International Journal of Security and Its Applications
Vol.7, No.6 (2013), pp.97-108
http://dx.doi.org/10.14257/ijsia.2013.7.6.10

Research of Data Security Model in Cloud Computing Platform for SMEs

Tao Sun[1,2] and Xinjun Wang[1]

[1]*School of computer science and technology, Shandong University*
Jinan, Shandong 250101, China
[2]*School of information, Qilu University of Technology*
Jinan, Shandong 250353, China
suntao0906@163.com,wxj@sdu.edu.cn

Abstract

In recent years, cloud computing has become a major mode to address SMEs inefficient and help improve their competitiveness. As cloud providers have priority access to data, so it is difficult to guarantee the confidentiality and integrity of users' data. For this reason this paper proposed a model to solve the problem of data security in cloud computing. The model adopts efficient encryption mechanisms to protect users' data. As the encrypted data is difficult to retrieve, we present a method of cipher text retrieval. Experimental results show that the model can better ensure data confidentiality and integrity.

Keywords: *data security; cloud computing; integrity verification; cipher text retrieval*

1. Introduction

The competition in business market has become increasingly fierce, technological innovation is a key factor in winning. Small and Medium enterprises (SMEs) have less advantages in market competition because of their limited resources, so that they rarely have opportunity to get new technologies for support [1]. But SMEs play a vital role in economic because they provide a lot of employment opportunities. How to help SMEs to reduce costs, process innovation and rapid implementation is an important direction for IT services.

In recent years, Cloud computing has become a major mode to address inefficient and help SMEs develop and improve their competitiveness. There are a large number of computing, storage and software resources in cloud computing, forms a huge shared virtual IT resource pool and it offers a variety of IT services for SMEs. These services can be accessed on-demand and provide support to new business requirement far faster than traditional mode[2].

Nevertheless, there are still many SMEs choose the traditional software architecture largely of the reasons is most likely that enterprise data security issues are unresolved in cloud computing. Some analysis of the survey results show that data security is one of the biggest obstacles to migrate enterprise applications to the cloud computing. At present, cloud computing security issues have been gotten more and more attention.

Data security is one of the basic security issue in cloud computing. Data encryption and access control are common methods to solve security issue. In cloud computing environment, the cloud service provider could not be trusted by SMEs users, the access control middleware run in untrusted environment, could not correctly implement user-defined access control policies. Data encryption could prevent unauthorized access, but difficult to achieve fast retrieval in encrypted data state. The application system with no corresponding encrypted data

ISSN: 1738-9976 IJSIA

retrieval methods and efficient implementation framework will not be accepted by SMEs users.

Data integrity is another important cloud computing security issue. Ordinary document or information integrity verification may take various hashing algorithm or signature algorithm, if the cloud service provider calculated in advance and retain the hash value of the document, then according to this hash value generated signature will not reflect the real situation of the current document, the owner of the document could not discover that the document had been tampered until fails to access.

Therefore the cloud computing platform needs a data security protection mechanism to provide reliable mass data security support. A novel cloud data security model presented by this paper has the following characteristics:

(1) Data security does not depend on cloud service provider. The core confidential data (such as cloud user key) stored in local environment, owned by SMEs users, avoid depending on cloud service provider. The component of the model will not disclose users' confidential data.

(2)The model presented in this paper with high security service efficiency. The model could help SMEs users find out encrypted data that have some key properties without data decryption, and help users determine the content that stored in cloud whether missing or damaged, support SMEs user transparent encryption and decryption, key management convenient and decryption efficiency.

(3)The model has high flexibility and scalability. The architecture of this model was independent of actual cloud storage system, could provide data security services for all kinds of cloud computing platform, the database of the model adopts scalable distributed database management system, rapid expansion along with the expansion of the cloud computing system.

