Data Security in Cloud computing using Fuzzy Logic and Mining

R.PushpaLakshmi
Department of Information Technology, PSNA College of Engineering & Technology
Dindigul-624622, Tamilnadu, India
pushparaman@rediffmail.com

***Abstract*— Today cloud computing is used in many fields such as finance, business, entertainment etc. One of the main challenges in the cloud environment is to ensure secure communication, as the user's data are stored on a remote computer provided by the cloud provider. Cloud user data must be protected against unauthorized access and modification of data. The proposed security mechanism mainly focuses on increasing the security level of data shared by cloud user. It applies fuzzy analysis to evaluate the trust level of cloud users. Cloud user's trust is evaluated using the following parameters: number of bytes of data from cloud service provider to cloud user, duration of access to the cloud system, timing of visit by the user, IP address used by the user for cloud access, and the number of threads occupied by the cloud user. Data transmitted between the cloud user and the provider or between the cloud users are encrypted using the session key generated by the end parties. The session key is generated based on trust value of the cloud users and their frequent access pattern. The proposed method applies data mining concept to extract the frequent access pattern of cloud users. The frequent access pattern is identified by mining the past cloud user's access pattern by using backtracking search algorithm. The proposed scheme provides a mechanism for key generation, distribution and revocation in cloud environment.**

***Keywords*— Security, Privacy, Trust, Cloud computing, fuzzy analysis, pattern mining.**

I. INTRODUCTION

Many organizations use cloud computing [1] today directly or indirectly to reduce the cost involved in deploying infrastructure and applications. One of the main applications of cloud is data storage. Cloud user can access their applications and data from anywhere at any time. The cost spent by the user in purchasing hardware is reduced by allowing the user to work using inexpensive terminal. Cloud allows the private sectors and normal users to access software without buying software license. It rent high storage devices for data storage thus removing the need for physical space on the front end. It reduces the cost of network and storage infrastructure.

As the data is stored remotely and virtually, it can be easily affected by hackers or attackers. So it is necessary to have data privacy protection in cloud environment. There is a considerable interest in the development of cloud security mechanisms for data protection. Several researchers have proposed different encryption techniques for cloud computing such as hardware based encryption, software based encryption, transparent file encryption, full disk encryption, whole disk encryption etc. The primary problem is to investigate the security issues exists in cloud computing and related security mechanisms. The proposed Fuzzy associated trust based data security mechanism improves the level of data security by mining user behavior in cloud environment. The method is based on evaluation of trust degree of cloud user based on their behavior. It uses encryption mechanism based on trust degree and user's behavior. Frequent pattern mining is applied to derive frequent cloud user behavior. The proposed framework can satisfy all security requirements such as confidentiality, integrity, availability, scalability, authentication, and authorization.

Most of the existing cryptography mechanism [2] used for information sharing mainly focused on key generation based on identities of cloud users. But in real time application, identities can be easily hacked by attackers and can be used to hack the information shared by other users. The proposed work evaluates trustworthiness of CU and CSP based on CU's behavior. Data is protected by using the key generated based on user trustworthiness and their frequent behavior in the cloud environment.

A. Cloud Security Issues

Cloud services and models have various security issues. It is necessary to satisfy the security requirements such as confidentiality, integrity, authenticity, accountability, non-repudiation. Confidentiality prevents unauthorized access and modification of data. As the cloud user access remote shareable resources in cloud environment, trust place a vital role in cloud computing. Trusting on-line services is harder than trusting off-line services because of lack of centralized control [3]. The distrust of on-line service will cause a negative effect on the truthness of the organization [4]. Cloud computing entails the following privacy issues: lack of user control, unauthorized resource usage, data flow restrictions, litigation and legal uncertainty. Fig. 1 represents the challenges of cloud computing. Top Threats identified in the Cloud computing are as follows:

- Multi-tenancy in cloud permits resource sharing across multiple cloud users. Sharing of resources affects data confidentiality that leads to data loss and increases number of attacks.
- Scalability of cloud allows cloud users to scale up and down as needed. The resources used by one user may later be assigned to other user. This leads to confidentiality issues.
- Due to absence of hiring standards for cloud employees, malicious insiders can easily hack the organization's data and sell it to competitors.

- Cloud APIs can be exploited by outsider hackers and attackers.
- Abuse and miscreant use of cloud resources.

