Privacy Preserving Third Party Public Auditing Scheme for Secure Cloud Storage

Swapnali S. More
Department of Computer Engineering
A.C.Patil College of Engineering
Kharghar, Navi Mumbai

Sangita Chaudhari
Department of Computer Engineering
A.C.Patil College of Engineering
Kharghar, Navi Mumbai

ABSTRACT

Cloud storage is one of the service provided by Cloud computing in which data is maintained, managed, backed up remotely and made available to users over a network (typically the Internet). The user is concerned about the integrity of data stored in the cloud as the user's data can be attacked or modified by outside attacker. Therefore, a new concept called data auditing is introduced which check the integrity of data with the help of an entity called Third Party Auditor (TPA). The purpose of this work is to develop an auditing scheme which is secure, efficient to use and possess the capabilities such as privacy preserving, public auditing, maintaining the data integrity along with confidentiality. Thus the new auditing scheme has been developed by considering all these requirements. It consist of three entities: data owner, TPA and cloud server. The data owner performs various operations such as splitting the file to blocks, encrypting them, generating a hash value for each, concatenating it and generating a signature on it. The TPA performs the main role of data integrity check. It performs activities like generating hash value for encrypted blocks received from cloud server, concatenating them and generates signature on it. It later compares both the signatures to verify whether the data stored on cloud is tampered or not. It verifies the integrity of data on demand of the users. Thus no additional burden is provided on the cloud server. It is used only to save the encrypted blocks of data. All the task for the scheme is performed by the TPA and data owner. This proposed auditing scheme make use of AES algorithm for encryption, SHA-2 for integrity check and RSA signature for digital signature calculation.

Keywords

Cloud Storage; TPA; Privacy Preserving; Public Auditing; Integrity;

1. INTRODUCTION

According to the NIST definition, “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” [12].

In Cloud computing, the term cloud is a metaphor for the Internet, so the phrase Cloud computing is defined as a type of Internet-based computing, where different services are delivered to an organization's computers and devices through the Internet [4]. Cloud computing is very promising for the Information Technology (IT) applications; however, there are still some issues to be solved for personal users and enterprises to store data and deploy applications in the Cloud computing environment. Data security is one of the most significant barriers to its adoption and it is followed by issues including compliance, privacy, trust, and legal matters. Therefore, one of the important goals is to maintain security and integrity of data stored in the cloud because of the critical nature of Cloud computing and large amounts of complex data it carries. The users concerns for security should be rectified first to make cloud environment trustworthy, so that it helps the users and enterprise to adopt it on large scale [4].

The foremost issues in cloud data security include data privacy, data protection, data availability, data location, and secure transmission. Threats, data loss, service disruption, outside malicious attacks, and multi tenancy issues are the security challenges included in the cloud. Data integrity in the cloud system means preserving the integrity of stored information. The data should not be lost or modified by unauthorized users. Cloud computing providers are trusted to maintain data integrity and accuracy of data. Data confidentiality is also important aspect from user's point of view because they store their private or confidential data in the cloud. Authentication and access control strategies are used to ensure data confidentiality. The data confidentiality could be addressed by increasing the cloud reliability and trustworthiness in Cloud computing. Therefore security, integrity, privacy and confidentiality of the stored data on the cloud should be considered and are important requirements from user's point of view [4]. To achieve all of these requirements, new methods or techniques should be developed and implemented.

Data auditing is introduced in Cloud computing to deal with secure data storage. Auditing is a process of verification of user data which can be carried out either by the user himself (data owner) or by a TPA. It helps to maintain the integrity of data stored on the cloud. The verifier's role are categorized into two: first one is private auditability, in which only user or data owner is allowed to check the integrity of the stored data. No other person has the authority to question the server regarding the data. But it tends to increases verification overhead of the user. Second is public auditability, which allows anyone, not just the client, to challenge the server and performs data verification check with the help of TPA. The TPA is an entity which is used so that it can act on behalf of the client. It has all the necessary expertise, capabilities, knowledge and professional skills which are required to handle the work of integrity verification and it also reduces the overhead of the client. It is necessary that TPA should efficiently audit the cloud data storage without requesting for the local copy of data. It should have zero knowledge about the data stored in the cloud server. It should not introduce any additional on-line burden to the cloud user [6].

