DOI: 10.19884/j.1672-5220.202601019

Minibatch-Enhanced Generative Adversarial Networks for IoT Network Intrusion Detection

YU Shoujian[1*], LIU Jiakang[1], ZHANG Hong[1], SHEN Shigen[2]
1. School of Information and Intelligent Science, Donghua University, Shanghai 201620, China
2. School of Information Engineering, Huzhou Normal University, Huzhou 313000, China

Abstract: With the rapid expansion of the Internet of Things (IoTs), network connectivity has exposed systems to increasingly complex and previously unseen security threats. Conventional intrusion detection systems (IDSs) often struggle to identify subtle anomalies in high-dimensional and dynamically evolving IoT traffic. To address this problem, we propose a generative adversarial network (GAN)-based IDS tailored for IoT networks. The proposed method integrates minibatch discrimination to enhance sensitivity to inter-sample anomalies, replaces the conventional L1 distance with cosine similarity to better characterize relationships among high-dimensional feature vectors, and incorporates L1/L2 regularization to improve training stability and generalization. These enhancements enable more accurate discrimination between benign and malicious traffic while reducing overfitting. Experiments on the TON_IoT dataset demonstrate that the proposed method achieves strong overall detection performance in terms of accuracy and F1 score while maintaining a low false positive rate.

Keywords: generative adversarial network; intrusion detection system; minibatch discrimination; cosine similarity; L1/L2 regularization
CLC number: TP183 **Document code**: A
Article ID: 1672-5220(2026)05-0090-09

0 Introduction

The rapid expansion of the Internet of Things (IoTs) has enabled large-scale data exchange across domains such as healthcare, smart cities, and industrial automation. However, the increasing connectivity of heterogeneous IoT devices also exposes networks to severe security threats, particularly malware propagation and previously unseen attacks[1]. Conventional intrusion detection systems (IDSs) struggle to detect such threats when faced with high-dimensional, dynamic, and noisy IoT traffic, and their effectiveness is further constrained by limited device resources and diverse communication protocols. These challenges make accurate and robust intrusion detection a critical requirement for practical IoT security[2].

To overcome the limitations of conventional IDS methods, learning-based methods have been increasingly explored in recent years. Recent advances in deep learning have provided promising alternatives for IoT network intrusion detection. Among them, generative adversarial networks (GANs) offer strong capabilities in modeling complex data distributions and identifying anomalous patterns through adversarial learning. Nevertheless, conventional GAN-based IDS models often suffer from unstable training, mode collapse, and reduced discrimination in high-dimensional feature spaces, which limits their effectiveness on real IoT traffic[3]. In particular, sample-wise discrimination, inappropriate distance metrics, and overfitting remain key obstacles to reliable anomaly detection.

To address these issues, this paper proposes an enhanced GAN-based IDS for IoT networks. The proposed method incorporates minibatch discrimination to capture inter-sample anomalies, replaces the conventional L1 distance with cosine similarity to better characterize relationships among high-dimensional feature vectors[4], and introduces L1/L2 regularization to stabilize training and improve generalization. Extensive experiments on the TON_IoT dataset[5] demonstrate that the proposed method consistently outperforms representative deep learning baselines in the detection accuracy and robustness while maintaining a low false positive rate (FPR). The method would provide an effective and practical solution for detecting subtle and emerging threats in IoT network.

1 Related Work

GANs have been extensively studied due to their strong ability to learn and replicate complex data distributions. Since the seminal work by Goodfellow et al.[4], GANs have been successfully applied to image generation, anomaly detection, and data augmentation[6-7]. Subsequent variants, such as Wasserstein GAN (WGAN) and least squares GAN (LSGAN), introduced alternative loss functions to

Received date: 2026-01-24
* Correspondence should be addressed to YU Shoujian, email: jackyysj@dhu.edu.cn
Citation: Yu S J, Liu J K, Zhang H, et al. Minibatch-enhanced generative adversarial networks for IoT network intrusion detection[J]. *Journal of Donghua University (English Edition)*, 2026, 43(5): 90-98.

address training instability and improve convergence. These developments significantly enhanced the robustness and effectiveness of GANs, particularly in tasks requiring high-quality data modeling and precise anomaly detection.

Various GAN architectures have been proposed to address challenges across different application domains. For instance, deep convolutional GAN (DCGAN) employs convolutional networks to improve representation learning[8], while StyleGAN and StyleGAN2 introduce style-based architectures to enable high-resolution and controllable data generation. Progressive growing GAN further improves training stability by incrementally increasing model complexity[9]. These architectural advancements demonstrate the adaptability of GANs and their potential for modeling complex data distributions in diverse domains.

