Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.055891

REVIEW

A Review of Generative Adversarial Networks for Intrusion Detection Systems: Advances, Challenges, and Future Directions

Monirah Al-Ajlan* **and Mourad Ykhlef**

Information Systems Department, King Saud University, Riyadh, 11495, Saudi Arabia

*Corresponding Author: Monirah Al-Ajlan. Email: maalajlan@ksu.edu.sa

Received: 09 July 2024 Accepted: 16 October 2024 Published: 18 November 2024

ABSTRACT

The ever-growing network traffic threat landscape necessitates adopting accurate and robust intrusion detection systems (IDSs). IDSs have become a research hotspot and have seen remarkable performance improvements. Generative adversarial networks (GANs) have also garnered increasing research interest recently due to their remarkable ability to generate data. This paper investigates the application of (GANs) in (IDS) and explores their current use within this research field. We delve into the adoption of GANs within signature-based, anomaly-based, and hybrid IDSs, focusing on their objectives, methodologies, and advantages. Overall, GANs have been widely employed, mainly focused on solving the class imbalance issue by generating realistic attack samples. While GANs have shown significant potential in addressing the class imbalance issue, there are still open opportunities and challenges to be addressed. Little attention has been paid to their applicability in distributed and decentralized domains, such as IoT networks. Efficiency and scalability have been mostly overlooked, and thus, future works must aim at addressing these gaps.

KEYWORDS

Intrusion detection systems; network security; generative networks; deep learning; dataset

1 Introduction

In recent times, security breaches by underground criminal enterprises have dramatically increased as cyber-criminals take advantage of the exponential growth in digital communications and its centrality to daily life and organizational operations. Protecting organizations from online security threats is becoming ever more important, as they are increasingly dependent on the Internet [1–4]. According to the CISCO network forecast report, it is speculated that 15.4 million denial of service (DoS) attacks took place in 2023 [5]. Consequently, communication networks must be secured as the number of daily attacks keeps increasing. One such mechanism for protecting organizational networks against malicious attacks is the intrusion detection system (IDS) [6,7].

In this article, we discuss two immensely significant research hotspots: (i) the development of robust and accurate IDSs that can react to ever-evolving threats; and (ii) the adoption of generative networks within IDSs to harness their ability to generalize and learn from input data. IDSs have been categorized into three main classes, namely signature-based, anomaly-based, and hybrid systems (a

summary of literature encompassing the three approaches is presented in Table 1). IDSs have been extensively studied and generally reported to exhibit excellent performance metric values. Machine learning (ML) and deep learning (DL) algorithms have been adopted within signature based IDSs to enhance detection accuracy. However, most works assume the existence of datasets that reflect actual network attack scenarios. However, difficulties arise when systems face previously unencountered attacks. This situation is becoming increasingly prevalent as attackers employ constantly evolving and rapidly changing techniques. In contrast, although anomaly based IDSs do not require labeled data, they rely heavily on certain data characteristics, such as the Euclidean distance [8]. However, this reliance tends to overlook the fact that real-life network traffic could follow alternate distributions and exhibit different characteristics. Hybrid IDSs harness both labeled and unlabeled data, but only a few studies have tested this approach [9]. Generative adversarial networks (GANs) have garnered increasing attention due to their breakthrough capabilities that have revolutionized major domains, such as natural language processing (NLP) [10] and computer vision (CV) [11] (especially medical imaging [12]). Additionally, they have been widely employed in computer network security [13]. They consist of two competing neural networks (NNs), namely a generator that generates fake data and a discriminator that distinguishes real data from the synthetic data generated by the generator. Then, the generator improves its output by the feedback from the discriminator. The aim here is to continue until the generator can generate realistic samples that the discriminator cannot distinguish [14]. Consequently, this remarkable workflow of GANs allows them to create and expand datasets, which are the basis of ML models. Training ML models requires vast amounts of data, which is hard to collect and label by human experts [15]. GAN's ability to generate and augment datasets was mainly exploited in the image field, to generate synthetic images for various reasons such as training ML models such as face recognition, and for generating realistic images [16]. However, the main role of GAN in IDS is to balance the dataset (since attack samples are mostly fewer than benign ones) by synthetically generating minority class samples. GANs have been adopted to address certain IDS limitations, largely focusing on data balancing [17]. However, little attention has been paid to other critical parameters, such as GAN efficiency and applicability in domains, such as Internet of Things (IoT) networks.

Existing IDS review articles mainly focus on IDS and the different methods to improve it, consequently overlooking the recent trend of GAN and its capabilities in strengthening the detection of intrusions and overcoming data imbalance issues. This paper differs from existing surveys in that it focuses on GAN value in strengthening IDS. This review offers several contributions to the field of cybersecurity:

- A comprehensive examination of GAN applicability within IDS field, an area previously under-explored.
- An analysis of existing datasets, methodologies, and frameworks employed in GAN-based IDS research.
- An illustration of GAN application into the three primary IDS categories.
- A comparative analysis of available datasets focusing on their strengths, weaknesses, and challenges.
- Suggestions for promising research problems for future research to bridge existing knowledge gaps.

The literature selection process was based on the inclusion criteria that papers should be peer-reviewed journal articles, written in English, published within the last five years, using keywords such as 'GAN', 'IDS' and 'Generative Networks' by exploring Google scholar database. The selected articles

were then analyzed using a thematic analysis approach, finding key themes and similarities across literature. Then, they were grouped into three IDS classes.

The paper is structured as follows: In Section 2, we review the applicability of GANs in IDSs. Further, we explore the available datasets, methods, and frameworks used up to date. In Section 3, we analyze the adoption of GANs within the three IDS classes and suggest possible use cases. In Section 4, comparisons between the available datasets and methods are presented along with their associated challenges and advantages. We also look into future research directions to address current research gaps. Finally, in Section 5, we present the primary conclusions of the study.

2 Literature Review

2.1 IDS

Numerous studies have extensively analyzed IDSs. Debar [18] established an IDS to be a system that collects details on and investigates the security status of information systems. Signature-based IDSs aim to detect new intrusions by finding signatures that match previously known intrusions [19]. In contrast, anomaly-based IDSs have the potential to identify novel intrusions by detecting abnormalities [20]. Hybrid IDSs are more comprehensive, combining the essence of anomaly- and signature-based IDSs to detect both novel and known attacks [14]. Recent developments in IDSs have heightened the need for collaboration among network nodes [15]. Thus, a subclass of IDSs, named collaborative IDS (CIDS), has emerged as a vital defense mechanism where a set of independent IDSs work together to add intelligence to an increasingly comprehensive network [16]. Researchers have also adopted deep learning methods to enhance IDS, in fact, numerous studies such as Reference [21] has concentrated on improving the architecture design of convolutional neural networks (CNNs) for accurately classifying intrusions. In their work, the aim was to build an automated method for choosing an appropriate CNN architecture for intrusion detection in industrial control systems (ICS), thus eliminating the need for manual architecture engineering. They adopted a differential evolution (DE) algorithm to automatically set key hyperparameters such as the number and type of CNN layers, learning rate, and batch size. The proposed approach was tested and they suggested that the proposed automated approach not only saved time but also outperformed the baseline model. IDS research has recently shifted into more specific domains, given that the IDS requirements vary according to the environment. Kumar et al. [22] have reviewed studies concerning IDS in Software Defined Networks (SDNs) and reported the application of ML and DL techniques within IDS. Results were promising with accuracy reaching 99%. However, despite the success of IDS in SDN, they reported several research gaps such as the high dependency of feature selection methods and outdated datasets.

The IDS architecture varies based on the case requirements. However, the general architecture of a signature-based IDS consists of the following components (Fig. 1) [6]. The backbone of the IDS is the dataset; it is preprocessed and cleaned for quality improvements, which may lead to sparse data dimensionality in certain cases and require reduction. Subsequent feature extraction allows one to reveal the characteristics that indicate the presence of intrusions. For classification purposes, the dataset is then split into the labeled training dataset for training the ML model and the unlabeled testing dataset for testing the designed model against predefined measures, such as accuracy and detection rate. Finally, network traffic is classified according to prediction probabilities for normal (benign) and malicious cases.