The three network entities viz. the client, cloud server and TPA are present in the cloud environment. The client stores data on the storage server provided by the cloud service provider (CSP). TPA keeps a check on client's data by periodically verifying integrity of data on-demand and notifies client if any variation or fault is found in client's data. Figure 1 shows the cloud data storage architecture.

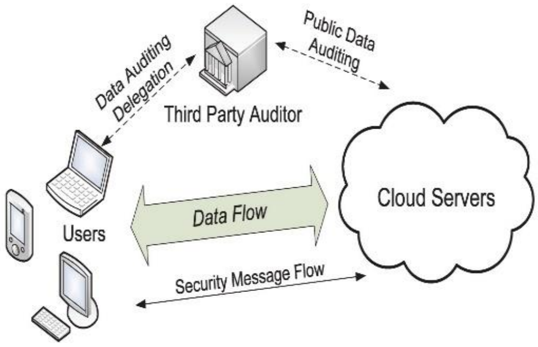

Figure 1. Cloud Data Storage Architecture [11]

The rest of the paper is organized as follows. Section 2 introduces Literature survey. The Proposed system is described in Section3 and Section 4 consists of Results and Discussion whereas Section 5 provides the Concluding remarks and Future work.

2. LITERATURE SURVEY

Cloud computing faces many problems on integrity and privacy of user's data stored in the cloud. Hence it requires some secure and efficient methods which can ensure the integrity and privacy of data stored in the cloud. Wang et al. [9] has proposed a privacy preserving public auditing protocol which makes use of an independent TPA to audit the data. It utilizes the public key based homomorphic linear authenticator (HLA) with random masking techniques. But this protocol is vulnerable to existential forgeries known as message attack from a malicious cloud server and an outside attacker. To overcome this problem, Wang et al. [6] proposed a new improved scheme which is more secure than the protocol proposed in [9]. It is a public auditing scheme with TPA, which performs data auditing on behalf of users. It uses HLA which is constructed from Boneh-Lynn-Shacham short signature referred as BLS signatures. It also uses random masking for data hiding. For the sake of data binding, this new scheme involves computationally intensive pairing operation thus making it inefficient to use. This proposed scheme has been implemented practically on Amazon EC2 instance which demonstrates the fast performance of the design on both the cloud and the auditor side. But the full-fledged implementation of this mechanism on commercial public cloud is not been tested. So it is difficult to expect it to robustly cope with very large scale data [7].

Wang et al. [10] proposed another protocol that supports both public auditing and data dynamics by using BLS based HLA along with Merkle Hash Tree (MHT). It achieves the integrity of data but fails to provide confidentiality to the data stored on the cloud. Wang et al. [8] has also proposed a design to detect the modified blocks easily using homomorphic token pre-computation and later erasure coded technique is used to acquire the desired blocks from different servers. Solomon et al. [11] proposed protocol uses the same security level as Wang et al. [7] but with better efficiency. It generates a signature set which is an ordered collection of signatures on each file block, thus incurring computation and communication overhead. Meenakshi et al. [2] has proposed a protocol which uses TPA to audit the data of the users using Merkle Hash Tree algorithm. It supports data dynamics but fails to provide confidentiality to the data stored in the cloud.

Tejaswani et al. [5] has achieved integrity of data using a Merkle hash tree by TPA and the confidentiality of data is achieved using RSA based cryptography algorithm whereas Jadhav et al. [3] have introduced an attacking module which continuously keeps track on data alteration in the cloud. The attacking module is a small code which resides on cloud server. Confidentiality of stored data is achieved by encrypting the data using AES algorithm. Arasu et al. [1] has proposed a method that uses the keyed Hash Message Authentication Code (HMAC) with homomorphic tokens to enhance the security of TPA. It is a technique for verifying the integrity of a data transmitted between two parties that agree on a shared secret key. HMAC's are based on a key that is shared between the two parties, if either party's key is compromised, it will be possible for an attacker to create fraud messages. Table 1. shows the Comparison of Existing Privacy Preserving Public Auditing Scheme.

