Journal of Computers Vol. 30 No. 4, 2019, pp. 159-166
doi:10.3966/199115992019083004015

A Distributed Data Security Storage Method for Intelligent Transportation

He-Fei Zhang[1*], Yun Liu[2], Dijun Liu[3], Jian Bai[4]

[1] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China
Science and Technology on Communication Security Laboratory, Chengdu, China
17120160@bjtu.edu.cn

[2] Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing 100044, China
liuyun@bjtu.edu.cn

[3] Science and Technology on Communication Security Laboratory, Chengdu, China
flyinox31@gmail.com

[4] Science and Technology on Communication Security Laboratory, Chengdu, China
jianbaiscience@163.com

Received 13 January 2019; Revised 10 March 2019; Accepted 12 March 2019

Abstract. Fog computing is proposed for overcoming the shortage of cloud computing in isomerism, low-latency, dense network access and service requirements. Intelligent transportation is an important application of fog computing. However, fog computing is threatened by the security of sensitive data and large data storage due to the complexity of its structure and the limited storage capability. In this paper, we will introduce a new method to solve these issues. This method combines the Information Dispersal Algorithm and Cipher-text Policy Attribute-Based Encryption to encrypt data and divides cipher-text into pieces through a random matrix generated by a trusted proxy server. The fragments of cipher-text data are then stored in storage nodes. When data is acquired, data pieces will be reconstructed into cipher-text and then be decrypted to original data. The proposed method employs double encryption to ensure security of data and takes the advantage of distributed storage to solve the problem of insufficient storage capacity in fog computing.

Keywords: CP-ABE, distributed storage, fog computing, IDA, intelligent transportation

1 Introduction

With the development of the automotive industry, people's lives have become more convenient. The increase in the number of cars has also caused a series of problems such as road congestion. Intelligent transportation is one of the most important ways to improve the environment of the vehicle. Intelligent transportation system (ITS) is the development direction of the future transportation system. ITS is a kind of real-time, accurate, efficient and integrated transportation management system which is built on information technology, data communication transmission technology, electronic sensing technology and so on. The 21st century will be the century of intelligent road traffic, and the intelligent transportation system will be adopted. The system is an advanced integrated transportation management system. In this system, the vehicle is free to drive on the road by its own intelligence, and the road adjusts the traffic flow to the best state by its own intelligence. There is no doubt that the intelligent transportation system will generate massive amounts of data, which put pressure on computing services [1]. Cloud computing cannot meet the requirements of intelligent traffic in terms of low latency, and network congestion is

[*] Corresponding Author

becoming more and more obvious. In order to solve the problems, fog computing, a distributed service computing model arises, which combines network, computing, storage, and application [2].

The fog computing extends the cloud computing to the edge of the network, and solves the problems of poor mobility, weakly aware of geographic information, and high latency. Since the number of fog computing nodes is large and varied, there are many deficiencies in maintenance and supervision. Unlike cloud computing, which has huge storage capacity, the storage capacity of nodes in fog computing tends to be small, and the generation of large amounts of data poses a challenge to the secure storage. The complexity of the fog computing environment also threatens the security of sensitive data.

At present, the research on storage security of fog computing has just started. Current research on security storage focuses on cloud computing and mobile cloud computing. There is little research on fog computing, especially the intelligent data storage of ITS. Data security is especially important for intelligent transportation systems. It is necessary to deeply study the key technologies in the secure storage of fog computing based on the analysis of the environmental threat of fog computing. The threat of fog computing security storage can be divided into two categories, physical device damage and human malicious attacks.

Physical device damage. When the service node in the fog computing system fails due to physical equipment damage, whether the fog computing system can provide services for users, and maintaining the reliability will directly affect the user's data security. Literature [3] proposes a scale-free network model based on network size growth and preferential selection of large node connections, which name is the BA scale-free network model. The literature [4] compares the robustness of the ER random graph and the BA scale-free network in the robustness of the network after the total node is removed. Based on the two strategies of random failure and deliberate attack, the research team has been carried out from the point of random access to the network part of the node and the conscious removal of the most moderate part of the network. The results show that the BA scale-free network has a stronger ability to tolerate faults than the ER random graph.

