A Modified Advanced Encryption Standard for Data Security

Lin Teng, Hang Li, Shoulin Yin, and Yang Sun
(Corresponding author: Hang Li)

Software College, Shenyang Normal University
No. 253, HuangHe Bei Street, Huang Gu District, Shenyang 110034, China
(Email: lihangsoft@163.com)

(Received June 2, 2018; Revised and Accepted Nov. 10, 2018; First Online June 26, 2019)

Abstract

With the continuous development of society and economic progress, when a large amount of data enters the cloud computing system, people will pay more attention to data security. In order to make the stored data in the cloud more secure, according to the characteristics of cloud computing, we study the modified data encryption algorithm in cloud computing. First traditional advanced encryption standard (AES) is analyzed. Then a modified advanced encryption standard for data security in cloud computing is proposed by introducing random disturbance information to improve the data security. What's more, column mix operation and key choreography in AES are improved. Finally, experiments are conducted on Hadoop. Formal security analysis and performance comparisons indicate that the proposed solutions simultaneously ensure attribute privacy and improve decryption efficiency for outsourced data storage in mobile cloud computing.

Keywords: AES; Key Choreography; Random Disturbance Information

1 Introduction

Hamming Codes are one of the EDC codes which are used to protect the registers and memories from soft errors. As technology scales, radiation particles can create more soft error likely to affect the more than one bit binary number. Big-data storage poses significant challenges to anonymization of sensitive information against data sniffing. Not only will the encryption bandwidth be limited by the I/O traffic, the transfer of data between the processor and the memory will also expose the input-output mapping of intermediate computations on I/O channels that are susceptible to semi-invasive and non-invasive attacks.

Cloud computing [6,9,12,17] is an emerging computing model applied to the Internet. It provides basic resource facilities, application systems, and software platforms as services to users.Cloud computing is also a virtualization-based architecture that virtualizes resources and builds large-scale resource pools and manages services externally.

With the development of cloud computing, amounts of user data and large-scale data are put into cloud computing systems. Because of the distributed and virtualized nature of cloud computing, users cannot intuitively determine the storage location and division of data, *etc.*, so the security of data becomes very important. In the cloud computing, data security is generally ensured through data encryption and identity management [3, 10]. At present, the common encryption algorithms are classified into symmetric encryption algorithm and public key encryption algorithm. Among them, DES algorithm and AES algorithm are two widely used algorithms in symmetric encryption algorithms [1, 4, 15, 19].

Therefore, many researchers proposed many new schemes to solve the above issue. Deng [2] proposed an effective PKC-based certificateless group authenticated key agreement protocol, the certificateless mechanism of the protocol simplified the complex certificate management problem and key escrow problem in ID-based protocols. The security of the scheme was proved and its computational cost was discussed. The result showed that the new protocol was secure and effective. Shan [13] proposed an improved protocol to append a signature in the second round to eliminate weakness of certificateless group key agreement protocol. The signature was related to the group identity, the broadcast messages in the first round and the computed message in the second round, to ensure the protocol freshness and the entity authenticity. The message in the second round guarantees that the adversary could not attack the protocol by corrupting neighboring entities. Zhang [20] studied authenticated AGKA in certificateless and identity-based public key cryptosystems. They formalized the security model of certificateless authenticated asymmetric group key agreement and realized a one-round certificateless authenticated asymmetric group key agreement protocol to resist active attacks in the real world. They also investigated the re-

lation between certificateless authenticated AGKA and identity-based authenticated AGKA. So a concrete conversion from certificateless authenticated AGKA was proposed to session key escrow-free identity-based authenticated AGKA. Yin [18] introduced the concept of distributed Searchable asymmetric encryption, which was useful for security and could enable search operations on encrypted data.

This paper proposes a modified data encryption algorithm in cloud computing. First traditional advanced encryption standard is analyzed. Then a modified advanced encryption standard for data security in cloud computing is proposed by introducing random disturbance information to improve the data security. What's more, column mix operation and key choreography in AES are improved. In terms of security, the protocol can prove safety in the random prediction model; For performance, the new protocol requires only one round to complete authentication and key negotiation.

