2024 International Conference on Smart Applications, Communications and Networking (SmartNets) | 979-8-3503-8532-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/SMARTNETS61466.2024.10577682

Lightweight Intrusion Detection System with GAN-based Knowledge Distillation

Tarek Ali
Department of Computing and Mathematics
Manchester Metropolitan University
Manchester, United Kingdom
tarek.ali@stu.mmu.ac.uk

Amna Eleyan
Department of Computing and Mathematics
Manchester Metropolitan University
Manchester, United Kingdom
a.eleyan@mmu.ac.uk

Tarek Bejaoui
Computer Engineering Department
University of Carthage
Tunisia
tarek.bejaoui@ieee.org

Mohammed Al-Khalidi
Department of Computing and Mathematics
Manchester Metropolitan University
Manchester, United Kingdom
m.al-khalidi@mmu.ac.uk

***Abstract*— In the rapidly changing realm of network security, creating efficient and flexible Intrusion Detection Systems (IDS) is crucial to combat the escalating complexity of network threats. Our study presents an innovative methodology that combines Generative Adversarial Networks (GANs) with knowledge distillation strategies to amplify the effectiveness of IDS, particularly in settings limited by computational resources like the Internet of Things (IoT) and Industrial Internet of Things (IIoT) networks. The fundamental novelty of our suggested system resides in its utilisation of GANs to produce varied datasets, which are subsequently employed to train deep learning models customised for intrusion detection. This approach empowers the IDS to adjust to different network setups and guarantees thorough defence against a broad range of attacks, emphasising adversarial assaults. By utilising knowledge distillation, we simplify the development of compact models that uphold the detection capabilities of their more intricate counterparts, presenting an optimal solution for settings with limited resources. Merging adversarial training with knowledge distillation enhances the IDS's resilience against adversarial threats. Empirical findings validate the efficiency of our method, showcasing its capacity to sustain high accuracy rates while ensuring resource effectiveness and adaptability in intricate and resource-constrained environments. This study signifies a notable progression in network security, providing a versatile, effective, and resilient IDS solution capable of functioning under diverse network conditions without compromising performance.**

***Index Terms*—Intrusion Detection System, Lightweight, Knowledge Distillation, Generative Adversarial Networks, Network Security, Deep Learning**

I. Introduction

The escalating complexity and frequency of network-based assaults necessitate advancing robust methodologies and instruments. These are essential to safeguard against service interruptions, unauthorised entries, and the exposure of confidential data [1, 2]. Within this context, Intrusion Detection Systems (IDS) emerge as a pivotal defence component against sophisticated and proliferating threats targeting network security [3]. However, the efficacy of these systems, particularly those predicated on Machine Learning (ML) algorithms, is contingent upon the availability of extensive, authentic, and pertinent network traffic datasets. Contemporary datasets have indeed made strides in encompassing a spectrum of network attack modalities and traffic configurations and providing insights into the attacking infrastructures. Nevertheless, modern network environments' dynamic and heterogeneous nature often renders these datasets insufficient for developing effective classification models [4, 5]. An expected shortfall of these datasets is their limited traffic diversity and volume and an incomplete representation of the entire gamut of known attack typologies. In order to effectively mitigate the ever-changing environment of network threats, it is critical to have datasets that possess greater dynamism. Including such datasets would bolster the intrusion detection system's capacity to detect and react to malicious operations. Deep learning techniques, including Generative Adversarial Networks (GANs), offer a potentially fruitful pathway. The utilisation of GANs enables the synthesis of additional data from pre-existing datasets. This enhancement can significantly improve the accuracy of an intrusion detection system's classification, especially when detecting rare types of network attacks.

IDS employs two primary methodologies: Signature-based Detection Systems (SNIDS) and Anomaly-based Detection Systems (ANIDS). SNIDS operates by identifying predefined patterns or "signatures" within network traffic, which could include specific byte sequences or sequences of instructions known to be malicious, thus proving effective against recognised threats[1]. In contrast, ANIDS leverages Machine Learning (ML) algorithms to scrutinise network traffic, aiming to unearth any activities that deviate from the norm, thereby offering a robust solution for detecting novel attacks [6].

