An Efficient Approach for Encrypted File Sharing and Anomaly Detection in Cloud Storage

[1]Miss. Neha A. Kale, [2]Miss. Snehal S. Kshatriya

[1,2]Assistant Professor, Computer Engineering Department, SOCSE, Nashik, India.

[1]nehak3781@gmail.com, [2]kshatriyasnehal1012@gmail.com

Abstract—Cloud Computing is the emerging technology that combines the concept of Software-as-a-Service and Utility Com-puting, provides the on-demand services to the end users. In cloud computing security is the important aspect and has various issues and problem. Nowadays many organizations are moving their data on the cloud, by using File Syncing and Sharing Services. End users uses their own devices to access the data and due to this there is rise in the new challenge for preventing the player/decoder abuse. In this paper a secure, efficient File Syncing and Sharing Service on cloud based on digital forensic mechanism is been developed. A group oriented cryptosystem is also developed called as PHE that is Partially-Ordered Hierar-chical Encryption which implements partial order key hierarchy. This hierarchy is similar to role hierarchy used in Hierarchical Role Based Access Control (HRBAC). This paper also introduces anomaly detection by using pattern matching. This anomaly detection will identify the suspected players and will trace and revoke the authorities of the suspected players. The performance of the system shows that this construction is more secure, highly efficient and have some features such as lower communication overload, lower computational and storage cost.

Index Terms—Audit, File Syncing and Sharing, Pattern Match-ing, Revocation, Risk Assessment, Traitor Tracing

I. INTRODUCTION

Cloud Computing is technology that provides different ser-vices to the users and it is used to manipulate, configure, and access the resources remotely. Cloud offers different services such as data storage, infrastructure, and application. We can access the data or services from any location at any point of time. There are different services of cloud that are available such as Box, Drop Box, Sky Drive, Sugar sync for individual and small to medium business. These cloud storage provide on services that are low cost, capacity on demand and long term archive. The cloud storage and services is very helpful to the users and convenient too because users can share and access the data from any location,at any point of time, by using their personal devices such as computers, mobile phones, etc. so many organizations and individuals have moved their personal data, large archive system into cloud storage. Cloud computing is been used by many organizations, institutes and also its been useful for government[1].

The cloud storage service is used by individual users who can store the data and download it to sync and collaborate. So cloud storage provide FSS that is file Syncing and Sharing Services. In File Sharing, users can access the files which are located anywhere, at any time from different end devices, and can also edit the files together. File Syncing is a backup mechanism that is used for syncing the data or information across various devices such as personal computer, smart phone etc [1].

Cloud computing model combines concept of software as a service(SaaS) and utility computing. It provides convenient and on demand services to the users. Cloud service providers provides different services to the end users. The providers and the customers that uses the services have to make it sure that data or information is kept secure from the external entities or attacks so that the data do not get lost [3].

Cloud is network of computers that are connected to each other in same or different geographical locations, that operates together to provide services to the customers that are having different need and workload on the demand basis. The services of cloud are provided to the users on the basis of usage that is user will pay as per the use and demand. These services are in the form of platform, Infrastructure or software [4].

The multi-tenant nature of cloud makes it vulnerable to data leakage, threats and malicious attacks. So there must some kind of policies that the enterprise must have such as Role Based Access Control or Attribute Based Access Control [5] so that the privacy and secrecy can be maintained.

As the data on the cloud is sensitive it needs to be encrypted before storing it on the websites, but there is one drawback of encrypting data, the data will be shared at

 DOI : 10.35291/2454-9150.2019.0540

coarse grain level ie giving your private key to another party. So, to overcome this problem, ABE with fine grain access control is introduced, in which the cipher text is given a label of attributes and private keys are associated with access structure [6].

In role based access control(RBAC) every user is assigned with the roles that he wants to perform and by assigning these roles to the users, the unauthorized access will be restricted. It is also called as role based security. This policy is defined around the roles and privileges. RBAC is used for administrating the security of various large organizations. The roles in RBAC represents the job function within the organization and authorization is granted to the role instead of single user. The authorization is restricted to data objects and resources needed by college. Users are authorized to play the appropriate role. This role in RBAC can support security policies of the organization. Role hierarchy is the means where the structuring roles to reflect the structure of the organization. For this a partial order relation is applied on the roles known as role hierarchy [7].

