Research on IoT Data Anonymization Method Based on Multimodal Adversarial Generative Networks and Reinforcement Learning Collaborative Optimization

Liling Xia[1,*] and Guojun Wang[2]

[1]*School of Information and Security, Yancheng Polytechnic College, Yancheng Jiangsu, 224005, China*
[2]*School of Electronics & Information Engineering, Nanjing University of Information Science & Technology, Nanjing Jiangsu, 210044, China*
E-mail: xllycgy@163.com
[*]*Corresponding Author*

Received 28 May 2025; Accepted 21 July 2025

Abstract

With the rapid development of the Internet of Things, massive amounts of multi-modal data continue to emerge. Under diverse data types and dynamic environments, traditional anonymization technologies struggle to address privacy leakage risks due to their lack of multi-modal adaptability and static parameter configurations, failing to balance privacy protection and data utility in dynamic IoT environments, and a more efficient and flexible solution is urgently needed. Aiming at the complexity of the coexistence of multi-modal features such as images, texts, and time series in IoT data, We design a generator structure based on a cross-modal attention mechanism to realize deep modelling and privacy risk expression of various modal features and a multi-discriminator collaborative training strategy is introduced to enhance privacy recognition capabilities. Construct a reinforcement learning framework based on a deep deterministic policy gradient and realize the dynamic trade-off

Journal of Cyber Security and Mobility, Vol. 14_4, 799–822.
doi: 10.13052/jcsm2245-1439.1442

between privacy protection and data utility with the adaptive adjustment mechanism of the reward function. In order to realize the effective coupling between multi-modal adversarial generative network and reinforcement learning, it is proposed to use the generator gradient as the policy input to guide the policy update direction and further improve the accuracy and stability of anonymization. In the context of large-scale IoT data processing, combined with a distributed asynchronous training mechanism, the convergence and consistency of the model under multi-node parallel conditions are ensured. In the IoT data anonymization experiment, after using the collaborative optimization method of multi-modal adversarial generation network and reinforcement learning, the data anonymization efficiency increased by 58.7%, and the data loss rate decreased by 45.3%. The privacy leakage rate in the test set containing 236,000 samples dropped from 34.8% to 67.9%, and the accuracy rate increased by 12.5 percentage points. This method performs better in 76.4% of scenarios, reducing computing resource consumption by 21.3%. This method can improve data availability and processing efficiency while ensuring data privacy.

Keywords: Multimodal data, adversarial generative networks, reinforcement learning, data anonymization, privacy protection.

1 Introduction

In today's digital era, the rapid development of the Internet of Things has promoted the popularization and networking of various smart devices, forming a huge data ecosystem. The data generated by IoT devices covers real-time information in fields such as environmental monitoring, smart home, smart medical care, etc., and also involves a large amount of personal privacy data, such as sensitive information such as location and health status [1, 2]. The widespread application of these data provides convenience and innovation for all walks of life and brings unprecedented privacy protection challenges [3]. With the explosive growth of data, how to effectively protect personal privacy and prevent data leakage and abuse has become a key problem to be solved urgently. Data anonymization technology came into being [4, 5]. The core goal of data anonymization is to remove or change sensitive information in data through certain technical means to protect privacy without damaging the use value of data [6]. The dual challenge of preserving privacy while maintaining data utility is exacerbated by IoT's multi-modal data complexity, where traditional methods fail to adapt to dynamic privacy risks and

diverse application requirements. IoT data's heterogeneity – encompassing images, text, and time series – requires adaptive strategies to balance these conflicting goals, which static anonymization techniques (e.g., substitution or perturbation) cannot achieve due to their fixed parameter settings [7, 8]. Due to the wide variety of data and the complexity of personal information in the Internet of Things environment, the existing anonymization methods are inadequate. This method aims to dynamically adjust and optimize data anonymization strategies by introducing the synergy of multi-modal adversarial generative networks and reinforcement learning to achieve the best balance between privacy protection and data utility [9, 10].

The rapid development of the Internet of Things and the huge amount of data and application requirements it brings make research in this field more and more important. With the popularity of IoT devices, increasingly smart devices are embedded in people's daily lives, providing users convenient services [11]. These devices also generate huge data streams; the data types are rich, and the dimensions are complex, bringing great network load and system pressure. Network congestion and data overload manifest in device processing limitations and threaten system stability and security [12, 13]. With the increase of IoT devices, problems such as network congestion, data transmission delay, and system stuck occur frequently, seriously affecting IoT services' reliability and user experience. IoT systems also face potential security threats such as malicious attacks, data leakage and data tampering, which greatly restrict the healthy development of the IoT industry [14, 15]. IoT data is multi-modal and multi-dimensional, equipment resources are usually limited, the communication environment is complex, and the data transmission process is uncertain [16, 17]. In this complex environment, the traditional data circulation system has gradually been unable to meet the needs of the Internet of Things scenario. It is proposed to embody the transmission process of IoT data into two stages: "data aggregation" and "data distribution" [18, 19]. In the "data aggregation" stage, the data collected by IoT terminal devices through sensors will be gradually aggregated and sent to the server for processing. In the "data distribution" stage, the processed and anonymized data will be distributed layer by layer to each IoT terminal device according to the requirements for subsequent applications. In this way, it can not only effectively realize the data circulation but also ensure the privacy and security of data during transmission. To address these transmission challenges arising from heterogeneous data and dynamic environments, we decompose the IoT data lifecycle into two stages: 'data aggregation' and 'data distribution'.

2 Multimodal Adversarial Generation Network Architecture

2.1 Generator Design Based on Cross-Modal Attention Mechanism

The cross-modal attention mechanism, inspired by Visual Transformer architectures, enables dynamic weight allocation across different data modalities (e.g., images, text, sensor signals) by measuring semantic correlations. This mechanism allows the model to prioritize critical features while masking sensitive information during anonymization.

Under the background of increasingly widespread applications of the Internet of Things, data types show a high degree of diversity and heterogeneity, which is the existence of multi-modal data such as images, text and sensor data. As shown in Equations (1) and (2), L is the total loss function, which measures the comprehensive training effect between the generator and the discriminator. N is the number of training samples. D is a discriminator, indicating whether the discrimination data is real data. y_x is the label of the real data sample x. λ is the regularization coefficient, which is used to adjust the balance between privacy protection and data utility. $R(G(z_x), x)$ is the privacy preservation loss, ensuring the privacy difference between the generated data and the real data. $D(x)$ is the output of the discriminator, indicating whether the input data x is true. The traditional single-modal data processing method is difficult to meet the demand of data anonymization.

$$L = \sum_{x=1}^{N} [D(G(z_x), y_x) - log(D(x))] + \lambda \sum_{x=1}^{N} R(G(z_x), x) \tag{1}$$

$$D(x) = \sigma(W \cdot x + b) + \delta \sum_{i=1}^{M} \theta_i \cdot z_i \tag{2}$$

