Advances in Computer, Signals and Systems (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2025.090220
ISSN 2371-8838 Vol. 9 Num. 2

IoT Security Architecture Design Based on Edge Computing

Yuting Yang*

Foreign Language Teaching Department, Hainan Vocational University of Science and Technology, Haikou, Hainan, 571126, China
**Corresponding author: 3197257865@qq.com*

Keywords: Edge Computing, IoT Security, Trusted Execution Environment, Threat Detection, Data Privacy

Abstract: This paper proposes an IoT security architecture based on edge computing, designed to address security challenges in heterogeneous device, resource-constrained, and dynamic threat environments. Through systematic analysis of the shortcomings of existing security mechanisms, a layered security system covering the device layer, edge layer, and cloud management layer is constructed. This system integrates distributed identity authentication, trusted execution environments, intelligent threat detection, and lightweight encryption technology to effectively enhance the overall attack resistance and response speed of the system. In terms of data protection, local data processing, hierarchical storage, differential privacy, and anonymization techniques are adopted to ensure the privacy and security of data in transmission, storage, and processing. In accordance with the requirements of compliance regulations, a comprehensive data management strategy is formulated to enhance the legal and security compliance of the system. This architecture realizes multi-dimensional, elastic, and adaptive security protection, providing a strong guarantee for the secure and trusted operation of the Internet of Things, and laying a theoretical foundation for future widespread application and continuous innovation.

1. Introduction

In the context of today's rapidly developing information technology, the Internet of Things (IoT), as an emerging architectural paradigm, is gradually permeating various fields such as smart cities, the industrial internet, intelligent transportation, and healthcare[1]. Its wide range of application scenarios and massive data scale have greatly promoted the prosperity of the digital economy, but the resulting security risks are also becoming increasingly prominent. IoT devices generally suffer from fundamental problems such as insufficient security protection measures, firmware vulnerability to breaches, and lack of effective protection for data transmission, which pose significant challenges to the overall security of the system. At the same time, the large number of devices, distributed locations, and strong heterogeneity in the IoT environment make it difficult for traditional centralized security strategies to effectively cover every node, significantly increasing the complexity of security management[2].

Against this backdrop, edge computing, with its advantages in reducing latency, alleviating core

network pressure, and improving real-time response capabilities, has become a key technology for improving IoT security architecture[3]. Edge computing migrates data processing, storage, and analysis capabilities to edge nodes closer to the data source, enabling more dynamic and granular security control. By deploying security mechanisms at the edge layer, dependence on the cloud can be effectively reduced, the system's response speed and continuity can be improved, and the risk of system paralysis due to single points of failure can be avoided[4]. At the same time, the edge layer can also implement localized privacy protection strategies during data transmission, which helps to cope with increasingly stringent data regulations and user privacy requirements.

Establishing an IoT security architecture based on edge computing not only solves the security bottlenecks existing in traditional centralized architectures but also realizes the distributed deployment of security policies, making security mechanisms more adaptable and resilient[5]. This type of architecture integrates various advanced technologies, including trusted computing, secure isolation, multi-layer authentication, and threat detection, providing multi-dimensional security protection for IoT devices from the hardware to the application layer. However, designing such an architecture requires a deep understanding of the unique needs and potential threats in edge computing environments, combined with the development of customized security policies for diverse application scenarios. Therefore, researching "IoT Security Architecture Based on Edge Computing" is not only of practical urgency but also promotes the innovative development of the IoT security system at the theoretical and practical levels.

2. Security Status and Challenges Analysis

2.1 Threats and Attack Models

The complexity of IoT security threats and attack models stems from the vast scale, heterogeneity, and wide range of application scenarios of IoT devices. IoT devices are often resource-constrained embedded systems with limited storage and computing capabilities, which creates inherent obstacles to implementing advanced security measures. Attack models cover multiple layers, including device-level attacks such as firmware tampering and hardware hijacking, as well as man-in-the-middle attacks, data eavesdropping, and tampering on communication links. Network-level DDoS attacks exhaust edge and cloud resources through a large number of forged requests, causing service interruptions. In addition, the weak links in device boundaries are vulnerable to remote intrusion and can even evolve into botnets, using compromised devices to launch larger-scale attacks, threatening the overall stability and security of the system.