II. LITERATURE SURVEY

Santosh Kumar and R. H. Goudar[3] described a survey of cloud computing which includes the security issues in cloud, various challenges, platform provided by cloud, its architecture and applications.

Prince Jain[4] presented a survey of various issues and challenges in cloud computing and the possible solutions to overcome these challenges.

Zhi Qiao, et al[5] presented a scheme called attribute based encryption(ABE) which uses the set of users attribute to generate the secret key and it also contains the access structure that performs the access control of the system. It is basically a combination of encryption and access control and used for secret sharing between users and groups and specially used in cloud environment.

John Bethencourt et al[6] developed Ciphertext Policy Attribute Based Encryption scheme. The CPABE is a technique in which the ciphertext are associated with the access policies and the attributes are shared with each other. CP-ABE has the fine grain access control over the data that is to be encrypted. But there is a concern about the CP-ABE method, the malicious user can share the attribute with other users which cause the leakage of decryption privileges for financial gain.

Zhen Liu, Zhenfu Cao[7] developed Traceable CPABE to overcome previous problem. It consists of 2 levels, White box traceability and black box traceability. Whitebox traceability scheme may not able to find the malicious users that make the decryption black box. But if the decryption blackbox is given and the tracing algorithm and decryption key is kept hidden then blackbox traceability can find the malicious user.

Michel Abdalla1 et al[8] described a scheme called Identity based traitor tracing which is the first tracing scheme. The security of this scheme is in the standard model, by making the assumption that the bilinear Decision Diffie-Hellman(DBDH) is difficult in the asymmetric bilinear pairing. The DDH defines the first co-ordinates of asymmetric pairing in the group. This scheme allows the tracing of the adaptive pairs.

Dan Boneh, Matthew Frankliny[9] developed a scheme called identity-based encryption (IBE). This scheme is fully functional IBE. The security of this scheme is defined in the random oracle model. This model is assumed to be a variant of the computational Diffie-Hellman problem. The working of this scheme is based on bilinear maps among the groups.

Dan Boneh et al[10] implemented a system called HIBE that is the hierarchical identity based encryption. This system has the constant ciphertext size which means that there are three elements only and only 2 bilinear map computations are required for decryption, regardless of the depth of hierarchy.

Dan Boneh Matthew Franklin[11] described a scheme which is efficient called Public Key Traitor Tracing. In this system there is one encryption key and many decryption keys. When-ever some contents such as audio or video clip is encrypted using public key and this contents are broadcasted over a channel, then every authorized user can read it by using their own decryption key. If an unauthorized user tries to create a decryption key to decrypt the contents then there is an efficient algorithm that will trace the creator of new key. This scheme is simple and provides efficient solution to the problem of traitor tracing which catches all the traitors and never accuses the authorized users.

Dan Boneh, BrentWaters[12] presented a scheme called Broadcast, Trace, and Revoke System which is fully resistant to collusions.In this system a Augmented Broadcast Encryp-tion[ABE] is developed.This scheme is helpfull to implement a traitor tracing, revoke and broadcast encryption. This ABE system is resistant to any numbers of collusions and it is secure agaisnt the abnormal or illegal users.

Dan Boneh, Amit Sahai et al[13] developed a traitor tracing system. This scheme is fully secure against collusion and have sublinear ciphertext and the size of the private key is constant. It generates O(N) size of ciphertext and O(1) size of private keys. In this model one primitive is developed called PLBE that is private bilinear broadcast encryption. This shows how to build a PLBE with O(N) ciphertext size and it also shows that any PLBE primitive gives a traitor tracing scheme with same parameters.

 DOI : 10.35291/2454-9150.2019.0540

III. KEY CONTRIBUTION

A more secure and efficient File Syncing and Sharing service(FSS) on cloud, based on digital forensic mechanism against abnormal player is developed. In the proposed system we implemented anomaly detection technique to detect the anomolous user in the system. Anomaly detection techniques uses pattern matchning scheme by using which we can detect the abnormal players. This pattern matching scheme will detect the suspicious players by analyzing the users behavoiur. Once the abnormal players are detected it will be traced and their authorities will be revoked. A threshold cryptosystem is also been implemented called Partial-Ordered Hierarchical Encryption(PHE) which provides two security mechanism called Traitor Tracing and Revocation.

IV. PROPOSED FRAMEWORK

The Figure shows the model that represents the proposed system and can be built using various types of components: End users: Users are the one who wants to access the files or the data stored on the cloud. Users can access the file from anywhere and from any device.