In anomaly detection, GANs have been shown to effectively capture subtle variations in data and identify abnormal patterns. Salimans et al.[7] introduced minibatch discrimination to enhance sensitivity to subtle differences among samples, which was particularly important for high-dimensional data. Large-scale GAN models such as BigGAN further improve sample fidelity, indirectly benefiting anomaly detection by enabling more accurate modeling of data distributions. These methods[10-11], especially minibatch discrimination, are closely aligned with the objectives of this study.

GAN-based architectures have also been explored in cybersecurity and IDSs, demonstrating their effectiveness in modeling complex network traffic patterns and identifying previously unseen attacks. Prior studies indicate that GANs possess strong potential for detecting novel and sophisticated intrusions[12], especially in high-dimensional network environments. However, most existing work either focuses on general anomaly detection or applies GANs to security tasks without fully addressing the specific challenges posed by IoT traffic[13].

The primary contribution of this study lies in its targeted enhancement of GANs for IoT network intrusion detection. Unlike prior work that mainly emphasizes image synthesis or general anomaly detection[14-15], this research integrates minibatch discrimination into the discriminator to improve sensitivity to subtle anomalies in network traffic. Although minibatch discrimination has been applied in other domains, its systematic use for IoT network intrusion detection remains largely unexplored.

Furthermore, this work replaces conventional L1 distance with cosine similarity in the loss formulation. While L1 distance is widely adopted in machine learning, it becomes less effective in high-dimensional feature spaces such as network traffic data. Cosine similarity measures angular relationships between feature vectors and is therefore more suitable for capturing fine-grained distinctions between benign and malicious traffic, enabling more accurate detection of IoT-specific anomalies[16].

To mitigate overfitting, which is a critical issue in IoT security, L1/L2 regularization terms are incorporated into the GAN loss function. These regularization strategies control the model complexity, enhance generalization to unseen data, and improve robustness in the real-world IoT network. The integration of regularization contributes to improved detection accuracy and reduced FPRs, further strengthening the practical applicability of the proposed framework[17-19].

In summary, by combining minibatch discrimination, cosine similarity, and L1/L2 regularization, this study advances existing GAN-based intrusion detection methods. The proposed enhancements enable more effective modeling of complex IoT network traffic and more precise identification of subtle malicious behaviors that conventional IDS methods[20] often fail to detect, thereby pushing forward the state of the art in IoT network security.

2 Methods

This section presents a GAN-based IDS architecture for IoT networks. To address the challenges of high-dimensional features[21], heterogeneous traffic, and evolving attack patterns, three key enhancements are introduced into the conventional GAN architecture: minibatch discrimination, cosine-similarity-based feature interaction, and L1/L2 regularization. These components could jointly improve anomaly sensitivity, training stability, and generalization in the complex IoT network[22-24].

2.1 System overview

As illustrated in Fig. 1, the proposed architecture consists of a generator, an encoder, and a discriminator. The generator synthesizes benign-like traffic, the encoder maps real samples to a compact latent representation, and the discriminator distinguishes benign from malicious behaviors. Compared with standard GANs, three mechanisms are integrated: minibatch discrimination to model inter-sample dependencies, cosine similarity to measure relationships among high-dimensional feature vectors, and L1/L2 regularization to stabilize training and reduce overfitting. These enhancements enable more reliable detection of subtle and previously unseen attacks in IoT traffic.

In this GAN-based architecture, a is the real input sample; E is the encoder yielding latent feature $E(a)$ for real data; b is the input of a generator G, which generates synthetic sample $G(b)$. The discriminator D receives both branches of features and makes binary decisions: "True" indicates samples originating from real data, while "Fake" denotes samples synthesized by the generator G. The encoder, generator and discriminator are jointly optimized under the adversarial loss to narrow the gap between real and synthetic feature distributions.

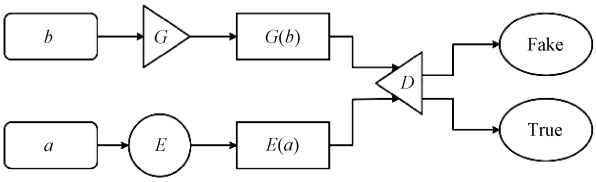


Fig. 1 Schematic of GAN-based IDS architecture

2.2 Model architecture

The generator maps noise vectors to synthetic traffic samples, approximating the distribution of benign data. The discriminator evaluates both real and generated samples and determines whether the traffic is malicious. Unlike conventional GANs, minibatch discrimination is embedded in the discriminator to analyze multiple samples jointly, enhancing sensitivity to subtle anomalies. The encoder compresses input traffic into a latent space, preserving essential features while reducing redundancy and facilitating deviation detection[25].

The adversarial objective $L(D, E, G)$ is defined as

$$L(D, E, G) = \underset{a\sim q_a}{E}[\log D(a, b_a)] + \underset{b\sim q_b}{E}[\log(1 - D(a_b, b))], \quad (1)$$

where b_a is sampled from the encoder; a_b is generated by the generator; q_a denotes the empirical distribution of real traffic data; q_b denotes the prior distribution over the latent space. The first term evaluates the discriminator's ability to identify real samples, and the second term measures the generator's effectiveness in producing realistic traffic.

