass

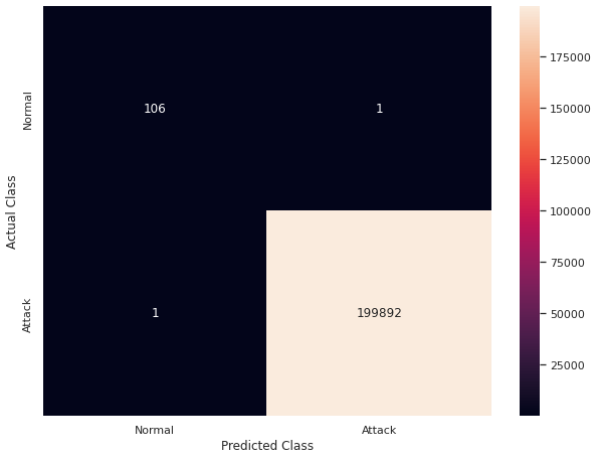

Figure 27. Confusion matrix of RF for the binary class

	precision	recall	f1-score	support
0	1.00	0.99	1.00	105182
1	0.99	1.00	1.00	89797
2	1.00	0.78	0.87	107
3	0.99	0.96	0.97	4911
4	1.00	1.00	1.00	3
accuracy			1.00	200000
macro avg	1.00	0.95	0.97	200000
weighted avg	1.00	1.00	1.00	200000

Figure 28. SVM Classification report

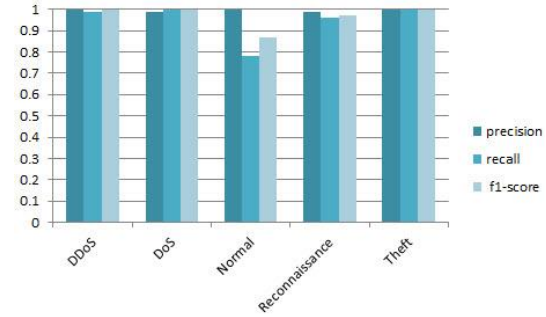

Figure 29. Graph of a classification report

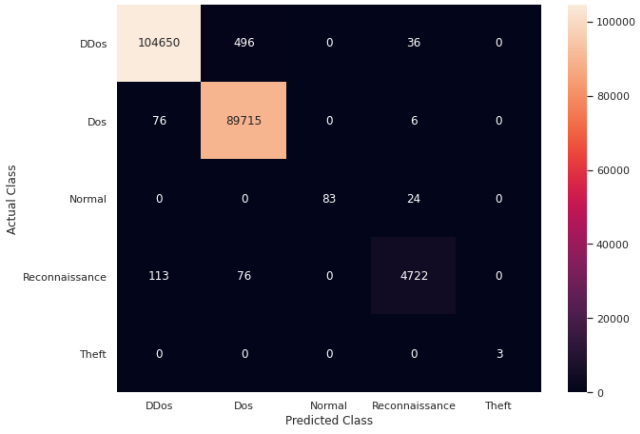

Figure 30. Confusion matrix of SVM for multiclass

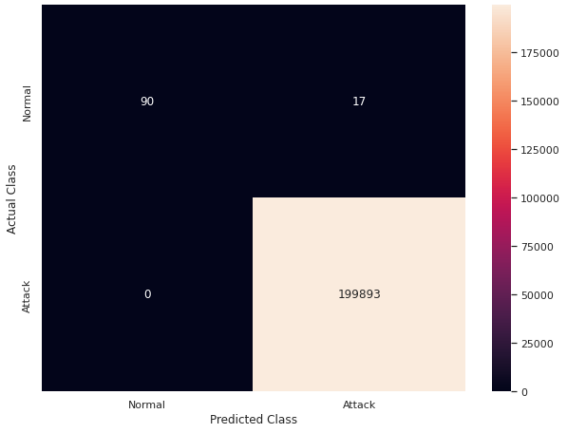

Figure 31. Confusion matrix of SVM for binary class

The SVM also gave almost 100% accuracy on the same data. The classification report and graphical description of this classification report in Figure 28 and Figure 29 express the outcomes. Figure 30 and Figure 31 portray the results in the form of a confusion matrix.

