IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Adversarial Intelligence: Leveraging GANs for Enhanced Autonomous Intrusion Detection in Cybersecurity Systems

Amarjeet Srivastava[1], Shiwangi Choudhary[2]

[1]M. Tech Scholar, Dept. of CSE, Rameshwaram Institute of Technology & Management (AKTU), Lucknow, India

[2] Assistant Professors, Dept. of CSE, Rameshwaram Institute of Technology & Management, (AKTU), Lucknow, India

Abstract— The ever-evolving threat landscape in cyberspace necessitates advanced and adaptive security mechanisms. Traditional Intrusion Detection Systems (IDS) often struggle to detect novel and stealthy attacks due to their dependence on static rules or labeled data. This paper explores the integration of Generative Adversarial Networks (GANs) into autonomous IDS frameworks to enhance their ability to detect complex and unknown cyber threats. GANs, through their adversarial learning paradigm, enable the simulation of sophisticated attack patterns, which can be used to augment training datasets and improve the robustness of detection models. We propose a novel adversarial intelligence-driven IDS that dynamically evolves by learning from synthetic attack behaviors generated by the GAN. The experimental results demonstrate significant improvements in anomaly detection accuracy and false positive rate reduction, showcasing the potential of GANs to revolutionize intelligent threat detection. This approach provides a scalable and proactive defense mechanism, aligning with the future of autonomous cybersecurity systems.

Keywords— Adversarial Intelligence, Generative Adversarial Networks (GANs), Intrusion Detection System (IDS), Cybersecurity, Autonomous Systems, Anomaly Detection, Threat Simulation, Deep Learning.

I. INTRODUCTION

In today's hyper-connected digital era, cybersecurity has emerged as a critical concern across industries, with cyber threats growing increasingly sophisticated and evasive. Intrusion Detection Systems (IDS) serve as the frontline defense mechanisms, monitoring network traffic and identifying malicious activities.

However, traditional IDS approaches, such as signature-based and anomaly-based detection, often face limitations when confronting novel or zero-day attacks due to their reliance on static signatures or handcrafted features [1][2]. These limitations underscore the pressing need for more adaptive, intelligent, and autonomous detection mechanisms.

Machine Learning (ML) and Deep Learning (DL) techniques have shown significant promise in addressing these challenges by enabling systems to learn patterns and detect anomalies in network traffic data [3][4]. However, even these models can suffer from data scarcity, imbalanced datasets, and overfitting issues, which may lead to poor generalization in real-world deployment [5]. This is where Generative Adversarial Networks (GANs) present a compelling opportunity.

Introduced by Goodfellow et al. [6], GANs consist of two neural networks—a generator and a discriminator—engaged in a competitive process. The generator creates synthetic data instances, while the discriminator evaluates their authenticity, effectively pushing both networks to improve.

In the context of cybersecurity, GANs can be utilized to generate realistic attack samples or augment training datasets, enhancing the learning capability of IDS models [7][8]. This form of adversarial intelligence allows IDS frameworks not only to detect known threats but also to anticipate and identify previously unseen attack vectors.

Recent studies have begun exploring the synergy between GANs and IDS for improved threat detection accuracy and resilience. For instance, Li et al. [9] used GANs to balance imbalanced datasets in intrusion detection, resulting in notable improvements in classification performance. Similarly, Lin et al. [10] applied adversarial training to make IDS models more robust against evasion attacks. These works affirm the growing interest in leveraging GANs for strengthening autonomous cybersecurity defenses.

This paper introduces a comprehensive framework for Adversarial Intelligence-Driven Intrusion Detection, leveraging GANs to simulate adversarial behaviors and enhance anomaly detection capabilities in autonomous IDS systems. By integrating GAN-generated data into the training pipeline, our system continuously evolves, becoming more resilient and effective in identifying diverse threat vectors.

II. LITERATURE SURVEY

The integration of deep learning techniques into Intrusion Detection Systems (IDS) has gained significant traction in recent years due to their ability to learn complex patterns in network traffic data. However, limitations such as data imbalance, adversarial vulnerabilities, and insufficient representation of attack patterns have prompted researchers to explore advanced models like Generative Adversarial Networks (GANs) to bolster IDS capabilities.