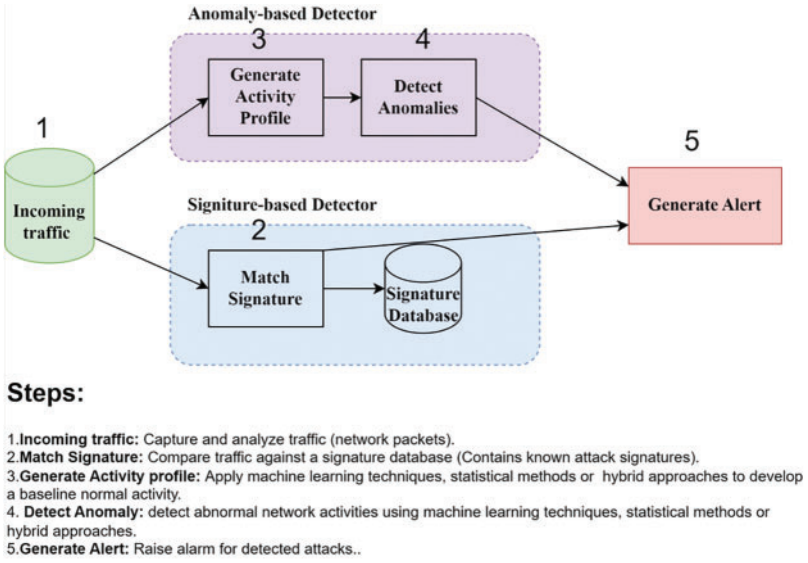

Figure 1: Signature *vs*. anomaly-based IDS "Adapted from Reference [23]"

Table 1: IDS literature review summary

IDS category	Paper	Year	Dataset	Algorithm	Accuracy
Signature-based	[24]	2017	KDDCUP99	Neural networks	97%
	[25]	2018	Private	Neural networks	98%
	[26]	2020	KDDCUP99	Random forests	94%
Anomaly-based	[27]	2018	NSL-KDD	C5	99.82%
	[28]	2021	NSL-KDD	Lightweight neural network	96.9%
	[5]	2018	MAWIFlow	Decision trees	85%
	[29]	2020	MIT Lincoln Lab datasets	SVM and Self-Organized Feature Map (SOFM)	94.19%
	[30]	2021	DS2OS	Decision trees	99.9%
	[31]	2022	Distilled-Kitsune-2018 and NSL-KDD	Ensemble methods, neural network, and kernel methods.	99.8%
	[32]	2022	NF-ToN-IoT-v2	Naive Bayes (NB), Random Forest (RF), Decision Tree (DT), and eXtreme Gradient Boosting (XGB).	99%
	[33]	2022	EDGE-IIOTSET 2022	Polynomial interpolation technique (Least Mean Squares)	97.27%
	[34]	2022	NF-UNSW-NB15-v2 and NF-CSE-CIC-IDS2018-v2	DeepSVDD	98.20%
	[35]	2022	NSL-KDD++	Deep forward neural network	99.46%

(Continued)

Table 1 (continued)

IDS category	Paper	Year	Dataset	Algorithm	Accuracy
Hybrid IDS	[36]	2021	MIT Lincoln laboratories repository	Neural networks and fuzzy logic	96.111%
	[37]	2019	Bot-IoT	C5 and one class support vector machine	99.7%
	[38]	2020	NSL-KDD and UNSW-NB15	Multi-objective genetic method (NSGAII), neural network and random forests	82%

IDS Datasets

IDS datasets are built by collecting data from verified sources, such as network traffic flows, which contain critical information on the host, destination, header, and user behavior [39]. Consequently, this information is vital in detecting abnormal activities and identifying network traffic patterns [40]. Several datasets have been constructed for intrusion detection, each with its pros and cons. Here, we discuss three such datasets, namely KDDCUP99, NetFlow, and CIC-IDS-2017. A summary of these datasets along with their advantages and disadvantages is shown in a table.

(I) KDDCUP99 dataset

KDDCUP99 is one of the earliest and most widely used datasets in the IDS domain [41]. It was introduced back by the Defense Advanced Research Projects Agency (DARPA), USA, using raw network traffic from 1998. It includes 41 features belonging to four classes: basic, content, time-based traffic, and host-based traffic features. It encompasses over 4,898,429 records covering four distinct attack categories: DoS, unauthorized node access (user-to-root; U2R), probe, and unauthorized access from an external machine (root-to-local; R2L). One of the limitations faced by this dataset is its redundancy which can potentially harm ML models by causing model bias, which can, in turn, lead to the neglect of key records that might be potential attacks [35].

(II) NetFlow-derived datasets

The problem with the previously published datasets in the security domain is the lack of a standard feature set, which does not allow fair comparisons of ML models. Therefore, to address this issue, Sarhan et al. [42] derived four sub-datasets, namely UNSW-NB15 network traffic dataset, BotNet and Internet of Things (IoT) Devices Dataset (BoT-IoT), Toys and Network Traffic of Internet of Things Devices Dataset (ToNIoT), and Canadian Institute for Cybersecurity (CIC) Intrusion Detection Evaluation Dataset 2018 (CSE-CIC-IDS-2018), from the original NetFlow dataset which had 12 features (Table 2) [42]. The BoT-IoT dataset is suitable for both binary and multi-class classifications where the type of attack is pre-specified. The Network Traffic and Features of BotNet and Internet of Things (IoT) Devices Dataset (NF-BoT-IoT dataset), which was generated using pcap files and labeled according to the attack type, can be used for IDSs in IoT environments. It contains a total of 600,100 samples, out of which 586,241 (97.69%) are attack samples (of four types) and 13,859 (2.31%) are benign (Table 2).

Table 2: Netflow's dataset features (Adapted from Reference [42])

Feature	Description	Significance
IPV4 SRC ADDR	IPv4 source address	Uniquely identifies source.
IPV4 DST ADDR	IPv4 destination address	Uniquely identifies destination.
L4 SRC PORT	IPv4 source port number	Uniquely identifies source's application.
L4 DST PORT	IPv4 destination port number	Uniquely identifies destination's application.
PROTOCOL	IP protocol identifier byte	Identifies the specific protocol for diagnosis of network.
TCP FLAGS	Collection of all TCP flags	Examines the state of the TCP's connection.
L7 PROTO	(Numeric) Application Layer 7 protocol	Information on the seventh layer (application layer) protocol.
IN BYTES	Incoming number of bytes	To understand the network's traffic pattern.
OUT BYTES	Outgoing number of bytes	Identifies the volume of data leaving the network, could be used to find anomalies.
IN PKTS	Incoming number of packets	Useful in finding anomalies, and to diagnose packets' loss.
OUT PKTS	The outgoing number of packets	Useful in understanding the packets' behaviour and identifying heavy applications.
FLOW DURATION MILLISECONDS	Duration of flow in milliseconds	Useful in identifying abnormal long sessions which could be attacked.

(III) CIC-IDS-2017 dataset

The CIC-IDS-2017 dataset incorporates the use of profiles to construct a dataset in a structured manner [37], as some of the older datasets were built to address specific organizational issues or research experiments. This dataset includes features of captured attacks and conceptual knowledge on several application models, network devices, and protocols. The network traffic used in this dataset was recorded using the CICFlowMeter, which appropriately labels the flows and lists the source and destination addresses, port numbers, timestamps, and any attacks encountered. The class distribution of this dataset is presented in Table 3. A newer version of this dataset, CIC-IDS-2018, boasts of several enhancements, such as (i) larger size, as it encompasses traffic data over 10 days, and (ii) 17 attack classes, as opposed to the 5-day traffic data and 15 attack classes of the previous version. Despite the strengths and applicability presented by this dataset, both versions suffer from several drawbacks. For example, the data being stored in separate files makes data processing highly time-consuming. Additionally, the two datasets also experienced redundancies and missing records [40]. A summary of the datasets reviewed showcasing their limitations is shown in Table 4.

Table 3: CIC-IDS-2017 dataset class distribution (Adapted from Reference [43])

Class	Description	Training examples	Testing set examples
Benign	Normal traffic flow	1,818,097	455,000
DoS hulk	Large-scale Distributed Denial of Service (DDoS)	185,027	46,046
PortScan	Identify open ports on a network	126,988	31,942
DDOS	Distributed Denial of Service	102,688	25,339
DoS Golden eye	Launch DDoS attacks efficiently and with several attack vectors	8289	2004
FTP Patator	Brute-force tool aiming at cracking FTP (File Transfer Protocol) passwords	6388	1550
SSH-Pataor	Brute-force tool aiming at cracking Secure Shell (SSH) passwords	4701	1196
DoS Slowhttptest	Launch HTTP requests that are intentionally slow	4410	1089
DoS Slowloris	Launch HTTP requests that are intentionally slow (different implementation)	4624	1172
Web attack	Set of malicious actions targeting web applications	1778	402
Bot	Automated software programs called bots	1568	398
Infilteration	Obtaining unauthorized access to a network	29	9
Heartbleed	A critical security vulnerability	9	2

Table 4: Summary of IDS datasets

Dataset	Year	Ref.	Attacks addressed	Limitation
DARPA	1999	[44]	DoS, Probe, R2L, U2R	Outdated does not reflect more advanced and current attacks' challenges
NSL-KDD	2012	[45]	DoS, Probe, R2L, U2R	Does not accurately detect current low-footprint attack scenarios
ADFA	2013	[46]	Adduser, Hydra_FTP, Hydra_SSH, Java_Metapreter, Meterpreter, Web shell	Primarily focuses on known attacks, and lack of further updates
UNSW-NB15	2015	[47]	Exploits, Fuzzers, DoS, Reconnaissance, Analysis, Backdoor, Shellcode, worms	Was built synthetically
CTU-13	2017	[39]	BotNet	Primarily focuses on BotNet, does not further classify the BotNet class attack

2.2 Generative Networks

GANs have rapidly garnered research interest both in industry and academia [48]. The general architecture is shown in Fig. 2. The workflow of the two competing NNs in GANs is as follows: (i) The generator initiates the process by receiving an input vector and attempting to generate an output that

appears to originate from the same data source; and (ii) the discriminator is responsible for determining if the input is from the original data source or constructed by the generator. Upon completion of this process, the weights of both NNs are adjusted according to the classification accuracy, and the next iteration begins [14,49].