2.3 Minibatch discrimination with cosine similarity

To improve anomaly separability in high-dimensional feature vectors, minibatch discrimination is combined with cosine similarity, as shown in Fig. 2. For a minibatch $B = \{y_1, y_2, \cdots, y_i, \cdots, y_n\}$, each sample feature y_i is projected via a learnable tensor $\boldsymbol{C}$ to construct a feature vector $\boldsymbol{c}(y_i)$ for capturing inter-sample relationships. Traditional minibatch discrimination D_{L1} computes similarity using the L1 distance:

$$D_{L1} = \sum_{i,j=1}^{n} |\boldsymbol{c}(y_i) - \boldsymbol{c}(y_j)|. \quad (2)$$

However, D_{L1} becomes less effective in high-dimensional vectors. We replace it with the cosine similarity $S_{\cos}(\boldsymbol{c}_i, \boldsymbol{c}_j)$

$$S_{\cos}(\boldsymbol{c}_i, \boldsymbol{c}_j) = \frac{<\boldsymbol{c}_i, \boldsymbol{c}_j>}{\|\boldsymbol{c}_i\| \cdot \|\boldsymbol{c}_j\|}, \quad (3)$$

and define the batch-level dissimilarity D_{batch} as

$$D_{\text{batch}} = \sum_{i,j=1}^{\beta} (1 - S_{\cos}(\boldsymbol{c}_i, \boldsymbol{c}_j)), \quad (4)$$

where $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ denote feature representations of two samples within the same minibatch; $<\boldsymbol{c}_i, \boldsymbol{c}_j>$ represents the inner product; $\|\cdot\|$ denotes the Euclidean norm; β is the batch size. This formulation emphasizes angular relationships between feature vectors, enabling the discriminator to capture fine-grained dependencies and improve detection of subtle anomalies in IoT traffic. d_1, d_2, $\cdots$, and d_n are the output features.

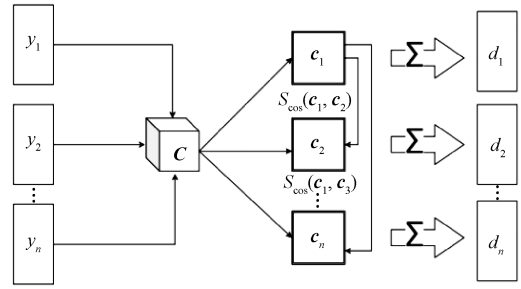


Fig. 2 Minibatch architecture

2.4 L1/L2 regularization

To mitigate overfitting and enhance generalization, L1/L2 regularization is incorporated into the GAN loss L:

$$L = L_{\text{GAN}} + \lambda_1 \|\boldsymbol{W}_1\|_1 + \lambda_2 \|\boldsymbol{W}_2\|_2^2, \quad (5)$$

where L_{GAN} denotes the original GAN loss; λ_1 and λ_2 are regularization coefficients; $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are the weight matrices of the generator and discriminator, respectively. These constraints reduce the model complexity, stabilize the adversarial training, and improve the robustness to noisy and high-dimensional IoT data.

2.5 Training and implementation

The overall system architecture is shown in Fig. 3. The generator, encoder, and discriminator are trained in an adversarial manner using the Adam optimizer. Hyperparameters such as the learning rate, batch size, and regularization coefficients are selected through cross-validation on the TON_IoT dataset to ensure stable convergence and balanced generalization[26].

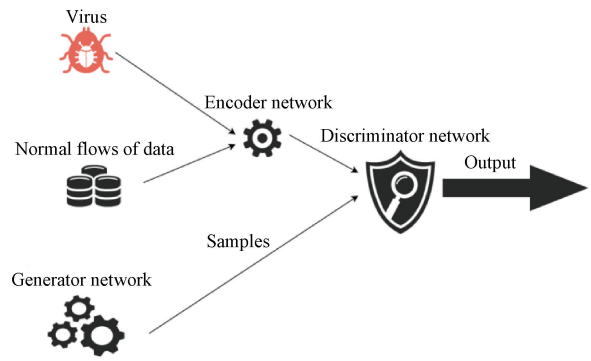


Fig. 3 Architecture of GAN-based IDS

3 Experiments

This section evaluates the minibatch-enhanced GAN using the TON_IoT dataset. We conduct hyperparameter sensitivity studies (learning rates and batch sizes), ablation analysis on model variants, and comparative evaluation against baseline deep learning models. Performance is assessed with the accuracy, precision, recall, F1 score, area under the receiver operating

characteristic curve (ROC-AUC), loss, true positive rate (TPR) and FPR[27-28].