The advent of deep learning technologies has significantly

979-8-3503-8532-8/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: INDIAN INST OF INFO TECH AND MANAGEMENT. Downloaded on March 06,2025 at 15:37:10 UTC from IEEE Xplore. Restrictions apply.

enhanced IDS capabilities, introducing advanced methods capable of identifying even the most sophisticated network threats. Deep Learning (DL) employs artificial neural networks to enable software agents, or ”learning entities,” to adapt and achieve objectives through function approximation and target optimisation. This approach facilitates mapping attack patterns to identify malicious activities within networks [3].

The dynamic nature of network environments is fundamental since their behaviours and patterns undergo evolution throughout time [7, 8]. Likewise, the dynamic nature of vulnerabilities influences the effectiveness of Intrusion Detection Systems (IDS) in maintaining a high level of categorisation accuracy. The difficulty is further exacerbated by the progressive expiration of current datasets, resulting in obsolescence, invalidity, and unreliability. Moreover, the dissemination of trustworthy data is sometimes impeded by privacy apprehensions, and publicly accessible datasets fall short of covering the whole range of recognised network attack categories, let alone new risks and weaknesses [9]. In order to tackle these problems, it is crucial to have a broader range of up-to-date datasets that precisely represent the characteristics of network intrusions, thereby improving the performance of Intrusion Detection Systems (IDS) [10, 11, 12].

We propose incorporating SNIDS with DL approaches as a solution to these difficulties. Our methodology entails using Generative Adversarial Networks (GANs) to create diverse datasets, which are then subjected to deep learning approaches for the development and training of Intrusion Detection System (IDS) models on these GAN-generated datasets. Our study utilises publicly available datasets, including CICIDS2017, IoT-23, and NSL-KDD [4], as shown in table I.

Using GANs, we aim to get a more equitable allocation of classes within the datasets. Afterwards, we train deep learning models using the original data and the data produced by deep neural networks (GANs). We then evaluate these models' classification situations. Our work evaluates the efficacy of deep learning models trained using synthetic datasets against adversarial attacks compared to conventional training methods.

II. Related Work

Network security defence significantly benefits from applying IDS. Initially, IDSs relied on predefined, static, or adaptable rules for identifying network intrusions. These systems were primarily effective in straightforward situations but struggled to address novel security threats [13]. The advent of ML brought about the incorporation of various algorithms into NID, including K-Nearest Neighbors (KNN) [14], Support Vector Machines (SVM) [15], and Light Gradient Boosting Machine (LightGBM) [16]. Despite these advancements, the evolving complexity and variety of network threats have outpaced the adaptive capabilities of these ML-based approaches [17].

DL has recently emerged as a potent tool for discerning underlying patterns and representing data characteristics. It has shown considerable success in feature extraction from anomalous traffic through an integrated process [18]. A notable study introduced a DL methodology utilising CNN for intrusion detection, achieving superior accuracy in identifying abnormal traffic compared to traditional ML-based IDSs. However, this approach revealed limitations in accurately classifying less frequent attack types in multi-class scenarios. To enhance the detection of these minority classes, Imrana et al. [11] introduced an IDS based on a bidirectional Long Short-Term Memory (BiDLSTM) network. Similarly, D'Angelo and Palmieri [19] integrated an autoencoder with CNN and Recurrent Neural Network (RNN) architectures to extract and leverage the relationships between spatial and temporal data features, thereby improving network traffic classification. Belarbi et al. [12] developed a multi-class classification IDS employing a Deep Belief Network (DBN) constructed from multiple Restricted Boltzmann Machines (RBMs) layers, demonstrating its efficacy with the CICIDS 2017 dataset.

While complex network models have elevated IDS accuracy, their deployment on devices with limited resources poses a significant challenge. A streamlined model was proposed to mitigate this, albeit at the expense of reduced detection capabilities [20]. Knowledge distillation has emerged as another strategy to simplify model complexity. Wang et al. [21] introduced a knowledge distillation framework aimed at enhancing model efficiency. This method transfers insights from a more complex ”teacher” model to a more straightforward ”student” model, though designing effective teacher-student model pairs remains a complex task.