Traditional IDS approaches often rely on signature-based detection methods, which are effective for known threats but fail against zero-day and novel attacks [1]. Anomaly-based IDS, though better at identifying unknown threats, often suffer from high false positive rates due to the unpredictability of normal network behavior [11]. To address these limitations, researchers began adopting machine learning (ML) and deep learning (DL) models for IDS, such as support vector machines (SVM), decision trees (DT), and convolutional neural networks (CNN), which significantly improved detection accuracy [12][13].

Despite these advancements, deep learning models require large and diverse training datasets to generalize well. Network intrusion datasets like NSL-KDD and CICIDS are often imbalanced, with a higher volume of normal data and a scarcity of labeled attack data [14]. This imbalance adversely affects the performance of supervised models, leading to bias and increased misclassification rates.
To mitigate this, Generative Adversarial Networks (GANs) have been introduced to synthetically generate realistic attack samples and balance datasets [15].

Goodfellow et al. [16] introduced GANs as a framework composed of two networks—a generator and a discriminator—trained in opposition to improve data generation realism. Applying GANs to cybersecurity, Lin et al. [17] proposed IDSGAN, which creates adversarial attack traffic to test IDS robustness. Their approach highlighted the ability of GANs to simulate intelligent adversarial behaviors and stress-test detection models.

Similarly, Li et al. [18] used a GAN-based approach to generate synthetic intrusion data, improving the diversity and quality of training datasets. Their experiments on the NSL-KDD dataset showed a measurable increase in detection performance and reduced false positives. Another study by Wang et al. [19] integrated Conditional GANs (cGANs) to generate class-specific attack traffic, aiding classifiers in identifying minority classes with greater accuracy.

Beyond data augmentation, GANs are also being used for adversarial training to make IDS models more resilient to evasion attacks. Shapoorifard et al. [20] explored adversarial examples generated by GANs to train IDS against sophisticated evasion techniques. This approach fortified IDS models by exposing them to adversarial samples during training, enhancing their robustness.

Moreover, recent works like that of Rahman et al. [21] introduced Semi-Supervised GANs (SS-GAN) for intrusion detection, which leverage unlabeled data—abundant in real-world networks—to improve model generalization. This is particularly useful for dynamic environments where labeling new attack patterns is not feasible in real time.

Despite these promising developments, challenges remain. Training GANs is computationally intensive and prone to instability, such as mode collapse or non-convergence [22]. Furthermore, the quality of synthetic data must be evaluated carefully to ensure it truly represents real-world attack patterns without introducing bias.

In summary, literature shows that GAN-enhanced IDS frameworks hold significant promise in addressing data imbalance, improving anomaly detection, and enhancing system robustness. However, ensuring stable GAN training and evaluating adversarial intelligence in real-time scenarios remain active research challenges.

TABLE 1: LITERATURE REVIEW TABLE BASED ON PREVIOUS YEAR RESEARCH PAPER METHODOLOGY AND KEY FINDINGS

S.No	Title	Authors	Year	Methodology	Key Findings
1	IDSGAN: Generative Adversarial Networks for Attack Generation	Lin et al.	2020	GAN-based attack generation	Generated adversarial examples bypassed IDS effectively.
2	Generative Adversarial Nets	Goodfellow et al.	2014	GAN architecture	Introduced the GAN concept for data generation tasks.
3	Enhancing Intrusion Detection Systems with GAN-based Data Augmentation	Li et al.	2021	Data augmentation using GAN	Improved classification performance and reduced imbalance effects.
4	Adversarial Attack and Defense in Deep Learning-based IDS	Lin et al.	2021	GAN + adversarial training	IDS robustness improved with GAN-generated attack samples.
5	Conditional GAN for Network Intrusion Detection	Wang et al.	2021	Conditional GAN	Generated class-specific attacks to enhance minority class detection.
6	Semi-Supervised GANs for Intrusion Detection in CPS	Rahman et al.	2021	Semi-supervised GAN	Leveraged unlabeled data to enhance detection accuracy.
7	Generating Adversarial Traffic Using GANs	Kim et al.	2020	GAN for synthetic traffic	Simulated complex traffic patterns to test IDS.
8	A Survey of GANs for Intrusion Detection	Shafiq et al.	2021	Literature survey	Reviewed GANs' application for IDS.
9	Adversarial	Huang et al.	2018	Survey +	Summarized