3.1 Dataset and preprocessing (TON_IoT)

TON_IoT is an IoT security benchmark dataset containing heterogeneous network features and attack categories (e.g., distributed denial of service (DDoS), scanning, and malware). Its realistic traffic patterns and class imbalance make it suitable for evaluating IDSs. Preprocessing steps include normalization, missing-value imputation, and consistent data splitting for training, validation, and test.

To reflect realistic IoT traffic distributions, we retain the original class proportions on the TON_IoT dataset without applying over-sampling or under-sampling techniques. No explicit class rebalancing strategy is introduced during training. Instead, imbalance-insensitive evaluation metrics such as the F1 score, ROC-AUC, and FPR are reported to provide a more comprehensive performance assessment under class-imbalanced conditions.

3.2 Implementation and training configuration

The dataset is randomly divided into 70% training dataset, 15% validation dataset, and 15% test dataset using stratified sampling to preserve class distribution and ensure unbiased evaluation. Only benign samples from the training dataset were used to train the GAN-based model, while the validation and test datasets contained both benign and malicious samples. The model is implemented in Python using TensorFlow. The generator synthesizes benign-like traffic, and the discriminator distinguishes benign and malicious traffic samples. Minibatch discrimination and L1/L2 regularization enhance anomaly sensitivity and mitigate overfitting. Models are trained for 1 000 epochs, and the training configuration is summarized in Table 1. All experiments are conducted with a fixed random seed of 29 for data partitioning and model initialization, and final results are reported on the held-out test dataset to ensure reproducibility.

Table 1 Summary of training configuration on TON_IoT

Indicator	Detail
Category	Adam (β_1 = 0.900, β_2 = 0.999)
Learning rate	0.000 01
Batch size/sample	64
Epoch	1 000
Regularization coefficient	$\lambda_1 = 1 \times 10^{-5}$, $\lambda_2 = 1 \times 10^{-4}$
Hardware	NVIDIA RTX 4060
Software	Python 3.11.5, TensorFlow 2.15
Dataset split	70% training dataset, 15% validation dataset, 15% test dataset

3.3 Hyperparameter sensitivity study on TON_IoT

We examine the impact of the learning rate α on model performance while keeping the batch size β fixed at 64. Learning rates of 0.001, 0.000 1, and 0.000 01 were evaluated, as shown in Fig. 4.

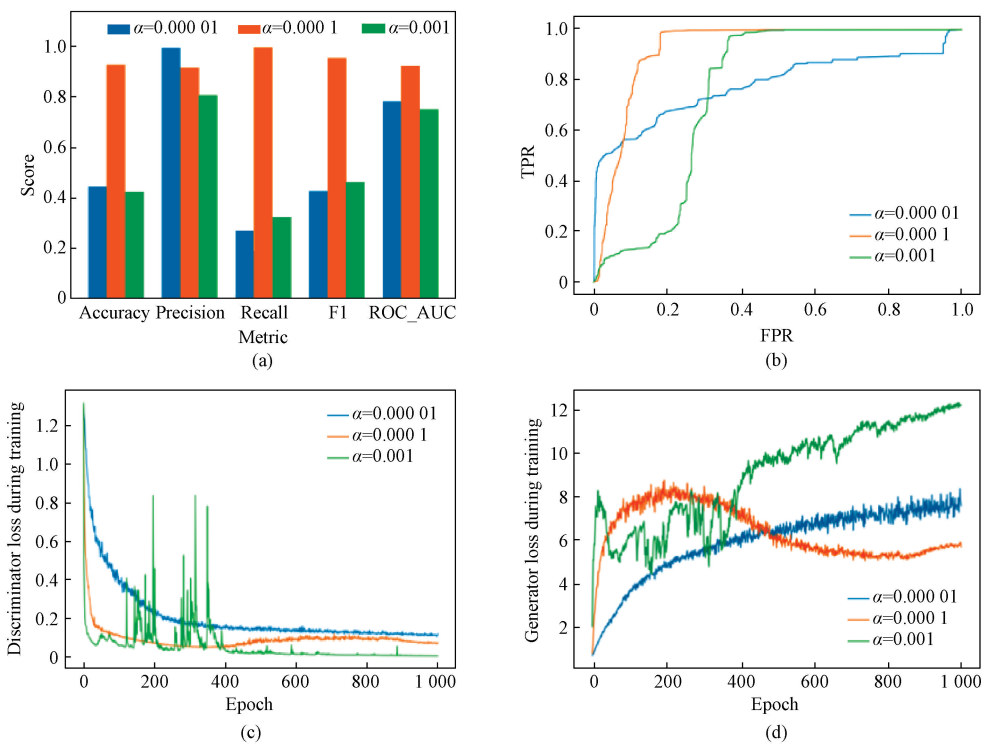


Fig. 4 Impact of learning rate on model performance; (a) performance metric comparison; (b) TPR-FPR analysis; (c) discriminator loss; (d) generator loss

At a high learning rate of 0.001, the discriminator exhibits pronounced oscillations during training, leading to unstable adversarial updates and degraded detection precision. In contrast, a small learning rate of 0.000 01 produces smoother loss curves but results in slower convergence and weaker anomaly discrimination capability due to insufficient exploration of the loss surface. A learning rate of 0.000 1 provides the best balance between training stability and convergence quality, yielding smooth generator-discriminator loss trajectories, stable TPR-FPR behavior, and the strongest overall performance, with an accuracy of 0.927, a precision of 0.915, a recall of 0.997, and a low FPR of 0.030. Overall, a moderate learning rate achieves the best trade-off between adversarial training stability and anomaly detection performance.

