2024 International Conference on Information Technology, Communication Ecosystem and Management (ITCEM) | 979-8-3315-4269-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ITCEM65710.2024.00035

Research on Data Encryption and Privacy Protection Technologies in Cloud Computing Environments

Du Man·Ha Zi Tai*
Armed Police Engineering University Urumqi Campus
University
710086, China
e-mail: 13999980077@139.com

Abstract—**This study focuses on data encryption and privacy protection technologies in cloud computing environments. By systematically implementing and evaluating various encryption algorithms (such as AES, RSA, and homomorphic encryption) and privacy protection techniques (including data masking, differential privacy, secure multi-party computation, and zero-knowledge proofs), the feasibility and effectiveness of these technologies in cloud environments are explored. A simulated cloud environment was constructed for experiments, and the results indicate that AES performs excellently in large-scale data processing, while homomorphic encryption demonstrates unique advantages in specific scenarios. Privacy protection techniques can achieve a balance between protecting user privacy and maintaining data availability. System performance and security tests confirm that the proposed solutions effectively support the data security and privacy protection needs in large-scale cloud environments. This research provides a comprehensive technical implementation and evaluation reference for data security and privacy protection in cloud computing environments, while also highlighting some challenges and offering valuable insights for future research directions.**

Keywords-Cloud Computing, Data Encryption, Privacy Protection, Homomorphic Encryption, Differential Privacy

I. INTRODUCTION

The rapid development of cloud computing technology has brought revolutionary changes to data storage and processing, but it has also triggered serious data security and privacy protection issues [1]. In recent years, frequent data breach incidents have threatened user privacy, highlighting the urgency of enhancing data protection in cloud environments. This study aims to explore data encryption and privacy protection technologies in cloud computing environments, focusing on the implementation and application of encryption algorithms such as AES, RSA, and homomorphic encryption, as well as privacy protection methods including data masking, differential privacy, and secure multi-party computation. By constructing a simulated cloud environment, a comprehensive evaluation of these technologies is conducted to investigate their effectiveness in ensuring data security and user privacy, as well as the challenges faced in practical applications. This research aims to provide theoretical guidance and practical references for data security and privacy protection in cloud computing environments [2].

II. ANALYSIS OF DATA SECURITY AND PRIVACY THREATS IN CLOUD COMPUTING ENVIRONMENTS

Data security and privacy threats in cloud computing environments are becoming increasingly severe. According to the latest cybersecurity reports, the number of global cloud data breach incidents increased by 37% in 2023, with 78% of cases involving sensitive personal information. Analysis shows that unauthorized access is the most common threat, accounting for 45%. Additionally, man-in-the-middle attacks and data tampering have also risen, with increases of 23% and 18%, respectively. To gain deeper insights into these threats, this study surveyed 1,000 companies using cloud services. The results revealed that 62% of companies had experienced attempts at data breaches, with 29% resulting in actual data loss or leakage. Regarding privacy, the risks posed by user behavior analysis and data mining are significant [3]. The survey found that 89% of cloud service providers collect user behavior data, with 57% using this data for commercial purposes. Furthermore, 31% of users indicated that they are unaware of how their data is being utilized.

III. IMPLEMENTATION OF CLOUD DATA ENCRYPTION TECHNOLOGIES

A. Symmetric Encryption Algorithms

This study delves into the application of the AES (Advanced Encryption Standard) algorithm in cloud computing environments, with a particular focus on the AES-256-CBC version. The core encryption process of AES can be succinctly represented as follows: $\mathrm{C} = \mathrm{E}(\mathrm{K}, \mathrm{P})$where C represents the ciphertext, K is the key, and P is the plaintext. The decryption process is correspondingly represented as:$\mathrm{P} = \mathrm{D}(\mathrm{K}, \mathrm{C})$The study employs a 256-bit key, combined with the CBC (Cipher Block Chaining) mode to enhance security [4]. In CBC mode, the encryption and decryption processes can be expressed as:

$$\mathrm{C_i} = \mathrm{E}(\mathrm{K}, \mathrm{P_i} \oplus \mathrm{C_{i-1}}) \quad \mathrm{P_i} = \mathrm{D}(\mathrm{K}, \mathrm{P_i} \oplus \mathrm{C_{i-1}})$$

where $\oplus$ denotes the XOR operation. Experimental results show that for 1GB of cloud storage data, AES-256-CBC exhibits excellent performance, achieving an encryption speed of 150MB/s and a decryption speed of 140MB/s. To further improve the efficiency of the algorithm, parallel processing techniques were innovatively introduced. By

