Design and Performance of Privacy Protection Model for Big Data Transmission Based on Mixed Encryption

Zhiqiang Chen[1], Zhihua Song[2], Tao Zhang[3,*] and Yong Wei[4]

[1]*Information Department, Zibo Institute of Vocational Education, Zibo, 255000, China*
[2]*Computer Applications, Zibo Electronic Engineering School, Zibo, 256100, China*
[3]*Zibo Education Service Center, Zibo, 255000, China*
[4]*Zibo Education Enrollment Examination Institute, Zibo, 255000, China*
E-mail: zhangtao_zb2024@163.com
[*]*Corresponding Author*

Received 09 April 2024; Accepted 15 October 2024

Abstract

With the advent of the big data era, data security and privacy protection have become particularly important. Big data has advantages such as large scale and diverse types, but it also brings risks of personal privacy leakage and data abuse. However, the symmetric or asymmetric encryption techniques alone have limitations in big data security and privacy protection. Therefore, a privacy protection model for big data transmission based on mixed encryption is proposed and experimentally validated. The research results indicated that the asymmetric encryption algorithm used had an encryption time of less than 20 ms, and the key space occupation was only 0.031 Kb to 0.063 Kb. After improving the symmetric encryption algorithm, it achieved a lower correlation of 0.16 within 18 ms and increased the number of ciphertext transformations to an average of 82 bits. In the performance verification of mixed encryption technology for 60 MB data packets, the proposed mixed encryption technology took 273.1 ms, the decryption took 254.7 ms, the

Journal of Cyber Security and Mobility, Vol. 13_6, 1425–1448.
doi: 10.13052/jcsm2245-1439.1369

correlation was as low as 0.12, and the average resistance time in resisting violent attacks exceeded 100 s. The model proposed in the study improves the encryption and decryption speed while ensuring data security, which has important practical application value and theoretical significance for data privacy protection in big data environments.

Keywords: Big data, symmetric encryption, asymmetric encryption, mixed encryption, privacy protection.

1 Introduction

Influenced by information technology, society has entered the era of big data. The development of big data starts from the wide application of the Internet [1]. Against the backdrop of exponential growth in global data volume, big data has a large scale, fast data transmission speed, and diverse data types, which has extremely high potential application value in various fields of society [2]. However, big data generally contains sensitive information such as personal identity, health, and financial records. Once a leak occurs, information owners will face unpredictable risks. The analysis results of big data will further expose the behavior patterns and preferences of individual users to a certain extent. If improperly used by criminals, personal privacy will be violated [3, 4]. In addition, the high liquidity in cloud computing and Internet environments also increases the risk of data leakage. Therefore, while reasonably developing the big data utilization value, the security and privacy of the data transmission are crucial. Encryption technology for big data refers to encryption methods and strategies specifically designed or adapted to the big data environment, which can satisfy the security requirements of big data storage, processing, and transmission processes [5]. Among them, symmetric encryption uses the same key for both encryption and decryption processes, while asymmetric encryption has a public key and a private key. However, when using symmetric encryption alone, there are issues such as key distribution and key leakage, while when using asymmetric encryption alone, there are drawbacks such as high computational resource consumption and slow speed [6]. In view of this, a mixed encryption technology combining symmetric and asymmetric encryption is proposed, and a privacy protection model for data transmission is proposed on the basis of this technology. The research aims to improve the speed of big data encryption and decryption by mixing symmetric and asymmetric encryption, presenting a more secure and reliable data transmission condition for the era of big data.

It is worth noting that among the existing hybrid encryption schemes, ECC combined with AES is also considered an efficient encryption method. However, ECC requires more computing resources and space for key generation and storage, which is a challenge in resource-constrained environments. In addition, CHACHA20-Poly1305 is a relatively new hybrid encryption scheme, which combines ChaCha20 stream cipher and Poly1305 authentication encryption algorithm to provide good security and performance. However, ChaCha20-Poly1305 presents challenges in key management and compatibility, limiting its use in some application scenarios.

The study contains four parts. The first part reviews relevant research on big data security and privacy protection worldwide. The second part provides a detailed explanation of the mixed encryption technology and privacy protection model proposed in the study. The third part conducts experiment to verify its effectiveness and superiority. The fourth part summarizes the entire article, pointing out its shortcomings and providing suggestions for future research.

2 Related Works

Affected by information technology, big data security and privacy protection are increasingly valued. Personal privacy data protection awareness is increasing. Therefore, many scholars have conducted comprehensive research on big data security and privacy protection. Lo'ai et al. conducted a detailed analysis of existing cloud architectures for data-driven applications in storage, processing, and privacy protection. A Peer-to-Peer (P2P) cloud system was proposed as an extended federated cloud architecture, which solved the big data storage and provided a new method for privacy protection of big data [7]. Awaysheh F M et al. proposed a novel security design framework for the scalability, flexibility, and cost-effectiveness of cloud deployment architecture in big data operations. In the design phase, cloud security domain knowledge was mapped to best practices, ensuring real-world security assembly and addressing vulnerabilities in specific domains [8]. Hu X et al. proposed a dynamic access control strategy based on trust evaluation to address the credibility generated by the interaction between nodes in medical big data cloud systems. The role-based bidirectional selection mechanism, and third-party real-time monitoring mechanism were used for dynamic access control, effectively solving the node trust problem in medical big data cloud systems [9]. Huang L et al. developed a framework based on semi-qualitative methods to address the personal

information leakage and the diversity of privacy protection awareness. This method integrated information extraction techniques such as topic modeling and key sentence extraction, and comprehensively collected user opinions, thereby revealing subjective patterns of information leakage and privacy protection, and providing practical guidance for stakeholders [10].