The impact of the batch size β was further studied while keeping the learning rate α fixed at 0.000 1. Batch sizes of 32, 64, and 128 were investigated, as illustrated in Fig. 5. A small batch size of 32 introduces noisy gradient updates, resulting in more volatile precision-recall behavior and slightly higher false alarms. Conversely, a large batch size of 128 accelerates early convergence but tends to overfit, as reflected by an increased FPR and reduced generalization performance. A batch size of 64 demonstrates the most stable learning dynamics and achieves the best overall detection performance, with a recall of 0.953, an F1 score of 0.945, and a reduced FPR of 0.028. These results confirm that moderate learning rates and batch sizes are preferable for GAN-based intrusion detection, where both adversarial training stability and sensitivity to rare attack patterns are critical.

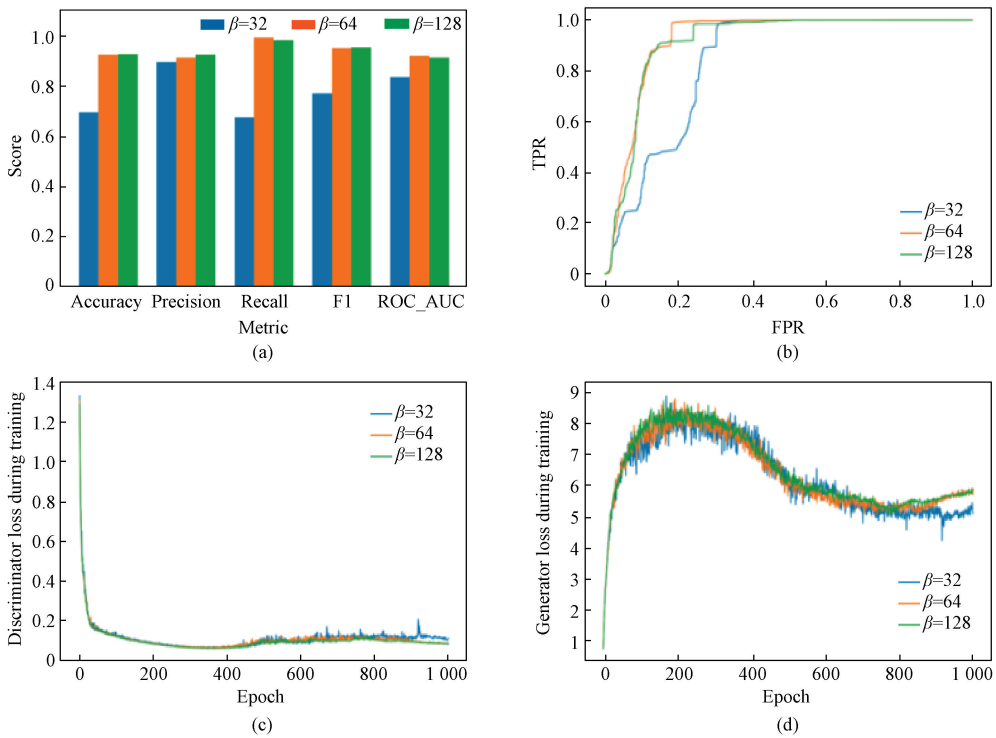


Fig. 5 Impact of batch size on model performance; (a) performance metric comparison; (b) TPR-FPR analysis; (c) discriminator loss; (d) generator loss

3.4 Ablation study on minibatch discrimination and regularization

We further compare three GAN configurations: standard GAN, GAN with minibatch discrimination (minibatch GAN for short), and the proposed regularized minibatch GAN with full enhancements, as illustrated in Fig. 6. The standard GAN exhibits unstable adversarial dynamics and a higher FPR, which is indicative of mode collapse during training. Incorporating minibatch discrimination alleviates this issue by exploiting inter-sample relationships, leading to improved recall and more stable TPR-FPR behavior; however, mild overfitting is still observed near convergence.