2. Related Work

Research fields of data security in cloud environment include: cipher text retrieval, integrity verification and privacy data protection *etc*. The study method of cipher text retrieval focused on two aspects, one is equivalent match retrieval, such as linear search algorithm, public key search based on keywords and security index algorithms etc.; the second is interval query on encrypted data, such as partaker barrel to achieve interval query and order preserving encryption algorithms. Song [3] presents an algorithm of search on encrypted data, uses an encryption information retrieval algorithm realizes adding authentication mechanism into encrypted result, the algorithm has a strong ability to resist the statistical analysis, but in cloud computing environment, complex validation process will consume too many resources. Goh [4] proposes a method of security index to solve attacker statistical analysis for keyword index, but "bloom filter" was introduced into this method result in a certain error, affecting retrieval efficiency. Hacigumus [5] put forward a method of interval division for the range of database fields, realizes the function of interval query on encrypted data. Agrawal [6] presents an algorithm of preserving encryption for digital data, this method implements digital data encryption protection.

In cloud computing environment, users need to check the integrity about the returned data that stored in cloud server. Integrity verification checks the consistency of the data that read back from cloud service provider with the original data, that is, determine whether the data has been tampered. Ateniese [7] proposes a PDP model, this model random sampling files

that stored on servers, with a certain probability to determine the server-side file is intact. Juels and Kaliski[8] presents a POR model, this model could combine detection point and error correction code to verify data integrity, meanwhile error data could recover.

In cloud storage security framework, the most widely used is the industry's cloud computing architecture. Amazon, Google and IBM adopt the access control security mechanism to protect data. However, enterprise data stored in cloud will bring security and privacy risks, making companies don't want to store their core business data in cloud storage. In order to solve cloud storage data security issues, Microsoft cryptography research group made virtual private storage solutions[9].Users upload data to the cloud was encrypted, using a token for key distribution and licensing, meanwhile using the cipher text retrieval and integrity verification methods for data retrieval and validation. In addition, this scheme using the certificate authority to authenticate documents, ensuring the confidentiality of data, while using access control mechanisms for data sharing. All operations are concentrated in cloud and there is no specific accelerate components to improve the efficiency of services.

3. Data Security Model in Cloud Computing

This paper presents a data security model for cloud computing, provides encrypted data services and reliable mass data security support. The platform used in this paper is second development software platforms. The interface required by the data security system provided by cloud computing, through it to achieve organic integration with other applications.

3.1. The Function of the Model

The security features of the model that proposed by this paper include encryption and decryption of data, cipher text retrieval and integrity verification.

(1) Data encryption and decryption. The cloud users could use this model to encrypt and decrypt data. The owner of the data holds a master key, could completely access data in the files, other users could not crack cipher text content because of no keys. In order to guarantee the security of the key, users don't need to manage or save a lot of keys' information. During decryption, key file should securely store and could not repeat. The model has the ability to transmit data and decrypt data simultaneously.

(2) Cipher text retrieval. The model could provide searching services for encrypted data that stored in cloud platform, and could support retrieval based on file properties and file contents. Retrieval based on file properties will search data through file name, owner, file creation time, file type and other general properties and user-defined attribute. Retrieval based on file content will search data through keywords that automatically generated from the file content. This model could support retrieval based on file attributes and cipher text attribute. The access control system controls the retrieval range that comprises the user's own files and the authorization files.

(3) Massive data integrity verification. Integrity verification of the model is mainly responsible for the users to verify the integrity of files. This task initiated by the users and could repeatedly verify. Comparison of authentication credentials information that read from the server and the results that calculated by key and the server data, then according to the comparison results to determine the integrity of the document.

3.2. The Basic Structure of the Data Security Model

The cloud computing data security model that present by this paper consists of three parts (Figure1 shows the specific composition):

(1) Be protected cloud computing (cloud storage system). The encrypted data stored in cloud storage system, and non-authorized data transmission could not directly cause leakage.

(2) Cloud data security service system. This system was responsible for the generation, storage, maintenance and management of cloud computing data attributes and other meta-data. The data security service system could provide data retrieval, data integrity verification and other services.

(3) Cloud security client. The client was responsible for locally save cloud user key and data encryption and decryption, to avoid key leakage causes data security risks.

3.2.1. Be Protected Cloud Computing (Cloud Storage System): Be protected cloud computing is a cloud storage system with the capabilities of cipher text management. The system is actual cloud user data storage location, data is encrypted via cloud security client. The meta information safely stored in cloud data security system.