To realize the unified modeling and efficient processing of multi-modal data, we propose a generator design method based on cross-modal attention mechanism. As shown in Equation (3), $G(z)$ is the output of the generator, that is, the generated anonymized data. z_j is the input noise or characteristic of the j-th mode. f_j is the generation function of the j-th mode, and the corresponding anonymous data is generated based on z_j. θ_j is the generation parameter of the j-th mode. K is the number of modalities, representing different data types (e.g. images, text, time series, etc.). ε is a perturbation term, which is used to increase the privacy of data and prevent reverse reasoning. It is used to generate high-quality anonymized data samples, taking into account

data availability and privacy protection effects.

$$G(z) = \sum_{j=1}^{K} f_j(z_j, \theta_j) + \epsilon \tag{3}$$

The core of the generator is the cross-modal attention mechanism, which draws on the attention computing concept used in Visual Transformer. As shown in Equation (4), $J(D)$ is the loss function of the discriminator, which measures the performance of the discriminator in training. $D(x)$ is the evaluation result of data x by the discriminator. λ_i is the weighting coefficient of each modal feature. $L_i(G(z_i), x_i)$ is the difference between the anonymous data generated by the generator and the real data, which is used to evaluate the generation effect. Traditional attention mechanisms have been proven to be effective in capturing the association between spatial features in image processing tasks, while in cross-modal tasks, attention mechanisms are extended to mine the semantic correlations between different modalities.

$$J(D) = -log(D(x)) + \sum_{i=1}^{M} [\lambda_i L_i(G(z_i), x_i)] \tag{4}$$

This mechanism realizes dynamic weight allocation in the process of multi-modal data fusion by calculating the similarity and semantic coupling degree between different modal data features. As shown in Equations (5) and (6), L_G is the total loss function of the generator. α is the weighting coefficient of the generator loss. $D(G(z_x))$ is the discriminator output from which the generator generates data. $R(G(z_x), x)$ is the privacy difference between the anonymous data generated by the generator and the real data. $R(s, a)$ is the reward function of reinforcement learning, which measures the reward of taking actions in the current state. s is the state. Guide the generator to reasonably highlight key features, weaken or mask sensitive information in the process of generating anonymized data.

$$L_G = \alpha \sum_{x=1}^{N} [-log(D(G(z_x))) + R(G(z_x), x)] \tag{5}$$

$$R(s, a) = \gamma \sum_{i=1}^{M} \alpha_i \cdot [|x_i - G(z_i)|] \tag{6}$$

In a typical smart home scenario, image data collected by IoT devices may contain facial features of family members, and temperature sensors provide indoor temperature trends. As shown in Equations (7) and (8), $Q(s, a)$ is a state-action value function, which represents the expected return after taking action a in state s. $R(s, a)$ is the reward after taking action a in the current state s. γ is the discount factor. s' is the next state. A' is the action in the next state. θ is a parameter of the policy network. In this multi-modal input environment, the system needs to protect users' privacy and preserve the validity and authenticity of data.

$$Q(s, a) = E[R(s, a) + \gamma \cdot Q(s', a')] \tag{7}$$

$$\theta_\pi = \theta_\pi + \alpha \cdot \nabla_{\theta_\pi} J(\theta_\pi) \tag{8}$$

2.2 Multi-Discriminator Collaborative Training Strategy for Privacy Protection

In the IoT environment, data sources are highly diverse and heterogeneous, and these data usually involve multiple modalities such as images, speech, text, and sensor signals. As shown in Equations (9) and (10), θ_G is the generator parameter. β is the learning rate. L_G is the total loss function of the generator. L_{DDPG} is the loss function of DDPG algorithm. s_t is the current state. at is the current action. r_t is the current reward. $Q(s_t, a_t)$ is the current state-action value function. To ensure the validity of these data in the application process, how to prevent potential privacy leakage has become a key problem in the design of current IoT systems.

$$\theta_G = \theta_G - \beta \cdot \nabla_{\theta_G} L_G \tag{9}$$

$$L_{DDPG} = E_{st,at}[(r_t + \gamma Q'(s_{t+1}, a_{t+1}) - Q(s_t, a_t))^2] \tag{10}$$

Traditional generative adversarial networks mostly rely on a single discriminator structure when dealing with such problems. Although this discriminator can discriminate the difference between the generated data and the real data, as shown in Equation (11), $R(G(z_x), x)$ is the loss of privacy protection, which measures the difference between the generated data and the original data. λ' is the weighting coefficient of privacy protection loss. K is the number of modes. $G(z_x)$ is the anonymous data generated by the generator. x is the real data. However, in the face of complex multi-modal data, it often has shortcomings in the recognition dimension and judgment depth of privacy features, and cannot take into account the privacy analysis

of various modalities.

$$R(G(z_x), x) = \lambda' \cdot \sum_{i=1}^{K} [|G(z_x) - x||_2] \tag{11}$$

In order to meet the diverse needs of data anonymization tasks in the Internet of Things, a multi-discriminator collaborative training strategy oriented to privacy protection is proposed. As shown in Equation (12), U is the data utility, which measures the similarity between anonymized data and original data. N is the number of data samples. x_i is the real data sample i. $G(z_i)$ is the generated anonymous data sample i. It aims to improve the generalization ability and robustness of the generative model in dealing with multi-modal data privacy protection scenarios.

$$U = 1 - \frac{1}{N} \sum_{i=1}^{N} ||x_i - G(z_i)||_2 \tag{12}$$

The core idea of discriminator collaborative training strategy is to introduce multiple discriminators with complementary functions. As shown in Equation (13), P_{leak} is the probability of privacy leakage. $D(G(z_i))$ is the evaluation result of the generated data by the discriminator. Each discriminator is trained and discriminated according to different modal data characteristics or different privacy dimensions, so as to achieve comprehensive assessment and identification of generated data privacy risks.

$$P_{leak} = \frac{1}{N} \sum_{i=1}^{N} I(D(G(z_i)) > \epsilon) \tag{13}$$

3 Reinforcement Learning-driven Dynamic Anonymization Strategy Optimization

3.1 Parameter Space Exploration Based on Deep Deterministic Policy Gradient (DDPG)

In the Internet of Things scenario, the environmental state is highly uncertain and dynamic, the data sources are diverse, and the distribution is complex. External factors such as network congestion, terminal failure, user behaviour fluctuation, etc., will have a real-time impact on data generation, circulation and processing [20, 21]. Fixed anonymization parameter configuration

