Proceedings of CONF-CDS 2025 Symposium: Data Visualization Methods for Evaluatio
DOI: 10.54254/2755-2721/2025.PO25408

Cryptography Techniques in Medical Data Privacy Protection: Applications and Challenges of Homomorphic Encryption, Differential Privacy, and Blockchain

Chenyuan Zhang

School of Information and Technology, PSB Academy of Singapore, Singapore
1558140400@qq.com

Abstract. With the rapid development of big data and artificial intelligence technologies, data security has become a critical bottleneck restricting the development of data science. This study systematically explores the innovative applications and implementation challenges of modern cryptographic techniques in the field of data science. The paper first reviews the fundamental theories of cryptography, such as symmetric encryption, asymmetric encryption, and hash functions. It then focuses on the cutting-edge applications of homomorphic encryption in privacy-preserving machine learning, differential privacy in user data analysis, and blockchain in data integrity verification. Through an in-depth analysis of typical cases such as medical data sharing and user behavior modeling, the study reveals the effectiveness and limitations of cryptographic techniques in practical deployment. The study further identifies the main challenges currently faced, including algorithmic computational efficiency, the transition to post-quantum cryptography, and the balance between data privacy and usability. Finally, this paper proposes future development directions for the deep integration of cryptography and data science from both technical evolution and policy-making perspectives. This study provides important theoretical references and methodological guidance for secure computing practices in the field of data science.

Keywords: Homomorphic Encryption, Differential Privacy, Blockchain Technology, Secure Multi-Party Computation, Medical Data Privacy Protection

1. Introduction

With the rapid development of information technology, the generation and circulation of data are growing explosively. From user data on Internet platforms to corporate trade secrets, from medical health data to financial transaction records, every link in the data lifecycle faces the risk of being stolen, tampered with, and misused. In recent years, frequent data leakage incidents have not only caused significant economic losses to individuals and enterprises but also triggered a severe crisis of social trust. For example, the large-scale leakage of Facebook user data in 2021 has highlighted the importance of data security in today's society. Cryptography, as the cornerstone of information security, provides robust protection for data confidentiality, integrity, and authenticity through

DOI: 10.54254/2755-2721/2025.PO25408

techniques such as encryption, authentication, and digital signatures. Integrating cryptography with data science can not only effectively protect data security during the processes of collection, storage, transmission, and analysis but also enable privacy protection and compliant use of data in data sharing and open scenarios. This is of great strategic significance for promoting the sustainable development of data science.

Although cryptography has achieved remarkable results in the field of information security, it still faces many challenges in the complex application scenarios of data science, such as achieving efficient data processing and analysis while ensuring data security, protecting privacy and confidentiality in data sharing, and dealing with the threat of quantum computing to existing encryption systems. Solving these problems is of great value for promoting the deep integration of data science and information security. This study aims to explore the current status, challenges, and future trends of modern cryptography in data science. By reviewing the basics of cryptography, analyzing its applications in data encryption, privacy protection, and secure multi-party computation, examining typical cases to reveal actual application effects and issues, and discussing challenges such as algorithm efficiency, the balance between data usability and security, quantum computing threats, and legal and ethical considerations, this paper proposes solutions and future research directions to provide references and guidance for data science practitioners.

2. Cryptographic fundamentals

2.1. Symmetric and asymmetric encryption

Encryption algorithms are the core technologies for ensuring data security and are mainly divided into symmetric and asymmetric encryption [1]. Symmetric encryption uses the same key for both encryption and decryption, which is fast but has complex key management. Common algorithms include AES (supporting 128/192/256-bit keys, efficient and secure) [2]. Asymmetric encryption uses a pair of public and private keys, which is highly secure but computationally expensive. Common algorithms include RSA (based on the difficulty of factoring large integers, reliable and widely used) and ECC (based on the elliptic curve discrete logarithm, short keys, and high efficiency) [2].