Player/Decoder: The player/decoder are the one whose provides the interface to access the data on the cloud.

FSS Service: It provides users with the ability to access the data remotely on the cloud. It includes file sharing and syncing. Anomaly Detection: It is used to detect the suspicious or abnormal players. In other words it is used to monitor the deployed resources and it returns the abnormal players as the output.

Traitor Tracing: This term is used for finding out the traitors from the suspected players that are detected using anomaly detection.

Revocation: This is responsible to revoke the authorities of the traitors found in traitor tracing step.

We present a new FSS model for player abuse prevention and enhanced protection against unauthorized access. The proposed model uses the hierarchical role-based access control (H-RBAC) model, which is recognized for its support for simplified administration and scalability of collaboration and working with teams. This model also supports other access control policies such as discretionary access control and multilevel security and hence it is more generic.The figure shows various components including FSS, Anomaly detection, Traitor Tracing and Revocation.

The anomaly detection technique will be applied on the players behaviour and by using pattern matching scheme, the anomaly or the abnormal player will be detected. Aho-Corasick algorithm which is a pattern matching algorithm is applied and by using this algorithm the abnormal players are detected. Once the abnormal players are detected, they are traced and after tracing the authorities or the license of the abnormal players are cancelled and the system is protected from the attack.

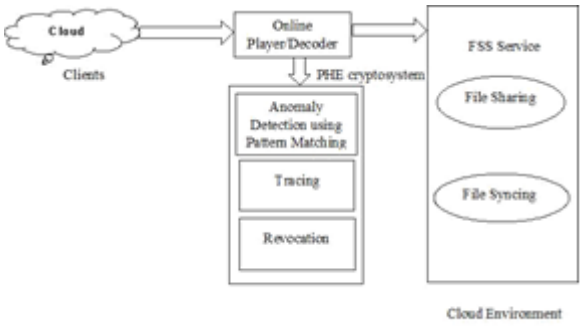

Fig. 1. System Overview

V. METHODOLOGY

A. Partially-Ordered Hierarchical Encryption (PHE)

It is a scheme that is similar to role hierarchy in role based access control and it implements partial order hierarchy. It is a 6 tuple scheme such as:

PHE(S,J,E,D,T,R).

Setup: Setup is the first tuple. It takes the input parameters such as partial order hierarchy that is omega, security parameter s, maximum collusion number t. It outputs the encryption key that is required for encrypting the data,set of public parameter p and a master key that will be the manager secret.

Join: It includes two algorithms. one for joining the individual user and the other one for joining group of users.

Encrypt: This tuple is used for encrypting the data stored on the cloud. It encrypts the message M and outputs the ciphertext.

Decrypt: The ciphertext is decrypted using the decryption key and output of this step is the original message

Trace: It is used to trace the abnormal players and detect the traitor from the players.

Revoke: This tuple will cancel all the authorities of the abnormal players and revoke them.

B. Elliptic Curve Digital Signature Algorithm

A Digital Signature is used to check that whether the information is been altered or not. It is a digital signature on electronic document that is generated by cryptographic technology. It checks if information is modified after it is signed. There are two algorithms for ECDSA.

1) Signing Process by Sender
2) Signature Verification Process by Receiver

In signing process the message digest is created of the message by using message digest algorithm. Encryption algorithm is applied to message digest by using senders private key. After the sender have created a digital

 DOI : 10.35291/2454-9150.2019.0540

signature of the message, the enrypted message digest is send to the receiver for verification process

In signature verification process the encrypted message digest is decrypted using senders public key and if the message digest is identical of sender and receiver then it means that message is not altered and received as it is send.

C. Hierarchical Role Based Access Control (HRBAC)

Access control is a policy which restricts access of data to authorised user. Many organizations uses this policy to restrict the access. In RBAC the role are assigned to the users and according to the role user can perform the operation on the data. There are three components of RBAC such as role-permission, role-role, and user-role relation. These components makes user assignments easier and simple.

The roles are assigned to the user by the system. A Hier-archy in RBAC is or automata called trie. The trie consists of links between different nodes. This links performs fast transition between the string matches that are failed to the other branches of the tree that have same or common prefix. Backtracking is not needed in this algorithm. It is like a dictionary that contains various the organizations role structure. It support the inheritance relationship of permissions. Hierarchy is the partial order relation between the different roles. The senior most role is on the top and the junior roles are at the bottom. The senior roles acquires the permission of juniors and the juniors get the membership of their seniors.