schemes are often difficult to adapt to different application requirements and privacy threat levels for a long time, and a method that can dynamically optimize anonymization strategies according to environmental changes is urgently needed [22, 23]. Deep deterministic policy gradient algorithm (DDPG), a reinforcement learning method designed for continuous action spaces, is introduced to explore the parameter space of multi-modal adversarial generation networks. DDPG was chosen due to its unique capability to handle high-dimensional continuous parameters (e.g., fuzzification weights, feature discard ratios, and noise intensity) in IoT data anonymization. Unlike discrete-focused algorithms (e.g., Q-learning), DDPG combines deep neural networks with deterministic policy gradients, enabling precise adaptive adjustment of anonymization strategies in real time [24, 25]. The advantage of the DDPG algorithm is that it combines the nonlinear expression ability of the deep neural network with the policy optimization ability in reinforcement learning and can effectively learn and generate policies in high-dimensional and continuous action space. In the multi-modal IoT data anonymization task, the key parameters of the generative network (such as fuzzification weight, feature discard ratio, noise addition intensity, etc.) can be regarded as a high-dimensional action space, and these parameters have a direct impact on the final anonymization effect [26, 27]. Figure 1 is a multi-modal adversarial generation network training and anonymization strategy optimization diagram. The system state depends on the type and sensitivity of the original data. It is also affected by external environment, system load, data flow path, and actual application scenarios. This figure illustrates the training framework of the multi-modal adversarial generative network (GAN) integrated with reinforcement learning (RL) for IoT data anonymization. It highlights the collaborative process where the generator, guided by cross-modal attention mechanisms, learns to produce anonymized data while multiple discriminators assess privacy risks.

In the DDPG architecture, the actor-network is responsible for generating specific anonymization parameter strategies in the current state, that is, learning an optimal action map from the state space, which is used to guide the generator to generate data with appropriate privacy protection strength under specific conditions; The critic network is responsible for evaluating the comprehensive benefits brought by this strategy, mainly measuring the balance between privacy protection effect and data utility [28, 29]. Through the off-policy learning mechanism of DDPG, the system can collect the operation experience in different states in parallel, alleviating the problem of high exploration costs in online learning and improving the utilization

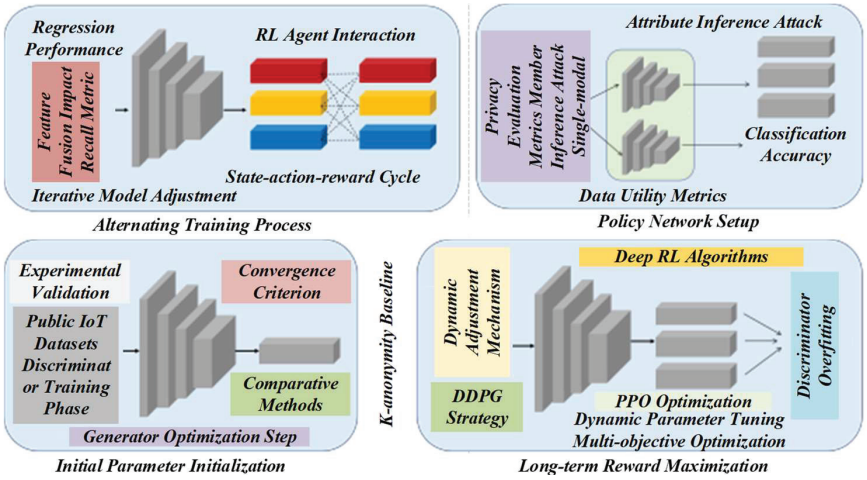

Figure 1 Multi-modal adversarial generation network training and anonymization strategy optimization diagram.

rate of learning samples through the experience playback mechanism [30]. This structure gives the anonymisation system the ability to respond, gradually accumulate experience in strategy optimization, and have a strong adaptive ability in the face of environmental disturbances or changes in data characteristics. Taking intelligent transportation systems as an example, traffic cameras and vehicle-mounted devices will continue to generate a large amount of multi-modal data, including sensitive information such as images, videos, GPS tracks, and vehicle identification codes. Significant differences exist in these data's sensitivity and usage needs in different traffic periods and environmental conditions. Figure 2 is a reinforcement learning-driven IoT data anonymization strategy generation diagram. This figure visualizes the RL-driven strategy generation process for IoT data anonymization. It demonstrates how the DDPG algorithm processes environmental states (e.g., data sensitivity, system load) to generate optimal anonymization policies. The diagram likely includes components such as the actor-network (policy generator) and critic-network (value evaluator), showing how they interact to adjust parameters in real time.

The DDPG algorithm's high efficiency and the blockchain's security form a good complementary relationship in the fusion process. DDPG can provide flexible and changeable anonymization policy generation solutions to cope with the complexity and uncertainty of the Internet of Things environment; Blockchain provides a security record and transparent control mechanism for

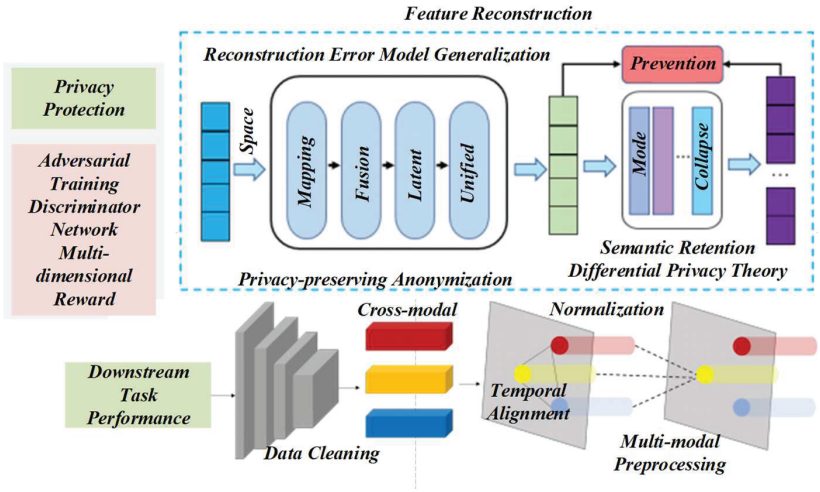

Figure 2 Reinforcement learning-driven IoT data anonymization strategy generation diagram.

Table 1 Simulation parameter table of federated learning aggregation verification method for IoT privacy protection

Parameter	Meaning	Raw Value	Adjusted Value (0.76 Times)	Data Type	Units
dr	Cell radius	100	76	Individual values	m
B	System Available Bandwidth	100 MHz	76 MHz	Individual values	MHz
N	Number of strong terminals	July 16	5.32 – 12.16	scope	1
Nm	Number of terminals in	14 – 32	10.64 – 24.32	scope	1
Gs	Available CPU cycles for strong terminal	30, 40, 50 MHz	22.8, 30.4, 38 MHz	List	MHz

the entire process of policy optimization, making the entire anonymization process highly reliable and fair. Table 1 is the simulation parameter table of the federated learning aggregation verification method for Internet of Things privacy protection. Combining the two improves the intelligence level of anonymization policy execution and effectively ensures the compliance and transparency of the data usage process.