2.2. Hash functions and digital signatures

Hash functions convert inputs of arbitrary length into outputs of fixed length and have characteristics such as determinism, fast computation, collision resistance, and the avalanche effect. Common algorithms include MD5 (128-bit, fast but low security), SHA-1 (160-bit, higher security but with collision risks), SHA-256 (256-bit, high security), and SHA-3 (variable output length, flexible and secure). Hash functions are used for data integrity verification, password storage, digital signatures, and other fields, but they need to be combined with measures such as salting to enhance security [2].

Digital signatures are used to verify the integrity of data and the identity of the sender, combining encryption algorithms and hash functions. The sender encrypts the hash value with their private key to generate a signature, and the recipient decrypts and verifies it with the public key. Digital signatures have non-repudiation, integrity, and identity verification functions and are applied in electronic documents, e-commerce, and other fields [2].

DOI: 10.54254/2755-2721/2025.PO25408

2.3. Public Key Infrastructure (PKI)

PKI is an important facility for ensuring users' secure access to resources, consisting of digital certificates, CAs, RAs, certificate repositories, etc. It verifies identities, ensures data integrity, and provides non-repudiation through digital certificates. Its working principle includes certificate issuance, use, and revocation, and it is applied in HTTPS, email encryption, and other fields, but it is also at risk of being attacked [3].

3. Applications of cryptography in data science

3.1. Data encryption and privacy protection

Data encryption has been widely used in database systems to protect sensitive fields. Oracle TDE technology encrypts data at the tablespace level through transparent encryption, with a performance loss of only about 5%. MySQL 5.7 uses the AES-256 encryption algorithm to protect highly sensitive fields such as user passwords, supporting hardware acceleration to reduce performance impact [4]. In practice, financial institutions generally use the national cryptographic algorithm SM4 to encrypt customer information, combined with a key management system to achieve hierarchical authorization. The new database Neo4j introduces an attribute-based encryption scheme, allowing different encryption strengths to be set according to node attributes. MongoDB defaults to using static encryption technology to protect disk data, combined with SSL/TLS to ensure transmission security. In medical database applications, homomorphic encryption technology supports direct calculations on encrypted data, such as average length of hospital stay and cost statistics [5].

The privacy protection issue in big data is essentially a data privacy issue. Data privacy refers to the sensitive data or characteristics represented by the data that data owners are unwilling to disclose [6]. In big data applications, securely computing user data and protecting user privacy is a fundamental issue. Since the computational problems in big data are very complex and diverse, privacy protection algorithms suitable for specific computational situations usually cannot meet the needs of big data. Therefore, a more comprehensive solution must be chosen to protect users' computational privacy. Fully homomorphic encryption (FHE) is a suitable choice [7]. Fully homomorphic encryption allows arbitrary computations on ciphertext data without decryption, suitable for scenarios where users store encrypted data on cloud servers and request cloud servers to perform computations. This solution includes four main algorithms: key generation (generating public and private keys), encryption (encrypting plaintext data into ciphertext), evaluation (computing on ciphertext and outputting new ciphertext), and decryption (decrypting ciphertext into plaintext with the private key). In big data secure computing, users can request cloud servers to call the evaluation algorithm to operate on ciphertext, and then users decrypt the results with their private keys. Current research mainly focuses on improving the operational efficiency and security of fully homomorphic encryption schemes to meet the requirements of big data computing for response time and security. Homomorphic encryption schemes are not new. In fact, before Gentry [8] proposed the fully homomorphic encryption scheme in 2009, there were already some homomorphic encryption schemes based on factorization and discrete logarithm problems.