In many real-world attack scenarios, attackers exploit inherent security vulnerabilities in IoT devices to carry out deceptive operations. For example, intruders can achieve unauthorized access by exploiting weak passwords, default configurations, and a lack of multi-factor authentication. Vulnerabilities in device firmware and software provide entry points, allowing attackers to inject malicious code, manipulate device behavior, or steal critical data[6]. Data in transit is susceptible to man-in-the-middle and replay attacks, especially in wireless communication environments where unencrypted or weakly encrypted transmission links face higher risks. Furthermore, as devices become interconnected, threats are also spreading to the cloud, including cloud database tampering, service configuration corruption, and abuse of cloud APIs, thereby causing global impact on the entire system.

2.2 Existing Security Mechanisms

Traditional security mechanisms exhibit significant limitations when addressing these threats. Cryptography-based authentication and access control mechanisms, while enhancing defense

capabilities to some extent, face multiple challenges in the IoT environment. Devices generally have limited resources, making it difficult to deploy complex encryption algorithms, and security protocols that consume significant bandwidth and computing resources are difficult to promote. Furthermore, untimely device firmware and software updates can easily become a breakthrough point for attacks. Virtualization technology and network isolation policies play a role in ensuring mutual isolation of systems, but their management difficulty and cost in large-scale deployments also become bottlenecks. The lag in vulnerability patching, the difficulty of key management, and the fragmentation of security policies make the overall security system lack continuity and consistency.

2.3 Specific Security Needs

In the edge computing environment, the security needs of IoT systems present new characteristics. Real-time performance has become an important indicator of IoT applications, and any delay or interruption may lead to serious consequences. The decentralized deployment of edge nodes exacerbates the complexity of security policy implementation, and device diversification increases management difficulty. The local processing and storage of data require ensuring privacy while ensuring data integrity and confidentiality to avoid sensitive information leakage. In addition, limited resources of edge devices make the construction of a trusted computing environment difficult, and traditional security technologies are difficult to directly migrate to edge nodes. The heterogeneity and dynamic variability of devices require security policies to be highly adaptable and flexible. The characteristics of edge computing also bring specific security challenges, such as the difficulty in ensuring the physical security of devices. Attackers may use lax access control or physical cracking methods to gain control of devices.

Security mechanisms in the edge environment need to ensure high flexibility while meeting the requirements of low latency and high response speed. A multi-level, multi-dimensional security protection system should be implemented, covering both autonomous security policies on the device side, security measures at the network layer, and strengthening the mutual trust relationship between edge nodes. Lightweight security protocols should be designed for the limited resources of edge nodes to ensure that security measures do not excessively affect system performance. Overall, the security threat model of the IoT is extremely diverse, and attack methods are becoming increasingly professional, while traditional security mechanisms have many shortcomings in the face of these threats. Exploring security strategies and technical systems that meet the characteristics of the edge computing environment is the key to ensuring the security of the future IoT.

3. Security Architecture Design

3.1 Overall Framework

The design of the IoT security architecture based on edge computing follows a layered design approach to fully leverage the advantages of edge computing, achieving efficient, resilient, and secure system operation. The entire architecture is divided into three main layers: the device layer, the edge layer, and the cloud management layer. Each layer undertakes different tasks and responsibilities, forming a closely coordinated and well-structured security system.