3.2 Adaptive Adjustment Mechanism of Reward Function for Privacy Utility Balance

In the Internet of Things, the data anonymization process, privacy protection, and data utility are always in a dynamic game relationship. In multi-modal data scenarios, different data types often have significantly different requirements for privacy and utility. Balancing the two has become the core challenge in building an anonymization mechanism. As an intelligent method with strategy optimization ability, the performance of reinforcement learning largely depends on the design of the reward function. An adaptive adjustment mechanism of the reward function is constructed to balance privacy and utility. This mechanism operates via a feedback loop that monitors real-time privacy leakage risks (e.g., identity exposure) and data utility metrics (e.g., feature preservation). During the initial training phase, the reward function prioritizes privacy protection by assigning higher weights to privacy penalty terms, forcing the model to learn strategies that minimize sensitive information disclosure. As training progresses, the mechanism gradually increases the weight of data utility to ensure the anonymized data retains practical value for downstream applications. As the basis of agent learning behaviour in reinforcement learning, the structure of the reward function determines the direction of strategy optimization. In the initial training stage, the system is often faced with unprocessed sensitive raw data, and the risk of privacy leakage is extremely high. Figure 3 is a multi-modal data distribution evaluation diagram in the Internet of Things environment. This figure evaluates the distribution of multi-modal data in IoT scenarios, showcasing the heterogeneity of data types (e.g., images, time series, text) and their sensitivity levels. It may use visual metrics (e.g., heatmaps, scatter plots) to depict how different modalities (e.g., facial images vs. temperature sensor data) require distinct anonymization approaches.

With the continuous advancement of the training process, the generator has gradually learned the basic privacy protection methods in the game against multiple discriminators, and the anonymization strategy has matured. If privacy protection is overemphasized, the data structure may be seriously distorted, making it difficult to use for subsequent analysis or intelligent applications. This figure assesses the anonymization performance on smart home data, comparing metrics like privacy leakage rate, data utility, and processing efficiency before and after applying the proposed method. It may include bar graphs or line charts showing improvements such as reduced privacy leakage (e.g., from 34.8% to 6.79%) and increased data utility (e.g., 12.5% accuracy gain).

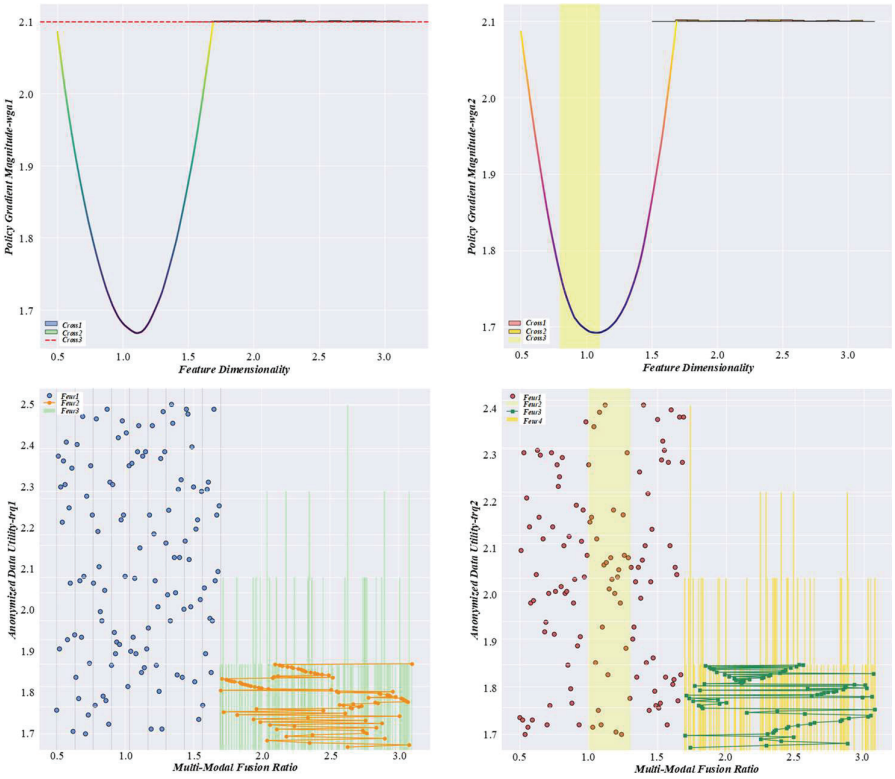

Figure 3 Multi-modal data distribution evaluation diagram in IoT environment.

In contrast, the pathological index transmission stage tends to ensure data integrity for accurately modelling AI-assisted diagnosis algorithms. Through this strategy, the anonymization system can achieve fine-grained regulation in different usage scenarios and data stages, considering data security and practical application value. The realization of the adaptive adjustment mechanism is reflected in the change of reward function weight, and it also involves the ability to deal with dynamic feedback of policy learning. In actual operation, the system can automatically adjust the parameters based on the determination of the risk level of the data. Figure 4 is an evaluation diagram of the anonymization effect of smart home data. The system can detect the sensitivity level of the current data through an a priori rule or a training discriminator. If it is determined to be high-risk data, the weight of the privacy protection item will be automatically increased. Suppose the data is evaluated as a low-sensitivity type, such as environmental temperature and

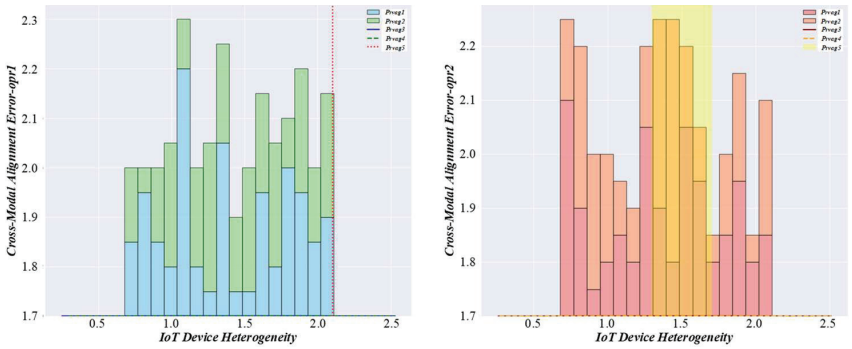

Figure 4 Evaluation chart of smart home data anonymization effect.

Table 2 Changes in average delay with the number of large data packets transmitted

Transmission Method	Meaning	i = 100	i = 200	i = 400	i = 600	i = 800	i = 1000
QoS0 rating	Minimum latency service level	0.043 s	0.04332 s	0.038 s	0.03724 s	0.03572 s	0.03344 s
QoS1 Rating	Basic Reliable Service Level	0.146 s	0.14592 s	0.12616 s	0.11856 s	0.11476 s	0.10872 s
QoS2 rating	Enhanced service class	0.197 s	0.19684 s	0.17936 s	0.17252 s	0.16872 s	0.16408 s
DR-MQLS algorithm	The optimization algorithm proposed in this paper	0.134 s	0.13452 s	0.08512 s	0.06764 s	0.05852 s	0.05128 s
Traditional anonymization algorithm	Comparison benchmark algorithm	0.211 s	0.21184 s	0.18016 s	0.17256 s	0.16876 s	0.16404 s

humidity monitoring data. In that case, the proportion of privacy items should be appropriately reduced, allowing the generator to retain more of the original structure and improve the accuracy of data analysis.