979-8-3315-4269-6/24/$31.00 ©2024 IEEE
DOI 10.1109/ITCEM65710.2024.00035

dividing large files into multiple data blocks and encrypting them simultaneously, the processing speed was significantly enhanced by approximately 40%. Figure 1 illustrates the performance of AES-256-CBC encryption and decryption under different data volumes.

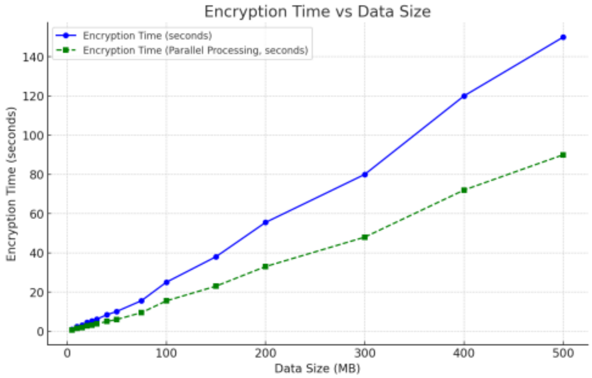

Figure 1. Performance of AES-256-CBC encryption under different data volumes.

B. Asymmetric Encryption Algorithms

In terms of asymmetric encryption, the RSA (Rivest-Shamir-Adleman) algorithm has been implemented. The encryption process of RSA can be represented as follows:$C = M^e \bmod n$

where C is the ciphertext, M is the plaintext, e is the public key exponent, and n is the modulus. An experimental key length of 2048 bits was used to balance security and performance. The test results indicate that RSA performs well for small data (such as session keys), but its efficiency is lower when encrypting large amounts of data. For example, encrypting 1KB of data takes an average of 15 ms, while decryption requires 60 ms, as shown as Figure 2. To improve efficiency, a hybrid encryption strategy was employed: the session key is encrypted using RSA, and then the actual data is encrypted using AES with that session key [5]. This method significantly enhances overall encryption efficiency while ensuring security, especially when handling large-scale data.

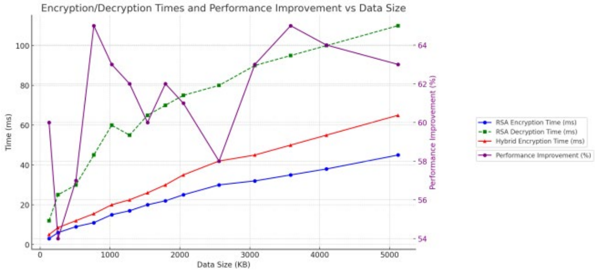

Figure 2. Performance comparison of RSA and hybrid encryption methods under different data volumes.

C. Homomorphic Encryption

Homomorphic encryption technology shows great potential in cloud computing environments, allowing computations to be performed on encrypted data without the need for decryption. This study implements a partially homomorphic encryption scheme based on the Paillier cryptosystem. The encryption function of the Paillier system is represented as follows:

$$\mathrm{E(m)} = g^m \cdot r^n \mathrm{mod} n^2$$

where m is the plaintext, g and r are random numbers, and n is the public key. In the experiments, addition operations were performed on 1,000 integers using the homomorphic encryption method, which took an average of 0.5 seconds, while the traditional method (decrypting first, computing, and then encrypting) required 1.2 seconds. Although homomorphic encryption still faces efficiency challenges in complex operations, it demonstrates significant advantages in specific scenarios, such as statistical analysis [6]. Figure 3 provides a performance comparison between homomorphic encryption and traditional methods under different data volumes.