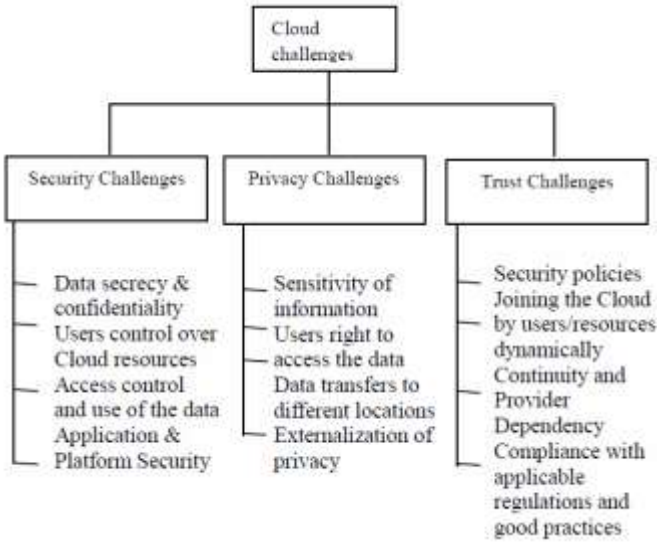

Fig. 1 Challenges of cloud computing

The proposed work mainly focuses on securing data in cloud environment. It ensures data confidentiality, integrity, authentication, and non-repudiation [5]. The proposed work mainly focuses on securing data in cloud environment. It ensures data confidentiality, integrity, authentication, and non-repudiation.

B. Background Techniques

Data Mining: Data Mining [6] techniques are commonly used to analyze data in the fields of marketing, finance, supply chain management (SCM), and customer relationship management (CRM) [7]. The proposed work ensures all security requirements by applying encryption mechanism in data transmission. Data transmission may take place between CSP and CU, or between CU and CU. The session key used for data transmission is generated based on frequent traffic pattern exists between the end parties. Backtracking based frequent pattern mining algorithm is applied to evaluate the frequent traffic pattern using data derived from past communications.

Fuzzy logic: Fuzzy logic is a problem-solving control system methodology that offers the ability to express uncertainty [8][9]. The proposed work applies a fuzzy logic system (FLS) to evaluate the trust degree of cloud users. Trust Degree of a user represents the degree of trustworthiness of the user computed based on user behavior.

C. Previous Works

Different security algorithms such as DES, AES, RSA, Blowfish are available to handle security issues and to provide data protection in cloud. DES, AES, Blowfish are symmetric key algorithms. They provide security for both providers and user. Whereas, RSA provide security only for user [10]. The work proposed by Kaur and Mahajan allows the user to dynamically select any of the required algorithms from DES, AES, Blowfish, and RSA [11]. The method only analyze technical privacy and encryption controls. Security mechanism proposed by Kalpana and Singaraju [12] is based on RSA algorithm. It only allows the authorized user to access the data. But the complexity involved in key management is high. The research work proposed by Tebaa et al., [13] uses homomorphic encryption method. It involves high implementation cost. The method proposed by El-Etriby et al., is suitable for applications which mainly focus on encryption duration. It uses different algorithms such as RC4, RC6, MARS, AES, DES, 3 DES, Twofish, and Blowfish [14]. Craig Gentry's homomorphic scheme and bootstrapping suggested by Meissen [15] unlimited usability of cipher text. Due to the limitation of the circuit, it cannot properly decrypt the cipher text. Key policy and cipher text policy introduced by Chung et al., [16] reduces the computation overhead of the data owner. The sensitive information cannot be revealed. But the method cannot fails on collusion attack. The mechanism discussed by song et al., uses chaos block encryption algorithm and homomorphic signature algorithm. It improves encryption/decryption speed. It maintains data confidentiality at higher level. Complexity involved in implementation is high. Gaurha and Shrivastava [7] presented an enhanced compete alu sequence algorithm that increases cloud efficiency. It's limitations involves less security and less data encryption process.