By jointly integrating minibatch discrimination with cosine similarity and L1/L2 regularization, the proposed regularized minibatch GAN consistently demonstrates superior stability and detection performance. It achieves a higher F1 score, reduced FPR, and smoother generator-discriminator loss convergence throughout training. These results confirm that the introduced enhancements effectively suppress overfitting, promote diversity in adversarial learning, and significantly improve discrimination capability on complex and high-dimensional IoT attack patterns. Specifically, minibatch discrimination primarily improves recall by modeling inter-sample dependencies, cosine similarity enhances feature separability in high-dimensional spaces, and L1/L2 regularization stabilizes training dynamics and reduces FPRs.

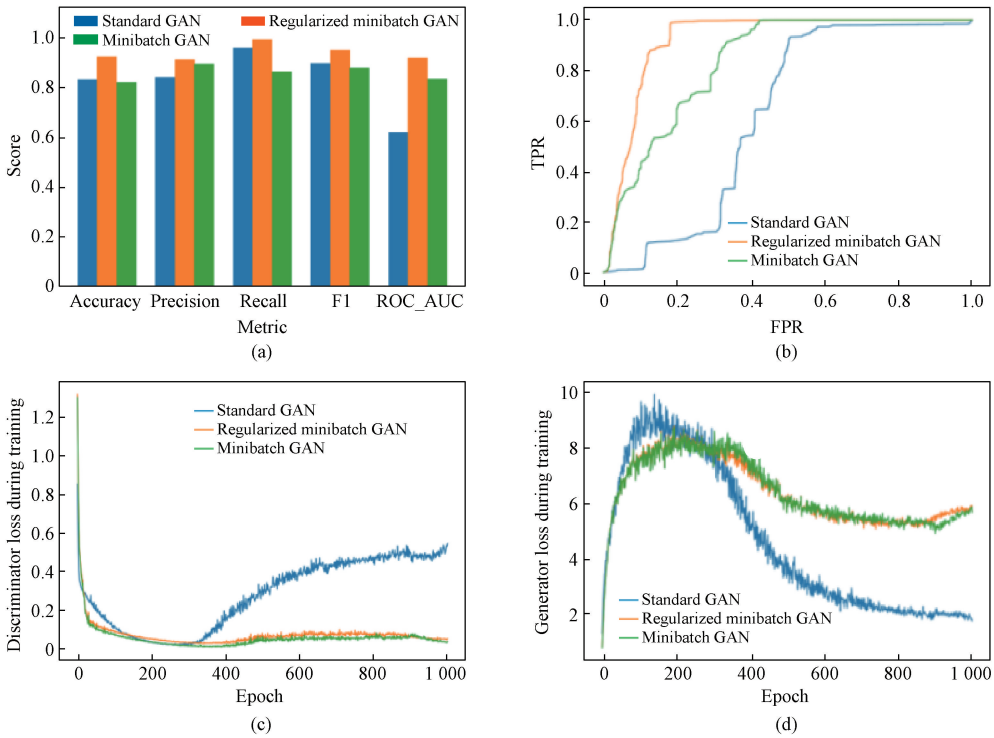


Fig. 6 Comparison before and after model improvement: (a) performance metric comparison; (b) TPR-FPR analysis; (c) discriminator loss; (d) generator loss

3.5 Comparison with other intrusion detection models on TON_IoT dataset

3.5.1 Comparison with common deep learning baselines

The proposed regularized minibatch GAN outperforms the baseline models such as an autoencoder (AE), a deep autoencoder (DAE), and bidirectional GAN (BiGAN) in the overall detection performance. Figure 7 shows a comparison of precision, recall, accuracy, and F1 score on the TON_IoT dataset. It can be seen that the proposed regularized minibatch GAN outperforms other models across all metrics.

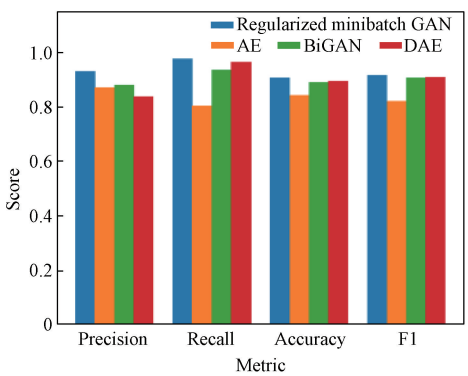


Fig. 7 Overall detection performance comparison among different models

3.5.2 Performance against state-of-the-art IDS models

Table 2 summarizes the performance comparison between the proposed regularized minibatch GAN and representative CNN-LSTM, GRU-AE, and Transformer-based IDS models. The proposed regularized minibatch GAN achieves the highest overall detection performance, with an accuracy of 0.927, a precision of 0.915, a recall of 0.997, and an F1 score of 0.954. Compared with all baseline models, it achieves superior and more balanced performance across the accuracy, precision, recall, and F1 score, demonstrating its strong capability in distinguishing benign and malicious traffic in complex IoT environments.

