Journal of Education, Teaching and Social Studies
ISSN 2642-2336 (Print) ISSN 2642-2328 (Online)
Vol. 8, No. 1, 2026
www.scholink.org/ojs/index.php/jetss

Original Paper

Research on Data Privacy Protection Mechanisms in the Internet of Things

Wanduo Wang

Xihua University, Chengdu, Sichuan, 610039, China

Received: February 9, 2026 Accepted: March 11, 2026 Online Published: March 19, 2026

doi:10.22158/jetss.v8n1p141 URL: http://dx.doi.org/10.22158/jetss.v8n1p141

Abstract

With the in-depth popularization of Internet of Things (IoT) technology, the access of massive terminal devices has brought unprecedented risks of data privacy leakage. This paper focuses on the issue of data privacy protection in the IoT environment, and systematically analyzes the privacy threats faced in four stages: data collection, transmission, storage and application, including illegal sniffing, man-in-the-middle attacks, cloud platform leakage and data mining attacks. On this basis, this paper elaborates on four core privacy protection mechanisms: data anonymization technologies based on K-anonymity and differential privacy, encryption technologies centered on lightweight encryption and homomorphic encryption, blockchain-based distributed trust mechanisms, and architectural innovations combining edge computing and federated learning. Through an in-depth analysis of the principles, application scenarios, advantages and disadvantages of these technologies, this paper reveals the limitations of single technologies and emphasizes the necessity of constructing a multi-level and collaborative protection system. Finally, the paper prospects the development trends of IoT privacy protection in terms of compliance requirements and the integration of emerging technologies, providing a theoretical reference for building a secure and trusted IoT environment.

Keywords

Internet of Things, privacy protection, data encryption, differential privacy, blockchain, federated learning

Introduction

By deeply integrating the physical world and the digital world, the Internet of Things is reshaping the operation mode of human society. From voice assistants in smart homes to traffic monitoring in smart cities, from wearable medical devices to smart factories in Industry 4.0, the number of IoT devices is growing exponentially. These devices generate massive amounts of data every moment, recording

users' location trajectories, health conditions, consumption habits and even daily life details. However, the ubiquitous and intelligent characteristics of the IoT have also made it a key area for privacy leakage. In recent years, frequent incidents such as remote control of smart cameras, leakage of recordings from smart speakers, and ransomware attacks on medical data have aroused deep public concern about IoT privacy security.

Compared with the traditional Internet, IoT data is characterized by massiveness, spatiotemporal correlation, real-time continuity and low value density, making it difficult to directly transplant traditional privacy protection methods. Therefore, in-depth research on data privacy protection mechanisms suitable for the IoT environment and the construction of a protection system covering the entire data life cycle are not only urgent needs to protect citizens' personal information security, but also the cornerstone for promoting the healthy and sustainable development of the IoT industry. Starting from threat analysis, this paper systematically sorts out the current mainstream privacy protection technologies and discusses the possible paths for their collaborative application.

1. Major Threats to IoT Data Privacy

In the entire life cycle of data collection, transmission, storage and application, IoT data faces diverse security threats, which are interwoven to form a complex risk map of privacy leakage.

In the data collection stage, threats mainly stem from physical attacks and software vulnerabilities at the perception layer. IoT terminal devices are usually deployed in unattended open environments. Attackers can tamper with sensor nodes through physical means, implant malicious code, or analyze the power consumption and electromagnetic radiation of devices through side-channel attacks to steal keys. More covertly, many IoT applications suffer from over-collection: applications continuously collect data beyond the scope necessary for services without the user's knowledge or explicit authorization, such as frequently reading location information and address books in the background. Such "data greed" lays hidden dangers for subsequent data abuse.

In the data transmission stage, the openness of wireless communications makes data highly vulnerable to interception. IoT devices widely use wireless communication protocols such as Wi-Fi, ZigBee, Bluetooth and LoRa, with signals broadcast in the air. Attackers only need to deploy receiving devices within the communication range to conduct illegal sniffing. More seriously, man-in-the-middle attacks allow attackers to intervene between communicating parties through techniques such as ARP spoofing or DNS hijacking. They can not only eavesdrop on content but also tamper with data packets in real time, or even inject false instructions to control devices. In addition, attacks against routing protocols are also common, where attackers forge routing information to induce data to flow through malicious nodes.