D. Aho- Corasick Algorithm

Aho-Corasick is a string searching algorithm. It is used to find out the string. This algorithm constructs a finite state machine strings, all these strings are matched simultaneously.

TABLE I PERFORMANCE OF SYSTEM IN TIME UNDER DIFFERENT THRESHOLD VALUES

Value of threshold	Setup algo	Join group algo	Encrypt algo
5	0.9	0.5	0.85
10	0.78	0.68	0.89
15	0.82	0.72	0.92
20	0.96	0.75	0.96

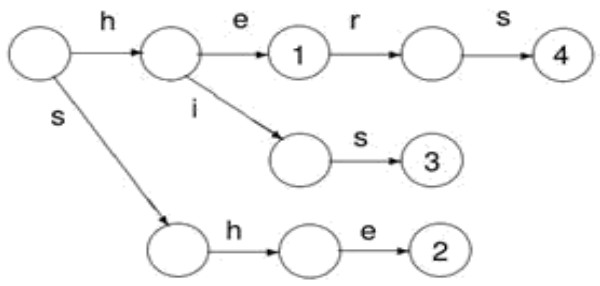

Fig. 2. Keyword tree

A keyword tree for P = fhe, she, his, hersg

Figure 2 shows the trie for patterns he, she, his and hers. A keyword tree is the construction of dictionary of strings. The construction of the tree begins with the root node. And go on inserting the characters or patterns one after the other.

There are 2 steps in the keyword tree one is preprocessing and second one is matching, In preprocessing there are 3 sub steps they are 1. Goto 2. Failure and 3. Output.

In go to step we build a trie and for all the characters that dont have edge at the root, an edge is added back to the root.

Failure: Failure function is used to store all the edges that are followed in trie when the current character do not have any edge in the tried Output: This function stores all the words that end at current state.

VI. METHODOLOGY OF EVALUATION

We have implemented our scheme in java and experiment was run on a small cloud based on open-source Open stack platform. Based on this platform a secure and efficient File Syncing and Sharing Service(FSS) is simulated. The perfor-mance of the system is evaluated for large number of user through some simple experiments.

A. Evaluation Parameter

In our experimental evaluation, we analyse the system performance under varying parameters, such as scalability, computational cost, Storage, anomaly detection, access control and time.

B. Experimental Setup

Our experiments were implemented in Java, Eclipse is used as integrated development environment and carried out on a PC with processor Pentium-IV and 2GB RAM.

C. Results and Discussion

We tested the time consumption of each function in PHE under the different value of thresholds (from 5 to 20). Table 1 shows the result of experiments of three functions, Setup, Join Group and Encrypt. From this table, it is easy to see that the time consumption of all functions grows large with the increase of threshold value, but the change is not obvious.

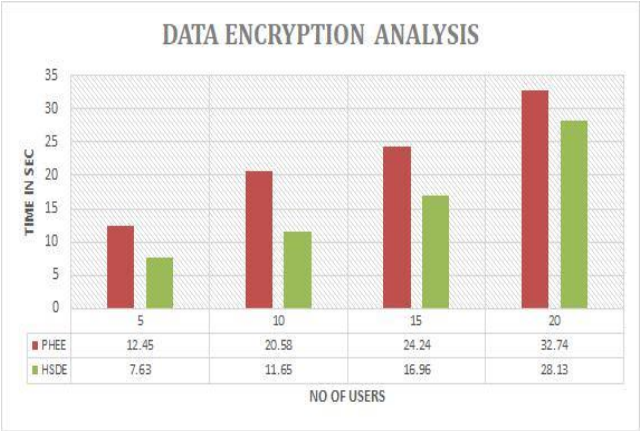

Fig. 3. Time of PHE and HSDE

Figure 3 shows the results based on time comparison of PHE system and the proposed system. The X axis shows the no of users and Y axis depicts the time in seconds. Different threshold values

 DOI : 10.35291/2454-9150.2019.0540

are taken and depending on that value the result is been calculated which shows that the time required is less as compared to the existing system. The average time of traitor tracing was less than 10 seconds for finding at least one traitor and the time of revocation was close to that of encryption.