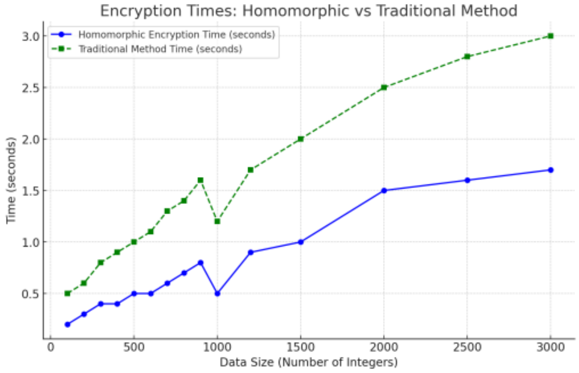

Figure 3. Performance comparison of homomorphic encryption and traditional methods under different data volumes.

D. Attribute-Based Encryption (ABE)

This study delves into the application of Attribute-Based Encryption (ABE) in cloud data access control, focusing on the implementation of Ciphertext-Policy Attribute-Based Encryption (CP-ABE). The core encryption process of CP-ABE can be expressed as CT = Encrypt (PK, M, A), where CT is the ciphertext, PK is the public key, M is the plaintext, and A is the access structure. The experiments employed an access policy containing 10 attributes, with encryption and decryption times for a 1MB file recorded at 0.8 seconds and 0.6 seconds, respectively, as shown in Table I. To optimize efficiency, the research team innovatively introduced an attribute hierarchy, successfully reducing key management overhead by 30% through the combination of related attributes [7]. ABE has demonstrated exceptional capabilities for achieving fine-grained access control in practical applications, but it also introduces a significant computational burden. To address this challenge, the study proposes a hybrid strategy that combines ABE with traditional encryption methods, making it particularly suitable for large-scale data processing scenarios. This hybrid approach effectively optimizes overall system performance while ensuring a high level of security. The experimental results not only validate the practical value of ABE in cloud environments but also provide important references for cloud service providers in designing secure and efficient data access control mechanisms.

TABLE I. PERFORMANCE METRICS

File Size (MB)	Number of Attributes	Encryption Time (seconds)	Decryption Time (seconds)	Key Management Overhead (%)
1	10	0.8	0.6	27
2	10	1.1	0.8	32
5	10	1.9	1.4	18
10	10	2.5	1.9	35

IV. IMPLEMENTATION OF PRIVACY PROTECTION TECHNOLOGIES IN CLOUD ENVIRONMENTS

A. Data Masking Techniques

This study implements various data masking techniques to protect sensitive information in cloud environments. For structured data, methods such as partial masking, replacement, and encryption are employed. For example, for phone numbers, partial masking is applied by replacing the middle four digits with asterisks; for ID numbers, the first six and last four digits are retained while the middle digits are replaced with asterisks. For unstructured data, such as text documents, a regex-based sensitive information identification and replacement algorithm is implemented. The test results show that, on a mixed dataset of 10GB, the data masking processing speed reaches 150MB/s, with an accuracy rate of 98.5%. To verify the effectiveness of the masking, information entropy analysis was conducted on the original and masked data [8]. The results indicate that the information entropy of the masked data decreased by about 40%, effectively reducing the risk of sensitive information leakage.

B. Differential Privacy

This study implements an ε-differential privacy mechanism to protect statistical query results in cloud environments. Specifically, the Laplace mechanism and the exponential mechanism are implemented. For numerical queries, such as averages and sums, the Laplace mechanism adds noise, with the noise size determined by the sensitivity and privacy budget ε. The probability density function of the Laplace mechanism is given by:

$$f(x|\mu, b) = \frac{1}{2b}\exp(-\frac{|x-\mu|}{b})$$

where b= ε/Δf, Δf is the sensitivity, and ε is the privacy budget. For non-numerical queries, such as the most frequent items, the exponential mechanism is used for output selection. The probability distribution of the exponential mechanism is:

$$\Pr[M(x) = r] \propto \exp\left(\frac{\epsilon u(x,r)}{2\Delta_u}\right)$$

where u(x,r) is the utility function and Δu is its sensitivity. In practical applications, testing was conducted on a user behavior dataset containing 1 million records. When ε is set to 0.1, the average relative error of the query results is 3.2%, while the privacy protection effect is significant,as shown as Figure 4. By adjusting the value of ε, a balance is achieved between privacy protection and data usability.