The device layer is located at the bottom of the IoT and covers various end nodes such as sensors, actuators, and embedded smart devices. The design of this layer emphasizes the autonomy and security of devices. Hardware security modules (such as Trusted Platform Modules, TPM) are used to implement device identity recognition and store keys, ensuring that devices operate without being

tampered with. The security policies of the device layer mainly involve intrusion detection, firmware integrity verification, anti-tampering measures, and security protocols in resource-constrained environments, such as lightweight encryption algorithms and fast authentication mechanisms. By embedding basic security policies inside the device, potential threats can be identified and corresponding measures taken at the first time, reducing the occurrence of security incidents.

The edge layer is located between the device layer and the cloud management layer, taking on the core responsibilities of preliminary data processing, filtering, and security control. Edge nodes are typically deployed on local gateways or edge servers close to the devices, possessing strong computing and storage capabilities. The security design of the edge layer emphasizes multi-layer isolation strategies, separating different devices or services through virtualization or container technology to reduce the risk of horizontal attacks. Trusted Execution Environment (TEE) is deployed at this layer, providing hardware guarantees for critical security operations. Edge nodes are also equipped with real-time threat detection and response systems, which can identify abnormal behavior through machine learning models and quickly take measures such as blocking and isolation, effectively preventing the spread of attacks. The security policies of the edge layer also include dynamic key management, device authentication, and secure communication protocols, providing a trusted transit space for the entire system.

The cloud management layer, as the highest level of the system, undertakes the responsibilities of formulating global security policies, unified management, and in-depth data analysis. With powerful computing resources, the cloud can implement complex security policy deployment, event linkage analysis, and security auditing. At the same time, the cloud management layer is equipped with a centralized key management system, policy control panel, and security monitoring platform, providing a high degree of controllability when configuring global security policies for the entire system. Through collaboration with the edge layer, the cloud can also perform real-time policy updates and patch pushes, achieving dynamic security defense. The security policies of the cloud layer are embedded in data encryption, access control, permission management, and audit tracking, ensuring that the entire system has a high degree of resilience and autonomy when facing complex threats.

Embedding security policies into these three layers enables each layer to cope with potential threats, while also achieving end-to-end security guarantees. The security mechanisms of the device layer ensure the integrity of individual endpoints, the policies of the edge layer strengthen local protection measures, and the cloud management layer provides global monitoring and policy coordination. This structured system design gives full play to the advantages of edge computing in low latency and elastic management, laying a solid foundation for building a secure and trustworthy IoT system. By reasonably configuring the security policies of each layer, the system can dynamically adapt to complex and changing network environments, effectively respond to the increasing security threats, and achieve secure, controllable, and trustworthy operation of the IoT.

3.2 Key Technical Solutions

In constructing an IoT security architecture centered on edge computing, the selection of key technical solutions is crucial for ensuring the system's security, resilience, and efficiency. The complex and ever-changing threat landscape requires multi-layered, multi-technology protection measures to address the limited resources of devices, dynamically changing application environments, and the need for real-time response and privacy protection.

A distributed identity authentication and authorization mechanism is the foundation for ensuring trusted device interaction in IoT systems. In large-scale, heterogeneous device environments,

centralized authentication models are difficult to respond to quickly and are prone to becoming targets of attack. Adopting a distributed key management and authentication system, using blockchain technology or distributed ledgers to implement decentralized identity verification, avoids single points of failure and improves the system's anti-tampering capabilities. When joining the network, devices are authenticated using hardware certificates or physically unclonable unique identifiers (such as "Physical Unclonable Functions, PUF") to ensure the authenticity of the devices. At the same time, precise permission management can be achieved through fine-grained access control policies combined with dynamic authorization mechanisms. The use of multi-factor authentication schemes, incorporating multi-dimensional information such as time, location, and behavior, enhances the robustness of device identification. This identity verification system, centered on distributed algorithms, not only reduces the pressure on centralized authentication servers, but also improves the efficiency and security of collaboration between devices[7].