Human malicious attacks: human malicious attacks can be studied from two aspects: active detection and passive prevention. In the aspect of active intrusion detection, the literature [5] pointed out that as a proactive security protection technology, the intrusion detection system can provide the network with the ability to prevent internal attacks and external attacks. Literature [6] proposed a multi-thread distributed intrusion detection system model. In this model, the cloud computing intrusion detection system can process large-flow data packets, analyze them, and then generate reports efficiently. Literature [7] proposed a cooperative intrusion detection system to improve the efficiency of intrusion detection in the face of coordinated attacks. Literature [8] proposed a network-based intrusion detection system in the cloud environment. By defining some columns of intrusion rules to judge the intrusion behavior, the system has a high detection rate when detecting external attacks.

In aspect of passive prevention, the literature [9] pointed out that password encryption technology is a common means to ensure the validity, integrity and confidentiality of data. There are many encryption methods, such as function encryption, searchable encryption, and so on. The definition of function encryption is given in literature [10]. In its definition, the key-generation algorithm generates the master public key and the master private key, wherein the master private key holder can generate a private function for any function f in the function family F. The given cipher-text and private key enable the private key holder to calculate the value of f(x) on the encrypted data. The security goal of function encryption is that for plaintext x, an attacker cannot obtain f(x) and any other information about x that is available from f(x). In literature [11], the confidentiality of the protection function f is proposed in the attribute-based encryption scheme, and an inner product encryption scheme that satisfies the confidentiality of the private key function is constructed.

2 Related Work

In this paper, we will introduce a new method to solve these issues. This method combines the Information Dispersal Algorithm and Cipher-text Policy Attribute-Based Encryption.

2.1 CP-ABE (Cipher-text Policy Attribute-Based Encryption)

Traditional encryption technologies often have clear acceptance targets, and it is difficult to satisfy one-to-many communication and flexible access control. In the Public Key Infrastructure (PKI) technology, publishers need to obtain the public key certificate of each recipient in order to encrypt the data. Although Broadcast-Encryption (BE) technology can realize one-to-many communication, it does not have flexible access control. If the number of recipients is too large, BE will generate a large overhead, and there is also a risk of leakage of the recipient identity information.

Attribute-based Encryption (ABE) can solve the key problems in fog computing: one-to-many communication and flexible access control. ABE uses a set of attributes to represent the user's identity, and attributes are treated as as the public/private key to encrypt data. The cipher-text can be decrypted only when the user's attributes meet requirements. ABE has the following characteristics [12]:

(1) The resource provider encrypts the data according to the attributes, and doesn't need to know the user, thereby protecting the privacy of the user;

(2) Users who meet the cipher-text attributes can decrypt the cipher-text, ensuring data confidentiality;

(3) The generation of key is related to a random polynomial or a random number, and the keys between different users cannot be combined to prevent the user's collusion attack;

(4) The mechanism supports flexible access control policies.

Attribute-based Encryption can be divided into Key Policy Attribute-Based Encryption (KP-ABE) and Cipher-text Policy Attribute-Based Encryption (CP-ABE) [13]. CP-ABE is closer to the real-world application scenes. It can be assumed that each user obtains a key from the attribute organization according to its own condition or attribute, and then the encrypted-person makes access control policy. As shown in the Fig. 1, the sender generates a cipher-text based on the access control policy, and the keys of the user 1 and user 2 which generated according to the attributes are {male, student, EN} and {female, teacher, MA}. User 1 satisfies the access control policy and then access the data. User 2 doesn't satisfy the access control policy then can't access the data.

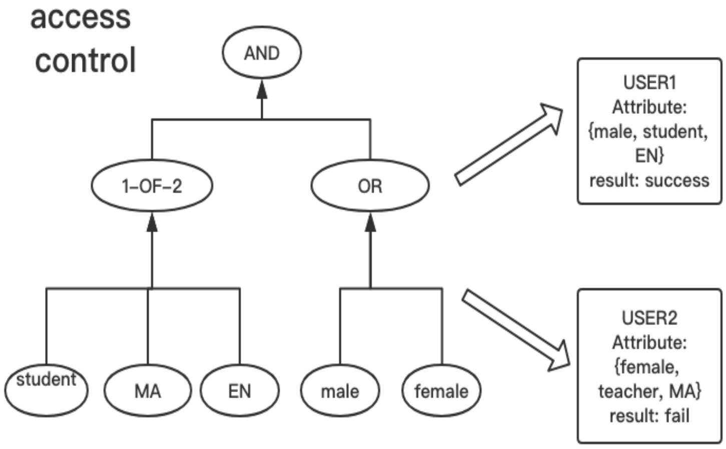

Fig. 1. CP-ABE scheme

A CP-ABE scheme usually consists the following algorithms: [14]

Setup. Initialization algorithm. Enter the security parameters and attribute domain parameters, then output the public key parameter PK and the master key MK.