In the data storage and processing stage, the centralized cloud model amplifies leakage risks. Massive IoT data eventually converges on cloud platforms. Once a cloud platform has configuration errors, unpatched vulnerabilities or suffers from advanced persistent threat attacks, large-scale data leakage

will occur. Internal leaks cannot be ignored either: operation and maintenance personnel with database access rights may illegally export user data for illegal transactions. Even if data is anonymized, attackers can still perform re-identification attacks by associating multiple public datasets. For example, combining anonymized health data with public information on social media can restore the identity of a specific individual. The superposition of these threats poses severe challenges to IoT privacy protection.

2. Privacy Protection Mechanisms Based on Data Anonymization

Data anonymization technology is an important means to prevent privacy leakage in the data release stage. Its core idea is to modify or replace identification information on the premise of ensuring data availability, so that attackers cannot associate data with specific individuals.

The K-anonymity model was one of the earliest widely used anonymization methods. It requires that in a released data table, each record is indistinguishable from at least K-1 other records in terms of quasi-identifiers. Specific implementation methods include generalization (replacing specific values with broader ranges, such as replacing age 25 with the interval 20–30) and suppression (directly deleting certain extreme values). K-anonymity can effectively resist link attacks based on quasi-identifiers, but its limitation lies in its inability to deal with homogeneity attacks (where attackers can still infer private information if all records in the same equivalence class have the same sensitive attribute value) and attacks based on background knowledge.

To address the shortcomings of K-anonymity, the L-diversity model emerged. It requires that the sensitive attributes in each equivalence class have at least L different values, increasing the diversity of sensitive data. For example, in medical data, each age group is guaranteed to contain at least L different disease types. This enhances privacy protection to a certain extent, but it may still face similarity attacks (where values are different but semantically similar). The T-closeness model further refines the requirement, stipulating that the distribution of sensitive attributes in each equivalence class differs from that in the entire data table by no more than a threshold T, effectively preventing the inference of privacy through statistical differences.

Differential privacy has become a research hotspot in academia in recent years. By adding calibrated random noise to query results, it makes the inclusion or deletion of any single record in the dataset have negligible impact on the output. Differential privacy has a strict mathematical definition: regardless of the background knowledge possessed by attackers, they cannot infer whether a certain record exists by observing the output. This makes it an ideal choice for statistical query scenarios, such as publishing demographic data and traffic flow analysis. However, the noise addition mechanism of differential privacy inevitably reduces data accuracy, and the computational overhead is high for massive real-time IoT data streams. How to balance privacy protection strength and data analysis utility remains a question to be continuously explored.

3. Privacy Protection Mechanisms Based on Encryption Technology

Encryption technology is the cornerstone of data confidentiality. By maintaining the ciphertext state throughout the data life cycle, it ensures that even if data is stolen, attackers cannot interpret its real content. For the resource-constrained characteristics of the IoT, the application of encryption technology must balance security and lightweight performance.

Lightweight encryption algorithms are basic protection means in the IoT environment. Traditional algorithms such as RSA and AES incur excessive computational overhead and energy consumption when running on resource-constrained sensor nodes, which may significantly shorten device battery life. To this end, academia and industry have developed a series of lightweight algorithms designed specifically for the IoT. The PRESENT algorithm adopts a permutation-substitution network structure with extremely low gate counts for hardware implementation, suitable for ultra-low-power devices such as RFID tags. The SPECK and SIMON series algorithms, designed by the US National Security Agency, have flexible software and hardware implementation characteristics. The ECC algorithm based on elliptic curve cryptography has a much shorter key length than RSA at the same security strength: 160-bit ECC is equivalent to the security level of 1024-bit RSA, significantly reducing storage and transmission overhead. The application of these lightweight algorithms enables terminal devices to achieve encrypted data transmission within acceptable energy consumption.