In this table, comparison is done on various factors such as method used, whether supports public auditing, privacy preserving, data dynamic and batch auditing. It also shows if the integrity and confidentiality of data stored in the cloud server is maintained or not. From the table, it is clearly seen that different methods have been implemented to check the integrity of the data. But each method has some issues connected with it. The existing methods succeeded in providing privacy preserving along with public auditing but failed to maintain the confidentiality of data. Public auditing allows not only the user but anyone to perform data auditing on user's stored data. Privacy preserving means that the TPA has zero knowledge about the data used for data auditing. It is one of the important factor need to be achieved. Because it is possible that TPA may itself be malicious and leak the user's data.So it necessary that it is unaware about user's data. It's not safe to save the data in its original form on the cloud server. It may tend to attacks from outside attackers. In order to provide better security to data, encryption techniques need to be used.

In the existing systems, the task of computing the proof for integrity check of data is carried out by cloud server who is also responsible for storing huge amount of user's data. Thus increasing the burden of storage as well as task of proof generating on the cloud server side. There is a need to propose a system which does not increase the load on cloud server side. As TPA is used for integrity check purpose, so it should be solely responsible for generating the verification proof and verifying it later. All the factor mentioned above are important and need to be achieved for a reliable scheme. Therefore it is necessary to develop an efficient and secure auditing scheme which can perform public auditing effectively by maintaining both the integrity and confidentiality of data.

Table 1. Comparison of Existing Privacy Preserving Public Auditing Scheme

Research papers	Method used	Supports Public auditing	Supports Privacy preserving	Supports Data dynamic	Supports Batch auditing	Maintains Integrity of data	Maintains Confidentiality of data
Wang et al.[9]	HLA with random masking	Yes	Yes	No	No	Yes	No
Wang et al.[6]	HLA with BLS signature	Yes	Yes	No	Yes	Yes	No
Wang et al.[7]	HLA with BLS signature	Yes	Yes	No	Yes	Yes	No
Wang et al. [10]	HLA+BLS signature + MHT	Yes	Yes	Yes	Yes	Yes	No
Wang et al.[8]	Homomorphic + erasure code	Yes	Yes	Yes	No	Yes	No
Solomon et al.[11]	HLA + BLS signature	Yes	Yes	No	Yes	Yes	No
Meenakshi et al. [2]	Merkle hash tree	Yes	Yes	Yes	No	Yes	No
Tejaswani et al.[5]	MHT + RSA Algo	Yes	Yes	No	No	Yes	Yes
Jadhav et al. [3]	HLA + AES Algo	Yes	Yes	No	Yes	Yes	Yes
Arasu et al. [1]	HMAC Algo	Yes	Yes	No	No	Yes	No

3. PROPOSED SYSTEM

It is necessary to develop an effective public auditing protocol which overcomes the limitation of the existing auditing scheme. The proposed system which is developed is used to verify the correctness of cloud data with the help of TPA, periodically or on demand without retrieving the entire data or without introducing additional online burden to the cloud users and cloud servers. It assure that no data content is leaked to TPA during the auditing process. It maintains storage correctness of data, integrity and confidentiality of stored data.

The proposed scheme consists of three basic entities; they are data owner, cloud server storage and TPA. The data owner or the user is responsible for splitting the file into blocks, encrypting those blocks using AES algorithm, generating a SHA-2 hash value for each, concatenating the hashes and generates a RSA signature on it. The cloud server is used only to store the encrypted blocks of files. Thus it has no additional burden of computing the verification proof. Here verification proof means generating of hashes for the encrypted blocks, concatenating them and generating a digital signature for verification. This task is performed by TPA itself.