Encryption. Encryption algorithm. Enter public key PK, plaintext M, access structure A, then output cipher-text CT.

Key generation. A key generation algorithm. Inputs the master key MK and the attribute set S and outputs the private key SK.

Decryption. Decryption algorithm. Input public key parameter PK, containing cipher-text CT accessing structure A, private key SK. If the S satisfies the access structure A, the algorithm completes the decryption and returns a plaintext message.

2.2 IDA (Information Dispersal Algorithm)

The Information Dispersal Algorithm is used to divide the data file into byte data slices. When the divided data pieces are transmitted or stored over the network, if the user or device does not have the legal key, they will not be able to access the data. Decentralized pieces of data can be recombined into originals using the legal key. File information dissemination algorithms have the ability to propagate data between different nodes in a very secure way, because an attacker may compromise a node but cannot compromise any data.

A file F, witch length is L, is divided into n pieces and each piece has the length of $\frac{L}{m}$. As long as m part of n is taken, the entire file can be reconstructed [15]. Suppose F is the original file, which is a byte queue of size N. The bytes in the F file can be divided into blocks containing m bytes.

$$F=(b_1,b_2,b_3,\ldots,b_m),(b_{m+1},b_{m+2},b_{m+3},\ldots,b_{2m}),(b_{m+1},b_{m+2},b_{m+3},\ldots,b_{2m}) \tag{1}$$

Suppose A is a matrix of n rows and m columns.

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \tag{2}$$

Assume that matrix B is a matrix of files F, and matrix C is an output matrix witch is transformed. Therefore, the following matrix can be obtained.

$$A\cdot\begin{bmatrix} b_1 & b_{m+1} & \cdots & b_{N-m+1} \\ b_2 & b_{m+2} & \cdots & b_{N-m+2} \\ \vdots & \vdots & \ddots & \vdots \\ b_m & b_{2m} & \cdots & b_N \end{bmatrix}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1\,N/m} \\ c_{21} & c_{22} & \cdots & c_{2\,N/m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{n\,N/m} \end{bmatrix} \tag{3}$$

Because c_{11} is equal to the product of the first row of matrix A and the first column of matrix B, the following equation can be obtained.

$$c_{11}=a_{11}b_1+a_{12}b_2+\cdots+a_{1m}b_m \tag{4}$$

Each line of C, witch is the file slice mapping relationship matrix, corresponds to one file piece.

Assume that the first to m piece of the file are used to reconstruct the original file. Then, the matrix A' has column 1 to column m as a submatrix of the matrix A. Assume that matrix A^{-1} is the transposed matrix of matrix A'. When reconstructing the original file, A^{-1} is used to convert the known file piece.

$$A^{-1}\cdot\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1\,N/m} \\ c_{21} & c_{22} & \cdots & c_{2\,N/m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{m\,N/m} \end{bmatrix}=\begin{bmatrix} b_1 & b_{m+1} & \cdots & b_{N-m+1} \\ b_2 & b_{m+2} & \cdots & b_{N-m+2} \\ \vdots & \vdots & \ddots & \vdots \\ b_m & b_{2m} & \cdots & b_N \end{bmatrix} \tag{5}$$

Finally, a partial file pieces can be successfully formed into the original file.

Some people have been studying the use of Information Dispersal Algorithm to replace traditional data encryption methods to ensure the security of transmitted data. Theoretical studies have shown that it can improve the security, integrity and usability of storage data.

3 Proposed Solution

The Distributed Data Security Storage Method proposed in this paper effectively solves the problem of sensitive data security and insufficient capacity of the fog storage nodes. The method introduced in this paper uses CP-ABE to encrypt data, solving the key problems in the two fog calculations of one-to-many communication and flexible access control. The IDA is used to slice the encrypted data, and the fragmented data is stored on different fog nodes respectively, which solves the problem of insufficient storage capacity of the nodes in the fog calculation. Due to the nature of the IDA algorithm, there is a guarantee in data security. A small amount of data loss can still restore the original data from the remaining data. A proxy server is introduced in the process of slicing, and the function of the proxy service will be explained in detail below.