And for computation, compared with state-of-the-art schemes, the calculation of new protocols is also significantly reduced. The rest of the paper is organized as follows. Section 2 introduces the Hadoop Framework in this paper. Traditional encryption is explained in Section 3. Section 4 outlines the proposed scheme to analyze detailed processes. Experiments and performance analysis are given in Section 5. Finally, Section 6 concludes this paper.

2 Hadoop Framework

Hadoop is a distributed computing framework developed by the Apache storage and calculations for massive amounts of data. The core design of Hadoop framework is distributed file system (HDFS) and parallel computing framework (MapReduce). HFDS is responsible for the distribution and storage of data, and MapReduce is responsible for the calculation of data [11].

2.1 HDFS System

The essence of HDFS [16] is a distributed file system, which can divide a large data into small data sets and back them up, distributed and stored on different nodes in the cloud environment. However, for a single user, HDFS is like a traditional hierarchical file system. When used, HDFS can operate on big data just like a single file.

The HDFS framework is built on a set of specific nodes, which includes a unique NameNode to provide metadata services, guide computing nodes and data nodes to handle assigned tasks. Multiple DataNode is mainly for HDFS to provide storage blocks, and to perform read and write operations for distributed files. The data redundancy in the Hadoop platform is three, and each piece of data is stored in three DataNodes.

In the cloud computing environment, HDFS ensures the reliable storage of massive data through the following measures. DataNode sends a "heartbeat" message to NameNode regularly and sends the data block list information to determine whether the node is normal to provide a secure mode, only read views in this mode, it does not allow for additional or deletions and modification operations, record detailed log files, and test the integrity of the data taken by the user.

2.2 MapReduce Framework

MapReduce is a software framework that processes large data sets in parallel [14]. The root of MapReduce is the *map* and *reduce* functions in functional programming, corresponding to the mapping and specification in the calculation process. The Map process accepts a set of data and converts it into a key/value pair list, then transmits and reorder it. The *Reduce* process takes a list generated by the Map and then shrinks the list of key/value pairs based on their keys (generating a key/value pair for each key). That is, the Reduce process processes the integration and sorting of the intermediate results generated by the Map process, and then forms the final result.

3 Traditional Encryption Algorithm

The amount of data in cloud computing is very large, and often scattered on different computing nodes. The security protection is very important. Encrypting and decrypting data through encryption algorithms is one of the most effective methods to ensure data security. This article mainly discusses and improves the symmetric encryption algorithm in traditional encryption algorithms. Symmetric encryption algorithms use the same key for encryption and decryption, such as DES and AES algorithms. A symmetric encryption system can be represented as $CS = M, C, K, e, d$, where: $m \in M$ represents a plaintext message set; $C = c$ represents a ciphertext message set; $K = k$ represents the key set; E represents the encryption mapping process, *i.e.* $E : K*M = C$; D represents the decryption mapping process, *i.e.* $D : K*C = M$.

During the execution of the DES algorithm, the plaintext is grouped in 64 bits, and the last group with less than 64 bits is patched according to a specific method. The key length is 8 bytes, but 8 bits are the check bits. In the encryption phase, the plaintext is first divided into 32 parts by initial replacement, represented by the left half and the right half. Then perform 16 rounds of operations to combine the data and the key. The key is shifted in each round of operations, 48 bits out of the 56 bits of the key are selected, the original 32 bits of the right half are replaced by 48 bits through expansion, and then the XOR operation is combined with the 48-bit key Then, the 48 bits are converted to 32 bits by the S box, and then XOR with the original 32 bits of the left half.Finally, the final ciphertext is obtained by inverse initial permutation. The algorithm flow chart is shown in Figure 1.

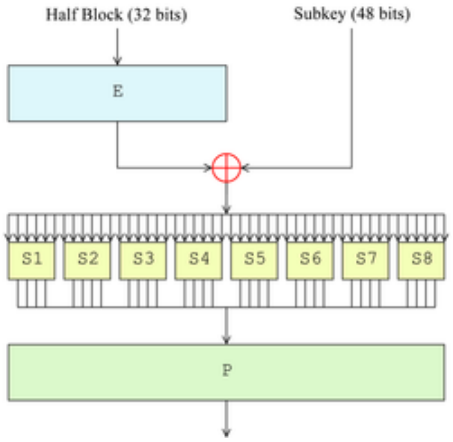

Figure 1: DES process

The AES algorithm is also a classical algorithm for symmetric key encryption. The length of the block in AES is 128 bits, and the key length can be 128 bits, 192 bits, or 256 bits.The AES encryption process operates on a 4×4 byte matrix.The AES encryption process mainly performs four types of operations:

1) SubBytes means that each byte is replaced by a lookup table through the S box.