III. Proposed Methodology

Our IDS solution is designed to enhance intrusion detection's adaptability and efficiency in environments characterised by limited resources, such as those found in IoT and IIoT networks. Our approach is distinguished by a three-pronged strategy that ensures comprehensive protection and adaptability across various network settings, as shown in Figure 1.

Firstly, we extensively test our IDS model across three diverse datasets to ensure broad coverage and effectiveness against various threats. The KDD Cup '99 dataset provides a foundational benchmark with its extensive collection of simulated network intrusions. The IoT-23 dataset, tailored explicitly to IoT environments, allows us to assess the IDS's efficacy in handling IoT-specific threats. Lastly, the CICD2017 dataset, developed by the Canadian Institute for Cybersecurity in 2017, presents modern and complex attack scenarios, ensuring our system is well-equipped to tackle contemporary cybersecurity challenges.

We leverage knowledge distillation techniques to facilitate the deployment of our IDS in resource-constrained environments. This involves training a compact, efficient ”student” model to emulate the performance of a more complex ”teacher” model by learning from the teacher's outputs. This process ensures that the student model retains high detection capabilities while optimising lightweight settings.

Moreover, our solution incorporates adversarial training and distillation to enhance resilience against adversarial attacks. Adversarial training strengthens the model by exposing it to

Authorized licensed use limited to: INDIAN INST OF INFO TECH AND MANAGEMENT. Downloaded on March 06,2025 at 15:37:10 UTC from IEEE Xplore. Restrictions apply.

TABLE I
CICIDS2017, KDD CUP 99, AND IOT-23 DATASETS

Feature/Aspect	CICIDS2017	KDD Cup 99	IoT-23
Year Released	2017	1999	2020
Source	Canadian Institute for Cybersecurity	DARPA	Avast AIC laboratory
Primary Use	IDS testing and research	IDS testing and research	IDS testing and research, IoT focus
Data Type	Network traffic flows	Network connections	Network traffic flows from IoT devices
Malware Focus	General network attacks	General network attacks	IoT-specific malware attacks
Volume	Large (2.5 million records)	Very large (5 million records)	Large (over 10 million records)
Class Imbalance	Yes	Yes	Yes
Real Devices Used	Simulated traffic	Simulated traffic	Real IoT devices
Annotation	Labeled attacks	Labeled attacks	Labeled IoT malware attacks
Usage in Research	Modern environments	Historical and baseline	IoT security