The methodology proposed by Kumar and Venkateswarlu [17] uses attribute based encryption, full homomorphic encryption, and linear programming. It satisfies basic security requirements. Wang et al., [18] suggest an mechanism by considering security issues like modification attacks, byzantine failures, and other attacks from cloud server. It maintains high efficiency and resilience. But it does not focus on dynamic data operations. Jachak et al., [10] proposed a framework of a very light-weight and provably secure provable data possession scheme. It supports dynamic operations on data. But it does not guarantee about privacy. The encryption technique proposed by Jain and Kaur [19] analyze the data security risk, its requirements, and deployment of security functions. It provides acceptable level of security. The encryption technique suggested by Kamara and Raykova [20] is suitable for massive datasets. It's limitation is the large-scale clusters. Bouti and Keller [21] increase the level of security by applying homomorphic properties of AES. There may be loss of confidentiality. Singh and Maini [22] used Blowfish, DES, and AES algorithm in their security mechanism. Inbarani et al., [23] suggest proxy re-encryption mechanism that is secure against chosen cipher text attack. Its performance is limited by collision problem and plaintext attack. Zhang et al., [24] handled practical message attack against all fully homomorphic encryption schemes. The suggested method is incapable of detecting the attack. None of the above has considered the user behavior in their encryption mechanism.

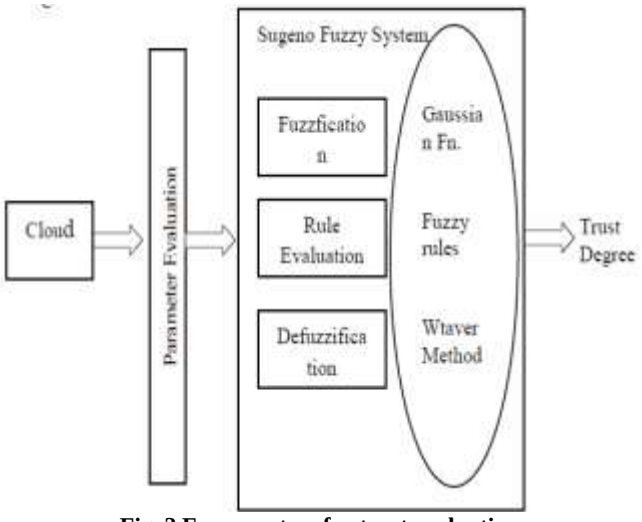

Fig. 2 Fuzzy system for trust evaluation

The main contributions of the proposed scheme are as follows:

1. The ideas of applying fuzzy system and user behavior mining into cloud data security and privacy are very novel.
2. The scheme will provide improved data security against different security attacks. As the keys are generated dynamically by the updated trust degree of cloud users, the attacker cannot reveal the key information of any uncompromised user. Moreover, the attacker cannot detect the session key involved in communication as the details of frequent behavior of the users involved are unknown to it.

II. PROPOSED WORK

A. *Fuzzy System for Trust Degree Evaluation*

Cloud user's trustworthiness is evaluated based on the following parameters: number of bytes of data from user to CSP, number of bytes of data from CSP to user, duration of access to the system, whether IP address is unusual, and whether timing of visit is abnormal. The fuzzy system shown in Fig. 2 is sugeno based system with 3 inputs and 1 output. This section describes the development of fuzzy logic controller for trust degree evaluation.

1) Parameters of the fuzzy system: The input parameters of the fuzzy logic controller are as follows:

- number of bytes of data from user to CSP
- number of bytes of data from CSP to user
- duration of access to the system
- whether IP address is unusual
- whether timing of visit is abnormal

2) Fuzzy rules and evaluation: The number of membership functions used for 5 inputs are 3, 3, 2, 2, and 2. The output uses 5 membership functions. The output of the sugeno based fuzzy controller is a constant (fuzzy singleton). Table 1 show the fuzzy sets and crisp ranges for all linguistic variables used in the research work.