As a core technology in data security and privacy protection, encryption technology is constantly evolving to meet the growing demand for data processing. Therefore, many scholars have conducted optimization research on encryption technology in different data application fields to improve its efficiency, security, and adaptability. Marqas R B et al. analyzed two commonly used encryption and decryption algorithms, namely the symmetric Advanced Encryption Standard (AES) algorithm and the asymmetric Rivest-Shamir-Adleman (RSA) algorithm, to response the security of resource sharing in data communication networks. The encryption and decryption execution time for messages of different lengths was compared. The conclusion was drawn that the AES exhibited better performance and faster speed than RSA [11]. Wang S et al. designed a privacy protection strategy on the basis of digital watermarking and elliptic curve cryptography asymmetric encryption to address the big data privacy leakage caused by the widespread installation of smart meters. Digital watermarking technology was used to hide sensitive data in collected readings, thereby enabling only authorized users to extract watermarks and decrypt confidential data using a private key [12]. Son L D et al. proposed an improved asymmetric key encryption technology and layered information transmission system architecture based on genetic algorithm to address the improvement needs of asymmetric key encryption and layered information transmission systems in information security. This further enhanced the security of asymmetric encryption algorithms, and ensured processing speed, which had practical application prospects [13]. Ghayvat H et al. developed a blockchain-based confidentiality-privacy protection strategy to response security risks and privacy in the storage and access of big medical data on cloud platforms. This scheme utilized elliptic-curve cryptography to establish secure communication session keys, and then enhanced security through an encryption algorithm authentication framework, thereby improving the security and efficiency of medical data cloud platforms [14].

In summary, in big data security and privacy protection, existing research has covered multiple aspects and made contributions in improving data security, strengthening privacy protection, and optimizing algorithm performance. However, the role of unilateral encryption technology is limited, and there are

problems, such as key leakage or slow speed. Therefore, a mixed encryption technology combining symmetric and asymmetric encryption is proposed and applied to privacy protection in big data transmission. The study aims to improve key security through mixed encryption technology. Then, the research innovatively optimizes key extension and column mixing operations, thereby enhancing the data transmission security and efficiency.

3 Methods and Materials

The hybrid encryption technique proposed in this section combines the high efficiency of the AES algorithm and the high security of the SM2 elliptic curve public key cryptography algorithm. The SM3 hash algorithm is introduced to ensure data integrity. A secure and efficient big data transmission privacy protection model is constructed, suitable for data privacy protection needs in cloud computing environments.

3.1 Asymmetric Encryption Techniques and Hash Functions in Mixed Encryption Technology

To address privacy leaks and other issues encountered during big data transmission, a mixed encryption technology is proposed by combining symmetric and asymmetric encryption techniques, and a privacy protection method for data transmission is constructed. In addition to symmetric and asymmetric encryption techniques, the hash function is used to ensure data integrity and consistency, and to detect whether data has been tampered with during transmission or storage. The study selects the SM2, an elliptic curve public key cryptography method, as the asymmetric encryption technology in mixed encryption. According to the results of Marqas R B et al. in reference [11], the AES symmetric encryption algorithm with better performance is selected as the symmetric encryption technology in mixed encryption. The SM2, as recommended by the National Cryptography Administration of China, has a shorter key length, higher security computational efficiency than general asymmetric encryption algorithms, and strong practicality [15]. The AES symmetric encryption algorithm also has fast speed and high flexibility [16]. For the SM2 algorithm, the affine coordinate diagram of its elliptic curve equation is shown in Figure 1.

In Figure 1, the study first assumes that the elliptic curve of the SM2 is shown in Equation (1).

$$y^2 = x^3 - x \tag{1}$$

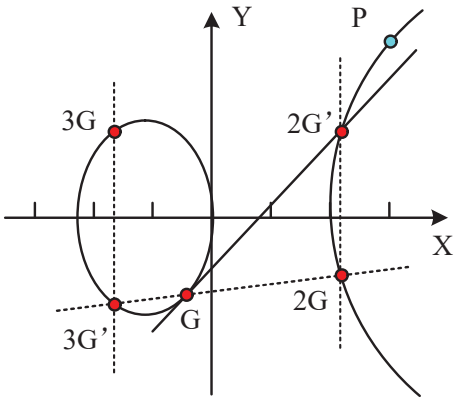


Figure 1 Affine coordinate diagram of elliptic curve equation.

On this elliptic curve, a base point G on the elliptic curve is selected or generated, whose coordinates satisfy the elliptic curve equation. The order of point G is a higher prime number that is difficult to factorize. That is, when the base point G is continuously added to itself on the elliptic curve, after a certain number of addition operations, a theoretical point is finally obtained, which is called an infinite point. This number is the order of the G. Higher prime orders can increase the difficulty of solving elliptic curve discrete logarithm problems, thereby providing stronger resistance to cracking.

After selecting point G, draw a line tangent to the ellipse at point G that intersects with the curve at point 2G', and then draw a line perpendicular to the X-axis at point 2G' that intersects with the curve at point 2G. At this point, connect point 2G to point G and intersect with the ellipse at point 3G'. Draw a straight line perpendicular to the X-axis and intersect with the ellipse at the 3G point. In the above process, 2G is defined as twice the base point G, while 3G point is defined as the three times the base point G. The above operation is repeated to obtain the d-fold point point of the base point G. At this point, the d-fold point is defined as the P point. Ultimately, the key pairs generated by the SM2 algorithm include the private key d and the public key P. Among them, the d is kept confidential by the user, while the public key P can be made public for encrypting messages or verifying digital signatures, and only individuals who own the private key d can deduce d times point of base point G from the public key point P. On this basis, Figure 2 displays the SM2 encryption process.