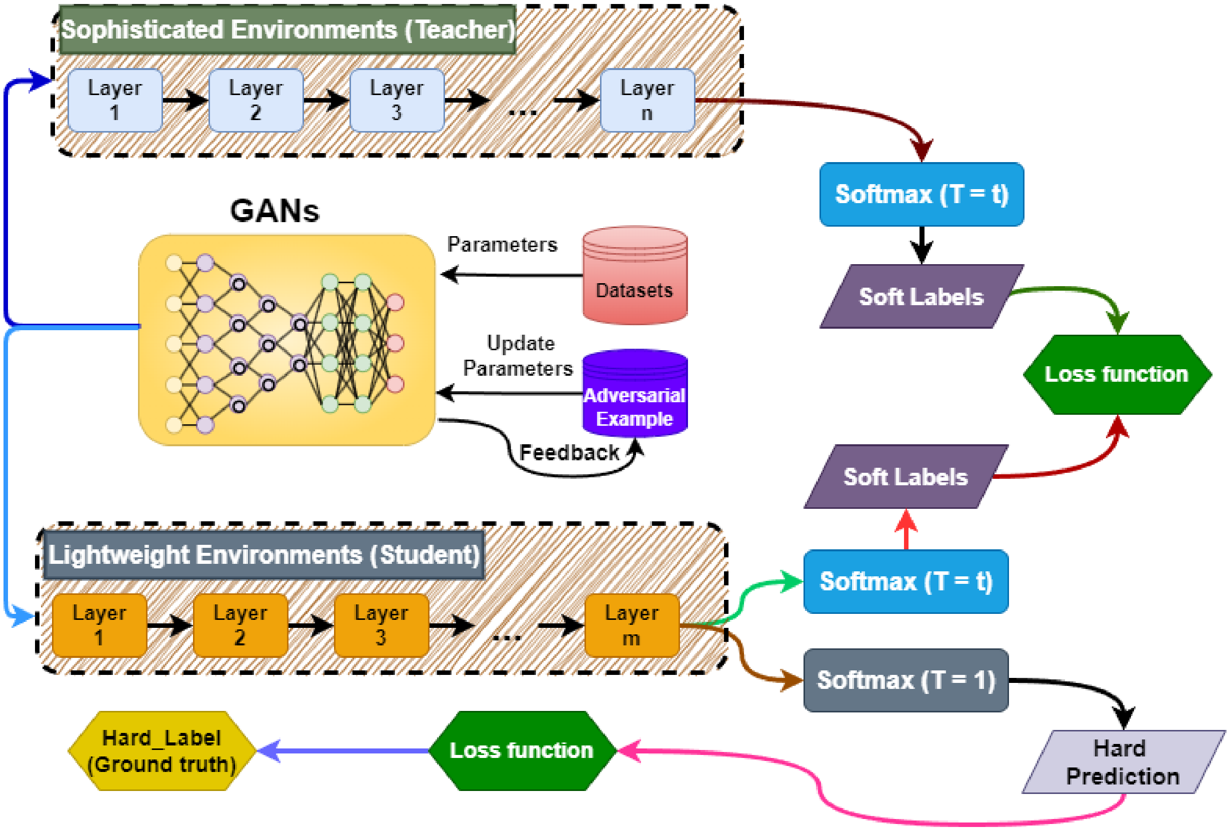

Fig. 1. Lightweight Intrusion Detection System with GANs.

malicious inputs during its training phase, thereby improving its ability to withstand sophisticated evasion techniques employed by attackers. The combination of adversarial training and distillation produces a student model that is resource-efficient and robust against adversarial threats.

Compared to other IDS solutions, our methodology offers several distinct advantages. Its ability to recognise a broad spectrum of attack vectors across different datasets demonstrates versatility. Knowledge distillation ensures resource efficiency, making our IDS particularly suitable for IoT and IIoT applications where computational resources are scarce. Additionally, integrating adversarial training with distillation enhances the security of our system, providing superior protection against adversarial attacks. Lastly, the adaptability of our solution to both complex and resource-limited environments without significant loss in performance distinguishes it from traditional IDS solutions, which may need help transitioning to such settings.

A. Advancements in Generative Adversarial Networks

A paradigm shift occurred in unsupervised machine learning with the introduction of Generative Adversarial Networks (GANs), an innovative study by Goodfellow et al. [22] and his team in 2014. Generating synthetic data that accurately mimics input data was an innovative objective behind developing GANs. This technology mitigates the difficulties arising from expanding datasets, which are frequently laborious, expensive,

Authorized licensed use limited to: INDIAN INST OF INFO TECH AND MANAGEMENT. Downloaded on March 06,2025 at 15:37:10 UTC from IEEE Xplore. Restrictions apply.

and time-intensive. The interaction between two distinct models, the discriminator and the generator, distinguishes the architecture of VCFs. The generator generates additional synthetic data by identifying and capitalising on similarities or patterns in the input data. In contrast, the discriminator operates as a classifier by differentiating between the initial and generated data and providing an evaluation for each instance of the data. The system generates a probability score between 0 and 1 to represent the degree of certainty regarding the data's authenticity (approaching 1) or artificiality (approaching 0). Figure 2 provides a comprehensive schematic overview of the workflow of GANs in order to clarify their operational dynamics see the Algorithm 2.

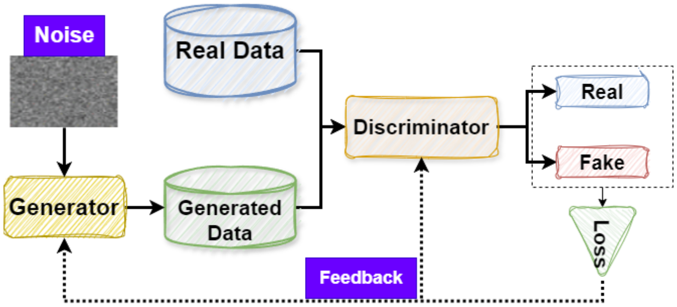

Fig. 2. Schematic Overview of Generative Adversarial Network Functionality.

