Published online: 31 August 2024 DOI: 10.54254/2977-3903/10/2024111

Exploration of innovative cryptographic application solutions under multimodal big data fusion

*Lin Yiheng [1, a, *], Zhang Junfei [1, b], Wei Shan [2, c], Fan Huimei [2, d]*

[1] China Telecom Corporation Limited Guangxi Branch
[2] Open Security Research

a. linyh15@chinatelecom.cn, b. zhangjf10@chinatelecom.cn, c. shan.wei@osr-tech.com, d. huimei.fan@osr-tech.com
[*] Corresponding author

Abstract. With the rapid development of information technology, data has become the cornerstone of digitalization, networking, and intelligence, profoundly impacting various sectors including production, distribution, circulation, consumption, and social service management. As the core resource of the digital economy and information society, the economic and social value of big data is increasingly prominent, yet it has also become a prime target for cyberattacks. In the face of a complex and ever-changing data environment and advanced cyber threats, traditional big data security technologies such as Hadoop and other mainstream technologies are proving inadequate in ensuring data security and compliance. Consequently, cryptography-based technologies such as fully encrypted execution environments and efficient data encryption and decryption have emerged as new directions for security protection in the field of big data. This paper delves into the latest advancements and challenges in this area by exploring the current state of big data security, the principles of endogenous security technologies, practical applications, and future prospects.

Keywords: big data, endogenous security technologies, cryptography, fully encrypted execution environment, efficient data encryption and decryption

1. Introduction

In the era of big data, data has become a key element driving economic and social development. However, with the explosive growth of data volume and the diversification of data types, data security issues have become increasingly severe. Frequent security incidents such as data breaches, tampering, and misuse not only harm the interests of individuals and businesses but also pose significant threats to the social security system [1]. Therefore, building a secure, efficient, and compliant big data protection system has become an urgent problem that needs to be addressed. This paper explores the potential and advantages of cloud-native distributed technologies in multimodal big data cryptographic applications and proposes an innovative solution. This solution aims to optimize and upgrade big data cryptographic applications by leveraging the flexibility and scalability of cloud-native technology in combination with the characteristics of multimodal data.

2. Analysis of the current state of big data security

2.1. Big data security threats

The security threats facing big data mainly include data breaches, data tampering, data misuse, and privacy violations. Every stage of data collection, storage, processing, transmission, and sharing can become a target for attackers. In particular, with the widespread application of technologies such as cloud computing, the Internet of Things, and artificial intelligence, the risk of data breaches has further increased [2].

2.2. Limitations of traditional security technologies

Mainstream technologies commonly used in the field of big data, such as Hadoop, have several limitations in data security protection. Firstly, these technologies often focus on data storage and computational efficiency while neglecting data security and

compliance. Secondly, traditional security technologies struggle to effectively cope with new, complex data environments and advanced cyber threats [3]. For example, while Hadoop's HDFS file system provides basic access control and data encryption functions, its protection capabilities are insufficient when faced with insider malicious operations or Advanced Persistent Threats (APT) [4].

3. Principles of multimodal big data cryptographic application technology

3.1. Cryptographic service platform

Big data systems handle large volumes of data with diverse types, including structured, unstructured, and semi-structured data, all coexisting as multimodal data. This diversity often requires various types of cryptographic devices to meet the cryptographic computation needs. The cryptographic service platform constructed in this case employs advanced microservice technology and a modular design structure, which is divided into the cryptographic resource pool layer, cryptographic resource management layer, cryptographic service layer, and big data cryptographic application layer [5] at the module level. By managing and dispatching various heterogeneous cryptographic devices (such as cryptographic machines, signature servers, timestamp servers, etc.), the platform provides multimodal cryptographic computation services for big data, including encryption and decryption services, signature and verification services, hashing algorithm services, timestamp services, and more. It establishes a "three-tier protection system" for the big data system, ensuring the security of regional boundaries, communication networks, and application computation, meeting the cryptographic protection needs of multimodal data throughout the entire lifecycle, including integrity, confidentiality, and non-repudiation [6]. The overall architecture of the cryptographic service platform is shown in Figure 1:

