GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Computer, Signal, and System Research

Article

AI Driven Payment System Security Improvement and User Privacy Protection Mechanism

Yue Qi [1,*]

[1] School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213, USA
* Correspondence: Yue Qi, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213, USA

Abstract: With the continuous expansion of electronic payment systems and the rapid evolution of sophisticated cyber-attack methodologies, traditional security measures are increasingly finding it difficult to address the multifaceted risk challenges of the modern era. Leveraging the high-dimensional feature extraction and adaptive learning characteristics of artificial intelligence, this paper establishes a comprehensive AI-driven payment security and privacy protection framework. In terms of system security, the proposed architecture utilizes a residual attention mechanism for precise anomaly detection, while incorporating graph neural networks to analyze and cluster complex account relationship topologies. Furthermore, reinforcement learning is integrated to dynamically adjust risk control strategies in real-time, facilitating the construction of a collaborative defense system through the fusion of multi-source information. Regarding data privacy and integrity, the system adopts homomorphic encryption to enable complex model operations within an encrypted state, which is further combined with blockchain technology to ensure the rigorous traceability and immutability of the entire data flow. The implementation of this integrated technological architecture significantly enhances the intelligent defense capabilities of payment systems, providing a robust and scalable solution for safeguarding digital transactions and sensitive information in high-risk environments. This research not only offers a theoretical advancement in payment security but also provides a practical implementation roadmap for developing next-generation resilient financial information systems.

Keywords: artificial intelligence; payment security; abnormal behavior recognition; privacy protection; Federated Learning

Received: 14 November 2025
Revised: 28 December 2025
Accepted: 10 January 2026
Published: 14 January 2026

1. Introduction

In the contemporary financial landscape, the rapid proliferation of diverse electronic payment systems and the continuous expansion of online transaction environments have led to an increasingly sophisticated array of adversarial attack methodologies. Traditional security frameworks, which primarily rely on static rule-based strategies and single-dimensional authentication methods such as conventional password protection, are becoming increasingly inadequate. These legacy systems struggle to meet the modern imperatives of high recognition accuracy, rapid detection speed, and robust model adaptability required to counteract evolving fraudulent behaviors. Consequently, the integration of advanced artificial intelligence (AI) technologies into security modeling has emerged as a transformative necessity. AI-driven approaches offer significant advantages in real-time threat detection, multi-dimensional risk evaluation, and the implementation of flexible policy-switching mechanisms, effectively serving as a primary driver in the construction of next-generation intelligent payment security systems.

Furthermore, the synergy between artificial intelligence and cutting-edge privacy-preserving technologies enables comprehensive protection across all stages of the data lifecycle, including collection, transmission, storage, and processing. By leveraging these

technologies, financial institutions can better safeguard sensitive information while maintaining operational efficiency. This article provides an in-depth study of the application of AI in enhancing the security of payment systems and fortifying user privacy protection. It elaborates on relevant architectural models and algorithmic strategies in detail, exploring the mechanisms behind intelligent anomaly detection and data encryption. Finally, the paper offers practical suggestions and a technical roadmap for the deployment of these systems, providing a solid theoretical and technical foundation for the sustainable and secure development of the digital payment ecosystem.

2. Advantages of AI Technology

Artificial intelligence technology has high adaptability and practicality in payment systems, and its core advantages are reflected in its ability to process complex data structures and construct dynamic models. The advantages of AI technology are shown in Figure 1.