Algorithm 1 Vanilla Generative Adversarial Network Training

```
Input: Real data distribution $D_{\text{real}}$, noise distribution $D_{\text{noise}}$, number of epochs $E$, batch size $m$, learning rate $\alpha$
Output: Trained generator $G$ and discriminator $D$
Initialize generator $G$ and discriminator $D$ with random weights
for $e = 1$ to $E$ do
  for $k = 1$ to $m$ do
    Sample minibatch of $m$ noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise distribution $D_{\text{noise}}$
    Sample minibatch of $m$ examples $\{x^{(1)}, \dots, x^{(m)}\}$ from real data distribution $D_{\text{real}}$
    Update the discriminator $D$ by ascending its stochastic gradient:
      $\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[ \log D(x^{(i)}) + \log(1 - D(G(z^{(i)}))) \right]$
  end for
  Sample minibatch of $m$ noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise distribution $D_{\text{noise}}$
  Update the generator $G$ by descending its stochastic gradient:
    $\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z^{(i)})))$
end for
return Trained generator $G$ and discriminator $D$
```

B. Distillation Technique for Defence

Defensive distillation is a technique in machine learning that was initially introduced by Hinton et al. [23] to transfer the knowledge from a large, complex neural network (referred to as the "teacher") to a smaller, simpler neural network (referred to as the "student"). This process was demonstrated to enable the student network to achieve performance levels comparable to the teacher network. Originally applied to classification problems within the teacher-student framework, this method was later adapted by Papernot et al. [24] for use in defending against adversarial machine learning attacks, showing that it could increase the robustness of models to such threats.

The defensive distillation process begins by training the teacher model using a high temperature parameter in the softmax function, which serves to soften the probability outputs of the deep learning model. The softmax function at a high temperature is defined as follows:

$$p_{\text{softmax}}(z, T) = \frac{e^{z/T}}{\sum_{i=1}^{n} e^{z(i)/T}} \quad (1)$$

In this equation, n represents the number of possible labels, and z is the output from the final layer of the deep learning model, which is calculated using the weight matrix $\mathbf{W}_n$, the activation from the previous layer $\mathbf{a}_{n-1}$, and the bias b_n.

Subsequently, the softened outputs from the teacher model are used to train the student model at a lower temperature setting. The objective function for training the student model is expressed as:

$$\mathcal{L}_{\text{student}}(T) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} \mathbf{y}_{ij} \cdot \log p_{\text{softmax}}(z_{ij}, T) \quad (2)$$

Here, N is the total number of training samples, $\mathbf{y}_{ij}$ is the true label, and z_{ij} represents the logits. The objective function for training the teacher model is similarly defined but with a negative sign, emphasising the minimization of the difference between the predicted probabilities and the true labels.

$$\mathcal{L}_{teacher}(T) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} \mathbf{y}_{ij} \cdot \log \frac{e^{z_{ij}/T}}{\sum_{i=1}^{n} e^{z_{ij}/T}} \quad (3)$$

By minimising these objective functions, the student model is trained using the soft targets provided by the teacher model as shown in Algorithm 2, thereby enhancing the model's robustness and its ability to withstand adversarial examples.

IV. Experimental Setup

The experimental setup for evaluating the performance of adversarial and conventional training approaches on IDS is meticulously designed to compare the effectiveness of these methods across various datasets and model complexities as demonstrated in algorithm 3. The setup involves two distinct training pathways: One is for conventional training, and the other is for adversarial training, as depicted in Figure 1. The experiment selects three datasets in the conventional training pathway: NLS-KDD, CICIDS2017, and IoT-23. These datasets

Authorized licensed use limited to: INDIAN INST OF INFO TECH AND MANAGEMENT. Downloaded on March 06,2025 at 15:37:10 UTC from IEEE Xplore. Restrictions apply.