As shown in the Fig. 2, it is a flow chart of Distributed Data Security Storage. The whole process is divided into an upload process and a download process. During the uploading process, firstly, the file is encrypted by CP-ABE to obtain the cipher-text. Then the trusted proxy server provides a random conversion matrix A, the cipher-text is divided into multiple pieces by the IDA. Finally, the file pieces are stored separately on the storage nodes. In the download process, no less than m pieces are acquired from the storage nodes. The legal user can obtain the corresponding conversion matrix from the proxy server, and synthesizes the cipher-text file. Finally, the legal user can decrypt the cipher-text according to the CP-ABE to obtain the plaintext.

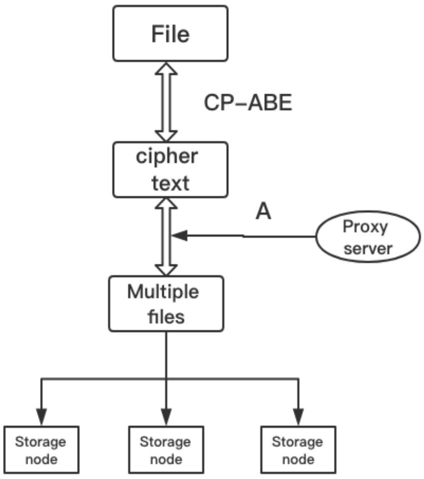

Fig. 1. Architecture of the system

The proxy server is a trusted server that has three main functions:

(1) Distributing random conversion matrix and maintaining the mapping relationship between the original data and the data pieces;

(2) Integrating the storage resources of the fog nodes and providing storage services for the nodes in the area;

(3) When obtaining data from the storage node, verifying the legality of the data requester according to the certificate. Only when the requester is legal, acquiring the cipher-text data pieces according to the mapping relationship, and distributing the conversion matrix.

The specific methods used in this paper are as follows:

(1) *The setup phase:* G_1 is a bilinear group whose order is prime p, g is the generator of G_1, bilinear mapping $e: G_1 \times G_1 \to G_2$, single function $H: \{0,1\}^* \to G_1$ is a predictor, and the size of the group element is determined by the system security parameters. Random selection $\alpha, a \in Z_p$, algorithm output public key parameter $PK = (g, e(g,g)^\alpha, g^a)$ and the master key $MSK = g^\alpha$.

(2) *The data encryption phase:* Algorithm input public key parameters, access structure (M, ρ) and plaintext $m \in G_2$, where ρ is a single shot function that associates each attribute in the access structure with a row that is shared into matrix M. Let matrix M have rows n columns, the algorithm first selects a set of random vectors $\vec{v} = (s, y_2, \ldots, y_n) \in Z_p^n$, selection $r_1, \ldots, r_1 \in Z_p$ randomly, compute $\lambda_i = \vec{v} \cdot M_i$, where $i = 1, \ldots 1$, and M_i represents the first row of matrix M. The algorithm outputs the cipher-text as follows:

$$CT = C(= me(g,g)^{\alpha s}, C = g^{s}, \{C_i = g^{\alpha\lambda_i} H(\rho(i))^{\gamma_i}, D_i = g^{\gamma_i}\}_{i=1,\ldots,l} \quad \textbf{(6)}$$

(3)*Data fragmentation stage:* the data is divided into n data blocks containing m bytes, and the matrix B is generated. The random conversion matrix A generated by the proxy server, and the matrix A and B are multiplied to obtain the data slice mapping relationship matrix C. Finally, the proxy server distributes the data slices on the storage nodes.

(4) *Private key generation phase:* The algorithm inputs the master key MSK and the attribute set S, selects $t \in Z_p$ randomly, and outputs the private key as follows:

$$SK = (K = g^{\alpha} g^{\alpha}, l = g^{t}, \{K_x = H(x)^{t}\}_{x \in s}) \quad \textbf{(7)}$$

The proxy server verifies whether the data requester identity is legitimate, and legally allows access to the data.

(5) *Data integration phase:* Select the sub-matrix A' of the matrix A corresponding to the file slice, and use A and the matrix file slice mapping relationship matrix C to obtain the matrix C that correspond to the cipher-text fragment, and finally obtain the cipher-text.