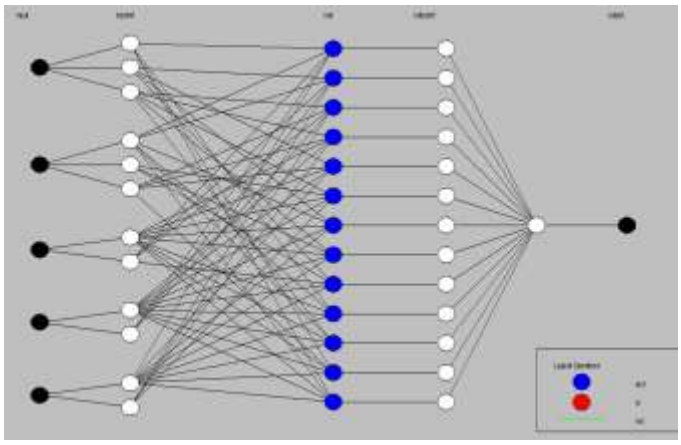

Fig. 3 Architecture of fuzzy system

B. *Data Security in Cloud Using Key Management Scheme Based On User Behavior Mining*

Frequent pattern mining [25] can be applied to extract the behavior of cloud user and hidden relationship exists between CU and CSP. This work follows the concept of Key Distribution through Traffic Mining (KDTM) [26]. The proposed work uses user behavior mining to generate the session key and to satisfy all basic security requirements. The Genmax, backtracking algorithm has been used in frequent traffic pattern identification and is applied later to generate session keys. Apriori, the most commonly used frequent pattern mining algorithm lists every single frequent itemset using bottom up, breadth first search. For dense data with long patterns listing all possible subsets of length pattern is computationally not viable. So the proposed work employs GenMax algorithm that list out all maximal patterns by applying a backtracking search. It uses progressive focusing technique for pruning non maximal itemsets and uses diffset propagation for quickly finding frequent itemset.

1)Public Key Generation: When a new CU enters a cloud, the cloud assigns a unique ID and a public key to the user. The public key of the user is initially computed based on user's ID.

$$Public\,key,\ P_N = H(ID_N),\ new\,cloud\,user \quad (1)$$

2) Private Key Generation: The private key of cloud user is generated by applying one-way hash function on identity and trust degree of the user.

$$S_n = H(ID_n \,\|\, TD_n) \quad (2)$$

3) Session Key Generation: For data transmission the source and the destination users or source CU and destination CSP will generate a session key. The concept of behavior mining is applied to attain frequent user's behavior Pattern (FBP) that exists between the CU's, or CU and CSP. Each CU maintains details about last few traffic carried on with other CU in the cloud. The session key shared by the cloud users CU1 and CU2 (SK_{CU1CU2}) is generated based on FBP that exists between CU1 and CU2. Only the users CU1 and CU2 have similar traffic pattern, which leads to the same FBP. Any other user who tries to extract the traffic will generate

different FBP. If CU1's and CU2's FBP are same, the nodes are considered to be valid. If FBP is different, it represents the malicious behavior of the node. The nodes must be verified while exchanging their FBP. The users CU1 and CU2 generate their session key SK_{CU1CU2} by using Equation (3).

$$SK_{CU1CU2} = H(FBP \parallel ID_{CU1} \parallel ID_{CU2} \parallel e(r_{CU1}P_{CU1} \parallel r_{CU2}P_{CU2})) \quad (3)$$

TABLE I
LINGUISTIC VARIABLES AND THEIR RANGES

Linguistic variable: Data_user_CSP(DUC) input	
Linguistic value	**Numerical range (normalized)**
LOW	[0 – 0.3]
MEDIUM	[0.2 – 0.7]
HIGH	[0.6 – 1.0]
Linguistic variable: Data_CSP_ user(DCU) input	
LOW	[0 – 0.3]
MEDIUM	[0.2 – 0.7]
HIGH	[0.6 – 1.0]
Linguistic variable: Access Duration(ACC_DUR) input	
NORMAL	[0 – 0.6]
ABNORMAL	[0.55 – 1.00]
Linguistic variable: IP Address (IP) input	
USUAL	[0 – 0.6]
UNUSAL	[0.55 – 1.00]
Linguistic variable: Timing_of_visit (TIME_VISIT) input	
NORMAL	[0 – 0.6]
ABNORMAL	[0.55 – 1.00]
Linguistic variable: Trust Degree (TD) output	
VERY LOW	[0 – 0.2]
LOW	[0.15 – 0.4]
MEDIUM	[0.35 – 0.6]
HIGH	[0.55 – 0.8]
VERY HIGH	[0.7 – 1.0]

III. EXPERIMENTAL RESULTS

Cloud environment is simulated using CloudSim and the CU's trust is evaluated using the fuzzy system designed using MATLAB. The fuzzy system is generated with 5 inputs and 1 output. The inputs DUC and DCU have 3 triangular membership functions, ACC_DUR, IP, TIME_VISIT have 2 triangular membership functions, and output has 5 membership functions. The sugeno based fuzzy system is trained using 250 data sets and its architecture is shown in Figure 3. The sample rule base of the fuzzy system is shown in Figure 4. Figure 5. Show the surface view of the fuzzy system.