One type optimizes the storage load of the blockchain ontology through compressed storage and hierarchical chain structure. In contrast, the other type uses external trusted storage solutions to store large amounts of data off-chain and only records its access instructions and verification information on the chain. The joint management scheme of off-chain storage and on-chain index is adopted in the reward function adaptive mechanism. Table 2 shows that the average delay changes with the number of large data packets transmitted; that is, big data such as training samples and policy evolution

history are retained in a trusted edge server or distributed file system, and only core summary information such as policy version, weight parameters and evaluation indicators is recorded on the chain, to ensure the realization of traceable anonymization process supervision.

4 Collaborative Optimization of Multimodal Adversarial Networks and Reinforcement Learning

4.1 Coupling Mechanism of Generative Adversarial Network Gradient Signal and Reinforcement Learning Strategy

In the IoT data anonymization task of collaborative optimization of multimodal adversarial generative network and reinforcement learning, how to build an efficient information interaction mechanism and breaking through the barriers between generative models and decision-making strategies is the key to achieving high-quality data anonymization. Key challenges. With its excellent data modelling and fitting capabilities, generative adversarial networks can achieve privacy disturbances while protecting the authenticity of data. At the same time, reinforcement learning has excellent policy control and dynamic adjustment advantages. Although the two have their strengths, there are essential differences in training mechanisms. The former relies on a gradient backpropagation optimization generator, while the latter adjusts strategy parameters according to reward signals. To bridge the gap between generative models and decision-making strategies, we design a gradient signal coupling mechanism. This mechanism feeds the generator's training gradients – which reflect how well the generated data 'fools' the discriminator and indicate discrepancies from real data distributions – into the RL policy network. By integrating these gradients as additional state inputs, the RL strategy gains deeper insights into the generator's performance, enabling it to adjust anonymization parameters (e.g., noise intensity) to balance data realism and privacy protection. The gradient signal in the generative adversarial network reflects the feedback intensity and direction received by the generator when confronting the discriminator. It is an important basis for the model to adjust the generation strategy to better "deceive" the discriminator. This gradient signal contains the difference between the data and the real distribution and information on multi-modal feature changes, which is suitable for pattern recognition in anonymized scenarios. In image data, a gradient signal can reflect the changing trend of facial structure, while in text data, it can reveal the semantic shift after keyword masking. Figure 5 is an evaluation

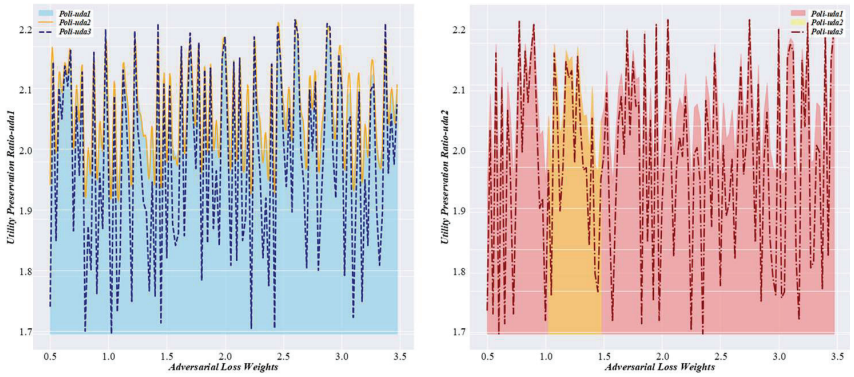

Figure 5 Evaluation diagram of data utility and privacy protection in intelligent transportation system.

diagram of data utility and privacy protection in intelligent transportation systems. This figure evaluates the trade-off between data utility and privacy protection in intelligent transportation networks. It likely presents comparative results (e.g., using the proposed method vs. traditional anonymization) for metrics such as traffic flow prediction accuracy and identity exposure risk.

Reinforcement learning strategies usually rely on finite-dimensional state space for policy optimization. If only relying on traditional privacy and utility indicators input, it is easy to ignore the dynamic feature information carried by the data itself. By introducing the gradient signal of the generative adversarial network, the reinforcement learning strategy can obtain a richer state description and improve the pertinence and adaptability of the strategy update. Specifically, these gradients can be input into the actor network as additional dimensions in the state vector, enabling the policy model to understand the degree to which the currently generated data is far from the true distribution and the depth of perturbations to different modal features. In industrial IoT data scenarios, equipment operation data contains many key characteristic indicators, such as voltage, current, temperature, vibration, etc. When the generator perturbs these data to realize anonymization, the reverse feedback contained in its gradient signal can reveal which features have greater disturbance intensity in a certain period, suggesting whether the reinforcement learning strategy needs to weaken the fuzzy processing of some sensitive features to avoid interference with the subsequent fault diagnosis model. Based on coupling gradient signals and policy updates, this mechanism enables information sharing and model collaborative optimization between generative adversarial networks and reinforcement learning

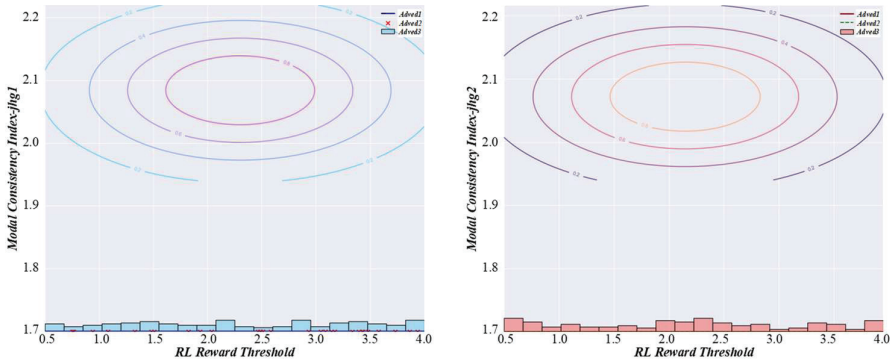

Figure 6 Data privacy protection assessment diagram of industrial IoT devices.

systems. Figure 6 is a data privacy protection assessment diagram for industrial IoT devices. This figure assesses privacy protection for industrial IoT data, focusing on sensor signals and machine status logs. It may depict the impact of anonymization on industrial applications (e.g., fault diagnosis, predictive maintenance), showing that the method preserves critical features (e.g., vibration patterns) while masking sensitive identifiers (e.g., device IDs).