3.2. Data integrity

Blockchain, as a decentralized distributed ledger, is composed of blocks, chains, and nodes [9]. Blockchain ensures data immutability through its chain structure: each block contains the hash value

DOI: 10.54254/2755-2721/2025.PO25408

of the previous block (Hn-1), current data (Datan), timestamp (Tn), and current hash value (Hn), with the hash value calculated by the formula Hn = Hash(Hn-1 || Datan || Tn) [10]. If any block data is tampered with, the hash values of all subsequent blocks will change, breaking the integrity of the chain. At the same time, the consensus algorithm of blockchain (such as Proof of Work, PoW) ensures that distributed nodes reach consensus without central control to verify the validity of new blocks [11]. PoW adds new blocks by calculating a hash value that meets a specific difficulty level (H = Hash(Block_Header || Nonce)). Tampering with data will change the subsequent hash values, making it almost impossible to succeed, thereby ensuring data integrity and system security [12].

4. Case studies

4.1. Analysis of medical data encryption

In the modern medical industry, the security of medical data is becoming increasingly prominent. Taking the Yale New Haven Health System data leakage incident in the United States as an example, on March 8, 2025, the personal information of more than 5.5 million patients was leaked, including names, dates of birth, race/ethnicity, home addresses, phone numbers, email addresses, social security numbers, medical record numbers, and types of visits, but it did not involve financial information, medical records, or treatment details [14]. This incident highlights the urgency and importance of strengthening the security protection of medical data. Faced with complex and changing data security threats, effective measures must be taken to comprehensively strengthen the security protection of medical data.

4.2. Differential Privacy protection of user behavior

Differential Privacy (DP) is a privacy protection technology that introduces controllable noise into the data analysis process to balance the relationship between data usability and individual privacy. Its core idea is to ensure that the changes in the analysis results are negligible when a record is added to or removed from the dataset, thereby protecting individual privacy. As an effective means of protecting user data privacy, differential privacy technology has attracted widespread attention from domestic and international researchers. In terms of improving data utility, HADIAN [15] et al. proposed a differential privacy mechanism based on Laplace noise, which protects privacy through bucket partitioning algorithms and Laplace distribution, significantly improving the accuracy of perturbed data. PREMA [16] et al. designed a differential privacy protection model for big data in body area networks, and experimental results showed that the scheme could maintain data usability while resisting background attacks. ZHANG [17] et al. proposed the Re-DPoctor algorithm, which combines Laplace noise mechanism with adaptive sampling, filtering, and budget allocation algorithms to achieve real-time differential privacy publishing of health data collected over multiple consecutive days.

In terms of reducing privacy attack risks, GUAN [18] et al. proposed the EDPDCS clustering scheme, which combines K-means clustering and Laplace noise in the Map-Reduce framework, effectively improving the accuracy of data publication. HAN [19] et al. proposed the PPM-HDA algorithm, which supports various aggregation operations (such as average value, variance, minimum/maximum value, median, etc.), is well adapted to cloud servers, and can effectively resist differential attacks. UKIL [20] et al. proposed a method of obfuscating sensitive data on-demand, which significantly reduces the risk of privacy leakage by meeting differential privacy requirements.

DOI: 10.54254/2755-2721/2025.PO25408

SALEHEEN [21] proposed the mSieve algorithm, which combines data-driven techniques and Laplace noise to further enhance privacy protection by obfuscating data.

In addition, researchers have also explored the combination of differential privacy with other technologies, such as the exponential mechanism [22], Fourier algorithms [23], and classification trees [24], to further optimize privacy protection effects and data utility.

5. Challenges and future directions

5.1. Technical challenges

With the rapid development of information technology, digital information has become a key element in driving the development of various fields such as economy and science and technology. However, while information is widely disseminated and used, information security issues have become increasingly severe. The "China Cybersecurity Industry Analysis Report (2024)" points out that since 2023, information security incidents such as ransomware, data leakage, and hacker attacks have emerged continuously and caused greater harm, seriously threatening the national security and healthy development of China's national economy [25]. Therefore, how to effectively protect sensitive information from being stolen, tampered with, and misused has become a challenge that has attracted widespread attention. Cryptographic techniques, as the cornerstone of information security, aim to protect sensitive information and data by using difficult mathematical problems. These mathematical problems are usually difficult to solve effectively in polynomial time on classical computers, thereby effectively ensuring the confidentiality, integrity, and availability of data. However, with the rapid development of quantum computing technology, traditional cryptographic systems relying on difficult mathematical problems, such as the problem of factoring large integers and discrete logarithms, have been proven to be efficiently solvable using quantum computers [26], which poses a severe challenge to traditional cryptographic systems.