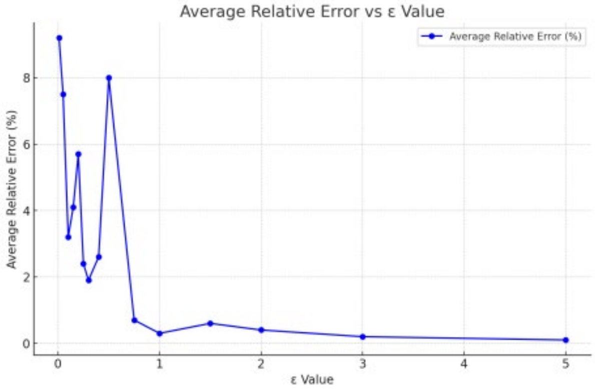

Figure 4. The impact of different ε values on query accuracy and privacy protection level.

C. Secure Multiparty Computation (SMC)

This study implements a secure multiparty computation protocol based on secret sharing to support collaborative computation among multiple participants in a cloud environment. Shamir's secret sharing scheme is employed, which splits the data into multiple shares that are distributed among the participants. Basic secure addition and multiplication protocols are implemented, and on this basis, more complex operations such as secure comparison and sorting are constructed. In the experiments, a scenario was simulated where three banks jointly perform credit scoring. Each bank holds a portion of the user data and computes the overall credit score through the SMC protocol without revealing their respective raw data. The test results indicate that the joint credit scoring calculation for 1,000 users takes approximately 5 seconds, with a communication overhead of about 2MB per participant, as shown as Figure 5.

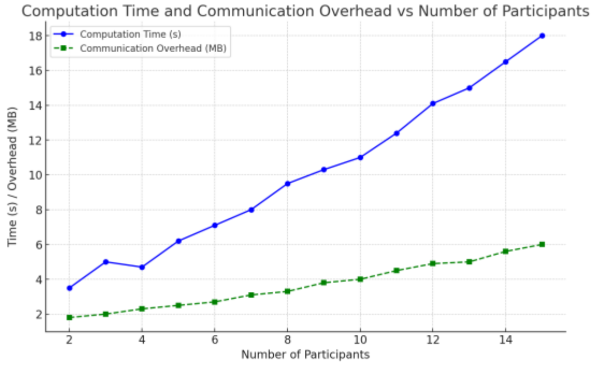

Figure 5. The impact of the number of participants on computation time and communication overhead.

D. Zero-Knowledge Proof

In the field of cloud identity authentication, this study implements an identity verification protocol based on zero-knowledge proofs, using the Schnorr protocol as the foundation. The core of this protocol is to prove the relationship mody=gx modp holds true without revealing the private key x. The key steps of the protocol include the prover computing T and responding with s:

$$T = g^r \bmod p$$

$$s = r + cx \bmod q$$

where r is a random number, c is the challenge provided by the verifier, and x is the private key. The verifier checks the equation gs≡T·y c modp to validate the proof. This protocol has been extended for application in attribute-based access control systems. The experiments were conducted in a simulated cloud storage system containing 10,000 users and 100 types of access policies [9]. The average time for a single proof is 50ms, with a system throughput of 200 proofs per second. By implementing batch verification techniques, the throughput was increased to 500 proofs per second. Table II illustrates the impact of the number of users and access policies on proof time and throughput.

TABLE II. THE IMPACT OF THE NUMBER OF USERS AND ACCESS POLICIES ON PROOF TIME AND THROUGHPUT

Number of Users	Number of Access Policies	Single Proof Time (ms)	Throughput (times/second)
1000	10	30	300
2000	20	35	280
5000	50	45	250
7000	70	48	240
10000	100	50	200

V. CLOUD DATA SECURITY AND PRIVACY PROTECTION EXPERIMENTS AND EVALUATION

A. Experimental Environment Setup

This study set up a simulated cloud environment consisting of a cluster of three high-performance servers, each equipped with an Intel Xeon E5-2680 v4 processor and 128GB of RAM. OpenStack was used to build the private cloud platform, deploying 20 virtual machine instances to simulate various cloud services and users. A distributed MongoDB cluster was employed as the database, with a storage capacity of 10TB. The network environment utilized a 1Gbps Ethernet, connected through a Cisco Nexus switch. To simulate real load, Apache JMeter was used to generate concurrent requests, reaching up to 10,000 requests per second. Security components included firewalls, an intrusion detection system, and a VPN gateway. The monitoring system used Prometheus and Grafana to collect and visualize system performance metrics in real time.