Table 2 Comparison results of different models

Model	Accuracy	Precision	Recall	F1 score
CNN-LSTM	0.921	0.887	0.984	0.933
GRU-AE	0.912	0.904	0.991	0.946
Transformer-based IDS	0.918	0.910	0.993	0.950
Regularized minibatch GAN	0.927	0.915	0.997	0.954

3.6 Computational efficiency

To further evaluate the feasibility of real-world deployment in resource-constrained IoT network, we analyze the computational complexity and inference

efficiency of the proposed model. All measurements are conducted on the same hardware platform (NVIDIA RTX 4060 GPU) to ensure the consistency.

Table 3 summarizes the number of trainable parameters, average inference latency, throughput, and so on. The results indicate that the proposed regularized minibatch GAN maintains a lightweight architecture and supports real-time inference, demonstrating its practicality for IoT deployment scenarios.

Table 3 Computational efficiency of regularized minibatch GAN

Metric	Value
Number of trainable parameters	12 254
Model size/MB	0.05
Average inference latency/(ms/sample)	0.85
Throughput/(sample/s)	1 050
Training time (1000 epochs)/min	2.3

4 Conclusions

This paper proposes an enhanced GAN-based IDS for IoT networks, integrating minibatch discrimination, cosine similarity, and L1/L2 regularization.

These innovations address the challenge of detecting subtle and evolving malicious behaviors in high-dimensional and heterogeneous IoT traffic. Minibatch discrimination improves sensitivity to subtle anomalies by leveraging inter-sample dependencies, while cosine similarity strengthens feature separation in high-dimensional spaces by focusing on directional relationships among feature vectors. L1/L2 regularization enhances training stability and robustness, particularly in the presence of class imbalance and noisy data. Experiments on the TON_IoT dataset show that the proposed model achieves strong performance in the accuracy, precision, recall, and F1 score, while maintaining a low FPR. These results highlight the effectiveness of combining batch-level discrimination, cosine similarity, and regularized adversarial learning for IoT network intrusion detection.

However, several challenges remain for real-world deployment. IoT environments often operate under resource constraints, motivating lightweight optimization techniques such as pruning and quantization-aware training to reduce inference costs. In addition, concept drift (time-varying traffic distributions) may degrade detection performance over time, necessitating adaptive and continual learning mechanisms. Enhancing adversarial robustness and integrating privacy-preserving strategies, such as federated learning, could further strengthen system security.

Future work will explore practical deployment in resource-constrained IoT scenarios, such as smart manufacturing systems and edge-based monitoring platforms. Additionally, extending the framework to industrial IoT and smart city infrastructures will be investigated to evaluate its robustness under large-scale heterogeneous environments. Overall, the proposed framework provides a solid foundation for adaptive and reliable IoT network intrusion detection.

References

[1] Bhale P, Chowdhury D R, Biswas S, et al. OPTIMIST: lightweight and transparent IDS with optimum placement strategy to mitigate mixed-rate DDoS attacks in IoT networks [J]. *IEEE Internet of Things Journal*, 2023, 10(10): 8357-8370.

[2] Kang X, Liu L, Ma H D. ESR-GAN: environmental signal reconstruction learning with generative adversarial network [J]. *IEEE Internet of Things Journal*, 2021, 8(1): 636-646.

[3] Jamshidi S, Amirnia A, Nikanjam A, et al. Self-adaptive cyber defense for sustainable IoT: a DRL-based IDS optimizing security and energy efficiency [J]. *Journal of Network and Computer Applications*, 2025, 239: 104176.

[4] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C]// Proceedings of the 28th International Conference on Neural Information Processing Systems. New York: ACM, 2014: 2672-2680.

[5] Li J, Jia J J, Xu D L. Unsupervised representation learning of image-based plant disease with deep convolutional generative adversarial networks [C]//2018 37th Chinese Control Conference (CCC). Piscataway: IEEE, 2018: 9159-9163.

[6] Lin R Q, Qiu H D, Wang J, et al. Physical-layer security enhancement in energy-harvesting-based cognitive Internet of Things: a GAN-powered deep reinforcement learning approach [J]. *IEEE Internet of Things Journal*, 2024, 11(3): 4899-4913.

[7] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs [C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: ACM, 2016: 2234-2242.

[8] Li Y T, Ouyang C K, Huang H Y. AEGANAuth: autoencoder GAN-based continuous authentication with conditional variational autoencoder generative adversarial network [J]. *IEEE Internet of Things Journal*, 2024, 11(16): 27635-27650.

[9] Mao X D, Li Q, Xie H R, et al. Least squares generative adversarial networks [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2813-2821.