2) ShiftRows performs a cyclic shift operation on each row in the 4x4 matrix.

3) MixColumns, uses linear conversion to mix 4 bytes per column.

4) AddRoundKey refers to the XOR operation between each round of input and round keys in the encryption process, and the XOR key is used in the decryption process.

With the continuous improvement of cloud computing capabilities and the rapid development of computer hardware, the shortest key for the DES encryption algorithm is too short. The key length is 64 bits and 8 check bits are removed. The actual effective number of bits is only 56 bits. If the brute-force method is used to crack, only 256 possibilities need to be calculated. Particularly, the computing power of the cloud platform is used to complete the cracking in a short time. Keys have become possible. For the AES algorithm, the key length is up to 256 bits, and the using of the brute force method is less likely to forcibly crack, but this algorithm is not absolutely secure, if an attacker designs different keys to measure the exact time required for the encryption process. Once the encryption routine is carelessly encoded, the execution time depends on the key value, and it is possible to derive information on the key.

4 Modified AES

AES adopts group iteration, patch size is 4×4 matrix. Each element is 8 bits. In order to make the algorithm applicable encryption, achieve better security and improve the encryption efficiency, we make the following improvements based on the AES algorithm framework.

4.1 Improved Key Sequence Generation Method

Chaotic dynamic system has pseudo randomness and is extreme sensitivity to initial conditions and system parameters. Therefore, it provides a good way for image information encryption. The improved algorithm adopts the following skew tent map to generate the key sequence.

$$F_a(x) = \begin{cases} x/a, x \in (0,a). \\ (1-x)/(1-a), x \in (a,1). \end{cases} \quad (1)$$

When $a \in [0,1]$, the system is in a chaotic state. The correlation of this mapping iterative trajectory sequence decreases exponentially, and the distribution of chaotic variables is uniform with good pseudo-random characteristics.

The method of generating pseudo-random sequences based on oblique tent mapping is as follows: one $M \times N$ image needs to encrypt R rounds. First, it iterates the oblique tent mapping and gets R sequence $X_r x_{r,0}, x_{r,1}, \cdots, x_{r,MN-1}, 1 \leq r \leq R$. X will be expanded to 0-255 integer sequence $K_r k_{r,0}, k_{r,1}, \cdots, k_{r,MN-1}$ according to following equation.

$$k_{r,i} = \llcorner x_{r,i} \times 255 \lrcorner.$$

where $\llcorner \lrcorner$ denotes round to-infinite.

4.2 Improved Encryption and Decryption

The encryption way of AES is that matrix D and key sequence K execute XOR operation. In order to increase the sensitivity to plaintext, the algorithm is improved. The encryption process is:

$$C[i][j] = \begin{cases} D[i][j] \bigoplus k_{r,i \times N+j}, i = M-1, j = N-1. \\ D[i][j] \bigoplus (k_{r,i \times N+j} \bigoplus D[i+1][0]), \\ \qquad i \neq M-1, j = N-1. \\ D[i][j] \bigoplus (k_{r,i \times N+j} \bigoplus D[i][j+1]), others. \end{cases} \quad (2)$$

Where $i \in [0, M-1]$, $j \in [0, N-1]$, $D[i][j]$ are plaintext pixels. $C[i][j]$ is obtained cipher pixel.

4.3 Key Tree

The plaintext matrix is encrypted from left to right. Ciphertext matrix is decrypted from top to bottom and from right to left. After XOR operation, key and plaintext are combined. Two different images even if use the same initial conditions, but the generated key sequences are different.