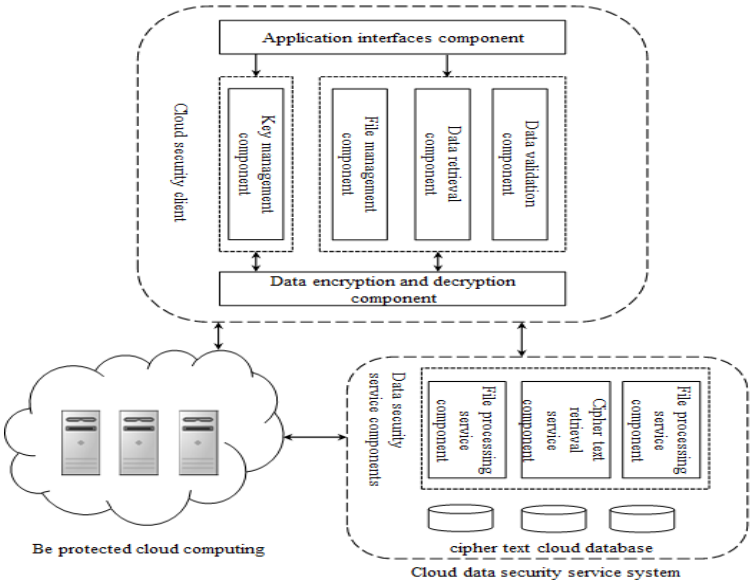


Figure 1. Structure of Data Security Model

3.2.2. Cloud Data Security Service System: The cloud data security service system comprises cipher text cloud database system and a number of data security service components, such as file processing service components, cipher text retrieval service components and data validation service components. All the components could respectively provide necessary support for users' documents operations, cipher text retrieval and data integrity validation.

The data stored in cipher text database system is meta information of users' data, key and authentication credentials. All the meta information that encrypted via the cloud security client was stored in the table of cipher text database. When users perform an update operation, such as insert, delete and update, not only the actual stored data is changed, but the meta information that stored in cipher text database also is changed. It is necessary to update the

cipher text database according to actual operation on users data, so that the information in actual storage systems and information in cloud security system is consistent.

When users perform cipher text retrieval, the cloud data security service system should transform the query command and then obtain the results from cipher text database. The results should be decrypted before return to SMEs users, the users obtain results and could view or download.

During integrity verification, it is necessary to compare the actual results from storage system against the results from cipher text database, then the authentication credentials was stored in cipher text database. Simple authentication need to compare the results from simple integrity verification table against the information from the actual data storage system. High level integrity verification needs to get authentication credentials from the high level integrity verification table, decrypt and calculate by cloud security client, then compare the information from authentication credentials against calculations to determine the integrity of a file.

Cloud users often need to search a particular file from massive data, the most direct way is to arrange file attribute according to certain organizational structure, then set up information index and store it in database. As cloud service provider is not credible, the plaintext index could leak the files' information, and will harm the data security and privacy, so it is necessary to store the encrypted index.

In order to improve the accuracy of data retrieval, multiple file attributes' meta information should be stored in cipher text database. The attribute information may be changed according to file operation, it requires more than one attribute can be extended. Column-oriented storage, highly scalable database is an ideal choice for cipher text database.

3.2.3. Cloud Security Client: Secure cloud client include application interfaces component, data security component, key management component and data encryption and decryption component. Among them, the application interfaces component provides the users cipher text files management, retrieval and integrity validation, and could choose appropriate function components based on users' operation.

Data security component includes files management component, data retrieval and data validation component. Among them, files management component implements files upload, download, copy, cut and delete functions. Data retrieval component implements constructing and uploading of the index meta information, and searching on cloud cipher text storage systems. Data validation component implements constructing and uploading of the verification meta information, and verifying integrity of cloud cipher text database.

In addition, keys management component is responsible for keys acquisition and management, data encryption and decryption component is responsible for data encryption and decryption.