In Figure 2, after obtaining the user's original information, the SM2 algorithm first generates a temporary random integer. This integer is used

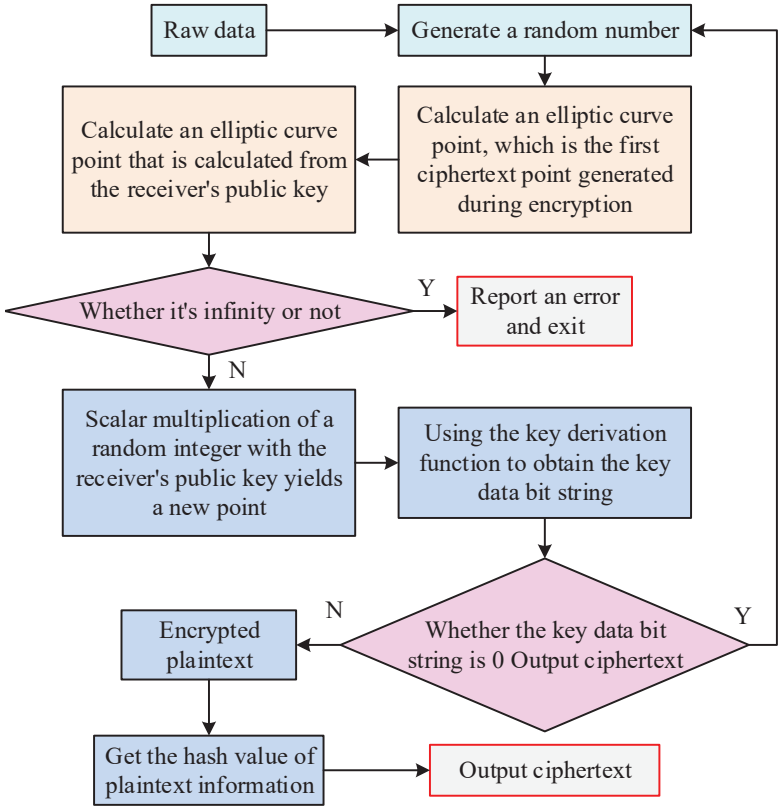


Figure 2 The encryption process of SM2.

for a one-time key to ensure that even if the similar plaintext is encrypted multiple times, the ciphertext will still be various. Subsequently, the first ciphertext point C_1 generated during the encryption process is solved. It is obtained by performing scalar multiplication on the randomly generated integer and the base point G. The coordinates of this ciphertext point are converted into bit string form for subsequent recovery of the private key during the decryption process. Further, based on the recipient's public key, another point S is calculated to generate the key data bit string during the decryption process. When S is an infinite point, an error message will occur and the entire encryption process will stop [17]. Then, scalar multiplication is performed on the random number and the recipient's public key to obtain point Q. The coordinates of point Q are also converted into a bit string and used in the key derivation function to generate the session key. The derived

key data bit string is used to encrypt the plaintext and generate ciphertext Q. The hash value of the plaintext information is solved to obtain C_3, and finally output the ciphertext information. The calculation in the above process is shown in Equation (2).

$$\begin{cases} k \in [1, n-1] \\ C_1 = [k]G \\ S = [h]P \\ Q = [k]P = (x_2, y_2) \\ t = DF(x_2 \| y_2, klen) \\ C_2 = M \oplus t \\ C_3 = \mathrm{SM3}(x_2 \| M \| y_2) \\ C = C_1 \| C_2 \| C_3 \end{cases} \tag{2}$$

In Equation (2), n represents the order of the G. h represents the hash function. P represents the public key of the decryptor. (x_2, y_2) is the coordinate of Q. t represents the key data bit string. DF represents the key derived function. M represents plaintext information. $klen$ represents the length of t. $\mathrm{SM3}$ represents the hash algorithm SM3 independently developed by China. C represents the final generated ciphertext information. Figure 3 displays the decryption process of SM2.

In Figure 3, the first extracts the C_1 from the obtained original ciphertext and verify whether C_1 is an effective point on the elliptic curve equation. If not, an error will occur and the decryption process will exit. On the contrary, the process proceeds to the next step. The next step is to use the hash function to calculate S, which is consistent with the encryption process. During the decryption, whether S is an infinite point is determined. If it is, an error message will be reported and the decryption process will exit. If not, the private key operation will be performed. When performing private key operations, the coordinates are calculated using C_1 and the private key, and the key data bit string is derived from the obtained coordinates. At this point, if all key data bit strings are 0, an error message may also occur. The final derived key data bit string is applied to decrypt the ciphertext and recover plaintext information. To verify the integrity and accuracy of the data, the hash value of the decrypted plaintext is compared with the hash value generated and transmitted during encryption. If both are the same, plaintext information can be output. On the contrary, it indicates that data information have been tampered with or invaded. In this case, the information can be

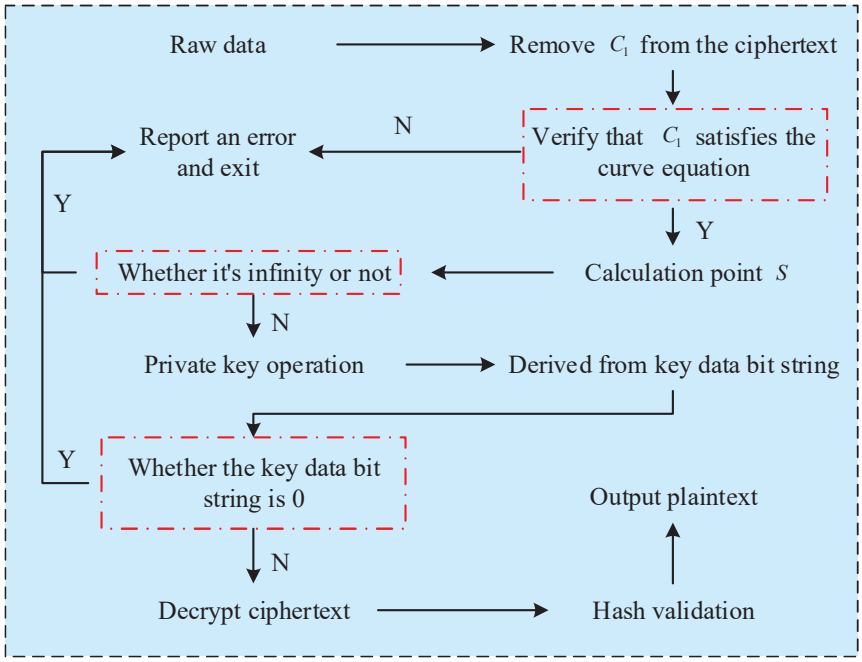


Figure 3 The decryption process of SM2.

refused. During the SM2 encryption and decryption, the hash algorithm SM3 is used. The steps are shown in Figure 4.