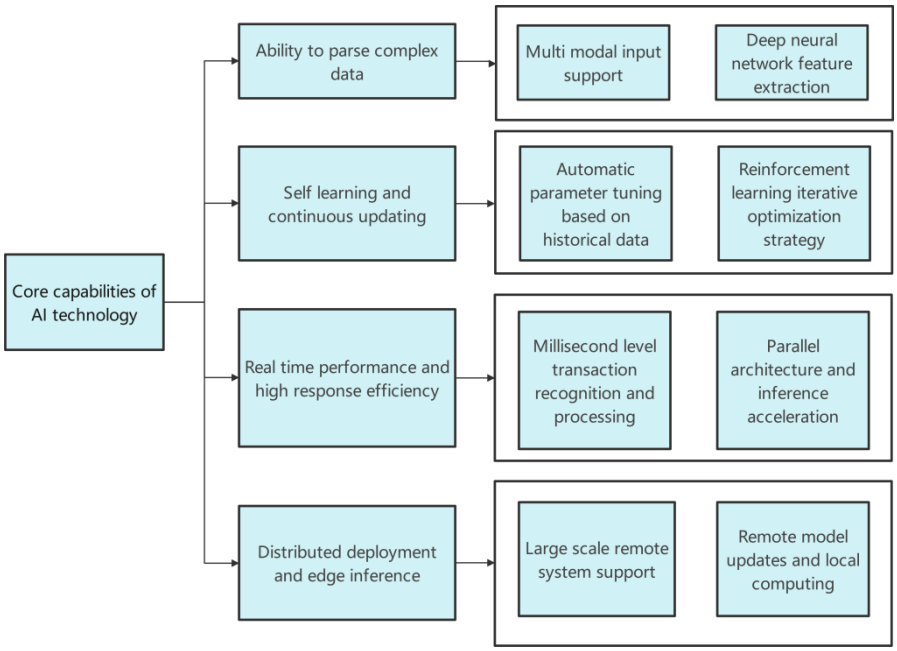

Figure 1. Advantages of AI Technology.

Firstly, AI is capable of effectively parsing unstructured data, making it suitable for processing various types of data such as language, images, sequences, etc. Based on deep neural networks, the system can autonomously process massive volumes of transaction data, extract key features, and detect abnormal trends beyond human perception. Secondly, AI models have self-learning and continuous updating capabilities. Unlike fixed rule sets, it is possible to iterate its own parameters through past data to quickly adapt to new types of attacks and build a dynamic risk identification loop [1]. Thirdly, AI offers high timeliness and responsiveness. Through distributed computing architecture and model acceleration technology, AI algorithms can complete transaction behavior recognition and classification at the millisecond level, fully meeting the requirements of high-frequency trading environments. Lastly, AI can support edge computing and distributed deployment, providing flexible computing resources and risk isolation capabilities for large-scale, remotely deployed payment systems.

3. Path to Enhancing the Security of AI Driven Payment Systems

3.1. Residual Attention Recognition of Abnormal Behavior

In payment systems, numerous transactions are generated in chronological order, with large volumes of data and diverse types. Normal and abnormal operations usually have high similarity in nature, making them difficult to distinguish. By introducing residual neural networks and attention mechanisms to achieve deep discrimination, the model's ability to identify abnormal operations is improved.

In the process of model construction, transaction activities arranged in chronological order are used as inputs, and convolutional layers extracts features from the raw data to obtain a preliminary encoded representation. Several residual modules, which are composed of two or more linear transformation stacks, are then added. Identity mapping is introduced to achieve hierarchical connectivity and prevent gradient vanishing in deeper neural network learning. On the basis of the above, an attention module is introduced, which can change in real-time to adapt to new recognition environments. For example, in detecting a specific type of abnormal transaction, changes in device fingerprint may be more discriminative than changes in transaction amount, prompting the model to adaptively allocate g attention weights to device characteristics. The calculation of attention weights is based on the mapping and normalization of embedded features, and the entire process can be briefly expressed using the following formula:

$$\text{Output} = \text{Attention}(X) + X \tag{1}$$

Among them, X is the original feature output by the residual module, and Attention (X) represents the weighted result of the attention mechanism on it. This structure retains the information of the original feature structure while enhancing the influence of key dimensions, ultimately improving the model's ability to recognize minor and sudden abnormal behaviors.

3.2. Graph Neural Network Clustering Account Relationships

The interaction relationship between accounts in payment system can be represented as a heterogeneous graph structure. In this approach, a GNN based clustering method is used to cluster the account risk structure, which includes three core steps.