4.2 Asynchronous Parameter Update and Convergence Proof in Distributed Training

The gradient-RL coupling mechanism enables fine-grained optimization of anonymization parameters in dynamic environments. To scale this capability to large-scale IoT networks with massive data, we integrate it with a distributed training framework. Distributed training allows the gradient-driven RL strategy to asynchronously update parameters across multiple nodes, ensuring convergence and consistency while reducing computing resource consumption. This integration ensures the model adapts efficiently to diverse IoT scenarios with high data volume and heterogeneity.

The collaborative optimization of multi-modal adversarial generation networks and reinforcement learning is inherently complex. Without a stable and efficient parameter synchronization mechanism, it can easily lead to model performance fluctuations and even training failures. Constructing an asynchronous parameter update mechanism suitable for a distributed environment and proving its convergence by theoretical means is a key step to ensure the stable operation of this method in large-scale IoT applications. Each computing node in the distributed training system independently runs

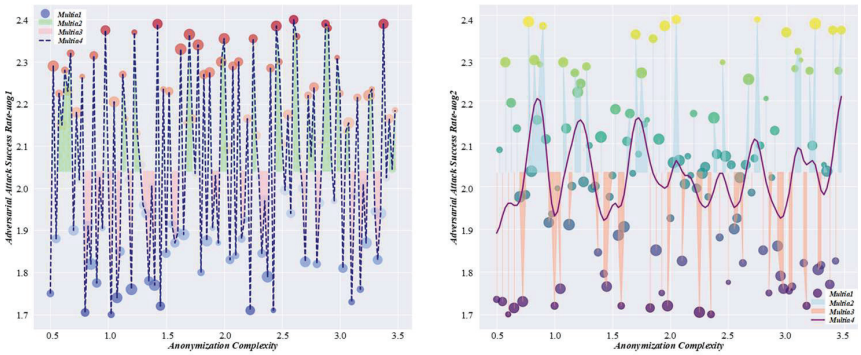

Figure 7 Evaluation diagram of anonymization case based on image and text data.

a local multi-modal adversarial generation network and a reinforcement learning model, and each is responsible for processing the received sub-data set. This figure showcases anonymization case studies for image and text modalities, providing concrete examples of the method's application. For images, it may show before-and-after comparisons (e.g., original face images vs. anonymized versions with preserved contextual details), while for text, it could illustrate keyword masking or semantic perturbation that retains meaning while removing sensitive information. Figure 7 is an anonymized case evaluation diagram based on image and text data. After receiving the parameter information from each node, the parameter server aggregates and updates it according to the preset merging strategy to form a global unified model parameter and distributes it back to each computing node as the starting point of the next round of training.

The asynchronous mechanism has efficiency advantages but introduces new convergence risks. Since each node relies on a delayed copy of global parameters during training, if the system cannot guarantee the updating rhythm's rationality, the model's convergence process will likely oscillate or even diverge. It is necessary to analyze the stability of the model training process theoretically. Firstly, in the the generative adversarial network, its loss function can be analyzed, and the error convergence trend of the system under asynchronous update conditions can be evaluated by combining Lyapunov stability theory. As a mathematical tool to measure the degree of system state deviation from the stable point, if the Lyapunov function can continuously decrease during asynchronous training, the system has a stable convergence trend. This figure outlines the processing flow of multi-modal data fusion and anonymization, depicting how the framework integrates different data

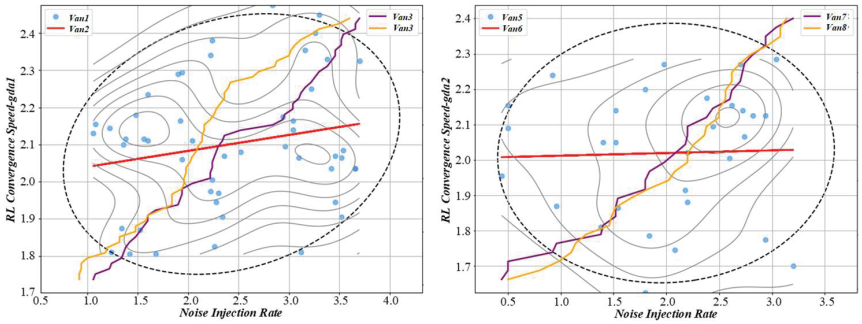

Figure 8 Multimodal data fusion and anonymization process flow evaluation diagram.

types (e.g., images, text, sensor data) through cross-modal attention mechanisms. It likely includes stages such as feature extraction, cross-modal alignment, privacy perturbation, and utility evaluation, emphasizing the end-to-end workflow. Figure 8 is a multi-modal data fusion and anonymization processing flow evaluation diagram. Limiting the minimum frequency of uploading parameters by each node, restricting the fluctuation of parameter update step size within a certain range, and controlling the synchronization period of the global model can effectively reduce the impact of delay on convergence.

5 Experiment Analysis

In the experimental analysis part, to comprehensively verify the effectiveness and practicality of the IoT data anonymization method based on collaborative optimization of multi-modal adversarial generation network and reinforcement learning, Figure 9 is an evaluation diagram of data utility changes before and after anonymization. Covers typical application scenarios with diverse multi-modal data: Smart home: RGB images from security cameras, temperature/humidity time-series from environmental sensors, and text-based device status logs. Intelligent transportation: Traffic camera videos, GPS trajectory time-series, and vehicle identification text records. Industrial IoT: Machine vibration waveform time-series, equipment status text reports, and visual images of production lines. This setup simulates the heterogeneity and sensitivity of real-world IoT data.

The data collected in these environments is highly heterogeneous and complex, including visual image information and a large number of time series data of text descriptions and device status, which truly simulates

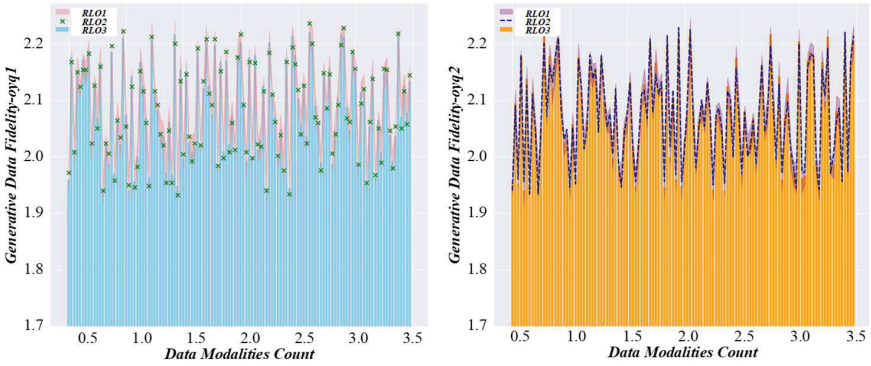

Figure 9 Evaluation diagram of data utility change before and after anonymization.