3.3. Implement of the Data Security Model

3.3.1. Data Encryption and Decryption: This paper adopt fully homomorphic encryption[11] for data encryption, could ensure cloud computing SMEs data security. The encrypted data was stored in cloud, improving the security of data. Even if the data was theft, the data could not be restored because there was no corresponding key, only the SMEs users know the key. Meanwhile, the algorithm of fully homomorphic encryption could operate cipher text on cloud, avoiding the low efficiency problem of traditional encryption methods caused. This paper uses the fully homomorphic sysmmetric encryption algorithm that presented by Gentry[10], the proposed data security solutions described as below:

Table 1. The Fully Homomorphic Sysmmetric Encryption Algorithm

Input: *m, p, q, r;* Output: *c*
Encryption parameters: $q, p, r; q \sim 2^{n^5}, p \sim 2^{n^2}, r \sim 2^n;$ Keys: odd number *p*; Encryption: plaintext $m, c = pq + 2r + m$;//*c* is cipher text; Decryption: $m = (c \bmod p) \bmod 2;$ Correctness verification: $\because pq >> 2r + m, (c \bmod p) = 2r + m; \therefore (c \bmod p) \bmod 2 = (2r + m) \bmod 2 = m;$

The following addition and multiplication example to verify the homomorphism:
Two cipher text $c_1 = q_1 p + 2r_1 + m_1, c_2 = q_2 p + 2r_2 + m_2$,
then $c_1 + c_2 = (q_1 + q_2) p + 2(r_1 + r_2) + m_1 + m_2$, thus the only condition is satisfied $2(r_1 + r_2) + m_1 + m_2$ far less than *p*, then $c_1 + c_2 = (q_1 + q_2) p + 2(r_1 + r_2) + m_1 + m_2$,that is ,the encryption meet the additively homomorphic conditions.
$c_1 * c_2 = [q_1 * q_2 p + (2r_1 + m_1) + (2r_2 + m_2)] p + 2(2r_1 r_2 + r_1 m_2 + r_2 m_1) + m_1 m_2$, therefore the only condition is satisfied $2(2r_1 r_2 + r_1 m_2 + r_2 m_1) + m_1 m_2$ far less than *p*, thus $c_1 * c_2 = [q_1 * q_2 p + (2r_1 + m_1) + (2r_2 + m_2)] p + 2(2r_1 r_2 + r_1 m_2 + r_2 m_1) + m_1 m_2$, that is, the encryption meet the multiplication homomorphic conditions.

When SMEs users want to encrypt data, the data security client-side encryption and decryption and keys management components should interact with cloud storage system.

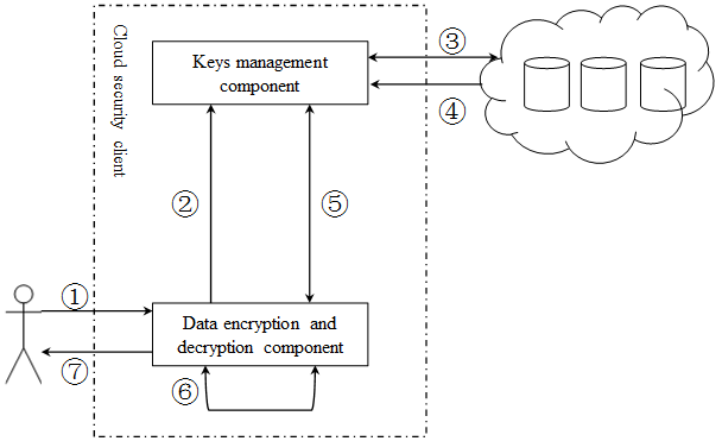
The concrete steps are as follows:

(1) Data encryption and decryption components request keys from keys management component;

(2) Keys management component retrieves information of keys and cipher text from cloud storage system;

(3)Get keys from data encryption and decryption component, and then encrypts the data.

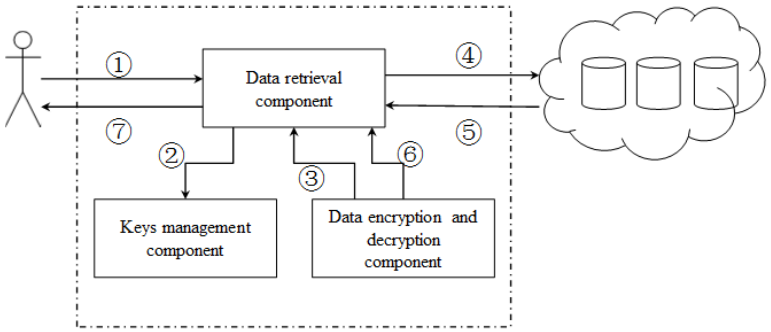
Data encryption and decryption are performed on client side, to avoid leakage information of users' data and key. With respect to the program of using encryption tools to encrypt data, the solutions proposed by this paper directly request keys and cipher text from the server, without need to download and decrypt all data simultaneously, thus saves network resources and improve the efficiency of computing resources. Figure 2 show the flowchart of the data encryption and decryption.