Finally, now implement the WGAN to develop a model capable of predicting this malicious data as normal. In WGAN, there are two neural networks (NN); one NN predicts the output and the second NN tells the first one how much fake or real output you generated. This process continues until the error rate becomes close to zero, or in the essence of this research, fake values become identical to actual values for the model. WGAN is learned and trained on the normal data that how it look like. So, subsequently, when trusted data that is almost 99.9% is malicious data test on WGAN, it predicts data as normal.

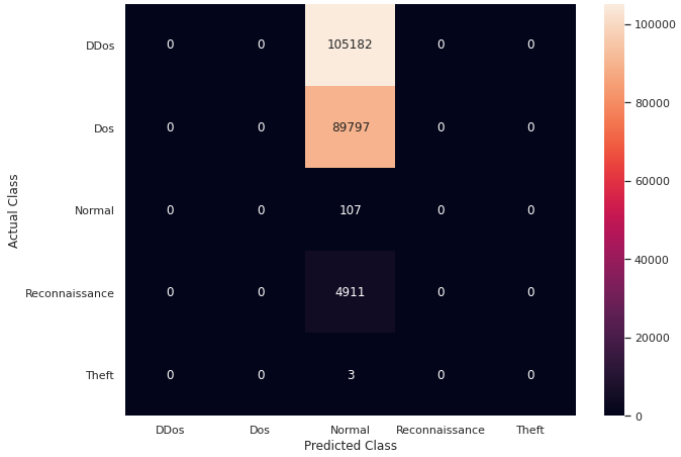

Figure 32. Confusion matrix of WGAN for multiclass

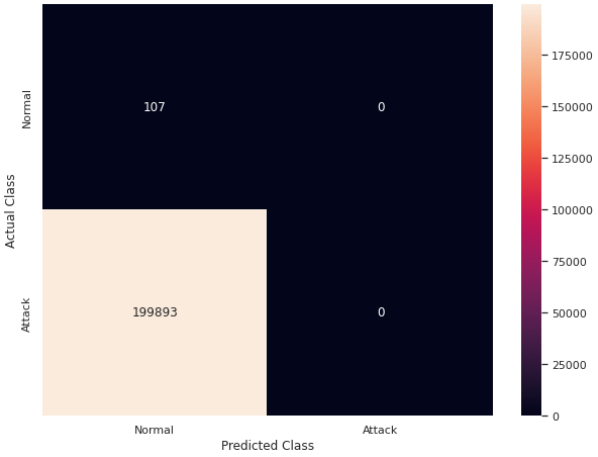

Figure 33. Confusion matrix of WGAN for binary class

Now, we tested the same 20% testing data on the WGAN used for the RF and SVM. The result was marvellous. Remarkably, WGAN predicted all the data as normal. Analysing the binary class confusion matrix in Figure 32 shows only 107 samples belonged to the normal class, but WGAN predicted all the as normal. Investigating the multiclass confusion matrix in Figure 33 indicates that 105,182 sample's actual class is DDoS, 89,797 belongs to DoS, only 107 is normal, 4,911 sample's actual class is reconnaissance and 3 thefts. Nevertheless, by looking into the predicted class, WGAN predicted all the data as normal.

6. CONCLUSION

Numerous studies have been published based on ML models for the recognition of attack traffic on the network. However, these studies were failed in detection when label flipped data or manipulated data passed to them. The cause of the failure of these studies was mainly the dataset and robustness of trained models. This study has provided an overview and detailed information about the GAN architecture in deep learning. It has discussed the various approaches of the GAN and IDS system in the network and the practical ways it can help the stakeholders to leverage the existing methodology to get more effective results.

In this study, the Bot-IoT dataset was used to train the models. Bot-IoT is collected in a simulated environment and is much flexible in detecting malicious traffic over the network. To our best knowledge, it

is used in only one study [14]. The architecture of models was also robust to classify the malicious samples over the benign traffic.