Firstly, construct an account behavior graph. Using the account as the central node, transaction behavior, shared devices, and similar login methods as edges, to define a multi-dimensional associated graphical topology structure [2]. Each node has basic attributes such as transaction volume, device number change rate, IP displacement amplitude, etc. A feature matrix X and adjacency matrix A are constructed as data inputs for the next graph neural operation.

Secondly, perform graph convolution to embed features. Based on graph convolutional network (GCN) for node embedding calculation, the joint modeling of structure and attributes is achieved through weighted aggregation of neighboring node features. Each layer of GCN can be embedded and updated through the following definition:

$$h_u = \frac{1}{|\mathcal{N}(u)|}\sum_{v\in\mathcal{N}(u)} h_v \tag{2}$$

Among them, h_u represents the output embedding of the target node, and $\mathcal{N}(u)$ represents its set of adjacent nodes. Through multiple iterations, each account node embeds its neighborhood structure and semantic information.

Thirdly, conduct vector space clustering analysis. After embedding the nodes, unsupervised clustering methods such as K-means are used to group accounts in the embedding space, and accounts with similar transaction patterns are automatically clustered into similar categories.

3.3. Reinforcement Learning Optimization of Risk Control Strategies

The core of risk control strategy lies in making real-time decisions on continuous trading behavior, reducing interference with users while avoiding risks. By utilizing reinforcement learning, an intelligent agent can be constructed to continuously learn optimal strategies through interactions with the environment, achieving flexible and contextualized risk control. This method is based on the ternary structure of state action reward, and simulates strategy functions through deep neural networks to optimize risk control operations in complex trading environments [3].

This system takes each transaction as a state model and has characteristic attributes such as transaction quantity, machine number, IP address, and user behavior trajectory. As a risk management system, intelligent agents generate corresponding actions based on the current situation, such as "allow", "refuse", or "require secondary verification". The results of these actions trigger system feedback, such as analyzing whether the transaction is a scam. Subsequently, the agent updates its policy function to make it more inclined to choose the optimal action when encountering similar states in the future.

The entire learning process uses the Q-learning algorithm to update the state action value function Q (s, a), with the following update rules:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \cdot (r + \gamma \cdot \max_{a'} Q(s',a') - Q(s,a)) \quad (3)$$

Among them, α is the learning rate, γ is the discount factor, and the proxy model is repeatedly trained based on this formula to optimize the strategy.

In practical applications, Q-values can be fitted through deep neural networks, which have the ability to handle complex state spaces. The system can dynamically adjust the reward function through the risk scoring function to better meet the needs of enterprise risk control.

3.4. Multi-Modal Fusion Construction of Linked Defense

Single mode data (such as simple transaction behavior or equipment feature information) cannot provide a comprehensive risk score, so it is necessary to construct the integrated models and joint defense decision-making across multiple categories of information through multiple fusion paths. Transaction behavior, device features, user profiles, geographic locations, and other data are classified and encoded, with feature registration and fusion performed using deep neural networks to construct a comprehensive defense mechanism with high consistency and robustness.

The system performs embedded encoding on various modal data. Time convolutional layers are used to extract momentum features in trading behavior; Device features are mapped into feature vectors through a fully connected layer; User profiles are unified in dimension through mapping layers. All modal feature vectors can be generated into a unified representation through stacking, averaging, and attention weighting.

This process can be expressed using the following concise fusion formula:

$$z = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \quad (4)$$

Among them, x_i is the feature embedding of the i-th modality, w_i is its fusion weight, and z is the fused output vector. By using automatic learning to obtain the corresponding weight allocation of different modalities under different attack conditions, the network can focus on key modal information.

The fused unified feature vector will be passed to the classification device or risk assessment unit to determine whether to implement joint defense measures, such as freezing account usage rights, sending mobile phone verification codes, or restricting high-risk actions.