① Encryption and decryption request; ② Request key
③Request key and cipher text; ④Turn back keys and cipher text
⑤Using the master key to decrypt cipher text;
⑥Encryption and decryption by file keys;⑦ Return results.

Figure 2. The Flowchart of the Data Encryption and Decryption

3.3.2. Cipher text retrieval: The cloud users could retrieve data according to the properties of data, such as data submission time, size, type, keywords, *etc*. The concrete steps are as follows:



①Queries;②Request password;
③Attribute encryption;④Inquiry;
⑤Return the intermediate results;
⑥Decryption result;⑦Return the results;

Figure 3. The Cipher Text Retrieval Processes

(1) The index components in data security client rewrite the cloud users' query to cipher text database query, and send the command to security service system.

(2) The security service systems find all records that satisfy the query, and show the results to users. The results include actual data storage location in storage system. If SMEs users want to access data, they could download it based on the link. Figure 3 show the cipher text retrieval processes.

3.3.3. Massive Data Integrity Verification: When SMEs users need to verify some data contents, data validation component should interact with cloud services system. The concrete steps are as follows:

(1) Cloud services system return associated authentication meta information that stored in cipher text server.

(2) The SMEs users could access files stored in actual storage system according to the parameter from step1, and calculate and return the current status information of files. After received the information, data index component comprehensive analysis to determine whether the data is abnormal.

When SMEs user chooses the cipher text files to verify in application interfaces component, calls the data validation component to request the server-side data validation service component. The data validation service component gets the validation number of the file and credibility, according to these parameters randomly generated files reading location, and extracts examples from the cloud storage system, calculate the authentication tag and return to SMEs users. Meanwhile, the data validation service component returns the state label of files to SMEs users. Through the relationship between the two types of validation tags, the SMEs users determine whether the content of files was changed. Users could also specify one or several directories that require authentication, verify all batches of files in that directory.

4. Performance Analysis

In order to test the data security model, we build a cloud data security system, cloud storage servers include Hadoop distributed files system for actual storage system and HBase as cipher text database. Signature function is HMAC-SHA1 and symmetric encryption algorithm is AES. Cloud security service systems and web servers are deployed on two blade server.

4.1. The Efficiency of Encryption and Decryption

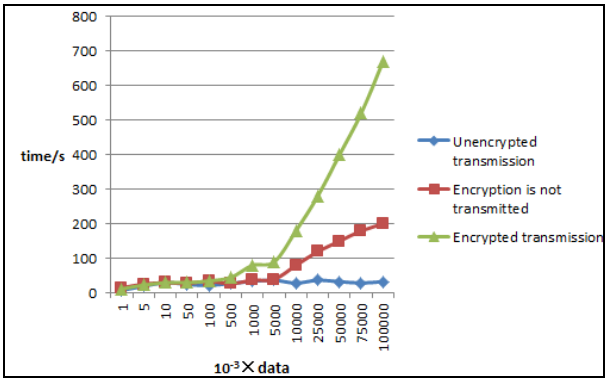


Figure 4. Data Encrypted Transmission and Unencrypted Transmission Time Comparison

Figure 4 shows that a different amount of data encrypted transmission and unencrypted transmission time comparison, unit is seconds (s). The experimental data is from $1\times10^{3} to 1\times10^{8}$.

Data transfer to cloud server from local site time-consuming include: attributes encrypted time, attributes uploaded time, data encryption and upload time. Unencrypted data transmission time does not include attribute encryption and data encryption time. Relative to the data, amount of data attributes is small and could be considered a constant value. The

network bandwidth of experiment, could be considered as a constant value. When the amount of data increases to a certain value, data encryption will spend most of the time, the total time consumed will be proportional with the amount of data.