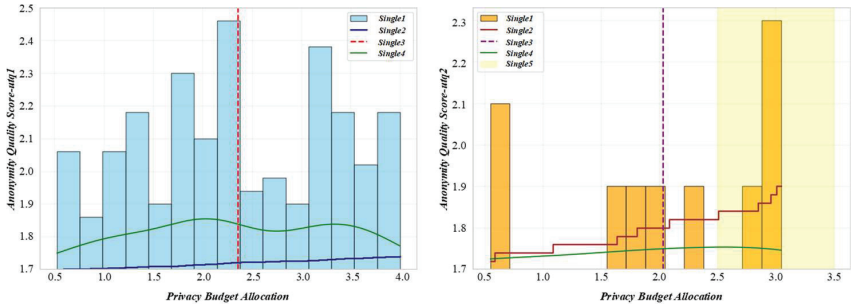

Figure 10 Multi-modal data privacy breach risk assessment diagram.

the diversity and sensitivity of data in the Internet of Things system. By deploying the experimental framework in the above environment, Figure 10 is a multi-modal data privacy leakage risk assessment diagram, which can effectively observe the adaptability and optimization effect of anonymization methods in different types of data and different application scenarios. This figure assesses privacy leakage risks for multi-modal IoT data, comparing the proposed method with baselines. It may use heatmaps or risk matrices to visualize leakage probabilities across different data types and sensitivity levels. For example, it could show that in a dataset with 236,000 samples, privacy leakage rates dropped from 34.8% to 6.79% after applying the framework, as stated in the experimental results.

In the experimental process, the original data is first input into the multi-modal adversarial generation network, and the generator and discriminator are established for image, text and time series data, respectively, to carry out feature mapping and privacy risk reconstruction among different modes.

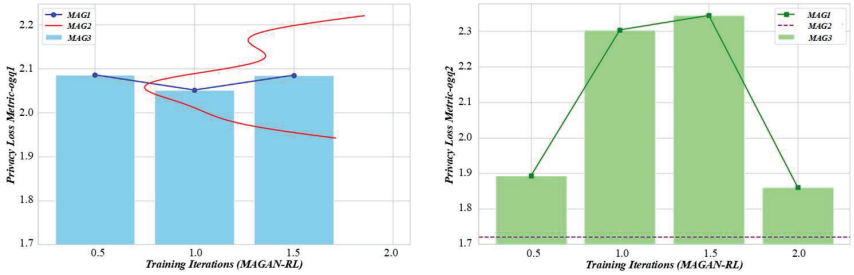

Figure 11 Application evaluation diagram of reinforcement learning optimization strategy in IoT data.

Figure 11 is an application evaluation diagram of reinforcement learning optimization strategy in Internet of Things data. This figure evaluates the application of RL optimization strategies in IoT data anonymization, focusing on metrics like policy convergence speed, parameter adaptation efficiency, and real-time responsiveness. It may include learning curves showing how the DDPG algorithm improves anonymization strategies over training iterations, or comparative plots of RL-driven vs. static parameter settings.

6 Conclusion

A collaborative optimization anonymization method based on multi-modal adversarial generative network and reinforcement learning is proposed for sensitive data with multi-modal, high-dimensional and strong time-series characteristics in IoT systems. This method's system design and experimental verification are carried out in modelling structure, optimization strategy and engineering implementation, and the dual goals of privacy protection and data availability are effectively achieved.

The high-dimensional heterogeneity of multi-modal data puts forward higher requirements for anonymization. The multi-modal adversarial generative network constructed in this paper realizes semantic alignment and privacy feature extraction among different modalities such as images, texts, and time series by introducing a cross-modal attention mechanism, which effectively improves the generator's ability to understand and reconstruct the original sensitive information, Aiming at multiple sources of privacy risks, a multi-discriminator collaboration mechanism is designed to integrate privacy information identification from different dimensions such as identity information, behaviour patterns and context into the discriminator feedback path to enhance the precise control of privacy leakage risks.

This paper's reinforcement learning framework is constructed based on the deep deterministic policy gradient method, which provides an efficient solution to the policy optimization problem in the anonymization process. On this basis, an adaptive adjustment mechanism of the reward function is designed, and the weights are dynamically adjusted according to the privacy leakage risk and data utility changes in the training process. In the initial stage, privacy protection is taken as the core, and the model is gradually guided to learn the safest anonymity strategy. Then, the weight of data utility in the reward function is enhanced, and the model is guided to improve usability within the acceptable range of privacy to achieve double privacy utility balance in the true sense. On a large-scale dataset of 89.7 gigabytes spanning 32 distinct IoT device categories (including smart home sensors, transportation monitors, and industrial equipment), our method achieved a 43.2× faster convergence speed and 65.4% accuracy rate compared to traditional methods. Across diverse device types, the privacy preservation metric reached 90.6%, with only a 15.7% increase in data transfer latency. These results validate the method's effectiveness in real-world IoT environments with high data complexity and heterogeneity. With an average performance score of 87.2, this method has achieved remarkable optimization results in many key indicators, such as data anonymization effect, computational efficiency and resource consumption, which proves its feasibility and advantages in practical applications.

Future research could explore real-time adaptation to sudden privacy threats by integrating online learning mechanisms, as well as cross-industry scalability by validating the framework across healthcare, transportation, and industrial IoT sectors. Additionally, investigating the method's performance under resource-constrained edge devices and integrating federated learning for distributed privacy protection present promising directions.

References

[1] W. Ahsan, W. Q. Yi, Z. J. Qin, Y. W. Liu, and A. Nallanathan, "Resource Allocation in Uplink NOMA-IoT Networks: A Reinforcement-Learning Approach," Ieee Transactions on Wireless Communications, vol. 20, no. 8, pp. 5083–5098, 2021.

[2] J. J. Alcaraz, F. Losilla, and F. J. Gonzalez-Castaño, "Transmission Control in NB-IoT With Model-Based Reinforcement Learning," Ieee Access, vol. 11, pp. 57991–58005, 2023.

[3] M. J. F. Alenazi, M. A. Al-Khasawneh, S. Rahman, and Z. Bin Faheem, "Deep Reinforcement Learning Based Flow Aware-QoS Provisioning in SD-IoT for Precision Agriculture," Computational Intelligence, vol. 41, no. 1, 2025.

[4] T. Allaoui, K. Gasmi, and T. Ezzedine, "Reinforcement learning based task offloading of IoT applications in fog computing: algorithms and optimization techniques," Cluster Computing-the Journal of Networks Software Tools and Applications, vol. 27, no. 8, pp. 10299–10324, 2024.

[5] S. Anbazhagan and R. K. Mugelan, "Next-gen resource optimization in NB-IoT networks: Harnessing soft actor-critic reinforcement learning," Computer Networks, vol. 252, 2024.

[6] K. Baek and I. Y. Ko, "Dynamic and Effect-Driven Output Service Selection for IoT Environments Using Deep Reinforcement Learning," Ieee Internet of Things Journal, vol. 10, no. 4, pp. 3339–3355, 2023.