This study was divided into two approaches. The first approach is used to generate new fake data that looks benign by changing the signature of the data using GAN. The Bot-IoT dataset trained RF, SVM, and ANN models for binary and multiple class classification. The train test split and cross-validation approach showed remarkable accuracy with other evaluation measures. GAN models generate the manipulated data after training on the attack samples of trusted data. The actual property of data was malicious, but GAN generated the fake data, which looks real to deceive the IDS. RF, SVM and ANN also classify the manipulated data generated by the GAN model. After testing the RF, SVM and ANN on manipulated data for binary classification, it is concluded that the accuracy of RF and ANN dropped to 50% and 38.63%. So, results demonstrate that RF and ANN compromised by using the adversarial model. On the contrary, the SVM model did not descent the accuracy and proved more robust against the adversarial model. The second approach is used to design and develop a model using WGAN, intelligent enough to predict malicious data as normal. The data is 99.99% was malicious, which was used to test the new model, and WGAN predicted all the data as benign. Overwhelmingly, it is evidenced from the observation of the investigational results adversarial example intimidate the ML/ DL based intrusion detection system. It is essential to train the IDS with GAN to improve the IDS's robustness.

6.1 Future Work

As for future work, to enhance the robustness of the IDS model, IDS models need to train alongside GAN. The performance of IDS can be increased by training these classifiers on the GAN generated fake data. It has been observed that the BotIoT data include very few normal data samples, which was approximately 0.1%. By increasing the normal samples in the training set can also help to improve the accuracy.

APPENDIX

Appendix A. RF for binary and multiclass classification.

```
def RandomForest(X=input_attributes, y=output_attribute):
    RF-model = RandomForestClassifier(random-state=0, n-estimators=20)
    predictions = cross-val-predict(RF-model, X, y, cv=4)
    confusion-matrix-summary = confusion-matrix(y, predictions)
```

Appendix B. SVM for binary and multiclass classification.

```
svcmodel = SVC(verbose=1,random_state=42)
svcmodel.fit(X,y)
score = svcmodel.score(X,y)
score
```

Appendix C. ANN for binary and multiclass classification.

```
def ANN_Model(X_train, X_test, y_train, y_test):
    model1 = Sequential()
    model1.add(Dense(20, input_dim=10, activation='tanh'))
    model1.add(Dense(60, input_dim=20, activation='tanh'))
    model1.add(Dense(80, input_dim=60, activation='tanh'))
    model1.add(Dense(90, input_dim=80, activation='tanh'))
    model1.add(Dense(5, input_dim=90, activation='sigmoid'))
    model1.compile(loss='sparse_categorical_crossentropy',optimizer='adam',
metrics=['accuracy'])
    model1.fit(X_train, y_train, validation_data=(X_test,y_test), epochs=10,
batch_size=256)
```

Appendix D. Testing of RF model for GAN generated data.

```
def RF_test(generatedData= X):
    modelName = 'trained_RF_model.sav'
    RF_trained_model = pickle.load(open(modelName, 'rb'))
    predictions = RF_trained_model.predict(generatedData)
```

Appendix E. Testing of ANN model for GAN generated data.

```
def ANN_test():
    ANN_trained_model = load_model('trained_dense_model.h5')
    ANN_trained_model.compile(loss='binary_crossentropy', optimizer='adam',
metrics=['accuracy'])
    model_score = ANN_trained_model.evaluate(X=input_features,
y=output_feature, verbose=0)
    print("%s: %.2f%%" % (ANN_trained_model.metrics_names[1],
model_score[1]*100))
```

Appendix F. Testing of SVM model for GAN generated data.

```
import pickle
# load the model from disk
filename = 'trained_SVM_model.sav'
loaded_model = pickle.load(open(filename, 'rb'))
result = loaded_model.score(X, y)
print(result)
```

Appendix G. Second approach RF for binary and multiclass classification.

```
RFmodel = RandomForestClassifier(verbose=1,random_state=42)
RFmodel.fit(train,y_train)
RFscore = RFmodel.score(test,y_test)
RFscore
RFprediction = RFmodel.predict(test)
print(classification_report(y_test,RFprediction))
```

Appendix H. Second approach SVM for binary and multiclass classification.

```
model = SVC(verbose=1,random_state=42)
model.fit(train,y_train)
score = model.score(test,y_test)
score
prediction = model.predict(test)
print(classification_report(y_test,prediction))
```

Appendix I. WGAN for Second approach

```
class WGAN(object):
  def __init__(self, options, n_attributes):
    self.n_attributes = n_attributes
    self.noise_dim = options.noise_dim
    self.generator = Generator(self.n_attributes + self.noise_dim, self.n_at
tributes)
    self.discriminator = Discriminator(self.n_attributes)

    self.device = torch.device("cuda" if torch.cuda.is_available() else "cpu
")
    self.generator.to(self.device)
    self.discriminator.to(self.device)
```

Appendix J. Google Colab all files and Dataset links

To run the code, first, download the dataset into your system and then give the dataset location while executing the code.