	Machine Learning in Network Security			taxonomy	adversarial ML techniques for IDS and defenses.
10	GAN-Based Oversampling for Intrusion Detection	Zhang et al.	2022	Data balancing using GAN	Alleviated imbalance in intrusion datasets using GAN oversampling.
11	GANIDS: GAN-based Intrusion Detection System	Li and Liu	2020	GAN + CNN classifier	GAN improved CNN's performance by realistic sample generation.
12	Improving Network Security Using GAN and LSTM	Singh et al.	2021	GAN + LSTM hybrid	Improved time-series anomaly detection accuracy.
13	Towards Robust Intrusion Detection Using GANs	Ahmed et al.	2019	GAN-enhanced anomaly detection	Increased detection rate against sophisticated attacks.
14	Detecting Adversarial Attacks on IDS Using GAN-generated Traces	Patel et al.	2020	Defense model testing with adversarial traces	GAN helped identify IDS vulnerabilities.
15	DeepIDS: Deep Learning-based IDS with GAN Data	Kumar et al.	2021	Deep CNN + GAN augmentation	Achieved higher accuracy on NSL-KDD dataset.
16	GAN for Imbalanced Cybersecurity Data	Chen et al.	2021	Data balancing technique	Reduced bias in detection of rare attacks.
17	Cyber Attack Generation Using GANs	Ali et al.	2019	GAN for synthetic cyber-attack creation	Enabled the simulation of unseen threats for IDS training.
18	Adversarial Training of Deep IDS Systems Using GANs	Tao et al.	2022	GAN-based adversarial training	Increased system resilience against evasion attacks.
19	NID-GAN: Network Intrusion Detection with GAN and Feature	Nair et al.	2022	GAN + custom feature engineering	Improved detection metrics through

	Engineering				hybrid feature generation.
20	A Novel Adversarial Framework for Robust IDS	Rahmani and Zhou	2020	Adversarial learning + feature transformation	Demonstrated improved detection of stealthy attacks.

III. METHODOLOGY

This study proposes a novel approach for enhancing Intrusion Detection Systems (IDS) using Generative Adversarial Networks (GANs) to simulate, detect, and defend against evolving cyber threats in an autonomous manner. The methodology is divided into five primary phases: Data Acquisition and Preprocessing, GAN-based Data Augmentation, Intrusion Detection Model Development, Adversarial Training, and Evaluation.

A. Data Acquisition and Preprocessing

The system utilizes benchmark intrusion detection datasets such as NSL-KDD, CICIDS2017, and UNSW-NB15 to ensure a diverse and realistic range of attack vectors. Preprocessing steps include:

- Data cleaning to remove null or inconsistent records.
- Encoding categorical variables using one-hot or label encoding.
- Feature scaling through min-max normalization.
- Splitting the dataset into training, validation, and test sets.

The imbalance in class distribution (attack vs. normal traffic) is addressed in subsequent phases using GANs.

B. GAN-Based Data Augmentation

To counteract data imbalance and improve IDS performance, a Conditional GAN (cGAN) is implemented. This consists of:

- Generator: Learns to generate realistic network traffic samples (normal or malicious) conditioned on class labels.
- Discriminator: Learns to distinguish between real and synthetic samples, improving the generator's output over time.

This synthetic data is then combined with the original dataset to provide a more balanced and diverse training corpus for the IDS.

C. Intrusion Detection Model Development

A deep learning-based IDS model is developed, trained on the enriched dataset. The model includes:

- Feature Extractor: A multi-layer deep neural network (DNN) or CNN to learn latent features from traffic data.
- Classifier: A Softmax or sigmoid output layer for multi-class or binary classification of intrusion types.
- Loss Function: Categorical cross-entropy or binary cross-entropy, optimized using Adam optimizer.

This model forms the baseline IDS before adversarial hardening.

D. Adversarial Training for Robustness

To enhance resilience against adversarial threats, adversarial training is conducted:

- GANs are used to create adversarial attack samples designed to fool the IDS.
- These adversarial examples are injected into the training set.
- The IDS model is retrained with these samples to learn robust detection boundaries.

This iterative adversarial training loop strengthens the IDS against evasion and mimicry attacks.

E. Evaluation and Performance Metrics

The proposed GAN-enhanced IDS model is evaluated using:

- Accuracy: Overall correctness of predictions.
- Precision, Recall, and F1-Score: To assess class-specific performance.
- ROC-AUC Score: For imbalanced classification performance.
- False Positive Rate (FPR) and False Negative Rate (FNR): To measure detection reliability.
- Training Time & Convergence Stability: To analyze GAN training efficiency.

Comparisons are made with baseline models (e.g., traditional ML, CNN-only IDS, RNN) to demonstrate performance improvement.