Secure isolation and a trusted execution environment for edge devices are designed to provide hardware-level security, blocking the spread of potential attacks. Deploying a trusted platform module (TPM) or hardware security module (HSM) on edge nodes provides a hardware foundation for critical operations, key storage, and secure booting. Using a Trusted Execution Environment (TEE) isolates sensitive code and data, ensuring a trusted computing environment with the limited resources of edge devices. This environment dynamically establishes a chain of trust, ensuring that device firmware and software have not been tampered with through strict boot processes and integrity verification. For multi-device collaboration scenarios, isolation measures can also restrict access rights between different devices or services, preventing horizontal attack spread. In addition, edge devices should have intrusion detection functions that combine software and hardware to monitor abnormal operating behavior in real time, and take prompt protective measures when potential threats are detected, minimizing the attack surface.

Intelligent threat detection and real-time response mechanisms rely on big data analysis and machine learning technology to continuously monitor the edge and the overall system environment. Combining strategies such as behavior analysis, abnormal traffic detection, and device history comparison, a dynamic threat model is established to improve the ability to identify emerging or complex attacks. Key indicators such as sudden traffic changes, abnormal device operations, or communication anomalies are detected in real time, and early warnings are implemented through rule engines or learning models. Once a potential threat is identified, the system can take automatic response measures, such as device isolation, communication disconnection, and policy adjustments, to curb the spread of the attack. The core of this mechanism is rapid decision-making and automated response, reducing human intervention time and improving the system's resilience in the face of diverse security threats[8].

In the IoT environment, resource-constrained devices have a greater need for lightweight encryption and privacy protection technologies. Adopting adaptive and low-computation-cost encryption schemes, such as hash-based or symmetric key algorithms, effectively ensures the confidentiality of data transmission and storage. Combined with a multi-layered key management strategy, sensitive information is divided into different security domains, and different levels of data are protected with differentiated measures. In terms of privacy protection, technologies such as differential privacy, anonymization, and data summarization are implemented to preprocess and obfuscate sensitive data at the edge, protecting user privacy while reducing the demand for transmission bandwidth. Dynamic privacy policies can adjust the protection level according to specific application scenarios, taking into account both user privacy and system security. In addition to reducing the load on devices, such technical solutions also comply with current data protection regulations, providing technical support for the safe and compliant development of IoT applications.

3.3 Data Security and Privacy Protection

3.3.1 Local Data Processing and Hierarchical Storage

Local data processing and hierarchical storage technologies are designed to reduce the risk of sensitive information leakage while optimizing system throughput and response speed. Preprocessing and analyzing data at the edge can effectively reduce the scope of sensitive information flowing in the network, thereby curbing potential leakage channels at the source. Implementing data filtering, compression, and preliminary analysis on edge devices or nodes helps to screen out non-sensitive information, avoid exposing detailed data during transmission, and reduce communication bandwidth requirements. Regarding storage strategies, a tiered system is designed based on the importance and sensitivity of the data. Core privacy data is stored using highly isolated and encrypted storage technologies, while non-sensitive information is stored with a lower level of protection. Through fine-grained storage strategies, the system can dynamically adjust storage resource allocation, balancing performance and security, while meeting the data protection needs of different application scenarios.

3.3.2 Differential Privacy and Data Anonymization Technologies

Differential privacy and data anonymization technologies play a crucial role in multi-layered protection systems. During the edge processing stage, data is blurred by introducing noise mechanisms, ensuring that individual data points cannot be identified without sacrificing analytical value, thereby achieving efficient privacy protection. The core of differential privacy technology lies in its quantifiable degree of protection, allowing developers to adjust the noise intensity according to actual needs, achieving a balance between privacy protection and data usability. Data anonymization schemes, such as multiple de-identification and data summarization, help to strip personal identity information during data release or analysis, preventing the data from being associated with specific users. Furthermore, combining differential privacy enhances the level of privacy protection, which not only meets data sharing needs but also ensures that personal privacy is not leaked, providing a strong guarantee for the secure flow of data in different business processes.