25	13	242	79
66	24	35	176
242	13	71	233
195	140	125	72
78	238	97	248

(a) 5×4 matrix

25	38	255	65
66	91	58	212
242	255	84	49
195	79	8	196
78	60	79	89

(b) Row diffusion result

25	13	242	79
41	10	49	97
201	3	22	136
250	137	102	191
83	102	252	58

(c) Column diffusion result

Figure 2: Diffusion enhancement result

4.4 Improved Column Mixing

In AES, Column mixing operation $MixColumns$ adopts the matrix operation, each pixel takes shift and XOR operation. In order to reduce the computational complexity and achieve good mixing effect, in the improved algorithm, we changed the $MixColumns$ matrix operations, the simple addition and subtraction are adopted to strengthen the relationship between pixels. The concrete measures are as follows: for each row, the first pixel remains the same, and the current pixel is updated with adjacent pixels from the second pixel (as shown in Equation (3)); For each column, the first pixel remains the same, and the current pixel is updated with the value of adjacent pixels starting from the second pixel (as shown in Equation (4)).

$$MixColumns = \begin{cases} D[i][j] = D[i][j], j = 0. \\ D[i][j] = (D[i][j] - D[i][j-1] \\ \qquad \bmod 256, others. \end{cases} \quad (3)$$

$$MixColumns = \begin{cases} D[i][j] = D[i][j], i = 0. \\ D[i][j] = (D[i][j] - D[i-1][j] \\ \qquad \bmod 256, others. \end{cases} \quad (4)$$

With an example of 5×4 matrix, the operation result is as shown in Figure 2. From this figure, we can know that when $D[0][0]$ is changed, it will affect all pixels. When $D[M-1][N-1]$ is changed, it does not affect other pixels in the same round. So in the row and column transformation operations, each line should be moved to the left, each column moves up. After several encryption round, it has obvious diffusion effect. Improved row and column mixed operations, it uses simple addition and subtraction, each pixel needs only two additive operations, the operation is not only reduces the computational complexity, but strengthens the connection between the pixels. After several rounds of encryption, it can achieve better mixing effect.

5 Experiment and Analysis

This experiments are conducted in MATLAB platform with SSH framework and clusters simulation cloud computing environment. Clusters are composed of six computers with one computer CPU I7, memory 8GHz, frequency 3.2GHz, this computer is as $NameNode$. To better simulate the cloud environment, the other five computers are selected five different machines as $Slave$ and $DataNode$.

Because the biggest advantage of cloud computing is that it can process large data's storage and calculation, this experiment chooses the size of 1.5GB text file as the experimental data. The data calculated by MapReduce in Hadoop platform, we test the performance of proposed algorithm and make comparison with other encryption methods including RSAE [5], CTME [8] and SUE [7]. Because the performance of AES is poorer than RSAE, we did not compare with AES.

5.1 Performance Analysis

We do the following performance analysis.

1) Plaintext sensitivity analysis. If no interference information is added, the plaintext sensitivity of the transformation is the same as AES algorithm if it falls into the M_D segment. If it falls into the M_A segment, the sensitivity is the same as AES algorithm too. However, new algorithm is more sensitive to plaintext than AES because random variable interference information is added to both sections of plaintext.

2) Key sensitivity analysis. The key sensitivity of new algorithm is determined by AES algorithm. If the key K_1 is changed, the middle plaintext M_{D1} is changed. If the key K_2 is changed, the middle plaintext M_{A1} will be changed too. When the key changes slightly, the final ciphertext will be greatly changed. This algorithm has better key sensitivity.

3) Against attack analysis. Attackers attack new algorithm which means that it needs to defeat AES algorithm. For the ciphertext attack, the obtained ciphertext just locates in the segment point, the probability is very small. Even getting the key to decrypt the ciphertext C_D and C_A, the plaintext M_A and M_D are more difficult to obtain, so the original plaintext M is difficult to acquire too. And that plaintext attack can also be difficult to speculate the original plaintext message.

5.2 Time Analysis

MapReduce calculates data by default in 64MB block. To intuitive display performance of the new algorithm, we set file block size with 64MB, 32MB, 16MB, 8MB, 4MB and 2MB. We conduct 8 encryption experiments and 8 decryption experiments. The average execution time of the algorithm was taken as shown in Figures 3 and 4. AE: average time.