F. Tools and Frameworks Used

- Python (TensorFlow / PyTorch for GAN implementation)
- Scikit-learn and Keras for model development
- Pandas, NumPy for preprocessing
- Matplotlib, Seaborn for result visualization.

IV. RESULTS ANALYSIS

The proposed GAN-enhanced Intrusion Detection System (IDS) was evaluated using three benchmark datasets—NSL-KDD, CICIDS2017, and UNSW-NB15—to assess its effectiveness in detecting both known and adversarial cyberattacks. Comparative analysis was carried out between Baseline Models, GAN-Augmented Models, and Adversarially Trained GAN-IDS Models.

A. Performance Metrics Used

Accuracy (ACC), Precision, Recall, F1-Score, False Positive Rate (FPR), and False Negative Rate (FNR)

Table 2: Model Accuracy Comparison on NSL-KDD Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FPR (%)	FNR (%)
Traditional CNN-Based IDS	89.72	86.11	85.35	85.73	10.21	14.65
RNN-Based IDS	91.43	89.26	87.98	88.61	8.45	12.02
GAN-Augmented CNN IDS	94.65	93.22	92.75	92.98	4.71	7.25
Adversarially Trained GAN-IDS	96.83	95.54	95.14	95.34	2.92	4.86

B. Key Observations

- Accuracy Improvement: GAN-augmented models showed a significant boost in accuracy (~4-7%) over traditional deep learning models.
- Robustness to Adversarial Attacks: Adversarially trained GAN-IDS demonstrated the highest resilience to evasion attacks, showing reduced false negative rates.
- Enhanced Minority Class Detection: Class imbalance, especially in multi-class intrusion scenarios, was better handled through synthetic sample generation using GANs.
- Generalizability: The model retained strong performance across diverse datasets, highlighting its adaptability to different network environments.

V. CONCLUSION

The rapid evolution of cyber threats necessitates equally adaptive and intelligent defense mechanisms. This study presents a novel framework that integrates Generative Adversarial Networks (GANs) into the lifecycle of Autonomous Intrusion Detection Systems (IDS) to effectively counter both known and adversarial attacks in complex network environments.

The findings clearly demonstrate that GANs significantly enhance the capability of IDS by:

- Augmenting training datasets with realistic synthetic attack traffic to address class imbalance,
- Hardening the model through adversarial training to make it resilient against evasion techniques,
- Improving classification performance, as evidenced by higher accuracy, precision, recall, and F1-scores across benchmark datasets (NSL-KDD, CICIDS2017, and UNSW-NB15),
- Reducing false positives and false negatives, a key challenge in traditional detection systems.

Moreover, the adversarially trained GAN-IDS model achieved over 95% accuracy and maintained low false negative rates, indicating its robustness in detecting both subtle and sophisticated intrusions.

This research highlights the potential of adversarial intelligence in transforming cybersecurity systems from reactive tools to proactive and adaptive defense mechanisms. By leveraging the generative power of GANs,

IDS models can be continually updated with evolving threat patterns, ensuring sustained security in dynamic and high-risk digital environments.

Future Work

Future research can focus on:

- Real-time deployment and optimization of GAN-IDS models in large-scale network infrastructures,
- Integration with Federated Learning to support privacy-preserving distributed training,
- Extending the approach to detect zero-day attacks using unsupervised or semi-supervised GAN variants,
- Exploring explainable AI (XAI) techniques to make adversarial decisions interpretable to cybersecurity analysts.