4. AI Driven User Privacy Protection Mechanism

4.1. Differential Privacy Protection

The differential privacy mechanism can protect personal information privacy by adding a certain amount of random noise during data analysis or model output. For example, in payment systems, differential privacy control modules can be designed into model training or query servers to add perturbations to the output of highly sensitive fields such as behavior frequency, location information, and transaction amount.

Assuming the analysis function is $f(x)$, calculate its global sensitivity Δf, which is the maximum difference between any two sets of data with only one record difference. Noise can be inserted into the obtained calculation results, and these noises originate from Laplace distribution, with parameters determined by sensitivity and confidentiality budget parameter ε.

The core mechanism of differential privacy is expressed as follows:

$$\mathcal{M}(x) = f(x) + \text{Lap}\left(\frac{\Delta f}{\varepsilon}\right) \tag{5}$$

Among them, $\mathcal{M}(x)$ is the final output result, $f(x)$ is the original analysis function result, Δf is the function sensitivity, ε is the privacy budget for controlling the disturbance amplitude, and Lap represents the noise value sampled from the Laplace distribution. These privacy budget values can be applied to query interfaces, model outputs, training gradients, or count statistics to perturb the original output.

4.2. Federated Learning Training

By distributing the model training process to multiple local clients, the centralized transmission of massive data is avoided. The system sends the pre-prepared model to each terminal, and the terminal inputs its own data for multiple rounds of learning and adjustment to obtain updated parameters. After completing the learning process, each terminal uploads the difference in model weights to the server. The server then fuses and updates the model based on the weight values uploaded by each node, and sends the updated model back to the client for the next stage of learning.

The training process can support switching between asynchronous or synchronous modes, based on device stability and network latency configuration. Under the synchronization mechanism, the server waits for all clients to complete training before summarizing; Under asynchronous mechanism, only the training results of a certain node need to be received and updated immediately. This system can filter out nodes that are currently being trained or have insufficient data. Expressed as follows:

$$w = \sum_{i=1}^{N} \frac{n_i}{n} w_i \tag{6}$$

Among them, w_i represents the model parameters trained by the i-th client, n_i is the number of local samples for that client, and w is the global model parameters generated after aggregation.

The training process of each client is managed by its own optimizer (such as SGD or Adam), and the aggregation process is centrally managed by the central node. The parameter transmission adopts encrypted channels, and the consistency of data is verified through the checksum and signature of the model. The system supports technologies such as terminal pruning and quantization compression to improve communication performance.

4.3. Homomorphic Encryption Inference

The model inference interface deployed in the payment system can be connected to a homomorphic encryption module, allowing users to perform feature extraction, model inference, and other operations on data without decrypting it. All encryption operations are completed in the terminal device. The server receives the encrypted data vector, performs simple calculation operations such as addition, subtraction, multiplication, and

division, and finally sends the encrypted result back to the terminal for decryption and output through the terminal device.

In the model deployment architecture, homomorphic encoders have been pre-placed before the input layer to perform numerical conversion and encrypted mapping on all data items. The inference module on the server side is an equivalent network that supports homomorphic operations, often constructed by limiting activation functions or using addition and multiplication gates to compute graphs. In order to meet the homomorphic environment, the model structure needs to avoid linear high-order operations or perform linear approximation processing on them. Taking addition homomorphism as an example, the calculation process between encrypted vectors is as follows:

$$E(a+b) = E(a) \oplus E(b) \tag{7}$$

Among them, $E(a)$ and $E(b)$ are the encrypted results of plaintext data a and b, respectively. $\oplus$ represents homomorphic addition operation, and the output is still in ciphertext form. This operation is completed in the ciphertext domain, consistent with plaintext addition, and satisfies reversibility.

The encryption scheme adopted by the system includes various architectures such as Paillier, BFV, and CKKS, and the specific choice depends on the accuracy requirements and computational efficiency. The server side only retains the public key and does not have the ability to decrypt raw data or intermediate states.