(6) *Decryption phase:* The algorithm inputs the cipher-text CT associated with the access structure (M, ρ), and the private key SK associated with the set of attributes S. S is assumed satisfying (M, ρ), defined $\{i : \rho(i) \in S\} C \{1, \ldots, 1\}$. According to the linear reconstruction characteristics of the access structure, we can find the constant set $\{\omega_i \in Z_p\}_{i \in I}$ satisfying $\Sigma_{i \in I} \omega_i \in \lambda_i = s$, where $\{\lambda_i\}$ is a group of secret s effective sharing. The algorithm first calculates:

$$e(C', K) / (\prod\nolimits_{i \in I} (e(C_i, L) e(D_i, K_{\rho(i)}))^{\omega_i}) = e(g,g)^{\alpha s} \quad \textbf{(8)}$$

Then the plaintext message m is restored from $C = me(g,g)^{\alpha s}$.

The above is the method of Distributed Data Security Storage which is introduced in this article. The CP-ABE guarantees the security of data generated in the intelligent transportation system, and also satisfies the two key problems of one-to-many communication and flexible access control in intelligent transportation. IDA uses the matrix A generated by the proxy server to fragment the encrypted data, and distributes the data that cannot be stored by a fog storage node to multiple fog nodes, which satisfies the requirements of massive data storage. When the data is acquired, even if the data partially stored in the node is lost or damaged, the original data can be restored. Since the data is encrypted and distributed, the illegal user cannot obtain mapping relationship of original data and data slice, also the random matrix generated by the proxy cannot be known. The original data cannot be obtained, even if a partially encrypted data segment can be obtained. As a result, the security is greatly improved.

4 System Evaluation

Intelligent transportation system has high requirements on the real-time and security of data. Cloud computing cannot meet the requirements of low latency and high quality of service in intelligent transportation systems. Applying fog computing to intelligent transportation systems solves the problems of high latency, low bandwidth, and poor service quality. However, the fog computing itself has its own shortcomings. Physical equipment damage and human attacks can affect the working conditions of the intelligent transportation system. The Distributed Data Security Storage Method solves the problem of the storage of massive data and the security of sensitive data. In terms of large-scale data storage, this solution uses IDA to be dispersed into data pieces, which solves the problem of insufficient storage capacity of nodes. At the same time, the encryption method of CP-ABE, which has one-to-many communication and flexible access control. In terms of sensitive data security, CP-ABE and IDA two data encryption methods achieve double encryption of data. Due to the particularity of the IDA algorithm, when some storage nodes are damaged such as physical device damage or human attacks, data can be recovered from other node storage nodes through the transformation matrix. The solution has certain fault tolerance.

In this paper, a 10426 bytes of data is encrypted and decrypted to verify the performance of the system, and the following figure is obtained.

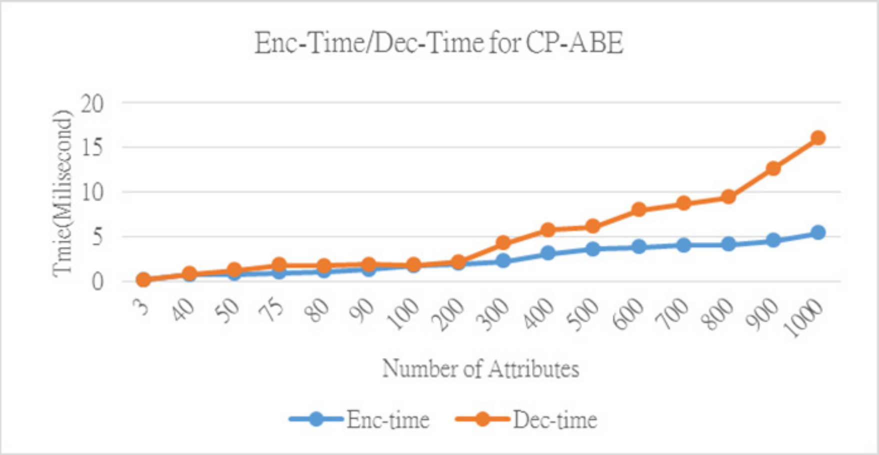

Fig. 3. Enc-time/Dec-time for CP-ABE

As can be seen from the figure, as the attributes increase, the security level of the data will become higher and higher. On the other hand, as the number of attributes increases, the computational overhead will also increase. In IDA, the ratio of the number of data fragments to the minimum number of pieces of data required to recover data also affects the security level and computational overhead.