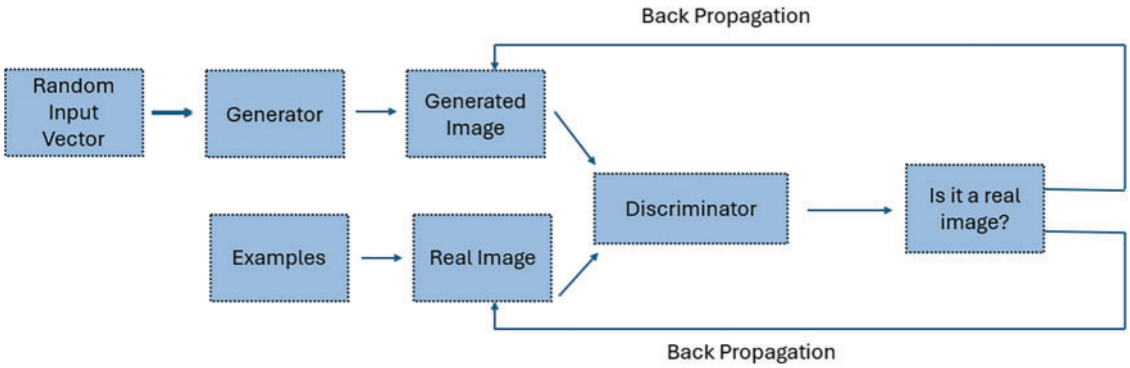

Figure 2: GAN general architecture "Adapted from Reference [50]"

GANs have been widely used to address key data mining problems, such as the lack of reliable datasets and data balancing issues [20]. For instance, Sauber-Cole et al. [49] applied GAN to correct tabular data imbalance and highlighted multiple issues. First, numerical input (discrete and continuous variables) can pose a difficulty to GANs, which may require several preprocessing steps to produce a GAN-readable input. However, thus far, there exist no standardized preprocessing pipelines for tabular data. Further, categorical variables require quantification methods; however, the choice for them remains mostly random (e.g., Gumbel SoftMax).

GAN Models

GAN performance is heavily dependent on the specific architectures of both generator and discriminator networks. Thus, several versions of the original GAN have been implemented to solve a variety of tasks. A timeline of various GAN models is illustrated in Fig. 3, while a summary of their strengths and weaknesses is shown in Table 5 [51].

Conditional GAN (CGAN)

Conditional image generation is one of the foundations of CV, and conditional GANs (CGANs) have been shown to demonstrate excellent performance [52]. CGANs were introduced by Mirza et al. [53] to generate images based on pre-specified class labels or textual descriptions. This is achieved by allowing a new input type, i.e., a conditional information vector, which subsequently governs the image generation process [53]. The primary difference between GANs and CGANs is the greater flexibility offered by the latter by generating images based on conditions and being more robust against noise. Despite its success in CV, the prevalent issue of high error rate caused by the regression of the generator encouraged revised versions (e.g., RoCGAN [52]), which study the target space structure to reduce the error rate.

CycleGAN

CycleGAN is another GAN variation specifically used for unpaired image-to-image translations that do not require paired examples in the training set [54]. It converts an image from domain A to domain B without requiring preexisting paired datasets. The general architecture of CycleGAN incorporates two generators and two discriminators at the same time. The first generator's main task

is to take a set of image examples as input from domain A and translate them into domain B, whereas the second generator does the opposite, translating the images from domain B to domain A. On the other hand, discriminators have the responsibility of distinguishing between real and synthetic images, which will result in gradients that will guide the next cycle of CycleGAN. A notable limitation of such networks is that they often fail to translate certain images in certain conditions, especially with geometric changes [54].

Deep Convolutional GAN (DCGAN)

In recent years, CNNs have been widely employed in several domains, especially CV [55]. Recent trends in CNN have led to a proliferation of studies attempting to combine them with GANs [11]. Deep convolutional GAN (DCGAN) was among the first such models, which was motivated by the fact that although CNNs are successful in supervised learning (classification in particular), their use in unsupervised learning tasks was limited. DCGANs incorporate CNNs trained on diverse image samples and generate a hierarchical representation of objects. Consequently, the extracted features can be used for both supervised and unsupervised learning [56]. Their use has extended the computer vision domain, and they have been implemented successfully in solving key machine learning issues such as data imbalance problems.

Progressive GANs (ProGANs)

One of the enhancements that have been implemented on top of GAN networks is the introduction of ProGANs [57]. The main idea behind progressive GANs (ProGANs) is that the generator and discriminator incrementally improve and increase in size during the training phase [56]. ProGANs are initiated with a low-resolution image that overlooks key features and then increase the resolution with subsequent training rounds. The resolution increases due to the added layers in each round that allow the ProGAN to capture smaller details. However, the downside of such approaches is that the training process becomes time-consuming due to the increased number of rounds.

StyleGAN

StyleGANs are one of the more recent GAN models that incorporate principles from the domain of style transfer [57]. Specifically, they introduce a new generator architecture that uses adaptive instance normalization, while preserving the expanding nature of ProGAN. This modification allows StyleGANs to identify high-level features, such as poses and face positions, in an unsupervised manner. This network has been tested on several datasets, along with novel diverse human face datasets to be used for benchmarking future StyleGAN improvements.

Gradient normalization GAN

Gradient normalization GANs (GranGANs) were developed to address the classic problem GAN problem of not being able to continue the generation process when gradients vanish. They introduce a novel gradient normalization technique that ensures that the gradient does not fall below a threshold (Lipschitz constraint). Generally, this step enhances the GAN and ensures its stability and performance while generating images [58].

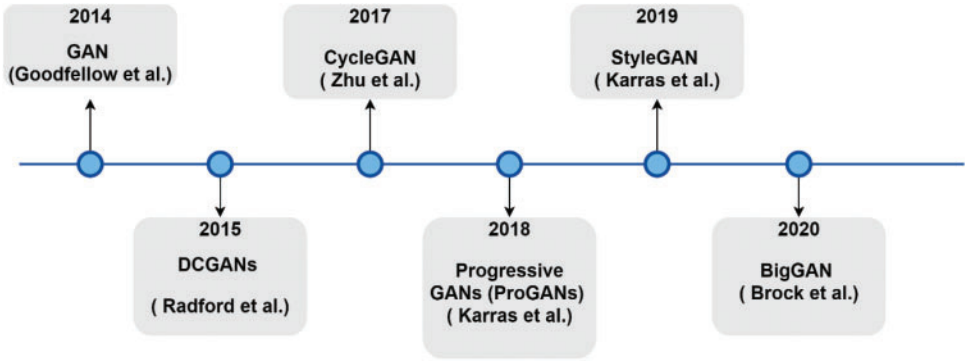

Figure 3: Research history of GAN [16,54,56,57–59]

Table 5: Comparison of GAN models

Model	Strength	Weakness
GAN	– High-quality image generation. – Used in supervised, unsupervised, and semi-supervised learning. – Continual learning through the training rounds.	– Privacy concerns. – Highly dependent on hyperparameters values. – Mode collapse: generating images with limited variations.
DCGAN	– Improved performance in unsupervised learning. – Hierarchical features learning. – Less prone to mode collapse.	– Excel with image data type, performance degrades with other data type such as text. – Hard to explain the generation process.
CycleGAN	– Performs well in image-to-image translation. – Produces realistic translations.	– Fails to translate images with geometric changes. – Longer training time. – Lack of suitable evaluation metrics in image-to-image translation.
ProGANs	– Captures fine details. – Improves image resolution gradually.	– Longer training time. – Computationally expensive.
StyleGAN	– Higher quality image generation. – Unsupervised model's ability to capture diverse styles and variations.	– Large memory requirement. – Computationally expensive. – Impractical for small devices with limited resources.

3 Generative Networks for Intrusion Detection

3.1 GAN for Signature-Based IDS

Recent developments in IDSs have led to a renewed interest in solving the prevalent class imbalance problem as, generally, attack samples are far fewer than benign samples in both established datasets and real-life scenarios [49]. Huang et al. [60] introduced imbalanced GANs (IGANs) with

a modified generator that includes a filter to generate samples only belonging to the minority class. They further incorporated this model into an IDS (IGAN-IDS) that worked in three phases: feature extraction, IGAN, and a deep NN (DNN). Experiments on three datasets along with ablation studies proved the effectiveness of this approach, reaching an accuracy of >99.0%, in terms of both F1 and AUC with the CIC-IDS-2017 dataset. Indeed, this study did not thoroughly discuss the specific GAN architecture used and rather followed the general architecture. It also did not test its robustness against adversarial attacks.

Park et al. [61] addressed the IDS class imbalance issue by implementing a model consisting of four phases: preprocessing, GAN training, autoencoder training, and classification. They combined the essence of GAN training, specifically the Boundary Equilibrium Generative Adversarial Networks (BEGAN) to generate minority class samples, with autoencoder training for anomaly detection. This approach involves transforming the raw input into a format compatible with DL models in the preprocessing step. BEGAN was employed given its stability during training over traditional GANs. They adopted the reconstruction error and Wasserstein-distance-based GANs to generate minority class samples which are then fed to the subsequent autoencoder phase. For this phase, they trained the autoencoder-driven DNNs and CNNs on both the original dataset and the synthetic data generated by the GAN, and generated features essential for the predictive phase. The predictive phase is a classification model implemented using DNNs, CNNs, and long short-term memory (LSTM) models. Experiments conducted on the NSL-KDD, UNSW-NB15, IoT, and real-world datasets show remarkable improvements in minority classes, particularly the R2L and probe classes in the NSL-KDD data set. The limitations of this study were the lack of experimentation on distributed networks and they looked robustness against adversarial attacks.