IV. SECURITY ANALYSES

A. Session Key Secrecy

The proposed work enforces the secrecy of the session key. Session key security specifies that the probability of obtaining the session key by any adversary user should be negligible. In the proposed work, the session key is defined using Eq (3). During the user verification phase of the proposed scheme, only the legal users CU1 and CU2 can derive a similar FBP by using the received information. Without knowing S_{CU1} of CU1 or S_{CU2} of CU2, the adversary cannot compute correct value of x_{CU1} or x_{CU2}. The end users CU1 and CU2 are authenticated by using their S and FBP. The condition is that the authentic users can derive correct FBP and compute SK. Therefore, only the authorized end users are able to satisfy the user verification and frequent pattern verification process. So the probability of deriving SK without FBP is negligible.

B. Mutual Authentication

The user verification process confirms the authenticity of the end nodes.

C. Non Repudiation and Confidentiality

The secrecy of the session key expresses confidentiality and non repudiation of the transmitted message.

D. Forward and Backward Secrecy

No previous session keys or subsequent session keys can be recovered, in a situation where the adversary is managed to obtain the current session key.

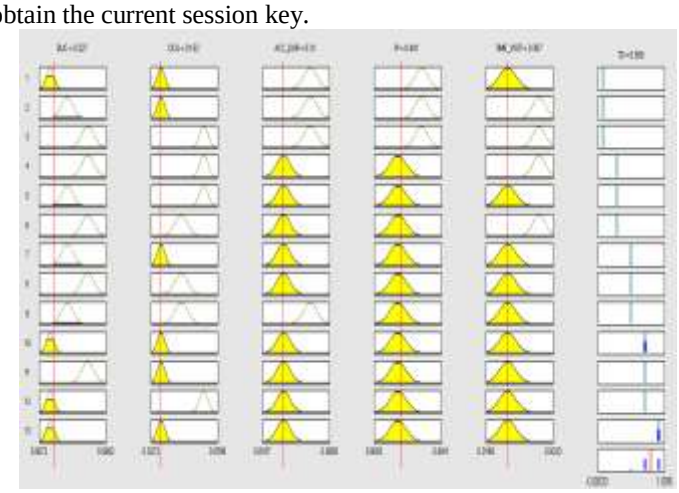

Fig. 4 Rule viewer

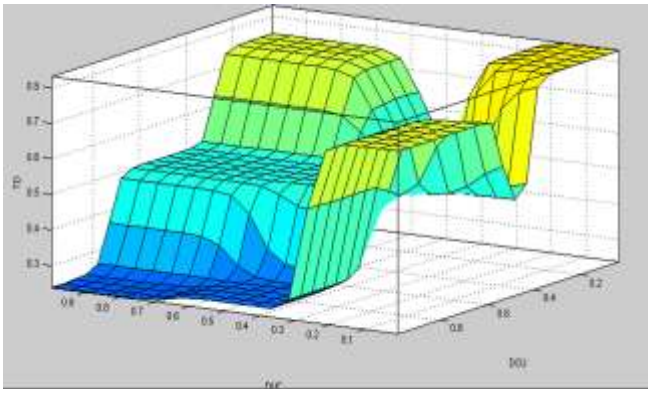

Fig. 5 Surface view of fuzzy system

V. CONCLUSION & FUTURE WORK

This paper presents a fuzzy associated trust based data security scheme by mining user behavior for cloud

environment. The proposed approach ensures authentication and confidentiality of CUs by applying encryption using the key generated based on CU's behavior. It computes direct trust of cloud users with fuzzy system using user behaviour parameters such as access duration, time of visit, IP address, and data transfer rate. Frequent access pattern of CUs is identified by mining their past cloud access pattern. The primary focus of future work includes analyzing the performance of proposed scheme under denial of service attack.

REFERENCES

[1] Mell P., and Grance T., "A NIST definition of cloud computing", National Institute of Standards and Technology, NIST Special Publication, pp. 800-145, 2009.

[2] Jadeja Y., and Modi K., "Cloud Computing - Concepts, Architecture and Challenges", Proc International Conference on Computing, Electronics and Electrical Technologies [ICCEET], Kumaracoil, India, pp. 877-880,2012.

[3] Best SJ., Kreuger BS., Ladewig J., "The effect of risk perceptionson online political participatory decisions", Journal of Information Technology & Politics, vol.4, no.1, pp.5-17, 2008.

[4] Jaeger PT., and Fleischmann KR., "Public libraries, values, trust, and egovernment", Information Technology and Libraries, vol.26, no.4, pp.35-43, 2007.