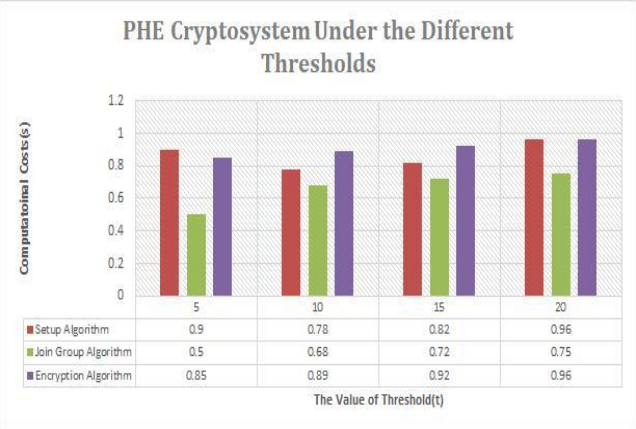

Fig. 4. Computational cost

Figure 4 shows the comparison result of computational cost between the PHE system and the HSDE system. The X axis of the graph indicates the value of threshold and Y axis shows the cost required for computation. The computational cost is calculated for three algorithm setup, Join group and Encrypt algorithm. By taking few threshold values, the cost is been calculated which shows that the computational cost is lower as compared to the existing system.

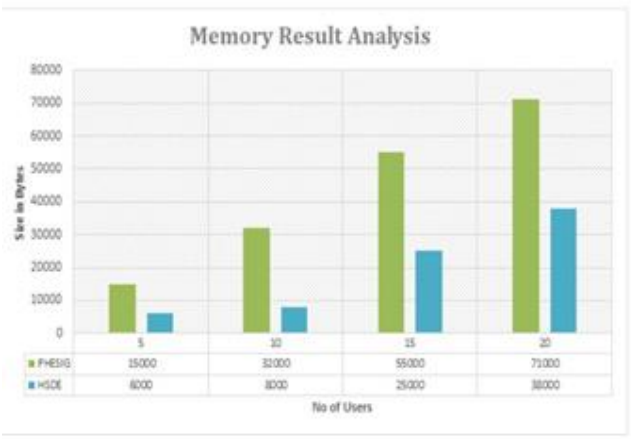

Fig. 5. Storage Cost

Storage overload is an important requirement for large scale information system. In this system we generate different keys for the users and these keys need to be stored in the system. The storage require to store the different keys is lower as compared to the existing system. The graph shows the comparison between the existing system and the Proposed one that is the PHE and the HSDE system. The X axis shows the no of users and Y axis depicts the size in bytes. The result clearly shows that the storage memory required for HSDE system is less as compared to the PHE system. The proposed system also provides scalability for the interoperability to the large no of organization as well as lower communication overload to the large no of users. Also the cost of traitor tracing and revocation was tested on small HRBAC system.

Anomaly Detection is the process of identifying the anoma-lous users in the system which do not conform to an expected pattern or other items in a dataset. It is the technique which will determine the anomalies in the given dataset by analyzing the behavioural pattern. The graph represents the anomaly detection accuracy of the existing system and the proposed system. By taking different values, the accuracy of attack detection is been shown. The proposed system gives high accuracy and attack detection as compared to the existing system.

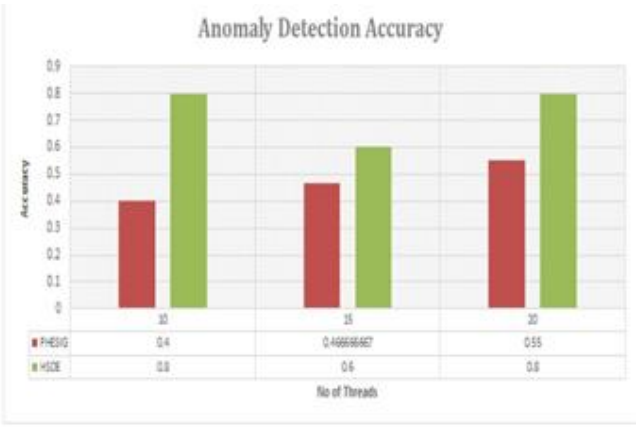

Fig. 6. Anomaly Detection Accuracy

VII. CONCLUSION

Cloud storage is increasing day by day so it needs security mechanism to protect the data from being misuse or being leak, so to provide security to the data, a group oriented cryptosystem called PHE(Partially-Ordered Hierarchical Encryption) is developed with digital forensics that focuses on protection and privacy of the outsourcing data and prevent the abuse of abnormal players in FSS by using methods such as tracing and revoking to ensure the security of the player/decoder. This model provides two security mechanism they are traitor tracing and revocation and it will enhance the performance of the system. Anomaly detection technique using pattern matching is developed, where the user's behaviour is identified by observing the pattern and if an abnormal player is trying to attack then false alarm is raised which will detect the abnormal player and will trace and revoke the authority of the anomalous user. For pattern matching Aho-Corasick Algorithm is used that will search for the patterns. This method is useful for providing security to the user's data over cloud.