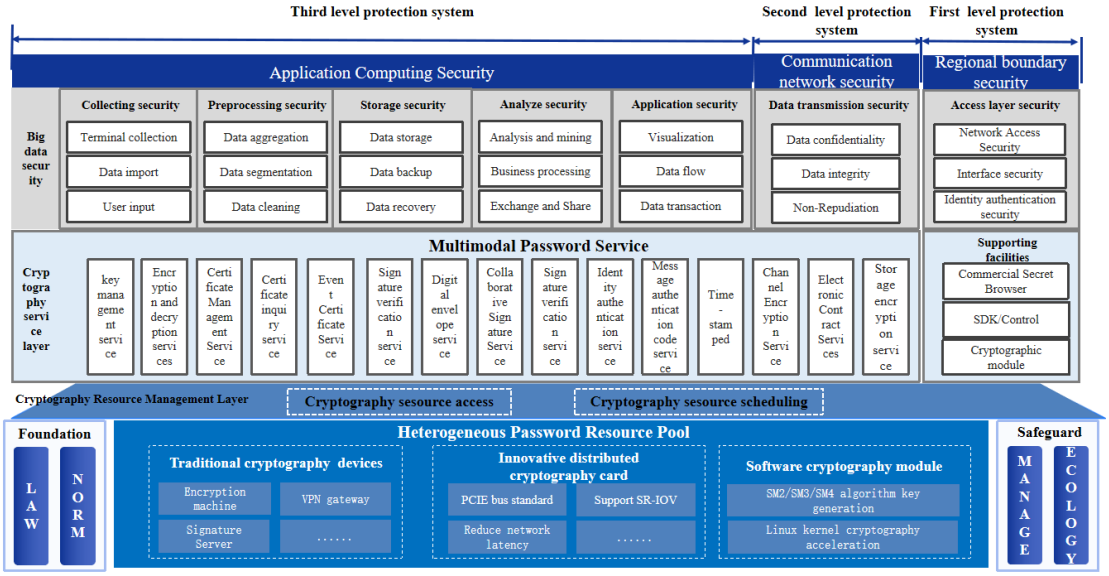

Figure 1. Overall Architecture of the Cryptographic Service Platform

3.1.1. Big data cryptographic application layer

The Big Data Cryptographic Application Layer primarily provides a unified cryptographic service interface for various application objects involved in big data systems, such as applications, communication networks, and regional boundaries. These objects can directly invoke the cryptographic service platform to implement certificate management, key management, data encryption and decryption, data signature verification, identity authentication, and other cryptographic services.

3.1.2. Multimodal cryptographic service layer

The Multimodal Cryptographic Service Layer is the critical intermediate layer within the overall framework of the unified cryptographic service platform. This layer, based on the cryptographic resource management layer, refines and summarizes the relatively fixed and abstract cryptographic requirements of applications, constructing the necessary products and systems to support various cryptographic services. It provides multimodal cryptographic services for big data, reducing the complexity of using cryptographic algorithms within big data systems. This layer can dynamically scale according to the needs of the big data

system, adding various products to meet the cryptographic application requirements of big data, and dynamically enhancing the types, capabilities, and levels of cryptographic services.

3.1.3. Cryptographic resource management layer

The Cryptographic Resource Management Layer implements the pooled management and service of cloud-based cryptographic resources, reasonably allocating and dispatching various cryptographic resources according to the cryptographic business needs of application systems. It ensures balanced scheduling of cryptographic resources and compatibility with various cryptographic devices. By employing container technology and the concept of a cryptographic security platform, it utilizes platform-based methods to discover, consolidate, and reuse cryptographic capabilities, achieving dynamic load balancing of cryptographic computational power. It provides efficient, stable, and reliable cryptographic computation services for business application systems. This layer also unifies the management of cryptographic devices connected to the platform, monitors the operational status of devices and cryptographic services, and issues timely status alerts.

3.1.4. Cryptographic resource layer

The Cryptographic Resource Layer pools cryptographic resources/devices and systems of various manufacturers, models, and functions, providing cryptographic computational support for diverse cloud-based cryptographic application needs. This includes resources such as cloud server cryptographic machines, server cryptographic machines, signature verification servers, distributed cryptographic cards, and other cryptographic devices.

3.2. Distributed cryptographic card

To address scenarios involving large data volumes and low latency in big data applications, a distributed cryptographic card can be embedded in the PICE card slot of a distributed node server [7]. Big data applications can directly invoke the distributed cryptographic card via the PICE bus for cryptographic operations, enabling local cryptographic computation, reducing cryptographic processing latency, and supporting real-time big data applications, as illustrated in Figure 2:

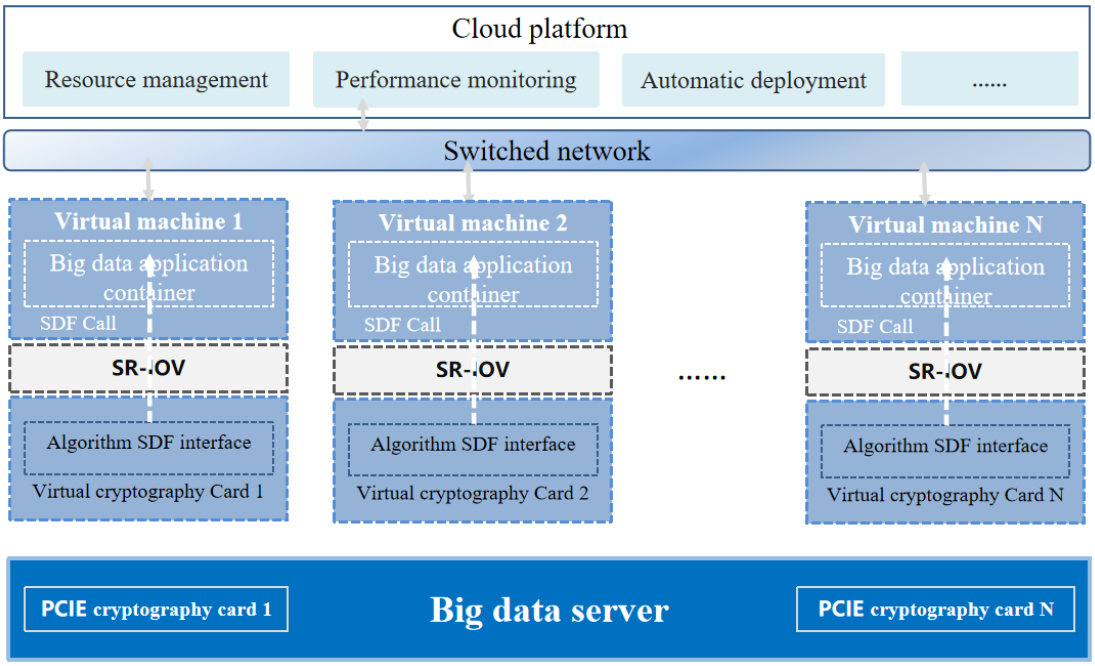

Figure 2. Overall Architecture of the Distributed Cryptographic Card

(1) The distributed cryptographic card uses SR-IOV technology to achieve virtualization, allowing container engines to access virtual cryptographic cards. Big data applications can access the local virtual cryptographic card through the container engine, directly invoking the SDF algorithm interface of the virtual card (in compliance with the "GM/T0018-2012 Cryptographic Equipment Application Interface Specification"). This approach avoids network delays associated with intermediate layers such

as virtual switches, effectively increasing cryptographic computation network throughput, reducing latency, better meeting real-time requirements, and enhancing processing speed.

(2) A single distributed cryptographic card can be virtualized into multiple (16/32) distributed virtual cryptographic cards, creating a cryptographic computing resource pool that works with the cloud management platform to achieve elastic scaling, on-demand scheduling, and load balancing, thereby supporting cryptographic operations for massive data in big data environments.

(3) The distributed cryptographic card supports multiple national cryptographic algorithms, including SM2, SM3, and SM4, and can perform functions of various specialized cryptographic devices, such as cryptographic machines, signature servers, and timestamp servers. This includes encryption and decryption using SM1 and SM4, generating and verifying message authentication codes with SM3, SHA1, SHA256, SHA384, and SHA512, generating/verifying digital signatures, and supporting digital envelope functions based on RSA/SM2 cryptographic algorithms. This "one card, multiple uses" approach reduces costs.

3.3. Distributed key management center

The distributed key management server for big data unifies the lifecycle management of keys for distributed nodes, including key generation, storage, distribution, updating, backup, and destruction. In coordination with the distributed cryptographic card, it provides real-time dynamic key synchronization for load balancing and elastic scaling of distributed nodes in big data, improving the efficiency of cryptographic operations for critical big data [8]. This supports real-time business scenarios, as illustrated in Figure 3:

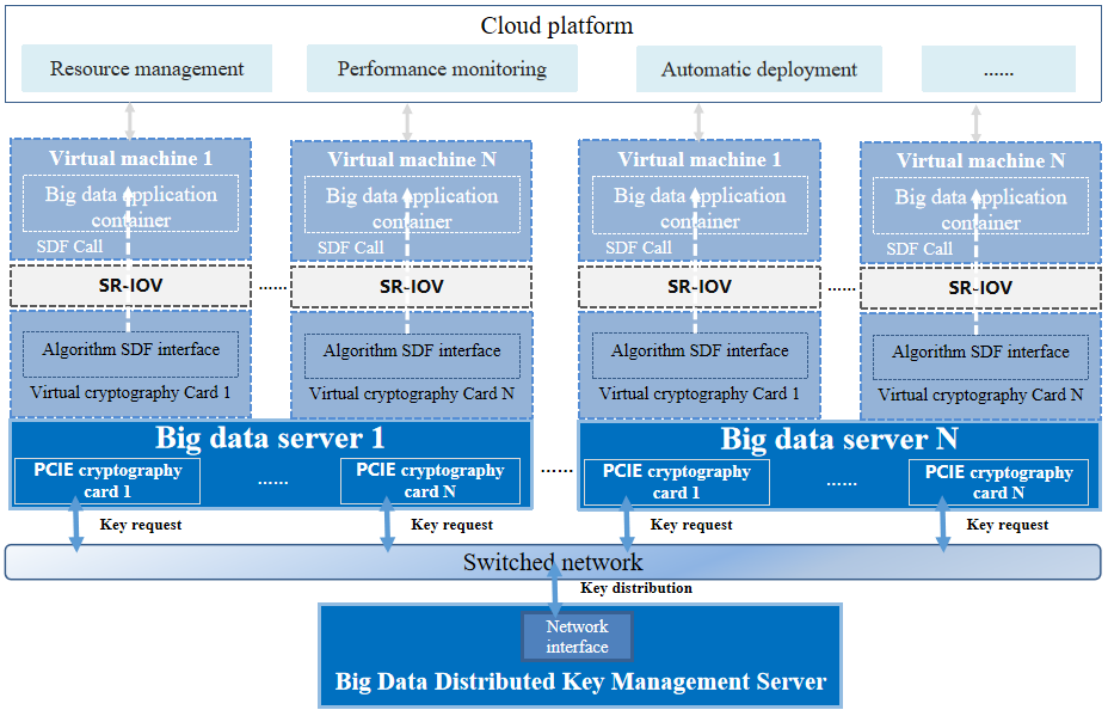

Figure 3. Architecture of the Distributed Key Management Center

The distributed key management server for big data unifies key lifecycle management for all distributed nodes. This includes key generation, storage, distribution and negotiation, usage, backup and recovery, revocation, and destruction. Keys are not stored within the distributed cryptographic cards; instead, the distributed cryptographic cards are interconnected with the distributed key management server via a network. When a key is needed, the distributed cryptographic card requests the key from the distributed key management server, which generates and distributes the key to the cryptographic card.

Each big data application that connects to the distributed key management server is assigned a unique application ID (AppID). The server establishes an association between the big data application ID (AppID), the distributed cryptographic cards, and the keys, implementing the concept of "keys follow the application." This enables the cryptographic operations of a single application to be distributed across multiple cryptographic cards, enhancing the efficiency of big data cryptographic operations.

The distributed key management server utilizes a physical noise source chip, certified by the National Cryptography Administration, to generate random numbers, which comply with the "GM/T0005-2012 Randomness Detection Specification." A root key is used to protect all stored keys, ensuring the security of keys during storage. The root key should be stored within the distributed key management server to prevent unauthorized access and tampering. All keys are encrypted during storage to ensure

their confidentiality, and integrity protection measures are applied simultaneously to safeguard the integrity of keys during transmission and storage. Through multi-layered security measures, the confidentiality and integrity of stored keys are ensured.

4. Application effectiveness

4.1. Feasibility

This case study focuses on the scenarios of big data migration to the cloud and distributed deployment, fully leveraging the advantages of cloud computing's elasticity and distributed architecture. It innovatively integrates a cryptographic service management platform, high-performance distributed cryptographic cards, and a comprehensive distributed key management system, establishing an efficient and secure distributed cryptographic service system on native big data servers. This system ensures that cryptographic operations such as encryption and decryption, signature verification, integrity checking, and timestamping can be swiftly and securely completed on native servers without the need for additional data transmission or invocation. This significantly reduces network latency, lowers the risk of data breaches and cyberattacks, and enhances the overall system's response speed and processing efficiency.

In terms of technical protection effectiveness, this case achieves distributed processing of cryptographic operations and secure management of keys through the integrated application of distributed cryptographic cards and the key management system. This ensures the confidentiality and integrity of data during transmission and storage. Additionally, through the unified scheduling and monitoring of the cryptographic service management platform, the system can monitor the status of cryptographic services in real time, promptly detecting and responding to potential security threats, thereby strengthening the overall cryptographic protection capabilities.

Regarding the comprehensiveness of demand coverage, this case thoroughly analyzes the security requirements of the big data platform throughout its lifecycle. It not only covers basic security needs such as identity authentication, confidentiality and integrity of data transmission and storage, and non-repudiation, but also extends to advanced security features such as data auditing, access control, and emergency response, ensuring comprehensive cryptographic security protection for the big data platform.