Links for the code are removed due to privacy reason

REFERENCES

[1] I. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz, and J. Lopez, "A survey of iot-enabled cyberattacks: Assessing attack paths to critical infrastructures and services," IEEE Commun. Surv. Tutorials, vol. 20, no. 4, pp. 3453–3495, Oct. 2018.

[2] M. A. Khan and K. Salah, "IoT security: Review, blockchain solutions, and open challenges," Futur. Gener. Comput. Syst., vol. 82, pp. 395–411, May 2018.

[3] M. Nawir, A. Amir, N. Yaakob, and O. B. Lynn, "Internet of Things (IoT): Taxonomy of security attacks," 2016 3rd Int. Conf. Electron. Des. ICED 2016, pp. 321–326, Jan. 2017.

[4] "ISTR Internet Security Threat Report Volume 24 |," 2019.

[5] F. Y. Sattarova, "Integrating intrusion detection system and data mining," Proc. - 2008 Int. Symp. Ubiquitous Multimed. Comput. UMC 2008, pp. 256–259, 2008.

[6] P. García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," Comput. Secur., vol. 28, no. 1–2, pp. 18–28, Feb. 2009.

[7] P. Amoli, T. Hamalainen, … G. D.-… (International J. of, and undefined 2016, "Unsupervised network intrusion detection systems for zero-day fast-spreading attacks and botnets," users.jyu.fi.

[8] M. J. Hashemi, G. Cusack, and E. Keller, "Towards evaluation of NIDSs in adversarial setting," Big-DAMA 2019 - Proc. 3rd ACM Conex. Work. Big DAta, Mach. Learn. Artif. Intell. Data Commun. Networks, Part Conex. 2019, pp. 14–21, Dec. 2019.

[9] C. Sinclair, L. Pierce, and S. Matzner, "An application of machine learning to network intrusion detection," Proc. - Annu. Comput. Secur. Appl. Conf. ACSAC, vol. Part F133431, pp. 371–377, 1999.

[10] Q. Niyaz, W. Sun, A. Y. Javaid, and M. Alam, "A deep learning approach for network intrusion detection system," EAI Int. Conf. Bio-inspired Inf. Commun. Technol., 2015.

[11] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," Proc. - IEEE Symp. Secur. Priv., pp. 305–316, 2010.

[12] Yinka-Banjo, C. and Ugot, O.A., 2020. A review of generative adversarial networks and its application in cybersecurity. Artificial Intelligence Review, 53(3), pp.1721-1736.

[13] J. Kaur, "Generative adversarial networks overview and applications," XenonStack, 23-Dec-2019. [Online]. Available: https://www.xenonstack.com/insights/generative-adversarial-networks. [Accessed: 12-Jun-2021].

[14] Papadopoulos, P., Essen, O.T.V., Pitropakis, N., Chrysoulas, C., Mylonas, A. and Buchanan, W.J., 2021. Launching Adversarial Attacks against Network Intrusion Detection Systems for IoT. Journal of Cybersecurity and Privacy, 1(2), pp.252-273.

[15] Wang, Z., 2018. Deep learning-based intrusion detection with adversaries. IEEE Access, 6, pp.38367-38384.

[16] Koroniotis, N., Moustafa, N., Sitnikova, E. and Turnbull, B., 2019. Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. Future Generation Computer Systems, 100, pp.779-796.

[17] Wang, L., Zhang, T. and Gong, B., Li, Y., Li, L., 2019, May. Nattack: Learning the distributions of adversarial examples for an improved black-box attack on deep neural networks. In International Conference on Machine Learning (pp. 3866-3876). PMLR.

[18] Yilmaz, I., Masum, R. and Siraj, A., 2020, August. Addressing imbalanced data problem with generative adversarial network for intrusion detection. In 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 25-30). IEEE.

[19] Nie, L., Wu, Y., Wang, X., Guo, L., Wang, G., Gao, X. and Li, S., 2021. Intrusion Detection for Secure Social Internet of Things Based on Collaborative Edge Computing: A Generative Adversarial Network-Based Approach. IEEE Transactions on Computational Social Systems.