The solution presented by this paper premises that the cloud service provider was incredible, adopts encryption mechanisms to protect the confidentiality of user data, meanwhile provides cipher text retrieval and integrity verification and other security services, efficient and flexible encryption mechanisms to ensure the confidentiality of users data system weak coupling, making access to a variety of cloud storage platform system. In addition, our model adopts open source technologies and services, using a symmetric encryption scheme. This model has more advantages when dealing with massive amounts of data.

4.2. The Efficiency of Cipher Text Retrieval

The test for cipher text retrieval, the horizontal axis is the number of files, the number of keywords that extracted from each files is five, and sets up 5 security index. The vertical axis is the search time, unit is second (s). Cipher text retrieval efficiency was shown in Figure 5.

As can be seen from Figure 5, with the increase in the number of files, the cipher text retrieval time increases linearly, could meet the users' requirements.

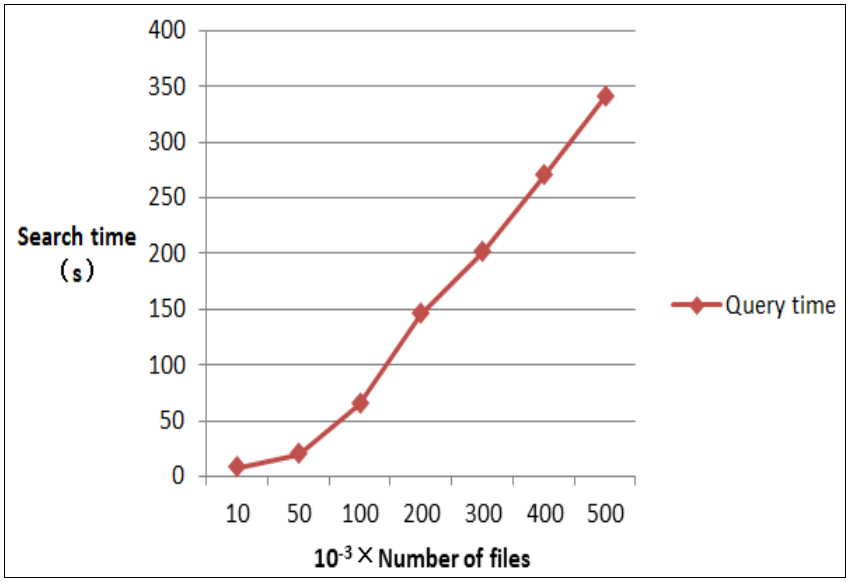


Figure 5. The Chart of Cipher Text Retrieval Time

4.3. The Efficiency of Integrity Verification

This part test for integrity verification, we verify the relationship of file sizes, the number of verification and the time required to generate credentials.

Figure 6 shows that the time required to generate labels almost was independent of the file size. When the file size increases from 10MB to 160MB, the time to generate a label increases less 1ms. This tiny growth was caused by the distance increases between different blocks. It is also shows that the selection of file blocks is random.

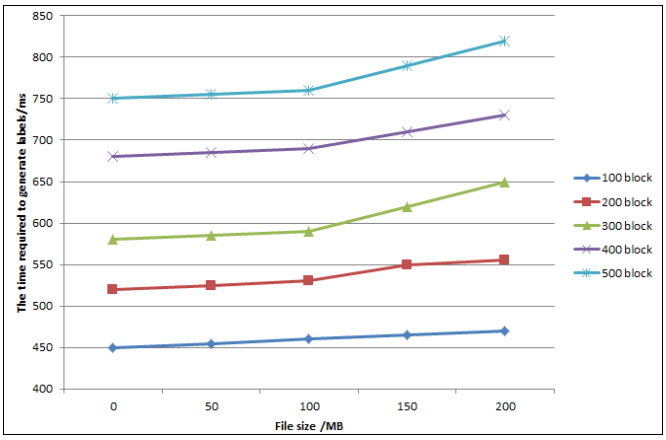


Figure 6. Relations between Files Size and Generate Authentication Credentials Time

Figure 7 shows that the generate labels time closely related with the number of file blocks. The time about different file takes may relates with the distance of file blocks.