[10] Arjovsky M, Chintala S, Bottou L. Wasserstein

generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ACM, 2017: 214-223.
[11] Song S H, Mukerji T, Hou J G. Geological Facies modeling based on progressive growing of generative adversarial networks (GANs) [J]. *Computational Geosciences*, 2021, 25(3): 1251-1273.
[12] Wu L, Xie Y L, Zhao S X, et al. Efficient intrusion detection via heterogeneous graph attention networks and parallel provenance analysis [J]. *Computer Networks*, 2025, 270: 111552.
[13] Li Z F, Huang C H, Qiu W Y. An intrusion detection method combining variational auto-encoder and generative adversarial networks [J]. *Computer Networks*, 2024, 253: 110724.
[14] Zhang H, Xu T, Li H S, et al. StackGAN++: realistic image synthesis with stacked generative adversarial networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(8): 1947-1962.
[15] Razavi A, Van Den Oord A, Vinyals O. Generating diverse high-fidelity images with VQ-VAE-2 [PP/OL]. arXiv (2019-06-02) [2026-01-02]. https://arxiv.org/abs/1906.00446.
[16] Chen Z Z, Fu A M, Zhang Y H, et al. Secure collaborative deep learning against GAN attacks in the Internet of Things [J]. *IEEE Internet of Things Journal*, 2021, 8(7): 5839-5849.
[17] Mao H Y, Shu J W, Song M C, et al. LrGAN: a compact and energy efficient PIM-based architecture for GAN training [J]. *IEEE Transactions on Computers*, 2021, 70(9): 1427-1442.
[18] Karras T, Aittala M, Laine S, et al. Alias-free generative adversarial networks [PP/OL]. V4. arXiv (2021-10-18) [2026-01-02]. https://arxiv.org/abs/2106.12423.
[19] Jiang Y F, Chang S Y, Wang Z Y. TransGAN: Two pure transformers can make one strong GAN, and that can scale up[C]//Advances in Neural Information Processing Systems 34 (NeurIPS 2021). San Diego: NeurIPS, 2021: 1-12.
[20] Menon S, Damian A, Hu S J, et al. PULSE: self-supervised photo upsampling via latent space exploration of generative models [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 2434-2442.
[21] Yu S J, Jin O W, Shen Y Z, et al. Availability evaluation of industrial Internet of Things under malware propagation: an extended reliability block diagram approach based on stochastic games [J]. *IEEE Transactions on Reliability*, 2025, 74(3): 4253-4267.
[22] Shen S G, Hao X B, Gao Z J, et al. SAC-PP: jointly optimizing privacy protection and computation offloading for mobile edge computing [J]. *IEEE Transactions on Network and Service Management*, 2024, 21(6): 6190-6203.
[23] Wang C, Jiang C X, Wang J J, et al. Blockchain-aided network resource orchestration in intelligent Internet of Things [J]. *IEEE Internet of Things Journal*, 2023, 10(7): 6151-6163.
[24] Liu H W, Qi L Y, Shen S G, et al. Microservice-driven privacy-aware cross-platform social relationship prediction based on sequential information [J]. *Software: Practice and Experience*, 2024, 54(1): 85-105.
[25] Fang Z X, Li Q, Liu J H, et al. Beamforming design for multi-antenna multi-tag symbiotic radio backscatter systems [J]. *AEU-International Journal of Electronics and Communications*, 2023, 170: 154820.
[26] Yu S J, Zhai R, Shen Y Z, et al. Deep Q-network-based open-set intrusion detection solution for industrial Internet of Things [J]. *IEEE Internet of Things Journal*, 2024, 11(7): 12536-12550.
[27] Shen S G, Cai C P, Li Z W, et al. Deep Q-network-based heuristic intrusion detection against edge-based SIoT zero-day attacks [J]. *Applied Soft Computing*, 2024, 150: 111080.
[28] Shen S G, Cai C P, Shen Y Z, et al. Joint mean-field game and multiagent asynchronous advantage actor-critic for edge intelligence-based IoT malware propagation defense [J]. *IEEE Transactions on Dependable and Secure Computing*, 2025, 22(4): 3824-3838.

基于增强小批量生成对抗网络的物联网入侵检测

于守健[1*]，刘佳康[1]，张　红[1]，沈士根[2]
1. 东华大学 信息与智能科学学院，上海 201620
2. 湖州师范学院 信息工程学院，浙江 湖州 313000

摘　要：随着物联网的快速发展，网络互联使系统面临愈发复杂、前所未有的安全威胁。传统入侵检测系统往往难以识别隐蔽在高维、动态变化的物联网流量中的异常。针对该问题，本文提出一种面向物联网网络的、基于生成对抗网络的入侵检测方法。该方法引入小批量判别机制以提升对样本间异常的敏感度；采用余弦相似度替代传统 L1 距离，以更好地表征高维特征向量之间的关联关系；同时，引入 L1/L2 正则化以提升训练稳定性与泛化能力。上述改进能够更精准地区分正常流量与恶意流量，并降低过拟合问题。基于 TON_IoT数据集开展的实验表明，所提方法在准确率、F1 分数方面具备优异的整体检测性能，同时可保持较低的误报率。

关键词：生成对抗网络；入侵检测系统；小批量判别；余弦相似度；L1/L2 正则化