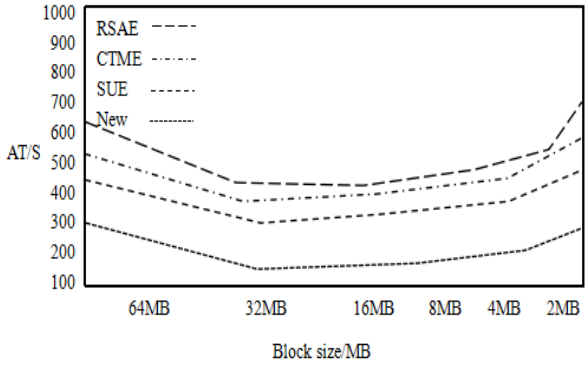

Figure 3: Average encryption time

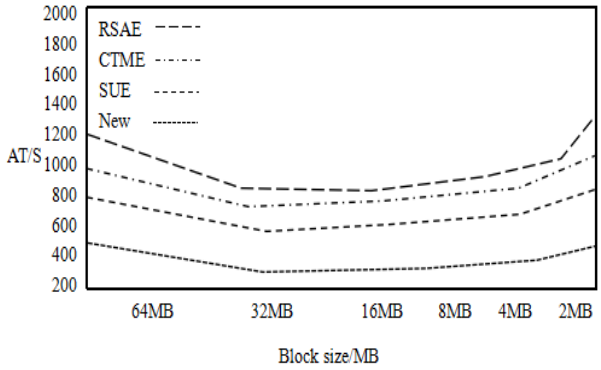

Figure 4: Average decryption time

In Hadoop platform, block size of the original data has the similar effect for the four algorithms. The influence of partitioned values changed from 64MB to 32MB or 16MB, the algorithms have the highest execution efficiency, if it continues to reduce the block value, especially when the block value is 2MB, four algorithms' execution time are rising sharply, this is because when the data block unit is too small, the block number will surge. In MapReduce calculation, Reduce process can consume time for the integrate ordering of Map. So in the cloud computing, the division of big data should take appropriate units, otherwise, it would affect the computation time. We also can get that new method has better performance in encryption and decryption process than other three methods. In summary, the proposed method has high security in cloud computing.

6 Conclusion

Cloud computing is a widely promising commercial calculation model based on virtualization of resources. Large huge amounts of data can be calculated and managed. It provides service according to the customer's demand. This paper analyzes the Hadoop technology and constructs experimental platform. Firstly, traditional data encryption algorithm is introduced. Aiming at the shortcomings of the raw algorithms in cloud computing environment, this paper puts forward a modified encryption algorithm by introducing random disturbance information to improve the data security. Finally, the experiments results prove that proposed algorithm is suitable for encryption in cloud computing environment.

Acknowledgments

This study was supported by the Natural Science Fund Project Guidance Plan in Liaoning Province of China (No. 20180520024). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] D. S. AbdElminaam, "Improving the security of cloud computing by building new hybrid cryptography algorithms," *International Journal of Electronics and Information Engineering*, vol. 8, no. 1, pp. 40–48, 2018.

[2] F. Deng, Y. Zhu, "Novel one-round certificateless group authenticated key agreement protocol," *Computer Engineering & Applications*, vol. 53, no. 5, pp. 111-115, 2017.

[3] S. T. Hsu, C. C. Yang, and M. S. Hwang, "A study of public key encryption with keyword search", *International Journal of Network Security*, vol. 15, no. 2, pp. 71–79, Mar. 2013.

[4] S. H. Islam, A. Singh, "Provably secure one-round certificateless authenticated group key agreement protocol for secure communications," *Wireless Personal Communications*, vol. 85, no. 3, pp. 879-898, 2015.

[5] S. Kaufmann, "RSA public-key encryption," *Journal of Biological Chemistry*, vol. 280, no. 5, pp .3636-44, 2018.

[6] S. Khatoon, T. Thakur, B. Singh, "A provable secure and escrow-able authenticated group key agreement protocol without NAXOS trick," *International Journal of Computer Applications*, vol. 171, no. 3, pp. 1-8, 2017.

[7] K. S. Lee, S. G. Choi, D. H. Lee, "Self-updatable encryption: Time constrained access control with hidden attributes and better efficiency," *Theoretical Computer Science*, vol. 667, pp .51-92, 2017.