As shown in Figure 4, the SM3 algorithm first initializes registers, then fills and compresses the input message, and finally divides the message into 64 byte groups. If the message length is not an integer multiple of 64 bytes, it needs to be extended. Each set of data is further applied with an extension function. This process is repeated until all data blocks are processed, ultimately generating a 256 bit hash value to ensure data integrity and security. The high security of the SM3 algorithm mainly benefits from its unidirectionality, which means that the original data can not be inferred from the hash value. Meanwhile, the SM3 algorithm is extremely sensitive to input data. Any small changes can lead to significant differences in hash values, greatly increasing the difficulty of data tampering. In addition, the SM3 algorithm effectively avoids hash conflicts in its design, ensuring that the probability of generating the same hash value from different data remains extremely low. Compared with the secure hash algorithm 256, SM3 is designed with certain quantum attack resistance in mind and may have

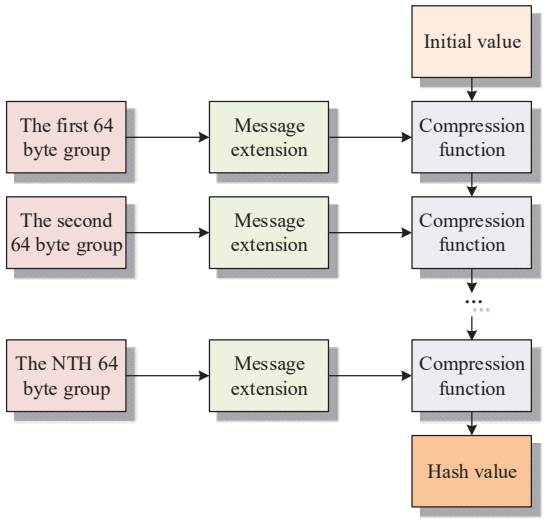


Figure 4 Steps of the hash algorithm SM3.

better performance on certain hardware platforms, which is more in line with Chinese cryptographic standards.

Finally, in order to improve the scalability of SM2 algorithm in large-scale data environment, a distributed network consisting of multiple compute nodes distributed on different physical or virtual servers is constructed to achieve load balancing and failover. In this paper, a load balancer is implemented to dynamically distribute encryption tasks to lighter nodes, and parallel processing is allowed by data sharding technology to improve encryption efficiency. In addition, asynchronous communication mechanisms reduce communication latency between nodes, while fault-tolerant and data recovery strategies ensure system stability and data security even in the case of partial node failures.

3.2 Symmetric Encryption Technology and Fusion in Mixed Encryption Technology

When choosing symmetric encryption techniques in mixed encryption, the research should focus on the AES algorithm, which is known for its high security, flexibility, and encryption efficiency. When using AES algorithm for encryption, the plaintext is first divided into fixed length blocks of 128 bits, and any gaps are filled in at the end. Subsequently, the key sequence is

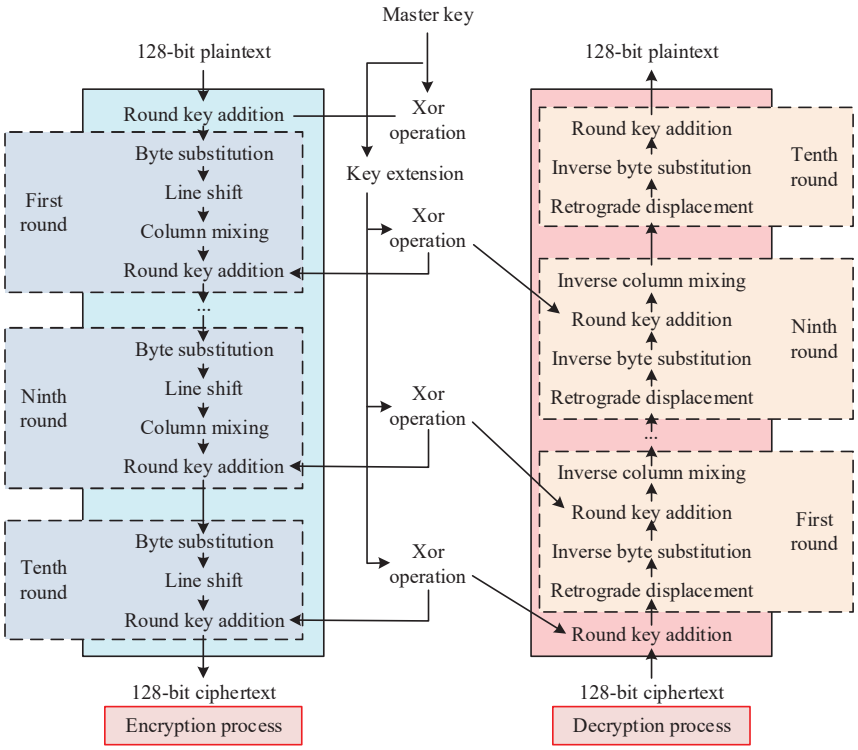


Figure 5 Encryption and decryption process of AES.

transformed using key extension technology, and the exclusive OR operation is performed with plaintext [18]. The data undergoes 10 rounds of encryption processing and finally outputs the encrypted ciphertext. The encryption and decryption process of AES is displayed in Figure 5.

In Figure 5, the decryption process of the AES is the reverse execution of its encryption steps. There are four basic operations involved in the encryption and decryption process. Among them, for byte substitution operations, the AES algorithm uses a specific lookup table (S-box) to replace each byte. S-box is used for encryption, and its inverse execution (S-1 box) is used for decryption. The top 4 bits and bottom 4 bits of each byte are used as row and column indexes in the S-box, respectively, to find the corresponding replacement byte. When performing a row shift operation, the rows of the matrix will be cyclically shifted based on their row numbers. When encrypting, the 0th line does not move, and the 1st line loops 1 bit. Subsequent encryption

follows this pattern. When decrypting, the opposite right loop shift is performed. The column mixing operation introduces nonlinear transformations through matrix multiplication. A fixed reversible matrix is used to multiply with the matrix of the current state, increasing data obfuscation. For the round key addition operation, in each encryption round, the state matrix is XOR operated with the current round key to mix key and data. When decrypting, due to the characteristics of XOR operation, if the same key is XOR operated again, this step can be revoked.