Class imbalance can be remedied by under-sampling the majority class. However, this approach can be highly time-consuming and costly and can overlook key samples belonging to the majority class, potentially affecting the classification accuracy. Rao et al. [62] highlighted the need for balancing data, while also avoiding random under-sampling. They proposed an IGAN that achieves both computational efficiency and detection accuracy. Initially, they applied data normalization and one-hot encoding to eliminate the effect of outliers. Then, they employed the IGAN model to generate synthetic attack samples. The actual intrusion detection aspect was implemented using a hybrid of LeNet 5 and an LSTM. Extensive experiments on the UNSW-NB15 and CIC-IDS-2017 datasets reveal an accuracy of >98.0% compared to previous works. Unfortunately, this study did not give detailed GAN architectural design or any justifications regarding the design choices.

Although several studies have applied GANs to overcome data imbalance issues, thus far, very few works have employed CGANs for this task [53]. Babu et al. [63] adopted the Modified Conditional GAN (MCGAN) to generate samples belonging to the minority class. Interestingly, this paper was one of the few that discusses the specific GAN architecture chosen, a modified CGAN. They further used the Nadam optimizer for feature extraction, followed by a linear-correlation-based feature selection. The final classification is carried out using a bidirectional LSTM (Bi-LSTM) algorithm. This model achieves an accuracy of 95.6% on the NSL-KDD+ dataset. Given the high complexity of CGANs, this study did not consider optimization algorithms to optimize computational complexity.

Most research studies have applied GAN to augment datasets, however, Rahman et al. [64] have used GAN network to construct synthetic datasets entirely. Their motivation was the lack of a benchmark dataset and the high cost of data collection in IoT environments. This research gap has not been addressed and requires no labelled datasets. Experiments on the synthetic data created from three datasets UNSW-NB15, NSL-KDD, and BoT-IoT showed that it could effectively train an IDS, achieving results close to those trained on the original datasets.

Data imbalance can pose particularly critical issues in military networks where attacks could have detrimental consequences. Chalé et al. [65] focused on the importance of constructing an accurate and up-to-date IDS that can flag attacks in real-time. They augmented the original dataset by applying statistical modeling, simulations, and a GAN, and further employed decision tree (DT) and random forest (RF) models for classification. The focus of this study was to find the minimum number of real samples required to yield high accuracy. They revealed that a combination of smaller (larger) percentages of real (synthetic) samples results in an underwhelming performance, with high false negative rates. Indeed, this was later remedied by training the model on an equal number of real and synthetic samples, which led to classification accuracy like a classifier trained on a completely real dataset.

Traditionally, optimization algorithms have been widely adopted for feature selection purposes [66]. However, several studies have examined the role of these algorithms in enhancing GAN-based IDSs [67]. Mouyart et al. [67] proposed a system focusing solely on insider threats in a multi-agent environment, consisting of an attacker and a defender. The actual intrusion detection was carried out using an adversarial environment-reinforcement learning (AE-RL) algorithm. Upon testing their approach on the publicly available CMU-CERT version 4.2 dataset, they observed poor results due to its imbalanced nature. Consequently, they utilized a conditional tabular GAN (CTGAN) to generate attack samples to overcome the data imbalance issue. Further, as hyperparameter values substantially affect model performance, they added the tree-structured Parzen estimator (TPE) optimization algorithm for automatic hyperparameter value selection. They finally trained and tested the AE-RL model on the combination of the original and synthetic (generated using optimized CTGAN) datasets, achieving a high recall of 86%. Indeed, this paper addressed the issue of high computational complexity associated with CGANs by exploiting optimization algorithms. However, more comprehensive thorough testing on several optimization algorithms would be beneficial.

Mary et al. [66] enhanced IDS accuracy for a large dataset, that inherently poses significant challenges due to its size, by adopting optimization algorithms for both feature selection and hyperparameter tuning. Specifically, they proposed a novel optimized DL-based IDS model with improved attack detection by selecting a feature subset using the Aquila optimizer (AO) and fuzzy entropy mutual information (FEMI) rather than considering the entire feature set. The selected features are then augmented using an enhanced canonical correlation-based technique and are later fed to an optimized ResNet152-based classifier. Hyperparameter tuning for the classifier is performed using the Wildebeest Herd optimization (WHO) algorithm. This approach reveals excellent values for several evaluation metrics, such as F1-score, specificity, selectivity, accuracy, and ROC curve, for the CICDDoS2019 and ToN-IoT datasets, along with an improved intrusion detection performance. The proposed model could be further investigated on other more recent datasets to prove its effectiveness.

There is an increasing concern that certain IDSs are vulnerable to adversarial training, i.e., exploitable by manually crafted examples that evade detection systems. Alhajjar et al. [68] explored this issue by incorporating evolutionary algorithms, namely particle swarm optimization (PSO) and genetic algorithms, and a GAN to create perturbations of the input examples that can potentially bypass the IDS. Through extensive experiments on two well-known datasets, they indeed demonstrated that all ML algorithms could be bypassed, however, with widely varying robustness. Specifically, support vector machines (SVMs) and DTs recorded the highest misclassification rates, with an evasion rate of >90%. Consequently, this result suggests that such algorithms are not suitable for critical domains that require accurate intrusion detection. The downside of this work is that it did not provide a new attack sample to bypass the IDS.

A broader view is supported by Zhao et al. [69] who not only generated adversarial examples but also introduced a new Wasserstein-GAN-based attack model (attackGAN), with the aim of bypassing NNs in IDSs. The objective of their model was to identify and prevent attacks against IDSs, while also providing feedback for subsequent detections. The adversarial attacks devised in this study achieved a higher success rate compared to existing GAN-based adversarial attack algorithms, such as the fast gradient sign method (FGSM) [3], project gradient descent (PGD) [8], and Carlini & Wagner attacks (CW).

Cloud computing and its security have been a research hotspot given its ever-growing usage. Several recent studies have addressed the class imbalance issue in cloud-based IDSs. Vu et al. [70] proposed a model combining a DNN for intrusion detection with two GANs for data balancing. The first GAN incorporates the conditional denoising adversarial autoencoder (CDAAE) to generate specific instances of the minority class, while the second GAN uses a combination of the CDAAE with the K-nearest neighbor (KNN) algorithm (CDAEE-KNN) to generate samples that lie near the boundary between classifiers. They accomplished intrusion detection on the newly augmented dataset using SVMs, DTs, and RFs and verified the effectiveness of the proposed approach via experiments on six IDS datasets. This work could be further investigated using DL algorithms which are widely used in this domain.

In contrast, Chkirbene et al. [71] addressed the class imbalance issue by optimizing a single GAN. They proposed a novel ML-based secure network model that automatically tunes GAN parameters, such as the number of inner learning steps for the discriminator, which plays a key role in balancing the dataset. Their experiments were conducted on the UNSW and NSL-KDD datasets, which reported increased classification accuracy even for rare classes. This model outperforms state-of-the-art models in identifying attacks in cloud environments.

3.2 GAN for Anomaly-Based IDS

Much of the available literature on IDSs considers them to be supervised learning problems that require considerable amounts of data, which is neither always available nor updated [8]. Consequently, several researchers consider unsupervised or even weakly supervised approaches to be far more effective and realistic and, therefore, better adapted to the rapidly changing network attacks [72]. From a weakly-supervised perspective, Iliyasu et al. [20] modeled benign traffic to construct a normality boundary, and thus, flagged samples falling outside the boundary as anomalies (attacks). However, establishing a normal boundary often results in a high false alarm rate (false positives) which is not useful in today's growing use of the networks. They introduced IDS-based N-GANs, which require minimal attack samples during training. The role of these samples is to allow the anomaly-detection algorithm to learn suitable representations instead of focusing solely on sorting through noise. Evaluation experiments conducted on the CIC-IDS-2017 dataset have revealed promising results (detection rate = 81.3% and area under the curve = 82%) for this model. However, testing on other benchmark datasets was not conducted.

Wang et al. [73] adopted GANs and vision transformers for an anomaly-based IDS. The GAN generates synthetic samples, and the augmented (min-max normalized) dataset is fed to the transformer for anomaly prediction. Network intrusion detection evaluation results for the CIC-IDS-2017 dataset suggest that this approach can be effective for data balancing, while also ensuring accurate predictions.

Questions have been raised about the applicability of the available anomaly-based IDSs for detecting anomalies in time-series data due to their unclear definition and the lack of labeled data in

critical domains, such as aerospace and military. Geiger et al. [74] developed the TadGAN framework which incorporates GAN for anomaly-detection in time-series data. Particularly, this model adopts an LSTM network to record the temporal correlations in time-series distributions. The model is trained with a cycle consistency loss to allow accurate time-series data reconstruction. Additionally, novel algorithms have been introduced for computing the reconstruction errors which are key to calculating the anomaly score. Extensive experiments on several datasets from NASA, Yahoo, Numenta, Amazon, and X (formerly Twitter), have shown that this approach vastly outperforms state-of-the-art methods in terms of the F1-score. Additionally, this model boasts a dramatically low false positive rate, a problem commonly faced by IDSs when dealing with time-series data. This is an open-source model which allows for future research in anomaly detection in time-series data. However, they compared their approach to only one GAN based IDS, therefore, further testing is essential.

The lack of appropriate and updated datasets is a particularly prominent issue in the IoT domain [40,74] which has consequently led to the adoption of anomaly-based IDSs for IoT environments [75]. Ullah et al. [76] proposed an anomaly-based IDS model using a CGAN. This study introduces three variations of CGANs for data augmentation, namely one-class CGAN (ocCGAN), binary-class CGAN (bcCGAN), and multi-class CGAN for generating samples belonging to one, two, and multiple classes, respectively. It further employs distance calculation to reflect the class label. This model exhibits an average accuracy of 98.01% for intrusion detection over seven datasets.