[20] Smys, S., Basar, A. and Wang, H., 2020. Hybrid intrusion detection system for internet of Things (IoT). Journal of ISMAC, 2(04), pp.190-199.

[21] Zhen-xiang, F.S.H., 2021. A Black Box Attack on a Network Traffic Intrusion Detection System. Computer & Telecommunication, 1(5), pp.46-51.

[22] Zixu, T., Liyanage, K.S.K. and Gurusamy, M., 2020, December. Generative adversarial network and auto encoder based anomaly detection in distributed IoT networks. In GLOBECOM 2020-2020 IEEE Global Communications Conference (pp. 1-7). IEEE.

[23] Zhang, J. and Zhao, Y., 2021, June. Research on Intrusion Detection Method Based on Generative Adversarial Network. In 2021 International Conference on Big Data Analysis and Computer Science (BDACS) (pp. 264-268). IEEE.

[24] Ferdowsi, A. and Saad, W., 2019, December. Generative adversarial networks for distributed intrusion detection in the internet of things. In 2019 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6). IEEE.

[25] Zheng, Y.J., Zhou, X.H., Sheng, W.G., Xue, Y. and Chen, S.Y., 2018. Generative adversarial network based telecom fraud detection at the receiving bank. Neural Networks, 102, pp.78-86.

[26] Shah, J. and Das, M., 2022. IGAN: Intrusion Detection Using Anomaly-Based Generative Adversarial Network. In Applied Information Processing Systems (pp. 371-379). Springer, Singapore.

[27] Viola, J., Chen, Y. and Wang, J., 2021. FaultFace: Deep convolutional generative adversarial network (DCGAN) based ball-bearing failure detection method. Information Sciences, 542, pp.195-211.

[28] Sedjelmaci, H., 2020. Attacks detection and decision framework based on generative adversarial network approach: Case of vehicular edge computing network. Transactions on Emerging Telecommunications Technologies, p.e4073.

[29] G. Developer, "The Discriminator | Generative Adversarial Networks", Google Developers, 2019. [Online]. Available: https://developers.google.com/machine-learning/gan/discriminator. [Accessed: 14- Jul- 2021].

[30] Usama, M., Asim, M., Latif, S. and Qadir, J., 2019, June. Generative adversarial networks for launching and thwarting adversarial attacks on network intrusion detection systems. In 2019 15th international wireless communications & mobile computing conference (IWCMC) (pp. 78-83). IEEE.

[31] Ying, H., Ouyang, X., Miao, S. and Cheng, Y., 2019, March. Power message generation in smart grid via generative adversarial network. In 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (pp. 790-793). IEEE.

[32] Dai, Y., Li, H., Qian, Y., Guo, Y. and Zheng, M., 2021. Anticoncept Drift Method for Malware Detector Based on Generative Adversarial Network. Security and Communication Networks, 2021.

[33] Siniosoglou, I., Radoglou-Grammatikis, P., Efstathopoulos, G., Fouliras, P. and Sarigiannidis, P., 2021. A unified deep learning anomaly detection and classification approach for smart grid environments. IEEE Transactions on Network and Service Management.

[34] Zhao, S., Li, J., Wang, J., Zhang, Z., Zhu, L. and Zhang, Y., 2021. attackGAN: Adversarial Attack against Black-box IDS using Generative Adversarial Networks. Procedia Computer Science, 187, pp.128-133.

[35] Cheng, Q., Zhou, S., Shen, Y., Kong, D. and Wu, C., 2021. Packet-Level Adversarial Network Traffic Crafting using Sequence Generative Adversarial Networks. arXiv preprint arXiv:2103.04794.

[36] Shu, D., Leslie, N.O., Kamhoua, C.A. and Tucker, C.S., 2020, July. Generative adversarial attacks against intrusion detection systems using active learning. In Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning (pp. 1-6).

[37] Huang, S. and Lei, K., 2020. IGAN-IDS: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks. Ad Hoc Networks, 105, p.102177.

[38] Song, L., Shokri, R. and Mittal, P., 2019, November. Privacy risks of securing machine learning models against adversarial examples. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (pp. 241-257).

[39] Seo, E., Song, H.M. and Kim, H.K., 2018, August. Gids: Gan based intrusion detection system for in-vehicle network. In 2018 16th Annual Conference on Privacy, Security and Trust (PST) (pp. 1-6). IEEE.