In terms of regulatory compliance, this case strictly aligns with relevant policies and standards, such as the "Cryptography Law" and cryptographic evaluations, ensuring that cryptographic applications are correct, compliant, and effective. By internalizing regulatory requirements into the core principles of system design and implementation, and through automated compliance checks, regular security audits, and risk warning mechanisms, the case ensures that the big data platform consistently meets regulatory requirements during ongoing operations.

4.2. Innovation

(1) The cryptographic service platform provides multimodal cryptographic services to big data applications by uniformly managing and scheduling heterogeneous cryptographic devices. It constructs a "three-layer protection system" that includes regional boundary security, communication network security, and application computing security, meeting the needs for authenticity, integrity, confidentiality, and non-repudiation protection functions throughout the entire lifecycle of multimodal big data.

(2) The distributed cryptographic card is embedded into the PICE card slot of big data servers, utilizing SR-IOV technology to enable local cryptographic operations for big data applications. This approach avoids intermediate layers like virtual switches, effectively increasing network throughput for cryptographic operations, reducing latency, and better meeting the real-time requirements of big data scenarios.

(3) The distributed key management platform establishes an association between big data applications, keys, and distributed cryptographic cards, supporting real-time key synchronization between different cryptographic cards. This "key follows the application" approach meets the distributed cryptographic operation requirements for massive big data.

5. Challenges and future outlook

Despite the numerous advantages of multimodal big data cryptographic applications based on cloud-native distributed technologies, several technical challenges still arise during practical implementation. The following are some major challenges and corresponding solutions:

(1) Data Consistency and Synchronization Issues: In multimodal data processing, inconsistencies and synchronization issues may occur between different data sources. To address this, distributed transaction management technologies and data consistency protocols (such as Raft and Paxos) can be employed to ensure synchronization and consistency of data across different microservices.

(2) Resource Scheduling and Optimization: Resource scheduling on cloud-native distributed platforms is complex, requiring the rational allocation of computing, storage, and network resources. To optimize resource utilization, intelligent scheduling algorithms (such as Kubernetes schedulers) can be used to dynamically adjust based on task requirements and resource status.

(3) Security and Privacy Protection: Multimodal data often contains sensitive information, making it crucial to ensure data security during transmission, storage, and processing. In addition to employing strong cryptographic algorithms and encryption technologies, it is necessary to establish comprehensive security management mechanisms and privacy protection policies to ensure lawful data use and privacy protection.

6. Conclusion

This paper proposes a solution for multimodal big data cryptographic applications based on cloud-native distributed technologies. The solution leverages microservices, containerized deployment, and automated operations to achieve efficient processing and secure protection of multimodal data. However, with the continuous development of technology and the expansion of application scenarios, many issues still require further research and resolution. In the future, we will continue to monitor trends in cloud-native distributed technologies, continuously optimize and improve multimodal big data cryptographic application solutions, and provide more secure and efficient data processing services across various fields.

References

[1] Wang, Y. (2023). Prospects for high-efficiency search privacy protection mechanisms in big data security. *Software, 44*(10), 140-142.
[2] Wang, Y., & Wang, S. (2024). Design of network information security protection under the background of big data. *Information and Computer (Theoretical Edition), 36*(07), 212-214.
[3] Wang, Y. (2024). An analysis of network security in the big data environment. *Communication Management and Technology,* (02), 50-52.
[4] Wu, M. (2024). An exploration of data security issues under the background of big data. In *Proceedings of the 38th China (Tianjin) 2024 IT, Network, Information Technology, Electronics, and Instrumentation Innovation Academic Conference* (p. 3). Tianjin TEDA Information Systems Engineering Supervision Co., Ltd. https://doi.org/10.26914/c.cnkihy.2024.014484
[5] Li, D., Zhu, Y., & Hai, J. (2023). Research on the application of cloud cryptographic services in information security. *Shanghai Informatization,* (08), 38-41.
[6] Zhang, J., Guo, Z., Wang, G., et al. (2023). Research on the construction of cryptographic service platforms based on national cryptographic algorithms. *China New Communications, 25*(05), 107-109.
[7] Cui, Y., Liu, L., Meng, W., et al. (2020). Implementation of high-speed cryptographic cards based on the SM2 signature algorithm. *Software Guide, 19*(07), 183-186.
[8] Xu, S., Guo, C., & Li, X. (2020). Improved key management scheme based on multi-party co-management and IBC. *Computer Applications and Software, 37*(08), 314-317.