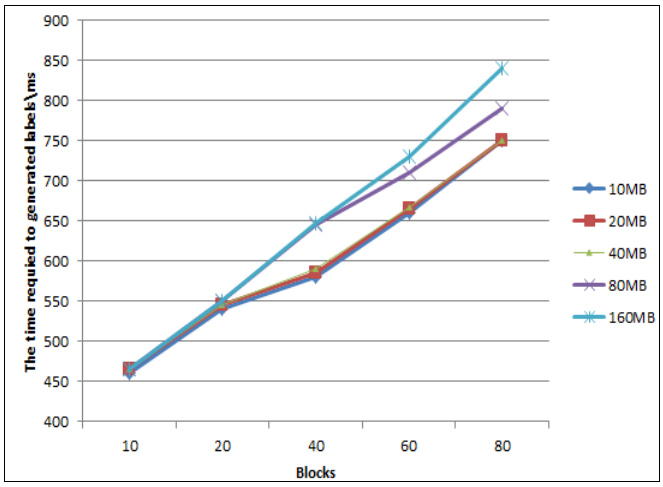


Figure 7. Relations between the Number of Verification and the Time of Authentication

5. Conclusions

In this paper, we propose a model to solve the problem of data security in cloud computing. The model adopts efficient encryption mechanisms to protect users' data. As the encrypted data is difficult to retrieve, we present a method of cipher text retrieval. To ensure the integrity of users' data, the users according to their own confidential information and the returned results could determine the massive data whether is integrity. Experimental results show that the platform can better ensure data confidentiality and integrity, and has good usability and scalability, in practice there are good technical value and application prospect.

Acknowledgements

This paper supported by the National Key Technologies R&D Program No.2012BAH54F04 and the Natural Science Foundation of Shandong Province of China under Grant No. ZR2010FM033.

References

[1] Primus. "Cloud computing for SMEs: bandwagon or good business strategy?", http://www.primustel.ca/en/business/news/articles/2011/2011-09-22_Cloud-computing-for-SMEs.php, **(2011)**.

[2] W. C. David, "Cloud computing Key Initiative Overview", http://www.gartner.com/resources/173600/173626/cloud_computing_key_initiati_173626.pdf, **(2010)**.

[3] D. Song, D. Wagner and A. Perrig, "Practical techniques for searching on encrypted data", IEEE Symp on Research in Security and Privacy. Los Alamitos, CA:IEEE Computer Society, **(2000)**, pp. 44-55.

[4] E. J. Goh, Secure indexes, 2003/216, IACR ePrint Cryptography Archive, **(2003)**.

[5] H. Hacigumus, B. Iyer, L. Chen and S. Mehrotra, "Executing SQL over encrypted data in the database service provider model", Proc of the 2002 ACM SIGMOD Int Conf on Management of Data (SIGMOD'2002). New York: ACM, **(2002)**, pp. 216-227.

[6] R. Agrawal, J. Kiernan and R. Srikant, "Order preserving encryption for numeric data", Proc of the 2004 ACM SIGMOD Int Conf on Management of Data(SIGMOD'04). New York: ACM, **(2004)**, pp. 563-574.

[7] G. Ateniese, R. Burns and R. Curtmola, "Provable data possession at untrusted stores", Proc of the 14th Conf on Computer and Communications Security (CCS'O7). New York: ACM, **(2007)**,pp. 598-609.

[8] A. Juels and B. A. Kaliski Jr., "PORs: Proofs of retrievability for large files", Proc of the 14th Conf on Computer and Communications Security (CCS'07). New York:ACM, pp. 584-597, **(2007)**.

[9] S. Kamara and K. Lauter, "Cryptographic cloud storage", LNCS 6054:Financial Cryptography and Data Security. Berlin: Springer, **(2010)**, pp. 136-149.

[10] M. van Dijk, C. Gentry, S. Halevi and V. Vaikuntanathan, "Fully Homomorphic Encryption over the Integers", Eurocrypt, **(2010)**.

[11] C. Gntry, "Fully homomorphic encryption using ideal lattices", Mitzenmacher M, ed. Proc. of the 2009 ACM Int'l Symp. On Theory of Computing. New York: Association for Computing Machinery, **(2009)**, pp. 169-178.

Authors

Tao Sun, is a PhD. Candidate of Shandong university and a staff of Qilu university of technology. His current research interests focus on data management for cloud computing.

Xinjun Wang, He is a professor and a doctoral supervisor of Computer Science and Technology Department, Shandong university. His research interests are database management, cloud computing and information integration. He has led more than 10 research projects.