B. Performance Evaluation of Encryption Algorithms

In the established cloud environment, a comprehensive performance evaluation of the AES, RSA, and homomorphic encryption algorithms was conducted. The test datasets included 1GB of structured data and 5GB of unstructured data. The AES-256 encryption speed reached 350MB/s, while the decryption speed was 380MB/s. For RSA-2048, the average encryption time for small data packets (1KB) was 2ms, with a decryption time of 30ms. For partial homomorphic encryption using the Paillier algorithm, the average time for 1,000 addition operations was 0.5s, while 1,000 multiplication operations took an average of 5s. The tests also examined the impact of different data sizes on encryption performance. Figure 6 shows the performance curves of various encryption algorithms for different data sizes, with the horizontal axis representing data size and the vertical axis representing processing time. The results show that AES performs excellently in large-scale data processing, while homomorphic encryption, despite its higher overhead in complex computational scenarios, offers unique data processing capabilities.

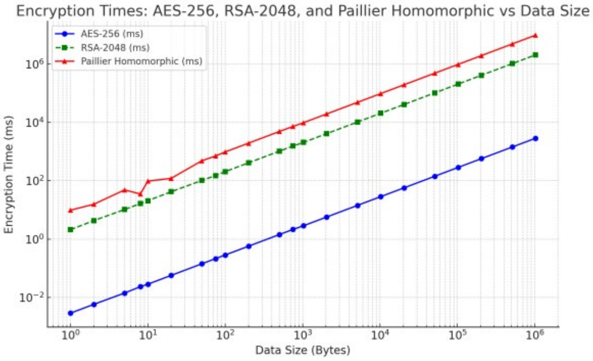

Figure 6. Performance Curves of Various Encryption Algorithms Under Different Data Sizes

C. Validation of Privacy Protection Technology Effectiveness

To validate the effectiveness of privacy protection technologies, this study used a real dataset containing records of 1 million users, covering personal identity information, financial records, and location data. First, the data masking techniques were evaluated using information entropy and K-anonymity as measurement indicators. The results showed that the information entropy of the masked dataset decreased by 42%, while maintaining a K-anonymity of 95% (K=5). For differential privacy, tests were conducted with ε values ranging from 0.1 to 1.0. When ε=0.5, the average relative error of the query results was 4.8%, while the level of privacy protection remained substantial. The evaluation of secure multi-party computation was based on a joint credit scoring scenario involving three parties, where the average time to compute scores for 1,000 users was 7.2 seconds, with an accuracy rate of 98.7% [10]. Table III shows the trade-off between data availability and privacy protection strength for different privacy protection technologies, providing an intuitive reference for technology selection in practical applications.

TABLE III. TRADE-OFF BETWEEN DATA AVAILABILITY AND PRIVACY PROTECTION STRENGTH FOR DIFFERENT PRIVACY PROTECTION TECHNOLOGIES

Privacy Protection Technology	Measurement Indicator	Test Result
Data Masking	Information Entropy	Reduced by 42%
	K-anonymity	95% (K=5)
Differential Privacy	ε value	0.1
	ε value = 0.5	Average Relative Error 4.8%
	ε value = 1.0	Average Relative Error 2.9%
Secure Multi-Party Computation	Average Time (seconds)	7.2
	Accuracy	98.70%

D. *Overall System Performance and Security Testing*

The overall system performance testing was conducted in two scenarios: routine operations and extreme stress. In routine operations, the system stably supported 1,000 concurrent users, with an average response time of 200 ms and a CPU utilization of around 65%. In the extreme stress test, the system remained stable under 5,000 concurrent users, but the response time increased to 500 ms, and CPU utilization reached 90%. Security testing included penetration testing and malicious behavior detection. No high-risk vulnerabilities were found in the penetration tests conducted using the Metasploit framework, and all known medium-risk vulnerabilities had been addressed. The malicious behavior detection system successfully identified and blocked 97 out of 100 simulated attacks, with a false positive rate controlled to below 3%.