When the client or data owner request for data auditing to the TPA, it immediately request for the encrypted data from the cloud server. After receiving the data, it generated the hash value for each block of encrypted files. It uses the same SHA-2 algorithm which was used by client. It later concatenate those hash values and generates a RSA signature for that file. In the Verification process, the signature generated by TPA and the one stored in the TPA which is provided by the data user are compared by the TPA. If it matches with each other it means that the data is intact and data is not been tampered by any outsider or attacker. If it does not matches then it indicates that the data integrity has been affected or tampered. The result for the data integrity check is provided to the data owner. Figure 2. shows the Architecture for the Proposed Auditing scheme.

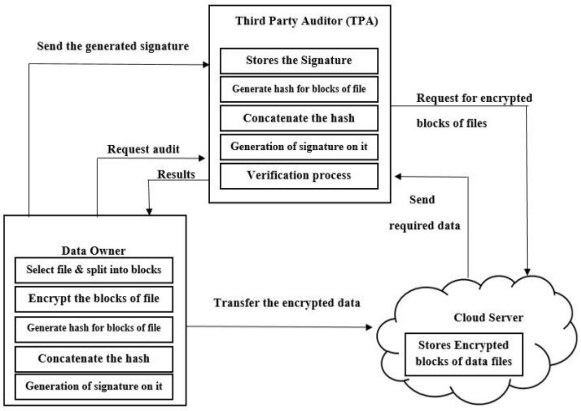
Third Party Auditor (TPA)
Stores the Signature
Generate hash for blocks of file
Concatenate the hash
Generation of signature on it
Verification process
Send the generated signature
Request for encrypted blocks of files
Request audit
Results
Send required data
Data Owner
Select file & split into blocks
Encrypt the blocks of file
Generate hash for blocks of file
Concatenate the hash
Generation of signature on it
Transfer the encrypted data
Cloud Server
Stores Encrypted blocks of data files

Figure 2. Proposed Auditing scheme Architecture

Data owner is an important part of the proposed system. It performs most of the responsibility related to the data. In the proposed auditing scheme, the data owner first performs login and registration with cloud server and TPA. The new user has to firstly register itself by filling the registration form and be the active member of the system. A message for successful registration will be provided. If a user is already the member of the system then he or she can perform login process. If the user name and password exist in the database, then they will be login successfully for being valid users or else they will receive an error message.

Once successfully login, the data owner will select the file he or she want to store on the cloud server. The file selected by user will be split into number of blocks. In order to carry out the splitting of the required file into blocks a FileSplitter algorithm is used. In this algorithm, it checks if the file exist or not. If exist then the file is split intouser specific size based upon the file size. For example if the file is of size 23kb then it will be split into 20kb and 3kb. Here in the example the size specified to split a file is 20kb. This file splitting algorithm is used to provide extra security to user's data. As the data is split into parts so if any attacker gets successful in accessing the data, will get data in parts and not the entire data. This algorithm is used to split files of any type such as .txt, .docx, .pdf, .jpg, .png, .zip etc. This file splitting technique is simple and easy to use. It quickly split large files into small chunks.

Advanced Encryption algorithm is the secure and strongest encryption algorithm that is frequently used today. So AES is used to provide confidentiality to the data. The blocks which are split now encrypted using AES algorithm by the data owner. Each blocks of file is encrypted and stored on the client. A copy of the encrypted file is transferred to cloud server for storage purpose. AES encryption algorithm uses same keys for encryption and decryption purpose. It supports three keys with different key lengths that is 128, 192 and 256 bits. Longer key provides the strongest encryption. But in terms of performance, shorter keys results in faster encryption time compared to the longer keys. So 128 bit AES is faster as compared to others. Therefore AES is used to encrypt data blocks of 128 bits using symmetric keys of 128 bits. Some other factors which make AES algorithm effective to use are given in Table 2 [13].