[40] Hasan Shahriar, M., Imtiazul Haque, N., Ashiqur Rahman, M. and Alonso Jr, M., 2020. G-IDS: Generative Adversarial Networks Assisted Intrusion Detection System. arXiv e-prints, pp.arXiv-2006.

[41] Chen, J., Wu, Y., Jia, C., Zheng, H. and Huang, G., 2020. Customisable text generation via conditional text generative adversarial network. Neurocomputing, 416, pp.125-135.

[42] Maetschke, S.R., Madhamshettiwar, P.B., Davis, M.J. and Ragan, M.A., 2014. Supervised, semi-supervised and unsupervised inference of gene regulatory networks. Briefings in bioinformatics, 15(2), pp.195-211.

[43] Krundyshev, V. and Kalinin, M., 2021. Generative Adversarial Network for Detecting Cyber Threats in Industrial Systems. In Proceedings of International Scientific Conference on Telecommunications, Computing and Control (pp. 1-13). Springer, Singapore.

[44] O. Ibitoye, O. Shafiq, A. M.-2019 I. Global, and undefined 2019, "Analysing adversarial attacks against deep learning for intrusion detection in IoT networks," ieeexplore.ieee.org.

[45] A. Nisioti, A. Mylonas, P. D. Yoo, and V. Katos, "From intrusion detection to attacker attribution: A comprehensive survey of unsupervised methods," IEEE Commun. Surv. Tutorials, vol. 20, no. 4, pp. 3369–3388, Oct. 2018.

[46] N. T. Van, T. N. Thinh, and L. T. Sach, "An anomaly-based network intrusion detection system using Deep learning," Proc. - 2017 Int. Conf. Syst. Sci. Eng. ICSSE 2017, pp. 210–214, Sep. 2017.

[47] E. Benkhelifa, T. Welsh, and W. Hamouda, "A critical review of practices and challenges in intrusion detection systems for IoT: Toward universal and resilient systems," IEEE Commun. Surv. Tutorials, vol. 20, no. 4, pp. 3496–3509, Oct. 2018.

[48] de Araujo-Filho, P.F., Kaddoum, G., Campelo, D.R., Santos, A.G., Macêdo, D. and Zanchettin, C., 2020. Intrusion detection for cyber–physical systems using generative adversarial networks in fog environment. IEEE Internet of Things Journal, 8(8), pp.6247-6256.

[49] Ma, W., Zhang, Y., Guo, J. and Li, K., 2021. Abnormal Traffic Detection Based on Generative Adversarial Network and Feature Optimization Selection. International Journal of Computational Intelligence Systems, 14(1), pp.1170-1188.

[50] Ang, J.C., Mirzal, A., Haron, H. and Hamed, H.N.A., 2015. Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. IEEE/ACM transactions on computational biology and bioinformatics, 13(5), pp.971-989.

[51] N. Donges, "A Complete Guide to the Random Forest Algorithm", Built In, 2021. [Online]. Available: https://builtin.com/data-science/random-forest-algorithm. [Accessed: 30- Sep- 2021].

[52] R. Meyers, Encyclopedia of physical science and technology, 3rd ed. Amsterdam: Elsevier, 2003, pp. 631-645.

[53] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Dec. 2014.

[54] Pisner, D.A. and Schnyer, D.M., 2020. Support vector machine. In Machine Learning (pp. 101-121). Academic Press.

[55] J. Brownlee, "A Gentle Introduction to Generative Adversarial Networks (GANs)", Machine

Learning Mastery, 2019. [Online]. Available: https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/. [Accessed: 22- Sep- 2021].

[56] J. Allingham, "Wasserstein GAN · Depth First Learning", Depthfirstlearning.com, 2019. [Online]. Available: https://www.depthfirstlearning.com/2019/WassersteinGAN. [Accessed: 29- Aug- 2021].

[57] J. Brownlee, "How to Develop a Wasserstein Generative Adversarial Network (WGAN) From Scratch", Machine Learning Mastery, 2019. [Online]. Available: https://machinelearningmastery.com/how-to-code-a-wasserstein-generative-adversarial-network-wgan-from-scratch/. [Accessed: 30- Aug- 2021].

[58] G. Research, "Google Colab", Research.google.com, 2021. [Online]. Available: https://research.google.com/colaboratory/faq.html. [Accessed: 22- Sep- 2021].