On this basis, to further enhance the security and decryption efficiency of the AES, two optimizations are performed for the AES. During the key extension process of the original AES, if an attacker knows a certain round of keys, it is possible to deduce the entire key sequence in reverse. Therefore, the study introduces a brand new key that is independent of the initial key during the first round of key expansion. Even if an attacker cracks a key in one round, they cannot deduce keys from other rounds, thereby improving the key security. In the AES algorithm, the column mixing operation is implemented by matrix multiplication of the data, which is to scramble the bytes in the data and increase the encryption strength. However, this operation brings additional challenges in decryption, as it requires the inverse of the original matrix to recover the original data, which typically means more complex computational steps. In order to simplify the process and increase the decryption speed, the matrix is specially designed. The matrix chosen for research has the same properties as its inverse matrix itself. This means that when decrypting, the research can operate directly with the same matrix used in encryption, without the need to compute a completely new inverse matrix. Therefore, the privacy protection model for big data transmission on the basis of mixed encryption proposed in the study is shown in Figure 6.

As shown in Figure 6, the encryption process of the model starts with the client. Firstly, the SM3 hash algorithm is used to calculate a hash value for the original private data plaintext. Then, this hash value is combined with the plaintext to form a new data string. Next, the client generates a pair of symmetric keys, and uses an improved AES algorithm to encrypt the new data string, generating ciphertext. To securely transmit these symmetric keys, the model further utilizes the SM2 public key provided by the cloud on the client side to perform asymmetric encryption on the symmetric keys and generate key ciphertext. Finally, the client combines the ciphertext and key ciphertext and sends them to the cloud.

At the decryption end, i.e. in the cloud, the key ciphertext is first extracted from the received ciphertext and key ciphertext, and the SM2 private key

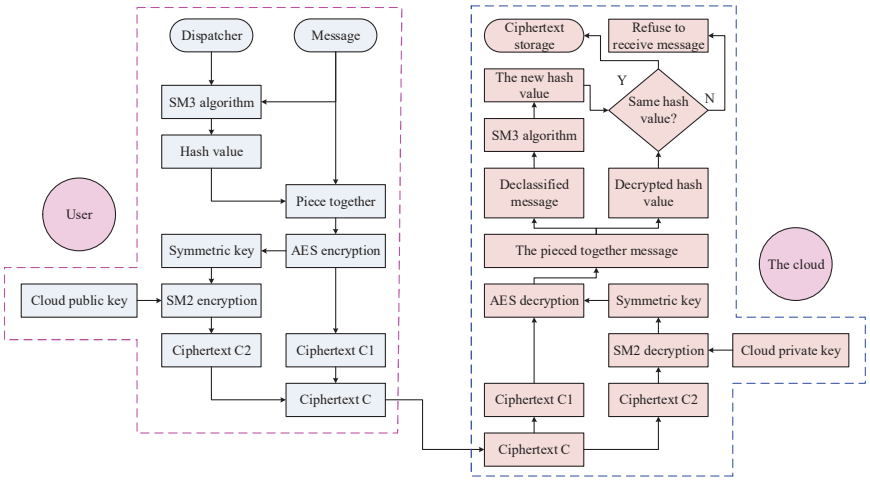


Figure 6 Overall process of big data transmission privacy protection model.

in the cloud decrypts it and successfully restores the original symmetric key. Subsequently, the model uses a symmetric key to perform symmetric decryption on the ciphertext in the cloud, obtaining the original data string. To demonstrate the data integrity during transmission, the model extracts the plaintext again in the cloud and calculates its new hash value using the SM3 algorithm. Finally, the model compares the newly calculated hash value with the initially calculated hash value in the cloud. If the two match, it can be confirmed that the data has not been tampered with, and the cloud will accept and securely store the ciphertext. If inconsistent, storage is refused and it indicates that the data may have been damaged or tampered with during transmission. The study combines efficient symmetric encryption with secure key exchange for asymmetric encryption, as well as hash verification for data integrity, providing a secure and efficient solution for privacy protection in big data transmission in cloud computing. Combining the above, the model proposed in the study not only achieves efficient data transmission privacy protection, but also adapts to the encryption needs of large datasets.

4 Results

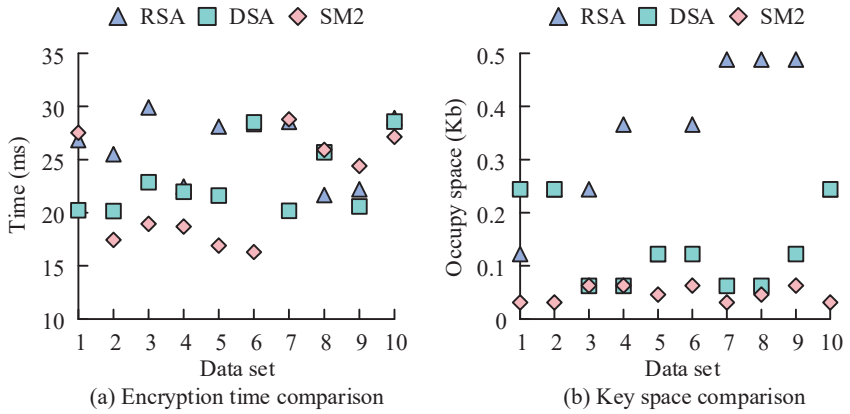
To demonstrate the effectiveness and superiority of the proposed model, a protection system is constructed based on this model. The model performance is discussed through experiments in the system. The specific system configuration is displayed in Table 1.