Table 2. Comparison of AES and DES algorithm

Factors	AES	DES
Block size	128 bits	64 bits
Key size	128,192,256 bits	56 bits
Security	Highly secure	Not secure enough
Encryption &decryption	Faster	Moderate
Power consumption	Low	Low
Stimulation speed	Faster	Faster

After encrypting the blocks, now a hash value for the blocks are generated separately. For this purpose a hashing algorithm SHA-2 is being used. A good hashing function must provide a one way hash function. In this, it is easy to compute on the input data but difficult to revert it back to normal input. It must also be collision resistant. In this features it's difficult to find hashes with same output for different inputs.SHA-2 algorithm supports both the property of one way hash function and collision resistant hash function which are the major features that a good hashing function must consist of. Therefore SHA-2 is used for hashing in the proposed system. SHA-2 is more effective and secure than SHA-1. Table 3. shows comparison of SHA-1 and SHA-2 [14].

Table 3. Comparison of SHA-1 and SHA-2

Factors	SHA-1	SHA-2
Output size	160	256
Block size	512	512
Rounds	80	64
Security	Collision found, theoretical attacks	No attack found

After the hashes are generated, the hashes for each blocks are concatenated and RSA digital signature is performed on it. Digital signatures are used to authenticate the source of messages. It uses private key to sign the message and public key to verify the signature. Later this signature is sent to the TPA, where it uses this signature to check the integrity of data stored in the cloud server storage is maintained or not. Data owner has the authority to request for data integrity check to the TPA. Figure 3. shows the working of the data owner in the proposed auditing scheme.

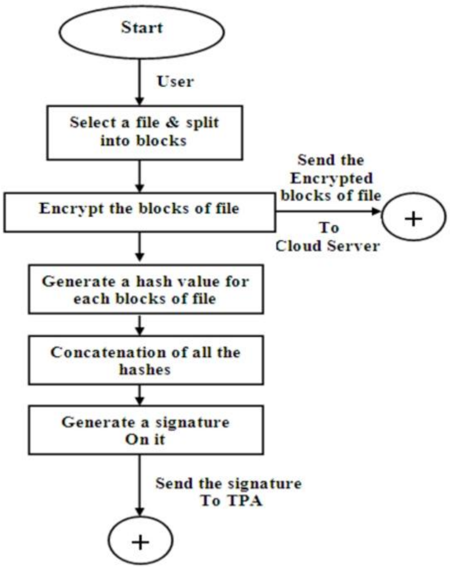
Start
User
Select a file & split into blocks
Encrypt the blocks of file
Send the Encrypted blocks of file
To Cloud Server
Generate a hash value for each blocks of file
Concatenation of all the hashes
Generate a signature On it
Send the signature To TPA

Figure 3. Flowchart for Working of Data Owner

The data owner make use of cloud storage to store the encrypted form of data. There is no additional computing burden on cloud server. This is one of the advantage of the proposed system. Because as discussed earlier most of the existing scheme involve cloud server for computing as well as storage purpose too. As the data is stored in encrypted form, so the cloud server has zero knowledge about the data. As well as if the cloud server turns into malicious server or is attacked by any outside attacker, the data will not be retrieved easily as it is in the encrypted form and it is not aware about the encryption algorithm implemented by the data owner.

In the proposed scheme, to perform the task of data auditing a TPA is been used for this purpose. TPA performs data auditing either periodically or on demand by the client. On receiving the auditing request from user or data owner, the TPA starts its auditing process. TPA also stores the signature which has been generated by data owner. The TPA follows the same process performed by data owner such as generating hash for encrypted blocks of data files, concatenating them and generating signature on it. Later it compares the two signature in verification process. If it matches then it means the integrity of data is maintained and otherwise not maintained. This means that data is not been tampered or changed. The results for the same is provided to the data owner by the TPA. The following Figure 4. shows the working of the TPA in the proposed auditing scheme.