REFERENCES

[1] M. Roesch, "Snort - Lightweight Intrusion Detection for Networks," in Proceedings of the 13th USENIX Conference on System Administration, 1999.
[2] K. Kendall, "A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems," MIT Technical Report, 1999.
[3] S. Revathi and A. Malathi, "A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection," International Journal of Engineering Research and Technology, 2013.
[4] T. B. M. Jansen and J. D. M. Rennie, "Machine Learning Approaches to Cyber Intrusion Detection," IEEE Security and Privacy, 2006.
[5] S. Wang et al., "A Survey of Imbalanced Learning Methods for Network Intrusion Detection," IEEE Access, 2020.
[6] I. Goodfellow et al., "Generative Adversarial Nets," in Advances in Neural Information Processing Systems (NeurIPS), 2014.
[7] H. Lin, W. Wang, and H. Lu, "IDSGAN: Generative Adversarial Networks for Attack Generation against Intrusion Detection," IEEE Access, 2020.
[8] N. Shafiq et al., "A Survey of Generative Adversarial Networks for Intrusion Detection," Computer Networks, 2021.
[9] J. Li et al., "Enhancing Intrusion Detection Systems with GAN-based Data Augmentation," Journal of Network and Computer Applications, 2021.
[10] T. Lin et al., "Adversarial Training for Robust Intrusion Detection Models," in Proceedings of the 2020 IEEE Symposium on Security and Privacy, 2020.
[11] H. Debar, M. Dacier, and A. Wespi, "Towards a taxonomy of intrusion-detection systems," Computer Networks, 2000.
[12] T. H. Kim et al., "A Survey of Machine Learning Techniques for Intrusion Detection Systems," Applied Sciences, 2020.
[13] H. Hindy et al., "A Taxonomy and Survey of Intrusion Detection System Design Techniques, Network Threats and Datasets," IEEE Communications Surveys & Tutorials, 2020.
[14] S. Revathi and A. Malathi, "Network Intrusion Detection System Using Support Vector Machine," Journal of Computer Applications, 2013.
[15] H. Lin, W. Wang, and H. Lu, "IDSGAN: Generative Adversarial Networks for Attack Generation against Intrusion Detection," IEEE Access, 2020.
[16] I. Goodfellow et al., "Generative Adversarial Nets," NeurIPS, 2014.
[17] H. Lin et al., "Adversarial Attack and Defense in Deep Learning-based Intrusion Detection Systems," IEEE Internet of Things Journal, 2021.
[18] J. Li et al., "Enhancing Intrusion Detection Systems with GAN-based Data Augmentation," Journal of Network and Computer Applications, 2021.
[19] Y. Wang, X. He, and X. Sun, "Conditional Generative Adversarial Network for Network Intrusion Detection," IEEE Access, 2021.

[20] H. Shapoorifard et al., “Adversarial Training for Improving the Robustness of IDS Models,” Computers & Security, 2022.
[21] M. Rahman et al., “Semi-Supervised GANs for Intrusion Detection in Cyber-Physical Systems,” Future Generation Computer Systems, 2021.
[22] M. Arjovsky and L. Bottou, “Towards Principled Methods for Training GANs,” ICLR, 2017.
[23] Y. Wang et al., "Intrusion detection in IoT using deep learning with GAN-based data augmentation," Wireless Networks, vol. 27, pp. 2125–2136, 2021.
[24] L. Huang et al., "Generating synthetic network traffic using GANs for intrusion detection training," in Proc. IEEE ICNC, 2020, pp. 167–171.
[25] Feng et al., "Enhancing intrusion detection systems with GAN-generated synthetic samples," Future Generation Computer Systems, vol. 127, pp. 112–121, 2022.
[26] J. Su et al., "Defending against GAN-based adversarial attacks in IDS," Computers & Security, vol. 87, p. 101568, 2019.
[27] N. Nguyen and L. Nguyen, "Smart grid intrusion detection using deep learning and GANs," in Proc. IEEE ISI, 2020, pp. 1–6.
[28] B. Zhang et al., "cGAN-based intrusion detection system for IoT networks," IEEE IoT Journal, vol. 8, no. 18, pp. 14234–14245, 2021.
[29] H. Yao, Y. Wang, and L. Zhang, "Feature augmentation for network intrusion detection using GANs," Journal of Network and Computer Applications, vol. 180, p. 103004, 2021.
[30] T. Reddy et al., "Comparative study of GAN variants for intrusion detection dataset generation," in Proc. ACM SAC, 2020, pp. 905–910.
[31] A. Singh et al., "Hybrid GAN-based intrusion detection system for cloud computing environments," IEEE Trans. Cloud Comput., Early Access, 2022.
[32] D. Goyal and M. Arora, "Using GANs for smart city cybersecurity intrusion detection," in Proc. IEEE SMARTCOMP, 2020, pp. 177–182.
[33] Q. Chen, W. Wang, and J. Xu, "TrafficGAN: Generative adversarial network based traffic simulation for IDS," IEEE Access, vol. 7, pp. 46098–46107, 2019.
[34] X. Wei et al., "Improving IDS performance using synthetic samples generated by GANs," Information Sciences, vol. 540, pp. 101–116, 2020.
[35] M. Islam et al., "Explainable AI for GAN-based IDS," Expert Systems with Applications, vol. 215, p. 119269, 2023.