Table 1 Protection system configuration

Category	Components	Details
Hardware Composition	Cloud Servers	Alibaba Cloud ECS
Software Composition	Operating System	Windows 7
	Database System	MySQL
	Programming Language & Tools	Java, JDK1.8, Maven Framework
	Integrated Development Environment	MyEclipse2017CI
	Database Interaction Tool	JPBC2.0
System Configuration	User Login Module	Identity Authentication, Registration
	File Upload Module	Supports Encrypted Uploads
	File Download Module	Supports Decrypted Downloads
	Computing Module	Data Processing, Secure Computation Results
	File Search Module	Lucene-based, Supports Filename Search
	Data Management Module	Manages Databases, Index Libraries, Cloud Services, and Cloud Storage Services

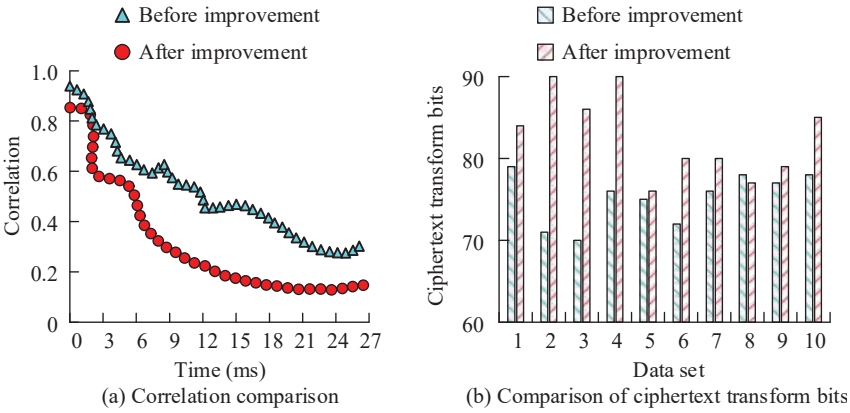
Based on the protection system, the study first checks the performance of the SM2 algorithm. Asymmetric encryption techniques, including RSA, and Digital Signature Algorithm (DSA), are used as comparison algorithms. Firstly, the encryption time and key occupancy space of the three algorithms are tested. Due to the small size of the selected encrypted data, the study compares the final encryption time by cycling the encryption process 200 times. The specific results are shown in Figure 7.

In Figure 7, a total of 10 groups of data are taken as encryption testing in the protection system. As shown in Figure 7(a), the encryption time of RSA, DSA, and SM2 were all in the millisecond range. Among them, the encryption time of RSA and DSA was within 20–30 ms, while the SM2 was sometimes less than 20 ms. From Figure 7(b), the SM2 algorithm had a smaller key occupancy space, with only three spans, namely 0.031 Kb, 0.046 Kb, and 0.063 Kb, corresponding to 256-bit keys, 384-bit keys, and 521-bit keys, while the RSA and DSA had generally larger key occupancy space. Therefore, considering all factors, choosing the SM2 algorithm as the proposed mixed encryption technology has more advantages. Furthermore, the improved AES algorithm is performed performance testing. Firstly, the



(a) Encryption time comparison (b) Key space comparison

Figure 7 Comparison of encryption time and key occupied space.



(a) Correlation comparison (b) Comparison of ciphertext transform bits

Figure 8 Comparison of diffusion and confusion of AES algorithm.

diffusion and confusion of the AES algorithm before and after improvement are compared, as shown in Figure 8.

From Figure 8(a), before AES algorithm improvement, the correlation between plaintext and ciphertext reached its lowest value of 0.34 after 22 ms. After improving the AES algorithm, the correlation between plaintext and ciphertext reached its lowest value after 18 ms, with only 0.16. The low correlation indicates better algorithm diffusion. As shown in Figure 8(b), before the improvement, the average number of ciphertext transformation bits for the AES algorithm in 10 rounds of encryption was 75 bits. After improvement, the average number of ciphertext transformation bits for the

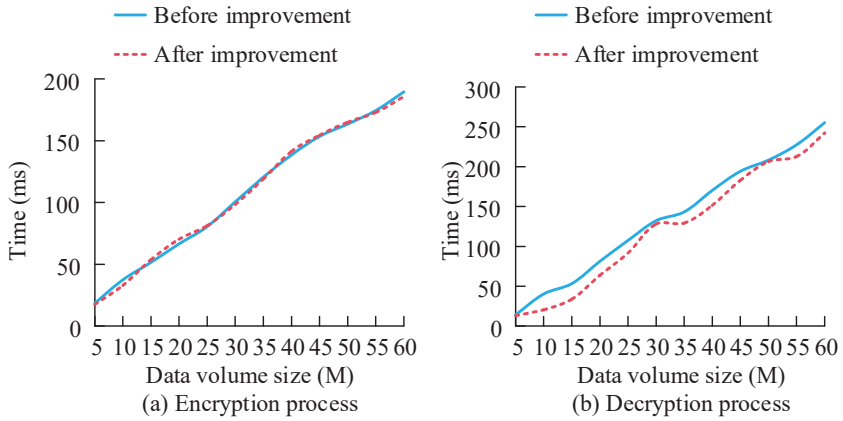


Figure 9 Comparison of encryption and decryption time before and after AES algorithm improvement.

Table 2 Verification results of mixed encryption technology

Method	Encryption Process (ms)	Decryption Process (ms)	Correlation
RSA+Primary AES	294.44	289.37	0.45
DSA+Primary AES	305.25	302.51	0.39
SM2+Primary AES	277.27	281.23	0.23
RSA+Improved AES	284.91	264.56	0.21
DSA+Improved AES	289.53	268.44	0.22
SM2+Improved AES	273.14	254.79	0.12

AES in 10 times of encryption was 82. This indicates that the confusion of the improved AES is also enhanced, thus improving the overall security of the AES algorithm. Furthermore, the encryption and decryption time of AES before and after improvement is shown in Figure 9.

From Figure 9(a), the time spent in the encryption process before and after the improvement of AES was not significantly different, with a maximum difference of only 3 ms. This indicates that the improved key extension and improved column mixing operation have not affected the original computational load. In Figure 9(b), during the decryption process, the improved AES algorithm took significantly less time than the unimproved AES algorithm. Therefore, the improved column mixing operation proposed in the study can effectively reduce the computational complexity and decryption time. On this basis, the performance of mixed encryption technology is verified based on a 60MB data packet, as displayed in Table 2.