5.2. Theoretical and policy considerations

The "Cryptography Law" and related regulations, by clarifying the four-level management system of national, provincial, municipal, and county levels, establishing the principle of classified management, prohibiting the misuse of cryptography, establishing comprehensive cryptography security requirements, regulating import and export controls, optimizing cryptography management, strengthening cryptography monitoring and crisis management, and improving supervision and inspection mechanisms in eight aspects, have improved China's cryptography management system. These measures not only rationalize the relationship between confidential and commercial cryptography management, clearly define the scope of commercial cryptography, but also, by connecting with related laws, strengthen the cryptography security management in key areas, reduce the occurrence of cryptography security incidents, and effectively control losses after incidents occur. At the same time, by canceling unnecessary licensing regulations, clarifying the scope of free trade in imports and exports, and optimizing the methods of supervision and inspection, they promote the rational application and development of cryptographic techniques and provide comprehensive security for cryptography [27].

6. Conclusion

This paper explores the applications of modern cryptography in data science, such as homomorphic encryption, differential privacy, secure multi-party computation, and blockchain technology, and

DOI: 10.54254/2755-2721/2025.PO25408

analyzes their feasibility and limitations through practical case studies. The study finds that although cryptographic techniques can significantly enhance data security, they still face many challenges: technically, it is necessary to balance computational efficiency with encryption strength and address the transition to post-quantum cryptography; in application, it is necessary to resolve the contradiction between privacy protection and data usability, as well as the compliance issues of cross-institutional data collaboration; in policy and ethics, it is necessary to deal with differences in regulations and the lack of an artificial intelligence ethics framework. In the future, the integration of cryptography and data science will show trends such as algorithm optimization, standardization processes, and interdisciplinary collaboration. Future research should focus on implementable technical solutions, strengthen policy adaptability, and achieve coordinated progress between security and development.