5 Conclusion

In intelligent transportation systems, fog computing solves the shortcomings of cloud computing in terms of low latency, high bandwidth and quality of service. The fog computing itself has its own shortcomings, complicated working environment and diverse business, which makes the data in the fog computing have some challenge in terms of security and storage. In terms of security, physical device damage and human-induced malicious attacks can cause data security in ITS to be threatened. Due to the business requirements in ITS, one data often needs multiple devices or users to access. The traditional encryption method is difficult to meet the two key requirements of one-to-many communication and flexible access control. CP-ABE can meet these two key requirements. After CP-ABE encrypted files, illegal users cannot decrypt the data. IDA divides the encrypted file into a number of data blocks and stores the data blocks in different fog nodes, which solves the problem of limited storage capacity of a single node. Moreover, distributed data storage has advantages in terms of security and robustness over storage in a single node.

In CP-ABE, as the attributes in the attribute set increase, the security of the data increases, but this also increases the computational overhead. In IDA, the ratio of the number of data fragments to the minimum number of pieces of data required to recover data also affects the security level and computational overhead of the data. The computational power of a fog computing node is limited, and a complex calculation will degrade the performance of a node and affect the operation of the entire system. In this paper, the proxy server is a relatively weak link. When the fog node in the intelligent transportation system acts as an attacker to break the proxy server and obtain the mapping relationship inside the server, there is a certain possibility to obtain the cipher-text data. Therefore, in the future research on distributed data security storage of intelligent transportation systems and fog computing, the main research objectives are divided into two: first, optimize CP-ABE and IDA, such as determined $\frac{n}{m}$ by security level or according to the security level determines the number of attributes in the attribute set. In different services, CP-ABE and IDA with different computational overheads are used for different security requirements. Secondly, the security of the proxy server is enhanced, and the node attributes are strictly divided for the nodes of different services to prevent illegal nodes from decrypting cipher-texts through access policies to obtain data.

Acknowledgements

This work is supported by the National Key R&D Program of China under Grant 2016YFC0801004, the Fundamental Research Funds for the Central Universities, grant number 2017JBZ107 and the National Natural Science Foundation of China under grant 61271308.

References

[1] D. George, D. Panagiotis, Intelligent transportation systems, IEEE Vehicular Technology Magazine 5(1)(2010) 77-84.

[2] E. Mohammad, M. Aazam, M.S. Hilaire, Using DEVS for modeling and simulating a fog computing environment, in: Proc. 2017 International Conference on Computing, 2017.

[3] A.L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286(5439)(1999) 509-512.

[4] R. Albert, H. Jeong, A.L. Barabasi, Error and attack tolerance of complex networks, Nature 406(2000) 378-382.

[5] H. Mohamed, L. Adil, T. Saida, A collaborative intrusion detection and Prevention System in Cloud Computing, in: Proc. 2014 IEEE Africon, 2014.

[6] H. Li, Q.-X. Wu, A distributed intrusion detection model based on cloud theory, in: Proc. 2013 IEEE International Conference on Cloud Computing and Intelligent Systems, 2013.

[7] C.-F. Zhou, C. Leckie, S. Karunasekera, A survey of coordinated attacks and collaborative intrusion detection, Computers and Security 29(1)(2010) 124-140.

[8] M. Claudio, R. Bifulco, R. Canonico, Integrating a network IDS into an open source Cloud Computing environment, in: Proc. 2010 International Conference on Information Assurance and Security, 2010.

[9] D. Maimut, R. Reyhanitabar, Authenticated Encryption: Toward Next-Generation Algorithms, IEEE Security and Privacy, 12(2)(2014) 70-72.

[10] D. Boneh, A. Sahai, B. Waters, Functional Encryption: Definitions and Challenges, in: Proc. 2011 Theory of Cryptography Conference, 2011.

[11] E. Shen, E. Shi, B. Waters, Predicate Privacy in Encryption Systems, in: Proc. 2008 Theory of Cryptography Conference, 2008.

[12] J. Bethencourt, A. Sahai, B. Waters, Ciphertext-Policy Attribute-Based Encryption, in: Proc. 2007 IEEE Symposium on Security and Privacy, 2007.

[13] B. Waters, Ciphertext-policy attribute-based encryption: an expressive, efficient, and provably secure realization, in: Proc. 2008 International Workshop on Public Key Cryptography, 2008.

[14] Y.-L. Ren, S.-Z. Wang, X.-P. Zhang, Z.-X. Qian, Fully secure ciphertext-policy attribute-based encryption with constant size ciphertext, in: Proc. 2011 Third International Conference on Multimedia Information Networking and Security, 2011.

[15] J.-L. Sian, H.-C. Wei, An efficient (n, k) information dispersal algorithm based on fermat number transforms, IEEE Transactions on Information Forensics and Security 8(8)(2013) 1371-1383.