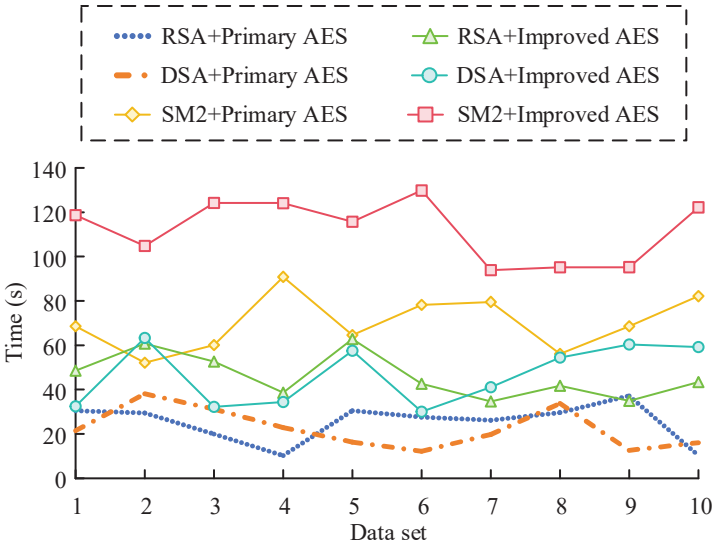


Figure 10 Model anti-attack test.

According to Table 2, the mixed encryption technology proposed in the study had the best performance, with an encryption time of 273.14 ms and a decryption time of 254.7 ms, and the lowest correlation of 0.12. This indicates that the designed method has the best security and efficiency compared with other mixed encryption technologies. Finally, the study verifies the anti-attack capability of the privacy protection model through violent attacks, including differential attacks on the model. The results are shown in Figure 10.

As shown in Figure 10, faced with violent attacks, the mixed encryption technology combining SM2 with improved AES proposed in the study had the longest resistance time, averaging over 100 s, and reaching 114.6 s. The worst performing mixed encryption algorithm was DSA combined with improved AES, with an average resistance time of no more than 40 s, only at 33.8 s. This indicates that the mixed encryption technology proposed in the study can more effectively resist violent attacks during actual deployment. Finally, in order to evaluate the impact of the improved AES algorithm on compute load, power consumption, and storage requirements in terms of key extension and column mix optimization, the original AES algorithm is used to encrypt and decrypt the selected dataset, and record compute load, power consumption, and storage requirements. The above tests are repeated with the improved AES algorithm. The results are shown in Table 3.

Table 3 Comparative experimental results

Dataset Size (KB)	Algorithm	CPU Usage (%)	Power Consumption (W)	Storage Requirement (KB)
1	Original AES	5.2	45	128
	Improved AES	5.5	46	132
1024	Original AES	12.3	60	128
	Improved AES	13.1	61	132
10240	Original AES	35.6	90	128
	Improved AES	37.4	92	132

From Table 3, although the improved AES algorithm had slightly higher CPU usage and power consumption than the original AES algorithm, this increase was small and remained consistent across all tested data set sizes. The additional 4 KB storage space required by the improved algorithm indicated that the key extension optimization had less impact on the storage requirements. Overall, the improved AES algorithm introduces only acceptable additional resource consumption while providing better security, suggesting that the improved algorithm is a valid choice when weighing security against performance.

5 Discussion and Conclusion

A mixed encryption technology was proposed to address privacy protection issues in big data transmission, which combined the advantages of symmetric and asymmetric encryption, and enhanced data integrity protection through hash algorithms. The experimental results showed that SM2 had significant advantages in key length and encryption efficiency, with an encryption time of less than 20 ms and a key space occupation of only 0.031 Kb to 0.063 Kb. Further performance testing was carried out the improved AES algorithm. After 18 ms, the correlation between plaintext and ciphertext reached its lowest value, with only 0.16. The average number of ciphertext transformation bits was 82, and the diffusion and confusion were enhanced. In the performance verification of mixed encryption technology for 60 MB data packets, the encryption time of SM2 combined with the improved AES algorithm was 273.1 ms, the decryption time was 254.7 ms, and the correlation was as low as 0.12, which was better than other comparison schemes. In addition, anti-attack tests showed that the mixed encryption technology had an average resistance time of over 100 s in resisting violent attacks, reaching 114.6 s,

which was much higher than the 33.8 s achieved by combining DSA with the unimproved AES algorithm. Therefore, the privacy protection model based on mixed encryption technology proposed in the study performs well in terms of security and efficiency, which can effectively resist violent attacks. It is suitable for the secure transmission in cloud computing environments.

Aiming at the actual deployment of the hybrid encryption model in large-scale cloud environment, the modular design of the model can effectively support distributed processing and concurrent scaling. By introducing load balancing and asynchronous communication mechanisms, the model can dynamically distribute encryption tasks and improve processing efficiency, especially when dealing with a large number of clients and different workloads. In addition, combined with data sharding and fault tolerance strategies, the model can maintain stable performance in resource-constrained environments. However, to further enhance the adaptability of cloud architectures, future deployments may require optimizing key management systems and exploring the application of hardware accelerators to ensure efficient operation in high concurrency scenarios. When dealing with noise or incomplete data packets in real data transmission, the model ensures data integrity through hash verification mechanism and has strong recovery ability. Combined with data redundancy and error correction techniques such as Reed-Solomon error correction codes, the model effectively addresses packet loss, latency, and data corruption. In addition, data fragmentation processing and redundancy check further enhance its robustness against network fluctuations, ensuring data security and consistency during transmission. Finally, although the model performs well in resisting brute force attacks, with the development of quantum computing technology, existing SM2 and AES encryption techniques may face new security threats. To this end, future research should consider combining post-quantum encryption algorithms, such as lattice-based encryption, to deal with potential attacks from quantum computing. The combination of dynamic key update mechanism and hardware security module will also further enhance the attack resistance of the model and ensure its security in the quantum era.

6 Fundings

The research is supported by Shandong Provincial Institute of Vocational Education and Industrial Talent, “Research on the Construction of Regional Industry-Education Integration Supply and Demand Service Platform” (No. 2023ZX033).