Algorithm 2 Defensive Distillation Procedure

```
Input: Training set $D_1$, a well-trained model $T$ (teacher), a less complex model $S$ (student), a defined loss function $\mathcal{L}$, learning rate $\eta$, total training cycles $E$
Output: Effectively trained student model $S$
Begin by setting initial weights for the student model $S$
for each epoch $e$ from 1 to $E$ do
    Shuffle the training set $D_1$ randomly
    for each sample index $i$ from 1 to the size of $D_1$ do
        Take the $i^{th}$ data point $(x_i, y_i)$ from $D_1$
        Pass the data point $x_i$ through the teacher model $T$ to get predicted probabilities $\hat{y}_i$
        Calculate the loss $\mathcal{L}$ with the predicted probabilities $\hat{y}_i$
        Apply backpropagation of the loss $\mathcal{L}$ through the student model $S$
        Adjust the weights of the student model $S$ using the learning rate $\eta$
    end for
end for
Return the trained student model $S$
```

are fed into two types of neural network models: a complex model representing a sophisticated environment akin to a ”teacher” and a lightweight model simulating a resource-constrained environment similar to a ”student.” The models are trained using a loss function, and their parameters are updated iteratively based on the loss. The network output is then compared to the true labels of the dataset, and the training process is repeated until the network achieves the desired level of accuracy. The setup for the adversarial training pathway is analogous to the conventional training but includes an additional step where an attacker generates adversarial examples to deceive the network. These adversarial examples are incorporated into the training process to enhance the model's resilience against such attacks. The adversarial training employs a loss function that accounts for the adversarial examples, and the models are trained to resist these examples effectively.

V. Results and Discussion

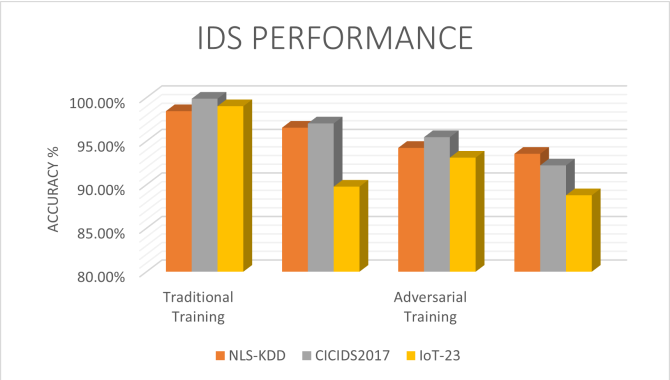

Fig. 3. The Experimental Results.

Algorithm 3 Evaluation of IDS Performance with Conventional and Adversarial Training

```
Require: Datasets (NLS-KDD, CICIDS2017, IoT-23), Sophisticated (Complex) Model, Lightweight Model, Loss Function, Adversarial Example Generation Technique
Initialize the complex and lightweight models with pre-trained weights if available
for each dataset in (NLS-KDD, CICIDS2017, IoT-23) do
    Split the dataset into training and testing sets
    Train the complex model on the training set using conventional training methods
    Evaluate the complex model on the testing set and record accuracy
    Train the lightweight model on the training set using conventional training methods
    Evaluate the lightweight model on the testing set and record accuracy
    Generate adversarial examples using the adversarial example generation technique
    Retrain the complex model on the training set with adversarial examples using adversarial training methods
    Evaluate the adversarially trained complex model on the testing set and record accuracy
    Retrain the lightweight model on the training set with adversarial examples using adversarial training methods
    Evaluate the adversarially trained lightweight model on the testing set and record accuracy
end for
Compare the accuracies of the complex and lightweight models under both training methods
Analyze the trade-offs between accuracy and robustness for each model and training method
Present the results in a table and bar chart for visual comparison
```

TABLE II
Comparison of IDS Performance: Traditional vs. Adversarial Training

Dataset	Traditional Training		Adversarial Training	
	Complex	Lightweight	Complex	Lightweight
NLS-KDD	98.44%	96.54%	94.22%	93.54%
CICIDS2017	99.87%	97.03%	95.45%	92.21%
IoT-23	99.01%	89.77%	93.10%	88.78%