Ezeme et al. [77] used an ocCGAN to construct the distribution of a given profile. The ocCGAN learns the pattern of the minority class and uses it to generate synthetic samples with a similar distribution. The resultant augmented data is then passed to a bcCGAN which constructs a knowledge basis for a cluster-based anomaly detector. Upon testing this model using various datasets that include logs and images against both (non-)GAN-based IDSs, this model has shown excellent values of precision, recall, and F1-score. Yao et al. [78] used BiGANs to build an anomaly-based IDS for IoT networks. This model is trained only on normal IoT data for learning the corresponding distribution. Further, Wasserstein distance is used to build a classifier. The model is further enhanced using a cycle consistency connection between data to avoid information loss, which plays a vital role in decreasing the false positive rate. This approach also addresses the limited capabilities of IoT devices by running and training the model in a fog computing environment. The fog computing environment helps achieve scalability; a prevalent issue faced by IoT-based IDSs. On two benchmark datasets, namely UNSW-NB15 and CIC-IDS-2017, this model achieves a 4% increase in intrusion detection accuracy, as well as a 4% decrease in false positive rates against state-of-the-art results. Their work considered unsupervised learning only, which could be problematic in IoT environment where benign traffic is hard to obtain. Semi-supervised methods were some labelled data is present could bridge this gap.

Existing literature on IoT anomaly detection is focused on developing robust and accurate models [79]. However, little attention has been paid to the construction of lightweight algorithms that take into consideration the limited computational capabilities of IoT devices. Interestingly, Boppana et al. [80] noticed that supervised approaches towards intrusion detection largely fail at detecting real-time and previously unseen attacks. They incorporated GANs and autoencoders (GAN-AE) for an anomaly-based IDS for IoT networks following the Message Queuing Telemetry Transport MQTT protocol and compared their results with other models, such as one-class SVM (OCSVM), autoencoders, and isolation forests (IFs). Tests conducted on a private and a publicly available MQTT dataset reveal a value of 97% in terms of both accuracy and F1-score for this model.

Numerous works have applied optimization algorithms in IoT-based IDSs. Balaji et al. [81] reported the urgent need to protect IoT devices that are vulnerable to a variety of security threats.

They proposed the novel dynamic distributed GAN (DD-GAN) architecture. This study made several advancements in terms of feature engineering, which enhanced classification accuracy and developed the improved Firefly optimization—hybrid DL-based CNN—adaptive neuro-fuzzy inference system (IFFO-HDLCNN-ANFIS) model. Initially, the data is preprocessed by applying the synthetic minority over-sampling technique (SMOTE) and the features are reduced via modified principal component analysis (MPCA). Implementation of IFFO improves classification accuracy by selecting the optimal feature subset from the set of all features. Subsequently, the GAN solves the class imbalance issue before the actual intrusion classification using a combination of CNN and DL algorithms. Despite the promising results reported, the dataset used is not publicly available, making comparisons and testing very challenging.

Shao et al. [82] applied optimization to federated learning IDSs in ICS, but from a different perspective. They proposed one of the earliest attempts toward automating the federated learning architecture using an evolutionary neural architecture search (ENAS) called Fed-GA-CNN-IDS. The advantage of this approach is that it eliminates the need for expertise in hyperparameter settings in the collaborative federated learning environment and relies solely on the results of the ENAS algorithm which automatically tunes several CNNs. This promising approach was tested on the SWAT dataset and suggested that it outperformed both centralized deep learning and machine learning approaches.

3.3 GAN for Hybrid IDS

Studies on IDSs have shown the importance of combining anomaly- and signature-based approaches to achieve the benefits of both. Sharma et al. [75] focused on increasing the detection rate of U2R and R2L attacks in the NSL-KDD dataset. Their approach involved adopting an autoencoder for feature reduction, employing a K-means clustering algorithm for cluster identification, followed by a generative local metric learning (GLML) algorithm to compute the distance between each cluster, a crucial step in anomaly detection. This model achieves superior results in detecting the two types of attacks, with a recall of >81%. Similarly, Xian et al. [72] adopted the hybrid IDS approach via a semi-supervised DL model based on local and non-local regularization. This work is largely motivated by the presence of far more unlabeled data compared to labeled data in real-life situations. This approach utilizes the discriminative deep belief network (DDBN) owing to its robustness in reducing error rates. Average distance is employed with unlabeled data to calculate a threshold to determine class labels. Experiments performed on the KDD Cup99 and NSL-KDD datasets reveal low training and testing error rates for this model.

Mari et al. [83] provided a broader perspective on the usage of GANs in hybrid IDSs by arguing that GANs could be used to generate automatically crafted adversarial attacks that bypass IDSs, which could then be used to train and strengthen the IDS. They tested their proposed approach using the NSL-KDD dataset using the artificial neural network (ANN), KNN, and RF algorithms. The GAN was used to generate realistic stack samples mimicking the actual attacks present in the dataset. Results reveal that this IDS detects artificial attacks with an accuracy of 90%. Moreover, they tested for each attack type present in the dataset to measure the effectiveness of the proposed IDS and found that DoS attacks were the easiest to detect owing to the large number of samples in the dataset.

Aldhaheri et al. [84] adopted a similar methodology, utilizing a GAN to respond to evolving and real-time attacks. They introduced the SGAN-IDS framework that combines a GAN with self-attention mechanisms for generating synthetic attack samples that bypass IDSs. Traditionally, attention is used to allow models to focus only on the part of the input that is relatively more important. The core of the attention component is the multi-head attention layer, which considers the query,

key, and value embeddings to calculate an attention score. It also computes the context features by utilizing the scaled dot product. Experiments conducted on the CIC-IDS-2017 dataset demonstrated that this approach constructs adversarial attacks that remain undetected, showing a 15.93% reduction in the detection rate of IDSs based on SVMs, KNNs, Naive Bayes (NB), logistic regression (LR), and LSTMs.

Strickland et al. [85] followed the IDS hybrid approach, where they incorporated synthetic data obtained from GAN) with Deep Reinforcement Learning (DRL) to classify attacks (binary and multiclass classification). The GAN was trained using NSL-KDD dataset and the resulting synthetic data was used to train the DRL model. Experiments were conducted in both binary and multiclass settings, where the original NSL-KDD dataset was used in the baseline model, whereas the augmented dataset was used in the proposed model. Results suggested that the use of GAN has enhanced the F1-score and improved the detection rate for minority classes. One limitation of this work is its dependency on the NSL-KDD dataset which does not reflect current real-time attacks and rather historical ones, therefore updated and more current datasets would bridge this gap. Table 6 presents a summary of the reviewed GAN-based IDS studies.

Table 6: GAN based IDS literature summary

Paper	Year	GAN model	Domain	Dataset	Results	Pros/cons
[60]	2020	DCGAN	Class imbalance	TheNSL-KDD dataset, the UNSW-NB15d dataset and the CICIDS2017 dataset	Accuracy: 99.79 F1: 99.79 AUC: 99.98	+ Did an ablation study to evaluate the proposed approach.
[61]	2022	Wasserstein distance-based GAN	Class imbalance	NSL-KDD, UNSW-NB15; IoT data set, and real-world data set	ACC: 95.6% F1: 95.8%	+ Evaluated using four datasets including a real-world one. + Considered both binary and multiclass classifications. – Did not cover the preprocessing step in much detail.
[70]	2022	CDAAE and CDAEE-KNN	Class imbalance in the cloud	Cloud IDS Dataset, NSL-KDD, UNSW-NB15, three malware datasets from CTU13s,	F1: 88.6% AUC: 84.2%	+ Considered hybridization of two gans. – Requires data to follow Gaussian distribution. – Lacks appropriate hyperparameter tunning algorithms.
[81]	2022	Dynamic Distributed-Generative Adversarial Network (DD-GAN)	Class imbalance IoT distributed	Daily activity recognition database	Accuracy: 94.45% F1: 93.6%	+ Applied metaheuristic optimization for feature selection. + Considered a dynamic and distributed environment. – Computationally expensive for IoT devices. – Did not explain the classification in much detail.

(Continued)

Table 6 (continued)

Paper	Year	GAN model	Domain	Dataset	Results	Pros/cons
[62]	2022	Imbalanced Generative Adversarial Network (IGAN)	Class imbalance	UNSW-NB15 and CICIDS2017	Accuracy: 98.96% TPR: 96.13	+ Used data normalization. + Considered a hybrid ensemble model for classification. – Manual hyperparameter tunning.
[20]	2022	N-GAN model (GAN and an encoder modules)	Class imbalance in anomaly-based IDS	CIC-IDS2017	DR: 81.3% AUC: 82%	+ Considered a weakly-supervised IDS approach. – Used only one dataset for testing.
[86]	2020	Generative Local Metric Learning (GLML)	Class imbalance in hybrid approach	NSL-KDD	DR: 81% Accuracy: 90.19%	– Focuses on only two attacks. + Could potentially identify unknown attacks.
[67]	2023	Conditional Tabular Generative Adversarial Network (CTGAN)	Class imbalance and optimization	e CMU-CERT	F1: 76% Recall: 86%	+ Applied optimization to balance data. + Considered both binary and multiclass classification. – Used only one dataset. – Limited multiagent environment (one attacker and one defender).
[65]	2022	CTGAN and TVAE	Class imbalance Protect military networks	Private	Recall: 86.1%	+ Found interesting relationship between the percentage of synthetic data to the classifier performance. – The dataset used was not explained thoroughly. – Experiments largely depended on recall.
[63]	2023	MCGAN	Class imbalance	NSL-KDD+	Accuracy: 6.6% F1: 91.88% Recall: 96.07%	+ Adequate comparison with other approaches. – Computationally expensive.
[76]	2021	CGAN	Class imbalance	KDD99, NSLKDD, BoT-IoT, IoT network intrusion, MQTT-IoT-IDS2020, MQTTset and IoT-23	Detection rate: 97% Precision: 94.92% Recall: 97.36%	+ Experiments conducted on several diverse datasets. + Avoids overfitting. – Did not study optimization.