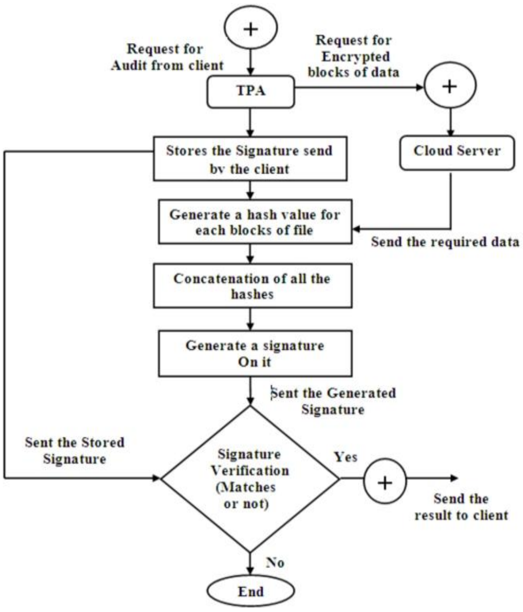

Figure 4. Flowchart for working of TPA

4. RESULTS AND DISCUSSIONS

The proposed scheme is implemented using java. For practical implementation use of three computers as data owner, cloud server and TPA is made. All the computers are connected in LAN. Database is used as a backend to store user related information.File size ranging from 10KB to 100MBis used for experiential purpose. In this scheme all the modules are implemented efficiently. An effective splitting algorithm which successfully split all the types of files such as .docx, pdf, txt, jpg etc. is developed. A strong AES encryption algorithm is used to maintain the confidentiality of data. Thus providing protection to data and making it difficult for the attackers to access and use. SHA-2 is used for hashing purpose.A 256 bit SHA is used to generate the hash values. Later all the hashes are concatenated and RSA digital signature is provided to it to make it more secure. RSA digital signature make use of two large prime numbers (p, q) to compute its module n=p*q. It is easy to multiply the two prime number but infeasible to obtain original prime number from total factoring. The value of n is used by both private and public key [14]. The Euler's totient function is obtained by (p-1)(q-1). User need to choose e such that e and (p-1)(q-1) are relatively prime numbers. Public key consists of module n and public exponent e whereas private key consists of module n and private exponent d. For encrypting $c = m^e \bmod n$ and decrypting: $m = c^d \bmod n$ is used. Whereas for Digital signature: $c = m^d \bmod n$ (signing) and $m = c^e \bmod n$ (verification) is used efficiently. Thus all the algorithms such as File splitter, AES, SHA-2 and RSA digital signature are studied and implemented efficiently. Figure 5, 6 and 7 shows the snapshots for the proposed system.