The outcomes depicted in Table II and Figure 3 provide an exhaustive comparison of adversarial and conventional training approaches across various datasets and model complexities. In conventional training scenarios, the performance of the lightweight models is consistently inferior to that of the complex models across all datasets. Notably, the complex model achieved the highest accuracy of 97.03% on the CICIDS2017 dataset, while the lightweight model managed 97.03%. This phenomenon highlights the efficacy of conventional training approaches in harnessing intricate models' computational prowess to attain exceptional precision. On the contrary, adver-

Authorized licensed use limited to: INDIAN INST OF INFO TECH AND MANAGEMENT. Downloaded on March 06,2025 at 15:37:10 UTC from IEEE Xplore. Restrictions apply.

sarial training results in a discernible reduction in accuracy for models of varying complexity and lightweight, encompassing all datasets. The accuracies of the complex model for the NLS-KDD, CICIDS2017, and IoT-23 datasets are 94.22%, 95.50%, and 93.10%, respectively. In the CICIDS2017 dataset, the accuracy of the lightweight models decreases most notably, from 97.03% to 92.21%. Notwithstanding this decrease, the findings underscore the robustness of adversarial training in sustaining comparatively elevated levels of precision, mainly when applied to lightweight models engineered for environments with limited resources. The marginal decline in precision may be ascribed to the trade-off between robustness and accuracy intrinsic to adversarial training. This trade-off emphasises safeguarding against adversarial attacks rather than achieving maximum accuracy. In general, the results of this study emphasise the significance of choosing the proper training approach to the particular demands of the operational setting and the intended trade-off between precision and protection.

VI. Conclusion

Our research investigates the pressing requirement for effective and flexible IDS, given the escalating complexity of network threats. We present an innovative methodology that integrates GANs with knowledge distillation methods to bolster the effectiveness of IDS in settings where computational resources are scarce, such as IoT and IIoT networks.

The unique capability of the proposed system, IDS, is its utilisation of GANs to generate diverse datasets, which are subsequently employed to train deep-learning models designed for intrusion detection. By adopting this methodology, IDS enhances its capability to adjust to diverse network configurations and guarantees all-encompassing safeguarding against an extensive array of attacks, particularly emphasising adversarial assaults. By enabling the development of lightweight models that retain the detection capabilities of more complex systems, knowledge distillation is an especially suitable solution for environments with limited resources.

In addition, incorporating adversarial training and knowledge distillation into the IDS strengthens its resilience against adversarial assaults. The experimental outcomes provide evidence of the proposed methodology's efficacy, showcasing its capacity to sustain elevated standards of precision while guaranteeing resource conservation and flexibility in intricate and resource-constrained environments.

To put it briefly, the IDS we have suggested signifies a substantial progression in network security. It effectively overcomes the constraints of conventional IDS by providing a flexible, effective, and resilient solution that can function under various network conditions without significantly degrading performance. Further investigation into various knowledge distillation techniques would enhance and refine the system's capabilities. The findings of this study possess the capacity to substantially influence the advancement of intrusion detection systems of the subsequent generation, thereby guaranteeing enhanced security measures for an ever more interconnected global society.

References

[1] A. AliAhmad, D. Eleyan, A. Eleyan, T. Bejaoui, M. F. Zolkipli, and M. Al-Khalidi, "Malware detection issues, future trends and challenges: A survey," in *2023 International Symposium on Networks, Computers and Communications (ISNCC)*, 2023, pp. 1–6.

[2] M. A.-K. Tarek Ali and R. Al-Zaidi, "Information security risk assessment methods in cloud computing: Comprehensive review," *Journal of Computer Information Systems*, vol. 0, no. 0, pp. 1–28, 2024. [Online]. Available: https://doi.org/10.1080/08874417.2024.2329985

[3] T. Ali, A. Eleyan, and T. Bejaoui, "Detecting conventional and adversarial attacks using deep learning techniques: A systematic review," in *2023 International Symposium on Networks, Computers and Communications (ISNCC)*, 2023, pp. 1–7.

[4] R. Yousef, M. Jazzar, A. Eleyan, and T. Bejaoui, "A machine learning framework & development for insider cyber-crime threats detection," in *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)*, 2023, pp. 1–6.

[5] T. Taylor and A. Eleyan, "Using variational autoencoders to increase the performance of malware classification," in *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, 2021, pp. 1–6.

[6] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, pp. 1–22, 2019.