Homomorphic encryption represents a higher-level form of encryption technology. It allows direct computation on ciphertexts, and the decrypted computation result is completely consistent with the result of the same computation on plaintexts. This feature is revolutionary: users can upload encrypted data to the cloud, where the cloud completes data analysis and statistics without knowing the plaintext content, and finally returns the encrypted result for the user to decrypt. This fundamentally solves the trust problem of cloud platform data processing.

Currently, homomorphic encryption is mainly divided into partially homomorphic encryption (supporting only addition or multiplication) and fully homomorphic encryption (supporting arbitrary addition and multiplication operations). However, the computational overhead of fully homomorphic encryption is extremely high, and it is still far from practical application in massive IoT data scenarios. At present, it is mainly used in high-value scenarios such as financial data auditing and joint analysis of medical data.

Trusted Execution Environments (TEEs) are hardware-level encryption protection mechanisms. By constructing an independent secure area inside the processor (such as ARM's TrustZone and Intel's SGX), the code and data loaded in this area are protected at the hardware level in terms of confidentiality and integrity. Even if the operating system is compromised, attackers cannot access data inside the TEE. This technology provides an end-side secure computing environment for IoT terminals, especially suitable for core sensitive operations such as key storage and identity authentication.

4. Distributed Trust Mechanisms Based on Blockchain

The rise of blockchain technology has provided a new distributed solution for IoT privacy protection. Its decentralization, immutability and traceability can effectively solve the problems of single point of failure and unclear data ownership in traditional centralized architectures.

In IoT scenarios, blockchain can construct a decentralized identity authentication system. Traditional IoT device authentication mostly relies on central servers; once the server goes down or is breached, the entire network will be paralyzed. Based on blockchain, each IoT device can obtain a unique distributed digital identity, and the authentication process reaches consensus among multiple nodes through smart contracts, eliminating single-point dependence. At the same time, communication keys between devices can be negotiated and generated through the blockchain network and recorded on the chain, ensuring the credibility and traceability of key distribution. Even if some nodes are maliciously controlled, the consensus mechanism can guarantee the reliability of the overall network authentication logic.

Data storage and traceability are another core application of blockchain. After collection, the hash value of IoT data can be stored on the chain. Due to the one-way nature of hash operations and the immutability of blockchain, this provides legal proof of data integrity and timestamps. In the event of data leakage or disputes, the entire process of data circulation can be traced through on-chain records to clarify the responsible entity. In data sharing scenarios, smart contracts can automatically enforce data usage authorization. After users set data access permissions, any third-party access to data must meet the preset conditions of smart contracts and leave access records, realizing "data available but invisible" with full-process traceability.

Blockchain-based data transaction and incentive mechanisms also inject an economic dimension into privacy protection. Users can sell their personal encrypted data to data demanders through the blockchain network, with transactions automatically settled via cryptocurrencies, making users the true owners of their own data. This not only incentivizes users to actively protect data privacy but also breaks the monopoly of Internet platforms on data.

However, the application of blockchain in the IoT also faces challenges, including high node computing and storage overhead, latency caused by consensus mechanisms, and energy consumption problems. Current research directions include introducing lightweight consensus protocols, adopting hierarchical sidechain technology, and combining edge computing to reduce the burden on the main chain, promoting the large-scale implementation of blockchain in IoT scenarios.

5. Collaborative Protection Architecture of Edge Computing and Federated Learning

Facing the dual demands of real-time processing of massive IoT data and privacy protection, the traditional centralized cloud computing model exposes problems such as long transmission delay, high bandwidth pressure and concentrated privacy leakage risks. A collaborative protection architecture combining edge computing and federated learning provides an innovative path to solve the above

problems.

Edge computing significantly reduces data transmission delay and bandwidth occupation by sinking computing and storage resources to the network edge close to data sources. More importantly, a large amount of sensitive data does not need to be uploaded to the cloud; instead, preliminary analysis and processing are completed at local edge nodes, reducing the scope of data exposure from the source. For example, in smart home scenarios, users' voice commands can be identified and processed on home gateway devices, with only abstract control instructions uploaded to the cloud, while original audio data is always stored locally. This localized processing mode prevents attackers from obtaining original privacy data through cloud attacks.