The use of GAN in the IDS context has improved the field remarkably, however, several limitations have not been addressed. First, while GANs are increasingly being employed to generate synthetic data for training IDS, the quality of this synthetic data is often overlooked. The reviewed papers did not report performance measures on the data generated. The effectiveness of an IDS heavily depends on the realism of the generated data, as this directly influences its ability to accurately identify intrusions.

This gap could be bridged by using and introducing new similarity measures that compare synthetic data to the real data. A significant gap in existing GAN-based IDS research is that most studies do not discuss the underlying GAN architecture. This lack of architectural specificity does not allow for reproducibility and limits the ability to evaluate different approaches. Furthermore, comparisons between different GAN architectures using the same experimental design were not conducted, thus, the answer to a question such as which architecture performs best for data balancing remains unclear. Another notable limitation is that only a few studies adequately address the threat of adversarial attacks [13,69]. The majority of papers overlooked the adversarial attacks that make such IDSs vulnerable.

4 Challenges and Future Directions

The rapid growth of Internet use has led to an increased need for improved security measures. Thus, IDSs play a vital role in ensuring the security and integrity of computer networks. These systems aim to monitor network flow, identify abnormal activities, and detect potential intrusions or attacks. While considerable research studies have aimed at harnessing the power of generative networks in IDS, they generally followed the same perspective. That is to utilize generative networks' remarkable ability to generate synthetic and realistic examples for dataset enlargement purposes. Furthermore, much of the available literature assumed centralized and computationally powerful environments, that can support the computational requirements required by the proposed approaches. There have been a few empirical investigations into the use of generative networks in other environments such as IoT networks, decentralized networks, and federated networks. Overall, the present review identifies several challenges and potential future research as listed below:

(1) *Lack of an updated benchmark dataset*: Thus far, there has been little agreement on the dataset to be used for investigating the potential of GANs in IDSs. Although several datasets are available, it has been noticed that most studies reviewed have used different parts of a dataset or even private datasets in certain cases. Unfortunately, this issue significantly affects the results, as one model can perform exceptionally well with one dataset, while failing dramatically with another one exhibiting a potentially varied data distribution. Sharafaldin et al. [87] empirically tested several benchmark datasets on a single IDS model and found many drawbacks such as lack of diversity and relevance. Furthermore, they noted that most datasets are outdated and do not reflect the evolving threat landscape. In the same vein, Kumar et al. [22] have highlighted the need for benchmark datasets in their systematic research review, and stated that the lack of benchmark datasets does not allow for accurate verification of proposed models. These factors indicate a need to propose a new benchmark dataset that allows for fair comparisons among models.

(2) *Lack of empirical studies on IoT environments*: A major issue with studies concerning IoT environments is that they only incorporate IoT datasets and assume the computational power of a computer rather than a limited IoT device. Consequently, much uncertainty remains about the ability of IoT devices, that are limited in memory and computational capabilities, to handle the heavy load of GAN-based IDS models. Memory use testing and real-life experiments with IoT devices have been largely overlooked. For instance, IDSs for home surveillance purposes should be designed carefully considering the limited capacities of IoT devices such as smart locks. They also should be tested realistically on the actual environment to avoid problems such as the system being flooded by vast amounts of data transmitted from the IoT services [88]. Thus, developing embedded lightweight models that consider actual IoT device capabilities would bridge this gap.

(3) *Centralized environment assumptions*: Traditionally, most work in the IDS field has been concentrated on centralized approaches where several network nodes depend on a trusted centralized server to govern the intrusion detection process. Although such approaches deliver promising results, they can be rather unrealistic. In most scenarios, it is difficult to maintain a centralized server that can be completely protected while efficiently handling a tremendous volume of network traffic. Moreover, data privacy laws do not allow organizations or even devices to transmit their data to a central server that is susceptible to exploitation by malicious users. Consequently, IDS research has shifted towards decentralized approaches. Unfortunately, current GAN-based IDS studies have not been responsive to the decentralization requirements and continue to largely focus on centralized approaches. IDS that exploit GAN to overcome data imbalance issues consume considerable amounts of computing resources which is often available in centralized environments. However, in decentralized environments such as IoT networks and edge computing, resources are limited and thus there is an urgent need to consider decentralized environments and their impact on communication cost, speed, and security.

(4) *Scalability and efficiency*: GAN and DL models are traditionally complex and consume vast amounts of computational resources. Moreover, with today's high-speed networks and the large volumes of traffic they generate, it is vital to develop algorithms that can scale well. Therefore, the applicability of the proposed models remains largely questionable. Efficiency concerns have been prominent in this field for several years. With the focus being mainly on performance metrics, such as accuracy, recall, and F1-score, computational efficiency and model scalability have been largely neglected. Without focusing on efficiency, it is hard to achieve scalability, and thus, future works must consider both. Possible research questions that should be answered by future research are: How to optimize GAN and DL models for resource-constrained environments? How to develop evaluation metrics that consider both performance and efficiency? How to exploit computational capabilities such as GPUs to enhance efficiency?

(5) *Lack of automatic hyperparameter tuning methods*: The performance of DL models heavily depends on hyperparameter tuning. Empirical research has shown that changing the hyperparameters dramatically influences the results. Surprisingly, only a few studies have attempted the inclusion of optimization algorithms. For instance, in a GAN model, numerous parameters must be set prior to the data generation, such as learning rate, batch size, and number of layers. Arriving at the setting of a parameter that generates good performance requires numerous tries. Thus, the development of automatic hyperparameter tuning methods, with a focus on the complexities associated with these algorithms, can help significantly improve the IDS models.

5 Conclusion

The widespread use of the Internet has resulted in an explosion of network traffic which requires adequate protection. Consequently, there has been an increased interest in strengthening IDSs and developing robust and advanced approaches. One such approach was the adoption of GAN to address several limitations, such as dataset availability and imbalance issues. This review aims to shine a light on the role of generative networks in IDSs, their use cases, benefits, and any research gaps that would guide future research. Overall, it is evident that the inclusion of GANs has enhanced IDSs, especially in terms of data balancing. Future works must focus on harnessing GANs in other areas, such as distributed environments and IoT networks, with a special focus on scalability and efficiency.

Acknowledgement: None.

Funding Statement: This research paper was supported by a grant from the "Research Centre of the Female Scientific and Medical Colleges", Deanship of Scientific Research, King Saud University.

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: Monirah Al-Ajlan, Mourad Ykhlef; data collection: Monirah Al-Ajlan; analysis and interpretation of results: Monirah Al-Ajlan, Mourad Ykhlef; draft manuscript preparation: Monirah Al-Ajlan. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. H. Farea, O. H. Alhazmi, and K. Kucuk, "Advanced optimized anomaly detection system for IoT cyberattacks using Artificial Intelligence," *Comput. Mater. Contin.*, vol. 78, no. 2, pp. 1525–1545, 2024. doi: 10.32604/cmc.2023.045794.

[2] O. R. Arogundade, "Network security concepts, dangers, and defense best practical," *Comput. Eng. Intell. Syst.*, vol. 14, no. 2, pp. 25–38, 2023.

[3] S. Duggineni, "Data integrity and risk," *Open J. Optim.*, vol. 12, no. 2, pp. 25–33, 2023. doi: 10.4236/ojop.2023.122003.

[4] M. A. Khan, A. Rehman, K. M. Khan, M. A. Al Ghamdi, and S. H. Almotiri, "Enhance intrusion detection in computer networks based on deep extreme learning machine," *Comput. Mater. Contin.*, vol. 66, no. 1, pp. 467–480, 2021. doi: 10.32604/cmc.2020.013121.

[5] E. Viegas, A. Santin, A. Bessani, and N. Neves, "BigFlow: Real-time and reliable anomaly-based intrusion detection for high-speed networks," *Future Gener. Comput. Syst.*, vol. 93, pp. 473–485, 2019. doi: 10.1016/j.future.2018.09.051.

[6] M. A. Khan and Y. Kim, "Deep learning-based hybrid intelligent intrusion detection system," *Comput. Mater. Contin.*, vol. 68, no. 1, pp. 671–687, 2021. doi: 10.32604/cmc.2021.015647.

[7] A. H. Mohammad, "Intrusion detection using a new hybrid feature selection model," *Intell. Autom. Soft Comput.*, vol. 30, no. 1, pp. 65–80, 2021. doi: 10.32604/iasc.2021.016140.

[8] R. Xu, G. Wu, W. Wang, X. Gao, A. He and Z. Zhang, "Applying self-supervised learning to network intrusion detection for network flows with graph neural network," *Comput. Netw.*, vol. 248, 2024, Art. no. 110495. doi: 10.1016/j.comnet.2024.110495.

[9] Z. Chkirbene, S. Eltanbouly, M. Bashendy, N. AlNaimi, and A. Erbad, "Hybrid machine learning for network anomaly intrusion detection," in *2020 IEEE Int. Conf. Inform., IoT, Enabling Technol. (ICIoT)*, IEEE, 2020, pp. 163–170.