4.4. Blockchain Audit Tracking

With the help of blockchain, it is possible to form an immutable access record of users' access information, enabling effective tracking of the use of personal privacy information in payment systems. The system records information such as data source identification, behavior type, and behavior results in the form of blockchain. Each record includes the timestamp of execution, the identity of the initiator, data source identification, behavior type, and result summary.

The system is built on a blockchain architecture, with each participating node jointly maintaining a distributed ledger. Use consistency protocols to authenticate data uploaded to the blockchain, ensuring that all nodes can access this data instantly and that record cannot be deleted. The audit logic is controlled by smart contracts, which automatically verify access permissions and operational legality, reject unauthorized behavior, and record violation attempts.

The record structure adopts a hash chain approach, where each new block contains the hash value of the previous block and the summary information of the current operation data. Its basic structure is as follows:

$$c = H(H_{n-1} \| D_n) \tag{8}$$

Among them, H_n represents the hash value of the current block, H_{n-1} is the hash value of the previous block, D_n is the summary of the current data access operation, and $\|$ represents the concatenation operation.

Table 1 shows that different mechanisms have constructed a comprehensive system including data input, model processing, result supervision, and log tracking in terms of functional division and implementation path. With the linkage and cooperation of security identification and privacy protection modules, intelligent protection and compliance control of the entire transaction process have been achieved.

Table 1. Comparison of Functions and Deployment Locations of Various AI Mechanisms in Payment Systems.

Module name	function	Deployment location	The required model type
Residual attention recognition	Detection of abnormal trading behavior	Risk control front-end model	Deep convolution + attention network

Graph neural network clustering	Account relationship modeling and gang identification	Account behavior graph analysis layer	Multi-layer graph convolutional network
Strengthen learning risk control strategies	Action decision and risk control	Real-time strategic decision level	Q-learning/DQN
Multimodal fusion recognition	Joint linkage judgment and behavior understanding	Core analysis module of risk control engine	Multi-input deep fusion model
Differential privacy disturbance	Output protection	Model output and query interface	Noise injectors
Federal learning training	Model is updated locally	Client and edge devices	SGD/Adam optimization model
Homomorphic encryption reasoning	Cryptography	Server-side inference engine	Paillier/BFV/CKKS
Blockchain audit tracking	Operation records are traceable	Data access control and audit links	Hash chain + smart contract system

5. Conclusion

The security and privacy control of the payment system is transitioning from a single technology application to a global collaborative approach. In the overall design, each functional unit is constructed based on the functional principles of division of responsibilities and closed-loop information flow. Anomaly recognition relies on the combination of residual networks and attention mechanisms; Account clustering employs graph neural networks for multidimensional correlation analysis; The risk control strategy utilizes reinforcement learning to achieve optimal action generation; Multimodal fusion promotes the structural unity and collaborative judgment of multi-source data. In the privacy protection process, four technologies including noise injection, joint modeling, encrypted inference, and traceability verification cover the entire process of data publishing, model training, and inference, forming an end-to-end protection system. This design emphasizes the practicality of algorithms, the quantifiability of indicators, and the traceability of processes, ensuring flexible deployment and stable operation in diverse trading scenarios, and meeting the technical requirements of a system that combines dynamic risk control and data protection.

References

1. H. W. Kim, and E. H. Song, "Abnormal behavior detection mechanism using deep learning for zero-trust security infrastructure," *International Journal of Information Technology*, vol. 16, no. 8, pp. 5091-5097, 2024.
2. L. Zheng, J. Zhang, X. Wang, F. Lin, and Z. Meng, "Multimodal-based abnormal behavior detection method in virtualization environment," *Computers & Security*, vol. 143, p. 103908, 2024. doi: 10.1016/j.cose.2024.103908
3. S. I. Smirnov, "A method for detecting abnormal behavior of a domain user based on intelligent analysis of security events," In *AIP Conference Proceedings*, August, 2023, p. 020047. doi: 10.1063/5.0161261