[7] S. Medileh, A. Laouid, E. M. B. Nagoudi, R. Euler, A. Bounceur, M. Hammoudeh, M. AlShaikh, A. Eleyan, and O. A. Khashan, "A flexible encryption technique for the internet of things environment," *Ad Hoc Networks*, vol. 106, p. 102240, 2020.

[8] S. Medileh, A. Laouid, M. Hammoudeh, M. Kara, T. Bejaoui, A. Eleyan, and M. Al-Khalidi, "A multi-key with partially homomorphic encryption scheme for low-end devices ensuring data integrity," *Information*, vol. 14, no. 5, 2023. [Online]. Available: https://www.mdpi.com/2078-2489/14/5/263

[9] F. Lalem, A. Laouid, M. Kara, M. Al-Khalidi, and A. Eleyan, "A novel digital signature scheme for advanced asymmetric encryption techniques," *Applied Sciences*, vol. 13, no. 8, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/8/5172

[10] K. Wu, Z. Chen, and W. Li, "A novel intrusion detection model for a massive network using convolutional neural networks," *Ieee Access*, vol. 6, pp. 50 850–50 859, 2018.

[11] Y. Imrana, Y. Xiang, L. Ali, and Z. Abdul-Rauf, "A bidirectional lstm deep learning approach for intrusion detection," *Expert Systems with Applications*, vol. 185, p. 115524, 2021.

[12] O. Belarbi, A. Khan, P. Carnelli, and T. Spyridopoulos, "An intrusion detection system based on deep belief

Authorized licensed use limited to: INDIAN INST OF INFO TECH AND MANAGEMENT. Downloaded on March 06,2025 at 15:37:10 UTC from IEEE Xplore. Restrictions apply.

networks," in *International Conference on Science of Cyber Security*. Springer, 2022, pp. 377–392.

[13] A. Drewek-Ossowicka, M. Pietrołaj, and J. Rumiński, "A survey of neural networks usage for intrusion detection systems," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 497–514, 2021.

[14] H. Ding, L. Chen, L. Dong, Z. Fu, and X. Cui, "Imbalanced data classification: A knn and generative adversarial networks-based hybrid approach for intrusion detection," *Future Generation Computer Systems*, vol. 131, pp. 240–254, 2022.

[15] J. Gu and S. Lu, "An effective intrusion detection approach using svm with naïve bayes feature embedding," *Computers & Security*, vol. 103, p. 102158, 2021.

[16] J. Liu, Y. Gao, and F. Hu, "A fast network intrusion detection system using adaptive synthetic oversampling and lightgbm," *Computers & Security*, vol. 106, p. 102289, 2021.

[17] R. Vinayakumar, K. Soman, and P. Poornachandran, "Evaluating effectiveness of shallow and deep networks to intrusion detection system," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2017, pp. 1282–1289.

[18] A. Dawabsheh, M. Jazzar, A. Eleyan, T. Bejaoui, and S. Popoola, "An enhanced phishing detection tool using deep learning from url," in *2022 International Conference on Smart Applications, Communications and Networking (SmartNets)*, 2022, pp. 1–6.

[19] G. D'Angelo and F. Palmieri, "Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial–temporal features extraction," *Journal of Network and Computer Applications*, vol. 173, p. 102890, 2021.

[20] R. Zhao, G. Gui, Z. Xue, J. Yin, T. Ohtsuki, B. Adebisi, and H. Gacanin, "A novel intrusion detection method based on lightweight neural network for internet of things," *IEEE Internet of Things Journal*, 2021.

[21] Z. Wang, Z. Li, D. He, and S. Chan, "A lightweight approach for network intrusion detection in industrial cyber-physical systems based on knowledge distillation and deep metric learning," *Expert Systems with Applications*, p. 117671, 2022.

[22] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, and Others, "Generative adversarial networks," 2014.

[23] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015.

[24] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, "Distillation as a defense to adversarial perturbations against deep neural networks," 2016.

Authorized licensed use limited to: INDIAN INST OF INFO TECH AND MANAGEMENT. Downloaded on March 06,2025 at 15:37:10 UTC from IEEE Xplore. Restrictions apply.