References

[1] Lu, H. W. (2024). Research on data encryption technology in computer network information security. Information Systems Engineering, (8), 132-135.
[2] Xu, H. L. (2025). Design of a security architecture for cloud computing: Application of symmetric and asymmetric encryption algorithms. China Broadband, 21(2), 109-111.
[3] Zhang, B., Zhang, Y., Zhang, W. Z., et al. (2025). Research and progress on PKI technology. Journal of Software, 36(6), 2875-2899.
[4] Chen, X. L. (2024). Data consistency mechanism in distributed database systems in cloud computing environments. Information and Computer (Theory Edition), 36(8), 137-139.
[5] Li, J. K. (2025). Analysis of data privacy protection technologies in database systems. Heilongjiang Science, 16(10), 159-161.
[6] Benjamin C. M. Fung, Ke Wang, Rui Chen, and Philip S. Yu. 2010. Privacy-preserving data publishing: A survey of recent developments. ACM Comput. Surv. 42, 4, Article 14 (June 2010), 53 pages.
[7] Huang, L. S., Tian, M. M., & Huang, H. (2015). A review of cryptographic techniques for big data privacy protection. Journal of Software, 26(4), 945-959.
[8] Gentry C. Fully homomorphic encryption using ideal lattices. In: Proc. of the 41st Annual ACM Symp. on Theory of Computing (STOC). New York: ACM Press, 2009. 169−178. [doi: 10.1145/1536414.1536440]
[9] Qiu, J., & Huang, M. H. (2023). Research on integrated medical security data management model based on blockchain technology. Popular Science, 25(2), 4-7.
[10] Yu, L., & Jin, Y. (2017). Research on data splitting technology of blockchain global ledger. High Technology Communications, 27(Z2), 875-888.
[11] Yang, C. J. (2023). Construction of trustworthy data service application for universities based on blockchain technology. Journal of Huaibei Normal University (Natural Science Edition), 44(4), 72-78.
[12] Lu, L. (2025). Data integrity verification method based on blockchain technology. Computer Programming Skills and Maintenance, (5), 85-87+122. https: //doi.org/10.16184/j.cnki.comprg.2025.05.047
[13] Zhou, J. L., Han, A. X., Liu, Y. W., et al. (2021). Research on data security in the medical and health industry. Chinese Health Service Management, 38(12), 916-917+921.
[14] Chen, J., & Wang, J. (2025). Innovative research on quantum encryption technology in the medical industry. In Chinese Medical Equipment Association. Proceedings of the Chinese Medical Equipment Conference and 2025 Medical Equipment Exhibition. Nanjing City Emergency Center; Affiliated Jiangning Hospital of Nanjing Medical University, 333-339.
[15] HADIAN M, LIANG Xiaohui, ALTUWAIYAN T, et al. Privacy-Preserving Health Data Release with Pattern Consistency [C]//IEEE. GLOBECOM 2016-2016 IEEE Communications Society. New York: IEEE, 2016: 1–6.
[16] PREMA K, SRIHARSHA A. Differential Privacy in Big Data Analytics for Haptic Applications [J]. International Journal of Computer Engineering & Technology, 2017, 8(3): 11-19.
[17] ZHANG Jiajun, LIANG Xiaohui, ZHANG Zhikun, et al. Re-DPoctor: Real-Time Health Data Releasing with W-Day Differential Privacy [C]//IEEE. GLOBECOM 2017-2017 IEEE Global Communications Conference. New York: IEEE, 2017: 1–6.
[18] GUAN Zhitao, LYU Zefang, DU Xiaojiang, et al. Achieving Data Utility-Privacy Trade off in Internet of Medical Things, a Machine Learning Approach [J]. Future Generation Computer Systems, 2019, 98: 60-68.

DOI: 10.54254/2755-2721/2025.PO25408

[19] HAN Song, ZHAO Shuai, LI Qinghua, et al. PPM-HDA: Privacy-Preserving and Multifunctional Health Data Aggregation with Fault Tolerance [J]. IEEE Transactions on Information Forensics and Security, 2016, 11(9): 1940-1955.
[20] UKIL A, JARA A J, MARIN L. Data-Driven Automated Cardiac Health Management with Robust Edge Analytics and Derisking [EB/OL]. (2019-06-18) [2024-12-01].
[21] SALEHEEN N, CHAKRABORTY S, ALI N, et al. mSieve: Differential Behavioral Privacy in Time Series of Mobile Sensor Data [C]//7 The 2016 ACM International Joint Conference. New York: ACM, 2016: 706–717.
[22] STEIL J, HAGESTEDT I, HUANG M X. Privacy Aware Eye Tracking Using Differential Privacy [C]//ACM. The 11th ACM Symposium. New York: ACM, 2019: 1–9.
[23] BOZKIR E, GUNLU O, FUHL W, et al. Differential Privacy for Eye Tracking with Temporal Correlations [EB/OL]. (2021-08-17) [2024-12-01].
[24] ZHANG Siqi, LI Xiaohui. Differential Privacy Medical Data Publishing Method Based on Attribute Correlation [EB/OL]. (2022-09-21) [2024-12-01].
[25] China Cybersecurity Industry Alliance. (2023). China Cybersecurity Industry Analysis Report (2023). https://www.vicsdf.com/doc/80e08981e09c8b0e
[26] SHOR P W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer [J]. SIAM Review, 1999, 41(2): 303-332.
[27] Huang, D. L., & Ma, M. H. (2024). The 5th anniversary of the promulgation of the Cryptography Law: Achievements in the rule of law, implementation challenges, and future directions. China Information Security, (10), 77-82.