[10] M. Alawida, S. Mejri, A. Mehmood, B. Chikhaoui, and O. I. Abiodun, "A comprehensive study of ChatGPT: Advancements, limitations, and ethical considerations in natural language processing and cybersecurity," *Information*, vol. 14, no. 8, 2023, Art. no. 462. doi: 10.3390/info14080462.

[11] Y. -J. Cao *et al.*, "Recent advances of generative adversarial networks in computer vision," *IEEE Access*, vol. 7, pp. 14985–15006, 2018.

[12] T. Iqbal and H. Ali, "Generative adversarial network for medical images (MI-GAN)," *J. Med. Syst.*, vol. 42, pp. 1–11, 2018. doi: 10.1007/s10916-018-1072-9.

[13] A. B. Singh, L. K. Awasthi, M. Shorfuzzaman, A. Alsufyani, and M. Uddin, "Chained dual-generative adversarial network: A generalized defense against adversarial attacks," *Comput. Mater. Contin.*, vol. 74, no. 2, pp. 2541–2555, 2023. doi: 10.32604/cmc.2023.032795.

[14] Y. Hong, U. Hwang, J. Yoo, and S. Yoon, "How generative adversarial networks and their variants work: An overview," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1–43, 2019.
[15] Q. Feng, C. Guo, F. Benitez-Quiroz, and A. M. Martinez, "When do gans replicate? On the choice of dataset size," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 6701–6710.
[16] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," 2018, *arXiv:1809.11096*.
[17] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. 4th Int. Conf. Inform. Syst. Secur. Priv. (ICISSP)*, 2018, vol. 1, pp. 108–116. doi: 10.5220/0006639801080116.
[18] H. Debar, "An introduction to intrusion-detection systems," in *Proc. Connect*, 2000, vol. 2000, pp. 1–18.
[19] C. Kruegel and T. Toth, "Using decision trees to improve signature-based intrusion detection," in *Int. Workshop Recent Adv. Intrusion Detect.*, Springer, 2003, pp. 173–191.
[20] A. S. Iliyasu and H. Deng, "N-GAN: A novel anomaly-based network intrusion detection with generative adversarial networks," *Int. J. Inform. Technol.*, vol. 14, no. 7, pp. 3365–3375, 2022. doi: 10.1007/s41870-022-00910-3.
[21] J. -C. Huang, G. -Q. Zeng, G. -G. Geng, J. Weng, K. -D. Lu and Y. Zhang, "Differential evolution-based convolutional neural networks: An automatic architecture design method for intrusion detection in industrial control systems," *Comput. Secur.*, vol. 132, 2023, Art. no. 103310. doi: 10.1016/j.cose.2023.103310.
[22] G. Kumar and H. Alqahtani, "Machine learning techniques for intrusion detection systems in SDN-recent advances, challenges and future directions," *Comput. Model. Eng. Sci.*, vol. 134, no. 1, pp. 89–119, 2023. doi: 10.32604/cmes.2022.020724.
[23] N. Niknami, E. Inkrott, and J. Wu, "Towards analysis of the performance of IDSs in software-defined networks," in *2022 IEEE 19th Int. Conf. Mobile Ad Hoc Smart Syst. (MASS)*, IEEE, 2022, pp. 787–793.
[24] A. Abubakar and B. Pranggono, "Machine learning based intrusion detection system for software defined networks," in *2017 Seventh Int. Conf. Emerg. Secur. Technol. (EST)*, IEEE, 2017, pp. 138–143.
[25] A. Shenfield, D. Day, and A. Ayesh, "Intelligent intrusion detection systems using artificial neural networks," *ICT Express*, vol. 4, no. 2, pp. 95–99, 2018. doi: 10.1016/j.icte.2018.04.003.
[26] H. Alqahtani *et al.*, "Cyber intrusion detection using machine learning classification techniques," in *Int. Conf. Comput. Sci., Commun. Secur.*, Springer, 2020, pp. 121–131.
[27] A. Khraisat, I. Gondal, and P. Vamplew, "An anomaly intrusion detection system using C5 decision tree classifier," in *Pacific-Asia Conf. Knowl. Discov. Data Min.*, Springer, 2018, pp. 149–155.
[28] Y. Otoum and A. Nayak, "AS-IDS: Anomaly and signature based ids for the internet of things," *J. Netw. Syst. Manage.*, vol. 29, no. 3, pp. 1–26, 2021. doi: 10.1007/s10922-021-09589-6.
[29] G. Pu, L. Wang, J. Shen, and F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," *Tsinghua Sci. Technol.*, vol. 26, no. 2, pp. 146–153, 2020. doi: 10.26599/TST.2019.9010051.
[30] P. Kumar, G. P. Gupta, and R. Tripathi, "Design of anomaly-based intrusion detection system using fog computing for IoT network," *Automa. Control Comput. Sci.*, vol. 55, no. 2, pp. 137–147, 2021. doi: 10.3103/S0146411621020085.
[31] Q. A. Al-Haija and A. Al-Badawi, "Attack-aware IoT network traffic routing leveraging ensemble learning," *Sensors*, vol. 22, no. 1, 2021, Art. no. 241. doi: 10.3390/s22010241.
[32] M. Awad, S. Fraihat, K. Salameh, and A. Al Redhaei, "Examining the suitability of NetFlow features in detecting IoT network intrusions," *Sensors*, vol. 22, no. 16, 2022, Art. no. 6164. doi: 10.3390/s22166164.
[33] P. Dini *et al.*, "Design and testing novel one-class classifier based on polynomial interpolation with application to networking security," *IEEE Access*, vol. 10, pp. 67910–67924, 2022. doi: 10.1109/ACCESS.2022.3186026.
[34] M. Sarhan, G. Kulatilleke, W. W. Lo, S. Layeghy, and M. Portmann, "DOC-NAD: A hybrid deep one-class classifier for network anomaly detection," 2022, *arXiv:2212.07558*.
[35] M. I. Alghamdi, "A hybrid model for intrusion detection in IoT applications," *Wirel. Commun. Mob. Comput.*, vol. 2022, pp. 4553502–4553511, 2022.

[36] S. Einy, C. Oz, and Y. D. Navaei, "The anomaly-and signature-based IDS for network security using hybrid inference systems," *Math. Probl. Eng.*, vol. 2021, pp. 1–10, 2021. doi: 10.1155/2021/6639714.
[37] G. Khraisat, K. Vamplew, and Alazab, "A novel ensemble of hybrid intrusion detection system for detecting Internet of Things attacks," *Electronics*, vol. 8, no. 11, 2019, Art. no. 1210. doi: 10.3390/electronics8111210.
[38] A. Golrang, A. M. Golrang, S. Y. Yayilgan, and O. Elezaj, "A novel hybrid IDS based on modified NSGAII-ANN and random forest," *Electronics*, vol. 9, no. 4, 2020, Art. no. 577. doi: 10.3390/electronics9040577.
[39] S. Garcia, M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," *Comput. Secur. J.*, vol. 45, pp. 100–123, 2014. doi: 10.1016/j.cose.2014.05.011.
[40] A. Thakkar and R. Lohiya, "A review of the advancement in intrusion detection datasets," *Procedia Comput. Sci.*, vol. 167, pp. 636–645, 2020. doi: 10.1016/j.procs.2020.03.330.
[41] S. Chaudhuri, D. Madigan, and U. Fayyad, "KDD-99: The fifth ACM SIGKDD international conference on knowledge discovery and data mining," *ACM SIGKDD Explor. Newsl.*, vol. 1, no. 2, pp. 49–51, 2000. doi: 10.1145/846183.846194.
[42] M. Sarhan, S. Layeghy, N. Moustafa, and M. Portmann, "Netflow datasets for machine learning-based network intrusion detection systems," in *Big Data Technologies and Applications*. Cham: Springer, 2021, vol. 371, pp. 117–135.
[43] L. Yu, L. Xu, and X. Jiang, "An effective method for detecting unknown types of attacks based on log-cosh variational autoencoder," *Appl. Sci.*, vol. 13, no. 22, 2023, Art. no. 12492. doi: 10.3390/app132212492.
[44] R. K. Cunningham *et al.*, "Evaluating intrusion detection systems without attacking your friends: The 1998 DARPA intrusion detection evaluation," *Proc. ID*, 1999, vol. 99, pp. 1–6.
[45] R. Thomas and D. Pavithran, "A survey of intrusion detection models based on NSL-KDD data set," in *2018 Fifth HCT Inform. Technol. Trends (ITT)*, 2018, pp. 286–291.
[46] G. Creech and J. Hu, "Generation of a new IDS test dataset: Time to retire the KDD collection," in *2013 IEEE Wireless Commun. Netw. Conf. (WCNC)*, IEEE, 2013, pp. 4487–4492.
[47] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *2015 Mil. Communi. Inform. Syst. Conf. (MilCIS)*, IEEE, 2015, pp. 1–6.
[48] M. Durgadevi, "Generative adversarial network (GAN): A general review on different variants of GAN and applications," in *2021 6th Int. Conf. Commun. Electron. Syst. (ICCES)*, IEEE, 2021, pp. 1–8.
[49] R. Sauber-Cole and T. M. Khoshgoftaar, "The use of generative adversarial networks to alleviate class imbalance in tabular data: A survey," *J. Big Data*, vol. 9, no. 1, 2022, Art. no. 98. doi: 10.1186/s40537-022-00648-6.
[50] J. Kalin, *Generative Adversarial Networks Cookbook: Over 100 Recipes to Build Generative Models Using Python, TensorFlow, and Keras*. Cincinnati, OH, USA: Packt Publishing Ltd., 2018.
[51] H. Alqahtani, M. Kavakli-Thorne, and G. Kumar, "Applications of generative adversarial networks (GANs): An updated review," *Arch. Comput. Methods Eng.*, vol. 28, pp. 525–552, 2021. doi: 10.1007/s11831-019-09388-y.
[52] G. G. Chrysos, J. Kossaifi, and S. Zafeiriou, "RoCGAN: Robust conditional GAN," *Int. J. Comput. Vis.*, vol. 128, pp. 2665–2683, 2020. doi: 10.1007/s11263-020-01348-5.
[53] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
[54] J. -Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
[55] V. Sampath, I. Maurtua, J. J. A. Martin, and A. Gutierrez, "A survey on generative adversarial networks for imbalance problems in computer vision tasks," *J. Big Data*, vol. 8, pp. 1–59, 2021. doi: 10.1186/s40537-021-00414-0.
[56] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.
[57] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," 2017, *arXiv:1710.10196*.