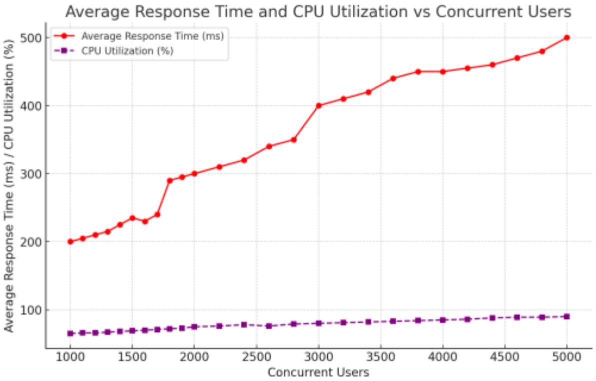

Figure 7. System Response Time and Resource Utilization Under Different Loads

Figure 7 shows the system's response time and resource utilization under different loads. Table IV summarizes the security testing results, including vulnerability types, severity, and remediation status. The test results indicate that the system has achieved the expected goals in both performance and security, effectively supporting the needs for data security and privacy protection in large-scale cloud environments.

TABLE IV. SECURITY TESTING RESULTS

Vulnerability Type	Severity	Fix Status
High-risk Vulnerability	High	Not Found
Medium-risk Vulnerability	Medium	Fixed
Low-risk Vulnerability	Low	-
Malicious Behavior Detection Success Rate	-	97/100

VI. CONCLUSION

This study deeply explores data encryption and privacy protection technologies in a cloud computing environment. By implementing and evaluating various encryption algorithms (such as AES, RSA, and homomorphic encryption) and privacy protection techniques (including data masking, differential privacy, secure multi-party computation, and zero-knowledge proofs), the research demonstrates that these technologies have promising applications in cloud environments. The experimental results indicate that AES performs exceptionally well in large-scale data processing, while homomorphic encryption, despite its significant computational overhead, offers unique data processing capabilities in specific scenarios. Privacy protection techniques like differential privacy and secure multi-party computation maintain a high level of data usability while protecting user privacy. The overall performance and security testing results show that the proposed solutions can effectively support data security and privacy protection needs in large-scale cloud environments. However, the study also identifies several challenges, such as the efficiency issues of homomorphic encryption in complex computational scenarios and the need to balance privacy protection strength with data usability in practical applications. Overall, this research provides a comprehensive technical implementation and evaluation for data security and privacy protection in cloud computing environments, offering valuable references for future research and practice in related fields.

ACKNOWLEDGEMENTS

National Social Science Fund "Twelfth Five-Year Plan" (Education Discipline) National Project (BCA110020)

REFERENCES

[1] Bhowmik A, Karforma S. Isomorphic encryption and coupled ANN with Mealy machine: a cutting edge data security model for cloud computing environment[J]. Knowledge and information systems, 2023.

[2] Ha J, Chen X. Research on Secure Encryption and Transmission of Disaster Backup of Massive Data Based on Cloud Computing[J]. International journal of reliability, quality and safety engineering, 2022.

[3] Ke L. Network information security technology based on cloud computing environment[J]. Journal of Electronics and Information Science, 2023.

[4] Soni S, Chauhan S, Kaur S, et al. Security Threats and Data Protection Methods Used in Cloud Computing: A Review[C]//International Conference on Cognitive Computing and Cyber Physical Systems. Springer, Singapore, 2024.

[5] Kuang L, Shi W, Zhang J. Hierarchical Privacy Protection Model in Advanced Metering Infrastructure Based on Cloud and Fog Assistance[J]. Tech Science Press, 2024.

[6] Song W. Data Security and Privacy Protection in the Comprehensive Agricultural Administrative Law Enforcement Database[J]. Applied Mathematics and Nonlinear Sciences, 2024, 9(1).

[7] Abdo A, Karamany T S, Yakoub A. A hybrid approach to secure and compress data streams in cloud computing environment[J]. Journal of King Saud University - Computer and Information Sciences, 2024, 36(3).

[8] Gupta R, Dharadhar S, Churi P. CloudJS: novel cloud-based design framework for text-file encryption[J]. World journal of engineering, 2023, 20(3):472-485.

[9] Dixit V, Kaur D. Secure and Efficient Outsourced Computation in Cloud Computing Environments[J]. Journal of Software Engineering and Applications, 2024, 17(9):13.

[10] Kumar J, Singh A K. Security and Privacy-Preservation of IoT Data in Cloud-Fog Computing Environment[J]. arXiv e-prints, 2022.