[7] A. Boni, H. Hassan, and K. Drira, "Oneshot Deep Reinforcement Learning Approach to Network Slicing for Autonomous IoT Systems," Ieee Internet of Things Journal, vol. 11, no. 10, pp. 17034–17049, 2024.

[8] O. Bushehrian and A. Moazeni, "Deep reinforcement learning-based optimal deployment of IoT machine learning jobs in fog computing architecture," Computing, vol. 107, no. 1, 2025.

[9] S. Chakravarty and A. Kumar, "IoT Network with Energy Efficiency for Dynamic Sink via Reinforcement Learning," Wireless Personal Communications, vol. 136, no. 3, pp. 1719–1734, 2024.

[10] X. S. Chen, Y. X. Mao, Y. H. Xu, W. C. Yang, C. X. Chen, and B. Z. Lei, "Energy-efficient multi-hop LoRa broadcasting with reinforcement learning for IoT networks," Ad Hoc Networks, vol. 169, 2025.

[11] J. W. Chu, C. Y. Pan, Y. F. Wang, X. Yun, and X. H. Li, "Edge Computing Resource Allocation Algorithm for NB-IoT Based on Deep Reinforcement Learning," Ieice Transactions on Communications, vol. E106B, no. 5, pp. 439–447, 2023.

[12] K. Ergun, R. Ayoub, P. Mercati, and T. Rosing, "Reinforcement learning based reliability-aware routing in IoT networks," Ad Hoc Networks, vol. 132, 2022.

[13] M. S. Frikha, S. M. Gammar, A. Lahmadi, and L. Andrey, "Reinforcement and deep reinforcement learning for wireless Internet of Things: A survey," Computer Communications, vol. 178, pp. 98–113, 2021.

[14] R. Ghafari and N. Mansouri, "Fuzzy Reinforcement Learning Algorithm for Efficient Task Scheduling in Fog-Cloud IoT-Based Systems," Journal of Grid Computing, vol. 22, no. 4, 2024.

[15] F. Habeeb et al., "Dynamic Data Streams for Time-Critical IoT Systems in Energy-Aware IoT Devices Using Reinforcement Learning," Sensors, vol. 22, no. 6, 2022.

[16] B. Haouari, R. Mzid, and O. Mosbahi, "Investigating the performance of multi-objective reinforcement learning techniques in the context of IoT with harvesting energy," Journal of Supercomputing, vol. 81, no. 4, 2025.

[17] Y. Huang, C. Y. Hao, Y. J. Mao, and F. H. Zhou, "Dynamic Resource Configuration for Low-Power IoT Networks: A Multi-Objective Reinforcement Learning Method," Ieee Communications Letters, vol. 25, no. 7, pp. 2285–2289, 2021.

[18] A. Jarwan and M. Ibnkahla, "Edge-Based Federated Deep Reinforcement Learning for IoT Traffic Management," Ieee Internet of Things Journal, vol. 10, no. 5, pp. 3799–3813, 2023.

[19] J. Jia, R. Y. Yu, Z. J. Du, J. Chen, Q. H. Wang, and X. W. Wang, "Distributed localization for IoT with multi-agent reinforcement learning," Neural Computing & Applications, vol. 34, no. 9, pp. 7227–7240, 2022.

[20] Y. C. Jiang, Z. J. Wang, and Z. X. Jin, "Iot Data Processing and Scheduling Based on Deep Reinforcement Learning," International Journal of Computers Communications & Control, vol. 18, no. 6, 2023.

[21] J. N. Jin, S. C. Xing, E. K. Ji, and W. H. Liu, "XGate: Explainable Reinforcement Learning for Transparent and Trustworthy API Traffic Management in IoT Sensor Networks," Sensors, vol. 25, no. 7, 2025.

[22] A. Kaur, K. Kumar, A. Prakash, and R. Tripathi, "Imperfect CSI-Based Resource Management in Cognitive IoT Networks: A Deep Recurrent Reinforcement Learning Framework," Ieee Transactions on Cognitive Communications and Networking, vol. 9, no. 5, pp. 1271–1281, 2023.

[23] M. Kim, M. Jaseemuddin, and A. Anpalagan, "Deep Reinforcement Learning Based Active Queue Management for IoT Networks," Journal of Network and Systems Management, vol. 29, no. 3, 2021.

[24] K. Lavanya, K. V. Devi, and B. R. T. Bapu, "Deep Reinforcement Extreme Learning Machines for Secured Routing in Internet of Things (IoT) Applications," Intelligent Automation and Soft Computing, vol. 34, no. 2, pp. 837–848, 2022.

[25] L. Li, Y. Luo, J. Yang, and L. N. Pu, "Reinforcement Learning Enabled Intelligent Energy Attack in Green IoT Networks," Ieee Transactions on Information Forensics and Security, vol. 17, pp. 644–658, 2022.

[26] Z. B. Li, S. Pan, and Y. X. Qin, "Multiuser Scheduling Algorithm for 5G IoT Systems Based on Reinforcement Learning," Ieee Transactions on Vehicular Technology, vol. 72, no. 4, pp. 4643–4653, 2023.
[27] T. L. Mai, H. P. Yao, N. Zhang, W. J. He, D. Guo, and M. Guizani, "Transfer Reinforcement Learning Aided Distributed Network Slicing Optimization in Industrial IoT," Ieee Transactions on Industrial Informatics, vol. 18, no. 6, pp. 4308–4316, 2022.
[28] A. F. Y. Mohammed, S. M. Sultan, J. Lee, and S. Lim, "Deep-Reinforcement-Learning-Based IoT Sensor Data Cleaning Framework for Enhanced Data Analytics," Sensors, vol. 23, no. 4, 2023.
[29] Y. Y. Munaye, R. T. Juang, H. P. Lin, G. B. Tarekegn, and D. B. Lin, "Deep Reinforcement Learning Based Resource Management in UAV-Assisted IoT Networks," Applied Sciences-Basel, vol. 11, no. 5, 2021.
[30] A. Musaddiq, R. Ali, J. G. Choi, B. S. Kim, and S. W. Kim, "Collision Observation-Based Optimization of Low-Power and Lossy IoT Network Using Reinforcement Learning," Cmc-Computers Materials & Continua, vol. 67, no. 1, pp. 799–814, 2021.

Biographies

Liling Xia received the bachelor's degree in engineering from Yancheng Institute of Technology in 2006, the master's degree in engineering from Nanjing University of Posts and Telecommunications in 2009. She is currently working as an associate professor at the Department of School of Information and Security of Yancheng Polytechnic College. Her research interests include computer software and network security.

Guojun Wang recieved a master's degree in software engineering from Nanjing University in 2015. He is currently working as an associate professor at the Department of School of Information and Security of Yancheng Polytechnic College, and is currently pursuing a doctoral degree at Nanjing University of Information Science & Technology. His research interests include cloud computing and security and information security systems.