[8] C. Li, G. Luo, K. Qin, et al, "An image encryption scheme based on chaotic tent map," *Nonlinear Dynamics*, vol. 87, no. 1, pp .127-133, 2017.

[9] J. Liu, S. L. Yin, H. Li and L. Teng, "A density-based clustering method for k-anonymity privacy protection," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 1, pp. 12-18, Jan. 2017.

[10] L. Liu, Z. Cao, C. Mao, "A note on one outsourcing scheme for big data access control in cloud," *International Journal of Electronics and Information Engineering*, vol. 9, no. 1, pp. 29–35, 2018.

[11] R. R. Parmar, S. Roy, D. Bhattacharyya, "Large-scale encryption in the hadoop environment: Challenges and solutions," *IEEE Access*, vol. 5, pp. 7156-7163, 2017.

[12] S. Rezaei, M. Ali Doostari, and M. Bayat, "A lightweight and efficient data sharing scheme for cloud computing," *International Journal of Electronics and Information Engineering*, vol. 9, no. 2, pp. 115–131, 2018.

[13] C. Shan, H. U. Kangwen, J. Xue, "Improved pairing-free constant round certificateless authenticated group key agreement protocol," *Journal of Tsinghua University*, vol. 57, no. 6, pp. 580-585, 2017.

[14] B. Sheintz, A. Chandra, R. K. Sitaraman, "End-to-end optimization for geo-distributed mapReduce," *IEEE Transactions on Cloud Computing*, vol. 4, no. 3, pp. 293-306, 2017.

[15] L. Teng, H. Li, S. L. Yin, "A multi-keyword search algorithm based on polynomial function and safety inner-product method in secure cloud environment," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8. no. 2, pp, 413-425, 2017.

[16] S. Wu, W. Zhu, B. Mao, "PP: Popularity-based proactive data recovery for HDFS RAID systems," *Future Generation Computer Systems*, vol. 86, pp. 1146-1153, 2018.

[17] S. L. Yin and J. Liu, "A k-means approach for map-reduce model and social network privacy protection," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 7, no. 6, pp. 1215-1221, Nov. 2016.

[18] S. L. Yin, L. Teng, J. Liu, "Distributed searchable asymmetric encryption," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 4, no. 3, pp. 684-694, 2016.

[19] Q. C. Zhang, T. L. Yang, X. G. Liu, Z. K. Chen, and P. Li, "A tucker deep computation model for mobile multimedia feature learning," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 13, no. 3, pp. 1-39:18, 2017.

[20] L. Zhang, Q. Wu, B. Qin, "Certificateless and identity-based authenticated asymmetric group key agreement," *International Journal of Information Security*, vol. 16, no. 5, pp. 559-576, 2017.

Biography

Lin Teng received the B.Eng. degree from Shenyang Normal University, Shenyang , Liaoning province, China in 2016. Now, she is studying for Master degree in Software College, Shenyang Normal University. Her research interests include Multimedia Security, Network Security, Filter Algorithm and Data Mining.

Email:910675024@qq.com.

Hang Li is a full professor in Software College, Shenyang Normal University. He received his B.S. and M.S. degrees from Northeastern University. He is also a master's supervisor. He has research interests in wireless networks, mobile computing, cloud computing, social networks, network security and quantum cryptography. Professor Li had published more than 30 international journal papers (SCI or EI journals) on the above research fields. Email: lihangsoft@163.com.

Shoulin Yin received the B.Eng. and M.Eng. degree from Shenyang Normal University, Shenyang , Liaoning province, China in 2016 and 2013 respectively. Now, he is a doctor in Harbin Institute of Technology. His research interests include Multimedia Security, Network Security, Filter Algorithm, image processing and Data Mining. Email:yslinhit@163.com.

Yang Sun obtained his master degree in Information Science and Engineering from Northeastern University. Yang Sun is a full professor of the Kexin software college at Shenyang Normal University. He is also a department head of network engineering. He has research interests in wireless networks, mobile computing, cloud computing, social networks and network security. Yang Sun had published more than 20 international journal and conference papers on the above research fields. Email:17247613@qq.com.