3.3.3 Regulatory Compliance and Data Compliance Strategies

In advancing data security efforts, regulatory compliance and data compliance strategies are particularly critical. Various countries have established strict regulatory requirements for the protection of personal information, such as the General Data Protection Regulation (GDPR) and China's Personal Information Protection Law (PIPL). These laws stipulate explicit provisions for data collection, storage, transmission, and use. Reasonably designing the system's compliance strategy requires defining the scope and permissions of data use from the very beginning of data collection, ensuring that the data complies with relevant regulations from the first stage of entering the system. Data access and processing must be controlled through multi-factor authentication, and the principle of least privilege should be strictly implemented, limiting sensitive information to authorized scopes. At the same time, a sound audit trail mechanism should be established to ensure that every data access and processing is recorded, providing a strong basis for compliance review. For IoT systems operating across regions, the differences in various laws and regulations should also be considered, and differentiated compliance strategies should be formulated to ensure the legal and compliant realization of data value utilization globally.

In the process of implementing a complete data security and privacy protection system, the combination of technical solutions and strategy systems is essential. Through local processing and

hierarchical storage at the edge, the flow path of sensitive data in the system is effectively controlled. The use of differential privacy and anonymization technologies enhances the data's privacy protection capabilities. In conjunction with laws, regulations, and compliance strategies, it ensures that data processing behaviors in different application scenarios comply with legal requirements. Multi-layered and multi-dimensional protection measures build a solid defense line, providing a guarantee for IoT systems to maximize data value while protecting user privacy.

4. Conclusion

This paper focuses on an IoT security architecture based on edge computing, systematically analyzing the diverse threats and attack models currently faced by IoT. It reveals the shortcomings of existing security mechanisms in resource-constrained and variable environments, and emphasizes the new security requirements in edge computing environments. On this basis, it proposes a hierarchical overall framework that organically integrates the device layer, edge layer, and cloud management layer to form a multi-dimensional security protection system. By introducing distributed identity authentication, trusted execution environments, intelligent threat detection, and lightweight encryption technologies, the overall security and resilience of the system are effectively improved. In terms of data security and privacy protection, local data processing and hierarchical storage, differential privacy and anonymization technologies are adopted, combined with regulatory compliance strategies, to establish a multi-level protection mechanism that meets actual application requirements. It integrates technological innovation and standardized strategies to provide a systematic solution for the secure and trustworthy operation of the IoT, and also provides a theoretical basis and practical guidance for future development. With the continuous expansion and complexity of IoT applications, a robust security architecture will be a key guarantee for its widespread application and continuous innovation.

References

[1] Zhang Hongrong, Zhang Feng. Internet of Things Data Architecture Design: Research on Theoretical Framework and Practical Application [J]. Internet of Things Technology, 2025, 15(11):111-114+118.
[2] Liu Min. Research on the Integration of Edge Computing and Industrial Internet of Things [J]. Modern Industrial Economics and Informatization, 2024, 14(05):55-56+87.
[3] Zhang Xinyue. Internet of Things Architecture Design and Optimization for Smart Cities [J]. China Broadband, 2025,21(05):136-138.
[4] Wang Yao. Research on the Security Architecture of the Internet of Things Based on Edge Computing and Blockchain [J] Information Record Materials, 2025, 26(03):187-190.
[5] Qin Xiaofei. Research on Secure Communication and Edge Computing for Space-based Internet of Things [D]. Dalian Maritime University, 2024.
[6] Li Sen. Research on Trusted Authentication of Internet of Things Devices in Edge Computing Environment [D]. East China Jiaotong University, 2024.
[7] Wei Jingli. Research on Intelligent Internet of Things Application Technology Based on Edge Computing [J]. Modern Industrial Economics and Informatization, 2023, 13(10):124-126.
[8] Zhou Qihao. Research on Distributed Security Mechanism and Key Technologies for the Internet of Things [D] Beijing University of Posts and Telecommunications, 2023.