REFERENCES

[1] Yan Zhu, Guohua Gan, Ruiqi Guo, and Dijiang Huang, "PHE: An Efficient Traitor Tracing and Revocation for Encrypted File Syncing-and-Sharing in Cloud", IEEE Transaction on Cloud Computing,2016.

[2] F. R. Institute, "Personal data in the cloud: A global survey of consumer attitudes", 2010.

[3] Santosh Kumar and R. H. Goudar, "Cloud Computing Research Issues, Challenges, Architecture, Platforms and Applications: A Survey", Inter-national Journal of Future Computer and Communication, Vol. 1, No. 4, December 2012.

[4] Prince Jain, "Security Issues and their Solution in Cloud Computing", International Journal of Computing Business Research,2012

 DOI : 10.35291/2454-9150.2019.0540

[5] Zhi Qiao, Shuwen Liang, Spencer Davis and Hai Jiang," Survey of Attribute Based Encryption", IEEE Conference,2014

[6] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data", in ACM Conference on CCS,pp.8998,2006

[7] A. Fiat and M. Naor, "Broadcast encryption", in Advances in Cryptology (CRYPTO93), vol. 773 of LNCS. springer-verlag, pp. 480491,1994

[8] Zhen Liu, Zhenfu Cao, "Traceable CP-ABE: How to Trace Decryption Devices Found in the Wild", IEEE Transaction on Information Forensics and Security, Vol. 10, N0. 1, January 2015

[9] Christian D.Peer, Dominik Engel, Stephen B.Wicker, "Hierarchical Key Management for multi-resolution Load Data Representation", IEEE In-ternational Conference on Smart Grid Communications, 2014

[10] Christian D.Peer, Dominik Engel, Stephen B.Wicker, "Hierarchical Key Management for multi-resolution Load Data Representation", IEEE International Conference on Smart Grid Communications, 2014

[11] David F. Ferraiolo, Janet A. Cugini, D. Richard Kuhn, "Role-Based ccess Control (RBAC): Features and Motivations".

[12] H. Chung, J. Park, S. Lee, and C. Kang, "Digital forensic investigation of cloud storage services", Digital Investigation, vol. 9, no. 2, pp. 8195, 2012.

[13] Chen, S. Nyemba, and B. Malin, "Detecting anomalous insiders in collaborative information systems", Dependable and Secure Computing, IEEE Transactions on, vol. 9, no. 3, pp. 332344, May 2012

[14] M. Blanton and K. B. Frikken, "Efficient Multi-dimensional key man-agement in broadcast services", in ESORICS, pp. 424 440,2010

[15] S. Garg, A. Kumarasubramanian, A. Sahai, and B. Waters, "Building efficient fully collusion-resilient traitor tracing and revocation schemes", in Proceedings of the 17th ACM Conference on Computer and Commu-nications Security, CCS 2010, Chicago, Illinois, USA, October 4-8, pp. 121130,2010.

[16] M. J. Atallah, M. Blanton, N. Fazio, and K. B. Frikken, "Dynamic and efficient key management for access hierarchies", ACM Trans. Inf. Syst. Secur., vol. 12, no. 3, 2009.

[17] E. Bertino, N. Shang, and S. Wagstaff, "An efficient time-bound hier-archical key management scheme for secure broadcasting", IEEE Trans. on Dependable and Secure Computing, vol. 5, no. 2, pp. 6570, 2008

[18] R. Ostrovsky, A. Sahai, and B.Waters, "Attribute-based encryption with nonmonotonic access structures", in ACM Conference on Computer and Communications Security, pp. 195203,2007.

[19] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption", in IEEE Symposium on Security and Privacy, pp. 321334,2007.

[20] A. D. Santis, A. L. Ferrara, and B. Masucci, "Efficient provably secure hierarchical key assignment schemes", in MFCS, pp. 371382,2007.

 DOI : 10.35291/2454-9150.2019.0540