[58] V. S. Bhaskara, T. Aumentado-Armstrong, A. D. Jepson, and A. Levinshtein, "GraN-GAN: Piecewise gradient normalization for generative adversarial networks," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2022, pp. 3821–3830.
[59] I. Goodfellow *et al.*, "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 27, pp. 27–36, 2014.
[60] S. Huang and K. Lei, "IGAN-IDS: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks," *Ad Hoc Netw.*, vol. 105, 2020, Art. no. 102177. doi: 10.1016/j.adhoc.2020.102177.
[61] C. Park, J. Lee, Y. Kim, J. -G. Park, H. Kim and D. Hong, "An enhanced AI-based network intrusion detection system using generative adversarial networks," *IEEE Internet Things J.*, vol. 10, no. 3, pp. 2330–2345, 2022. doi: 10.1109/JIOT.2022.3211346.
[62] Y. N. Rao and K. S. Babu, "An imbalanced generative adversarial network-based approach for network intrusion detection in an imbalanced dataset," *Sensors*, vol. 23, no. 1, 2023, Art. no. 550. doi: 10.3390/s23010550.
[63] K. S. Babu and Y. N. Rao, "MCGAN: Modified conditional generative adversarial network (MCGAN) for class imbalance problems in network intrusion detection system," *Appl. Sci.*, vol. 13, no. 4, 2023, Art. no. 2576. doi: 10.3390/app13042576.
[64] S. Rahman, S. Pal, S. Mittal, T. Chawla, and C. Karmakar, "SYN-GAN: A robust intrusion detection system using GAN-based synthetic data for IoT security," *Internet of Things*, vol. 26, 2024, Art. no. 101212. doi: 10.1016/j.iot.2024.101212.
[65] M. Chalé and N. D. Bastian, "Generating realistic cyber data for training and evaluating machine learning classifiers for network intrusion detection systems," *Expert. Syst. Appl.*, vol. 207, 2022, Art. no. 117936. doi: 10.1016/j.eswa.2022.117936.
[66] D. S. Mary, L. J. S. Dhas, A. R. Deepa, M. A. Chaurasia, and C. J. J. Sheela, "Network intrusion detection: An optimized deep learning approach using big data analytics," *Expert. Syst. Appl.*, vol. 251, 2024, Art. no. 123919. doi: 10.1016/j.eswa.2024.123919.
[67] M. Mouyart, G. M. Machado, and J. -Y. Jun, "A multi-agent intrusion detection system optimized by a deep reinforcement learning approach with a dataset enlarged using a generative model to reduce the bias effect," *J. Sens. Actuator Netw.*, vol. 12, no. 5, 2023, Art. no. 68. doi: 10.3390/jsan12050068.
[68] E. Alhajjar, P. Maxwell, and N. Bastian, "Adversarial machine learning in network intrusion detection systems," *Expert. Syst. Appl.*, vol. 186, 2021, Art. no. 115782. doi: 10.1016/j.eswa.2021.115782.
[69] S. Zhao, J. Li, J. Wang, Z. Zhang, L. Zhu and Y. Zhang, "Attackgan: Adversarial attack against black-box ids using generative adversarial networks," *Procedia Comput. Sci.*, vol. 187, pp. 128–133, 2021. doi: 10.1016/j.procs.2021.04.118.
[70] L. Vu, Q. U. Nguyen, D. N. Nguyen, D. T. Hoang, and E. Dutkiewicz, "Deep generative learning models for cloud intrusion detection systems," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 565–577, 2022. doi: 10.1109/TCYB.2022.3163811.
[71] Z. Chkirbene, H. B. Abdallah, K. Hassine, R. Hamila, and A. Erbad, "Data augmentation for intrusion detection and classification in cloud networks," in *2021 Int. Wireless Commun. Mobile Comput. (IWCMC)*, IEEE, 2021, pp. 831–836.
[72] G. Xian, "Cyber intrusion prevention for large-scale semi-supervised deep learning based on local and non-local regularization," *IEEE Access*, vol. 8, pp. 55526–55539, 2020. doi: 10.1109/ACCESS.2020.2981162.
[73] Z. Wang, J. Zhou, and X. Hei, "Network traffic anomaly detection based on generative adversarial network and transformer," in *Int. Conf. Natural Comput. Fuzzy Syst. Knowl. Discov.*, Springer, 2022, pp. 228–235.
[74] A. Geiger, D. Liu, S. Alnegheimish, A. Cuesta-Infante, and K. Veeramachaneni, "Tadgan: Time series anomaly detection using generative adversarial networks," in *2020 IEEE Int. Conf. Big Data (Big Data)*, IEEE, 2020, pp. 33–43.
[75] B. Sharma, L. Sharma, C. Lal, and S. Roy, "Anomaly based network intrusion detection for IoT attacks using deep learning technique," *Comput. Electr. Eng.*, vol. 107, 2023, Art. no. 108626. doi: 10.1016/j.compeleceng.2023.108626.

[76] I. Ullah and Q. H. Mahmoud, "A framework for anomaly detection in IoT networks using conditional generative adversarial networks," *IEEE Access*, vol. 9, pp. 165907–165931, 2021. doi: 10.1109/ACCESS.2021.3132127.

[77] O. M. Ezeme, Q. H. Mahmoud, and A. Azim, "Design and development of AD-CGAN: Conditional generative adversarial networks for anomaly detection," *IEEE Access*, vol. 8, pp. 177667–177681, 2020. doi: 10.1109/ACCESS.2020.3025530.

[78] W. Yao, H. Shi, and H. Zhao, "Scalable anomaly-based intrusion detection for secure Internet of Things using generative adversarial networks in fog environment," *J. Netw. Comput. Appl.*, vol. 214, 2023, Art. no. 103622. doi: 10.1016/j.jnca.2023.103622.

[79] K. Albulayhi, A. A. Smadi, F. T. Sheldon, and R. K. Abercrombie, "IoT intrusion detection taxonomy, reference architecture, and analyses," *Sensors*, vol. 21, no. 19, 2021, Art. no. 6432. doi: 10.3390/s21196432.

[80] T. K. Boppana and P. Bagade, "GAN-AE: An unsupervised intrusion detection system for MQTT networks," *Eng. Appl. Artif. Intell.*, vol. 119, 2023, Art. no. 105805. doi: 10.1016/j.engappai.2022.105805.

[81] S. Balaji and S. S. Narayanan, "Dynamic distributed generative adversarial network for intrusion detection system over internet of things," *Wirel. Netw.*, vol. 29, no. 5, pp. 1949–1967, 2023. doi: 10.1007/s11276-022-03182-8.

[82] J. -M. Shao, G. -Q. Zeng, K. -D. Lu, G. -G. Geng, and J. Weng, "Automated federated learning for intrusion detection of industrial control systems based on evolutionary neural architecture search," *Comput. Secur.*, vol. 143, 2024, Art. no. 103910. doi: 10.1016/j.cose.2024.103910.

[83] A. -G. Mari, D. Zinca, and V. Dobrota, "Development of a machine-learning intrusion detection system and testing of its performance using a generative adversarial network," *Sensors*, vol. 23, 2023, Art. no. 1315. doi: 10.3390/s23031315.

[84] S. Aldhaheri and A. Alhuzali, "SGAN-IDS: Self-attention-based generative adversarial network against intrusion detection systems," *Sensors*, vol. 23, no. 18, 2023, Art. no. 7796. doi: 10.3390/s23187796.

[85] C. Strickland *et al.*, "DRL-GAN: A hybrid approach for binary and multiclass network intrusion detection," *Sensors*, vol. 24, no. 9, 2024, Art. no. 2746. doi: 10.3390/s24092746.

[86] M. Saharkhizan, A. Azmoodeh, H. HaddadPajouh, A. Dehghantanha, R. M. Parizi and G. Srivastava, "A hybrid deep generative local metric learning method for intrusion detection," in *Handbook of Big Data Privacy*, NY, USA: Springer, pp. 343–357, 2020.

[87] I. Sharafaldin, A. Gharib, A. H. Lashkari, and A. A. Ghorbani, "Towards a reliable intrusion detection benchmark dataset," *Softw. Netw.*, vol. 2018, no. 1, pp. 177–200, 2018. doi: 10.13052/jsn2445-9739.2017.009.

[88] A. Thakkar and R. Lohiya, "A review on machine learning and deep learning perspectives of IDS for IoT: Recent updates, security issues, and challenges," *Arch. Comput. Methods Eng.*, vol. 28, no. 4, pp. 3211–3243, 2021. doi: 10.1007/s11831-020-09496-0.