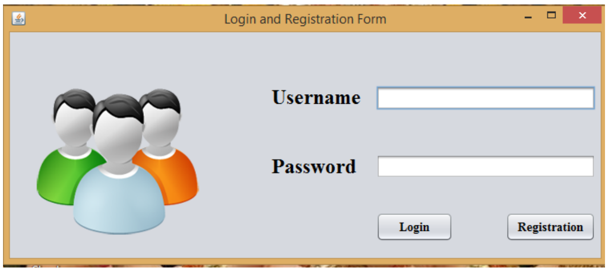

Figure 5. Snapshot for Login and Registration form

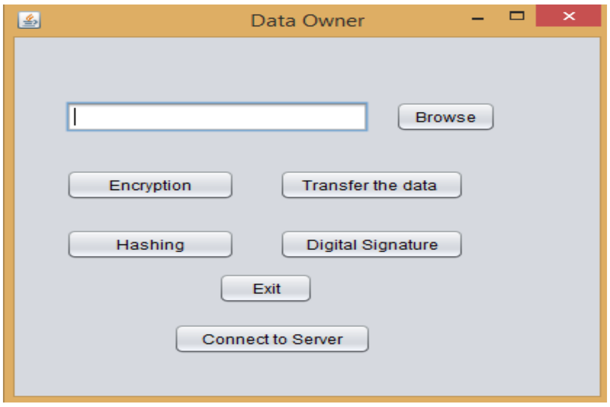

Figure 6. Snapshot of Data Owner

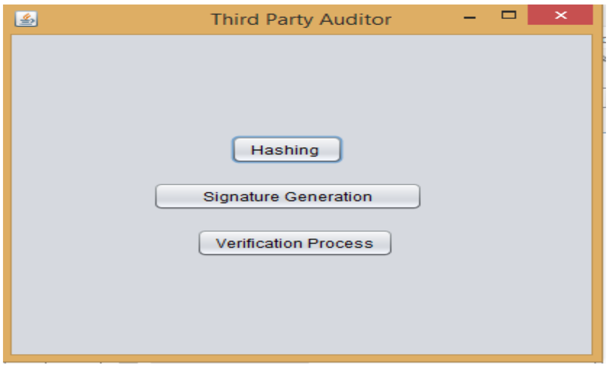

Figure 7. Snapshot of TPA

5. CONCLUSION AND FUTURE WORK

A secure and efficient privacy preserving public auditing scheme is been proposed. It achieves privacy-preserving and public auditing for cloud by using a TPA (Third Party Auditor), which does the auditing without retrieving the data copy, hence privacy is preserved. The data is split into parts and then stored in the encrypted format in the cloud storage, thus maintaining the confidentiality of data. The data integrity is verified by TPA on request of the client by verifying both the signatures. It only check whether the stored data is tampered or not and informs about it to the user. The cloud server is used only to store the encrypted form of data. Thus providing no online computing burden on it. An attempt is made to overcome the limitations of the existing auditing scheme. All the modules in the system are implemented to develop an effective auditing scheme. The partial results for the proposed system is been shown. In future, data dynamic operations such as updation, deletion and insertion of data would be performed.

6. REFERENCES

[1] S Ezhil Arasu, B Gowri, and S Ananthi. Privacy-Preserving Public Auditing in cloud using HMAC Algorithm. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277, 3878, 2013.

[2] IK Meenakshi and Sudha George. Cloud Server Storage Security using TPA. International Journal of Advanced Research in Computer Science & Technology (IJARCST) ISSN: 2347-9817, 2014.

[3] Jadhav Santosh and B.R nandwalkar. Privacy Preserving and Batch auditing in Secure Cloud Data Storage using AES. Proceedings of 13th IRF International Conference, ISBN: 978-93-84209-37-72014.

[4] Zissis, Dimitrios, and Dimitrios Lekkas. Addressing cloud computing security issues. *Future Generation computer systems* 28.3 (2012): 583-592.

[5] Tejaswini, K. Sunitha, and S. K. Prashanth. Privacy Preserving and Public Auditing Service for Data Storage in Cloud Computing. *Indian Journal of Research PARIPEX*, 2(2), 2013.

[6] Cong Wang, Sherman SM Chow, Qian Wang, Kui Ren, and Wenjing Lou. Privacy Preserving Public Auditing for Secure Cloud Storage. *http://eprint.iacr.org/2009/579.pdf*

[7] Cong Wang, Sherman SM Chow, Qian Wang, Kui Ren, and Wenjing Lou. Privacy Preserving Public Auditing for Secure Cloud Storage. *Computers, IEEE Transactions on,* 62(2):362–375, 2013.

[8] Cong Wang, Qian Wang, Kui Ren, Ning Cao, and Wenjing Lou. Toward secure and dependable storage services in cloud computing. *Services Computing, IEEE Transactions on*, 5(2):220–232, 2012.

[9] Cong Wang, Qian Wang, Kui Ren, and Wenjing Lou. Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing. *In INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

[10] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, and Jin Li. Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing. *Parallel and Distributed Systems, IEEE Transactions on*, 22(5):847–859, 2011.

[11] Solomon GuadieWorku, Chunxiang Xu, Jining Zhao, and Xiaohu He. Secure and efficient privacy-preserving public auditing scheme for cloud storage. *Computers & Electrical Engineering,* 40(5):1703–1713, 2014.

[12] Mell, Peter, and Tim Grance. The NIST definition of cloud computing. (2011).

[13] Ritu Tripathi and Sanjay Agrawal. Comparative study of symmetric and asymmetric cryptography techniques. International Journal of Advance Foundation and Research in Computer (IJAFRC), 2014.

[14] Understanding Cryptography: A textbook for Students and Practitioners, Christof Paar, lzleP naJ