References

[1] Widiasanti I, Zanuara A A F, Maulidina D F, Atmaja F A R, Putri S R. Implementation of Big Data in the Zoom and Google Classroom Applications as Online Learning Media. Indonesian Journal of Education and Mathematical Science, 2023, 4(2): 86–92.

[2] Srihith I D, Donald A D, Srinivas T A S, Thippanna G, Anjali D. Locking down big data: a comprehensive survey of data encryption methods. International Journal of Advanced Research in Science, Communication, and Technology, 2023, 3(2): 84–93.

[3] Guan S, Zhang C, Wang Y, Liu, W. Hadoop-based secure storage solution for big data in cloud computing environment. Digital Communications and Networks, 2024, 10(1): 227–236.

[4] Saravanan S, Poornima N. Big Data analytics for privacy through ND-homomorphic encryption. Journal of Control and Decision, 2023, 10(1): 64–71.

[5] Chelladurai S P, Rajagopalan T. Intelligent Digital Envelope for Distributed Cloud-Based Big Data Security. Comput. Syst. Sci. Eng., 2023, 46(1): 951–960.

[6] Ajala O A, Arinze C A, Ofodile O C, Okoye C C, Daraojimba O D. Reviewing advancements in privacy-enhancing technologies for big data analytics in an era of increased surveillance. World Journal of Advanced Engineering Technology and Sciences, 2024, 11(1): 294–300.

[7] Lo'ai A T, Saldamli G. Reconsidering big data security and privacy in cloud and mobile cloud systems. Journal of King Saud University-Computer and Information Sciences, 2021, 33(7): 810–819.

[8] Awaysheh F M, Aladwan M N, Alazab M, Alawadi S, Cabaleiro J C, Pena T F. Security by design for big data frameworks over cloud computing. IEEE Transactions on Engineering Management, 2021, 69(6): 3676–3693.

[9] Hu X, Jiang R, Shi M, Shang J. A privacy protection model for health care big data based on trust evaluation access control in cloud service environment. Journal of Intelligent & Fuzzy Systems, 2020, 38(3): 3167–3178.

[10] Huang L, Zhou J, Lin J, Deng S. View analysis of personal information leakage and privacy protection in big data era—based on Q method. Aslib Journal of Information Management, 2022, 74(5): 901–927.

[11] Marqas R B, Almufti S M, Ihsan R R. Comparing Symmetric and Asymmetric cryptography in message encryption and decryption by using AES and RSA algorithms. Xi'an Jianzhu Keji Daxue Xuebao/

Journal of Xi'an University of Architecture & Technology, 2020, 12(3): 3110–3116.
[12] Wang S, Chen H, Zhao Q, Q. Y., Guo L Y, Deng X Y, Si W G, Sun Z Q. Preserving scheme for user's confidential information in smart grid based on digital watermark and asymmetric encryption. Journal of Central South University, 2022, 29(2): 726–740.
[13] Son L D, An T V, Thuy N N. Improving the asymmetric encryption algorithm based on genetic algorithm, application in online information transmission. International Journal of Electronic Security and Digital Forensics, 2021, 13(6): 612–629.
[14] Ghayvat H, Pandya S, Bhattacharya P, Zuhair M, Rashid M, Hakak S, Dev K. CP-BDHCA: Blockchain-based Confidentiality-Privacy preserving Big Data scheme for healthcare clouds and applications. IEEE Journal of Biomedical and Health Informatics, 2021, 26(5): 1937–1948.
[15] Atadoga A, Farayola O A, Ayinla B S, Amoo O O, Abrahams T O. A comparative review of data encryption methods in the USA and Europe. Computer Science & IT Research Journal, 2024, 5(2): 447–460.
[16] Hebbi C, Mamatha H. Comprehensive Dataset Building and Recognition of Isolated Handwritten Kannada Characters Using Machine Learning Models. Artificial Intelligence and Applications, 2023, 1(3): 179–190.
[17] Widiasanti I, Puteri A A A, Savitri D, Anjani D, Nathania N. Big Data in E-Learning: Review Study on Google Classroom Usage and Big Data Challenges. Scripta: English Department Journal, 2023, 10(1): 121–127.
[18] Fan Y, Zhang W, Bai J, Lei X, Li K. Privacy-preserving deep learning on big data in cloud. China Communications, 2023, 20(11): 176–186.

Biographies

Zhiqiang Chen, Male, May 1984, Zibo, Shandong, Han nationality. In 2010, he obtained a bachelor's degree in computer science and technology from Shandong Normal University.

Work experience: From 2010 to 2015, he was a teacher at the Second Primary School of Aluminum City, Zhangdian District, Zibo City. From 2015 to present, he has been the section chief and senior engineer of Zibo Institute of Vocational Education.

He has published 3 academic papers, 2 academic textbooks, participated in 7 scientific research projects, and obtained 4 software copyrights.

Zhihua Song, Female, March 1992, Zibo, Shandong, Han. In 2016, obtained a bachelor's degree in computer science and technology from Shandong University of Traditional Chinese Medicine. From September 2023 to present, studying for a graduate degree in public administration at Liaoning Normal University.

Work experience: From 2018 to present, she was a teacher at Zibo Electronic Engineering School.

Tao Zhang, Male, April 1983, Zibo, Shandong, Han. In 2006, obtained a bachelor's degree in Electronic Information Engineering from Weifang University. In 2011, obtained a master's degree in computer technology from Shandong University of Technology.

Work experience: From 2011 to 2017, served as technical director at Zibo Zhanggang Co., Ltd. of Shandong Iron and Steel Group. From 2017 to present, served as a senior engineer at Zibo Education Service Center.

He has Published 5 academic papers, published 1 academic work and textbook, participated in 1 research project, and obtained 1 patent.

Yong Wei, Male, May 1990, Heze, Shandong, Han. In 2014, obtained a bachelor's degree in computer science and technology from Linyi University.

Work experience: From 2014 to 2016, served as a staff member in the maintenance section of Linzi Branch of Zibo Highway Administration Bureau. From 2016 to present, served as deputy chief of the general college entrance examination section of Zibo Education Enrollment Examination Institute.