Federated learning further realizes a collaborative learning paradigm of "data stationary, model mobile". Traditional machine learning requires centralizing data from all parties to a central server for training, while federated learning allows participants to retain data locally and only upload encrypted model gradient parameters. The central server aggregates these gradients to update the global model and then distributes it to each node. During the whole process, original data never leaves local devices, effectively avoiding the privacy leakage risks caused by data centralization.

In IoT scenarios, this means that smart devices distributed in various regions can jointly train a shared prediction model (such as input prediction for smart keyboards and disease early warning models for health monitoring devices) without uploading users' personal data to the cloud.

The deep integration of edge computing and federated learning can construct a multi-level collaborative protection architecture. Edge nodes undertake local model training and gradient aggregation, which not only shares the computing pressure of the central server but also further narrows the scope of data transmission. In the architectural design, privacy protection runs through the whole process: lightweight encrypted communication is adopted between terminal devices and edge nodes, and differential privacy technology is used to add noise to gradients between edge nodes and the cloud center to prevent attackers from inferring original data through gradients. At the same time, blockchain technology can be used to record the complete history of model updates, ensuring trusted auditing of the training process.

This hybrid architecture gives full play to the synergistic effects of various technologies, realizing the real-time and availability requirements of IoT intelligent applications on the premise of ensuring data privacy, representing an important evolution direction of future IoT privacy protection.

Conclusion

While the rapid development of the IoT brings convenience to life and industrial upgrading, it also makes data privacy protection an unavoidable major issue. This paper systematically analyzes the diverse threats to data privacy in the IoT environment, and deeply explores core protection technologies such as data anonymization, encryption technology, blockchain mechanisms, and the collaborative architecture of edge computing and federated learning.

Research shows that a single technology is difficult to cope with all privacy challenges in complex IoT scenarios. Data anonymization protects the release stage but sacrifices accuracy; encryption technology ensures confidentiality but incurs computational overhead; blockchain ensures traceability but faces performance bottlenecks. Therefore, it is an inevitable choice to construct a multi-level and collaborative protection system covering the entire data life cycle, integrating the advantages of multiple technologies to seek a dynamic balance among privacy, availability and efficiency.

Looking forward, with the continuous improvement of data privacy regulations (such as the in-depth implementation of the Personal Information Protection Law), privacy protection will shift from a technical compliance requirement to a core element of market competition. At the same time, the development of quantum computing may pose a subversive challenge to the existing encryption system, requiring advance layout in the research of post-quantum cryptography algorithms.

IoT privacy protection is a systematic project intertwined with technology, law and management, requiring joint efforts from academia, industry and policymakers to build a secure and trusted digital future.

References

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, *54*(15), 2787-2805.

Dwork, C. (2008). Differential privacy: A survey of results. International Conference on Theory and Applications of Models of Computation. *Springer*, *2008*, 1-19.

Feng Dengguo, Zhang Min, & Li Hao. (2014). Big Data Security and Privacy Protection. *Chinese Journal of Computers*, *37*(1), 246-258. (in Chinese)

Gubbi, J., Buyya, R., Marusic, S., et al. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*(7), 1645-1660.

Li Fenghua, Li Hui, Jia Yan, et al. (2016). Research Progress and Technology Prospects of Privacy Protection. *Journal on Communications*, *37*(4), 1-11. (in Chinese)

Liu Jianwei, Li Weiyu, & Sun Yu. (2021). Internet of Things Security Technology. Beijing: China Machine Press, 2021:156-189. (in Chinese)

McMahan, B., & Ramage, D. (2017). Federated learning: Collaborative machine learning without centralized training data. *Google Research Blog*, *3*(1), 1-9.

Xu Ke, Wang Yong, & Li Qi. (2018). Blockchain Network Security: Technologies and Applications. *Journal of Information Security*, *3*(3), 1-17. (in Chinese)

Zhang Yuqing, Zhou Wei, & Peng Anni. (2017). A Survey of Internet of Things Security. *Journal of Computer Research and Development*, *54*(10), 2130-2143. (in Chinese)

Zyskind, G., & Nathan, O. (2015). Decentralizing privacy: Using blockchain to protect personal data. 2015 IEEE Security and Privacy Workshops. *IEEE*, *2015*, 180-184.