[5] Behl A., Behl K., "An Analysis of Cloud Computing Security Issues", Proc. 2012 IEEE World Congress on Information and Communication Technologies, Trivandrum, India, pp.109-114, 2012.

[6] Zhang C., Tiwari R., and Chen W., "A Data Mining Method to Extract and Rank Papers Describing Coexpression Predicates Semantically", IEEE International Conference on Data Mining Workshops , vol, no.6,pp. 483-488, 2009.

[7] Gaurha N., and Shrivastava M., "Data Security in cloud computing using linear programming", Int. J.Emerging Technology. Adv. Eng., vol.2,no.7, pp.28-30,2012.

[8] Wang, L.X., "A Course in Fuzzy Systems and Control", Prentice-Hall, 1997.

[9] Takagi T., and Sugeno M., "Fuzzy Identification of Systems and Its Application to Modeling and Control," IEEE Transactions on System, Man and Cybernetics, vol.15, no. 1, pp. 116-132, 1985.

[10] Jachak K.B et al, "Homomorphic Authentication with Random Masking Technique Ensuring Privacy and Security in Cloud Computing", J.Bioinfo Secur. Inform., vol.2, no.2, pp. 49-52, 2012.

[11] Kaur M., and Mahajan M., "Implementing Various Encryption Algorithms To Enhance The Data Security Of Cloud In Cloud Computing", VSRD International Journal of Computer Science & Information Technology, vol. 2, no.10, pp.831-835, October 2012.

[12] Kalpana P., and Singaraju S., "Data Security in Cloud Computing using RSA", International Journal of Research in Computer and Communication technology, vol.1, no.4, pp.143-146, 2012.

[13] Tebaa M., Hajji S.L., and Ghazi A.E., "Homomorphic Encryption Applied to the Cloud Computing Security", Proc. on Engineering, London, vol 1, 2012.

[14] El-etriby S., and Mohamed E., "Modern Encryption Techniques for Cloud Computing Randomness and Performance Testing", 3rd International Conference on Communications and Information Technology (ICCIT), Hammamet, Tunisia, pp. 800-805, 2012.

[15] Meiseen R., "A Mathematical Approach to Fully Homomorphic Encryption", Project Report, WPI, 2012.

[16] Chung P., Liu C., and Hwang M., " A Study of Attribute-based Proxy Re-encryption Scheme in Cloud Environments", International Journal of Network Security, vol.16, no.1, pp.1-13, 2014.

[17] Kumar S.K., and Venkateswarlu S., "Efficiently Providing Data Security and Linear Programming in Cloud Computing", International Journal Of Computer Science & Technology , vol.4, no.2, pp.1569-1570, 2013.

[18] Wang G. et al, "Hierarchical attribute-based encryption and scalable user revocation for sharing data in cloud servers", computers & security, vol.30, no.5, pp. 320-331, 2011.

[19] Jain N., and Kaur G., "Implementing DES Algorithm in Cloud for Data Security", VSRD International Journal of Computer Science & Information Technology, vol. 2, no.4, pp.316-321, 2012.

[20] Kamara S., and Raykova M., "Parallel Homomorphic Encryption", Workshop on Applied Homomorphic Encryption (WAHC '13), Okinawa, Japan, pp. 213-225,2013.

[21] Bouti and Keller,"Securing cloud-based computations against malicious providers", ACM SIGOPS Operating System review, vol.46, no.2, pp.38-42, 2012.

[22] Xia J., and Wang Y., "Secure Key Distribution for the Smart Grid", IEEE Transactions On Smart Grid, vol.3, no.3, pp 1437-1443, 2012.

[23] Inbarani W.S., Shenbagamoorthy G., Paul C.K.C., "Proxy Re-encryption Schemes for Data Storage Security in Cloud- A Survey", International Journal of Engineering Research & Technology, vol.2, no.1, pp.1-5, 2013

[24] Zhang Z., Plantard T.,and Susilo W., "Reaction attack on outsourced computing with fully homomorphic encryption schemes", Information Security and Cryptology, Springer Berlin Heidelberg, pp. 419–436, 2012.

[25] Dikaiakos MD. et al, "Cloud Computing: Distributed Internet Computing for IT and Scientific Research" , IEEE Internet Computing, IEEE Computer Society, vol.13, no.5. pp. 10-13, 2009.

[26] Stallings W., "Cryptography & Network Security: Principles & Practices", Prentice Hall, 3rd edition, 2003.

