Third Party Public Auditing Scheme for Storage

Priyanka P Koli[1] Megha S Karosiya[2] Damini P Kangate[3] Sonu R Rathod[4] Prof. Dr. Girish K. Patnaik[5]
[1,2,3,4]Research Scholar [5]Head of Department
[1,2,3,4,5]Department of Computer Engineering
[1,2,3,4,5]SSBT College Engineering and Technology Jalgaon, India

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. The cloud server is used only to save the encrypted blocks of data. This proposed auditing scheme make use of Cipher algorithm for encryption and AES algorithm for digital signature calculation.

Key words: Cloud Storage, TPA, Privacy Preserving, Public Auditing, Integrity

I. INTRODUCTION

"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" .Making use of the cloud saves both users time and money. 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 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. 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 . 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 users.

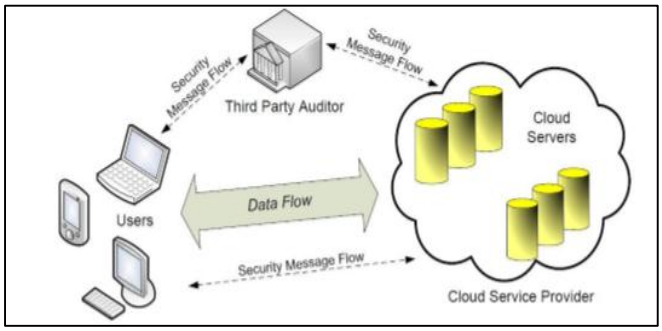

Fig. 1: Data Storage Structure

II. PROBLEM STATEMENT

One of the biggest challenges concerning data users is how to ensure the integrity and the confidentiality of their data while it is in the cloud. Users may be concerned about this issue for various reasons. First of all, data could be lost due to hardware or software failures on the Cloud. Secondly, the CSP could hide data loss incidents from data users in order to maintain their reputations. Thirdly, the CSP may deliberately delete rarely accessed data for saving more storage capacity. Another reason is that the CSP or unauthorised entities might access the content of users' data. User's data needs to be checked regularly while it is in the cloud. However, as users may have many out sourced data files, they will find it difficult to check the integrity of their data files themselves due to their limited computing resources. In addition, data integrity check could be expensive on the user's side in terms of computation and storage. To tackle these issues, users can resort to a Third Party Auditor (TPA) to check the correctness of the remote data on their behalf. Resorting to a TPA can have various benefits. The first benefit is to eliminate the burden of checking the data integrity on the user's side, resulting in saving users' computation resources. Another benefit is that TPA is usually a specialist who has more capability and computational resources than ordinary users.

The focus of this system on achieving data integrity and data confidentiality. Specifically, the project propose a solution to ensure that users' data is maintained in its original state without alteration or corruption, and without requiring the local copy of the data. The solution will also ensure that users' data cannot be accessed except by authorised users. A secure and efficient solution will be implemented to enable data users to check the correctness of their data while it is in the cloud by resorting to a TPA. The solution should enable users to detect any data integrity drift such as data alteration or data loss. Also, the solution should preserve the confidentiality of users' data by prohibiting unauthorised people or even the TPA from accessing the content of the user's data.

III. OBJECTIVE

There are two main objectives that are needed to achieve the main aim of this system, which are as follows:

- Propose an efficient protocol for checking the integrity of remote data files. To ensure efficiency, the protocol must have the following features:
  1) Detect any data integrity drift such as data loss or data alteration.
  2) Support public auditability by relying on a trusted third party auditor (TPA) for checking the data integrity on behalf of the users.
  3) Support dynamic data operations such as data insertion, data alteration, and data deletion.
- Extend the data integrity checking protocol to support data confidentiality by employing encryption techniques. To ensure the confidentiality of the data, the protocol must meet the following requirements:
  1) Prohibit the CSP from accessing the content of users' data files or even learning any knowledge regarding those files.
  2) Prohibit the TPA from accessing the content of users' data files during the auditing task.
  3) Prohibit malicious parties or attackers from accessing the content of users' data files while they are in the cloud.
- Several specific tasks need to be done in order to end up successfully with the entire system. These tasks are as follows:
  - Understand the concept of cloud computing in terms of storing users' data. Thus, a database will be created to store users' data files.
  - Read and analyse different data integrity checking methods such as MAC and Hash methods. Then, more reading is required to fully understand the existing protocols, such as Provable Data Possession, that can be used to detect data integrity drifts.
  - Create and design a new protocol that can detect remote data integrity drifts as well as preserving the confidentiality of users' data.
  - Create a web-based application that allows users to perform the following tasks:
  1) Upload their data files to the cloud.
  2) View or download their remote data files at any time.
  3) Check the integrity of their remote data files by resorting to a third party auditor for this task.
  4) Perform data operations on their data files such as insertion, deletion, or alteration while it is in the cloud when needed. Analyse, test, and evaluate the new protocol.

IV. EXISTING SYSTEM

Cloud improves due to centralization of data, increased security focused resources, etc., but concerns can persist about loss of control over certain sensitive data, and the lack of security for stored kernels. Security is often as good as or better than other traditional systems, in part because providers are able to devote resources to solving security issues that many customers cannot afford. To securely introduce an effective third party auditor (TPA), the following two fundamental requirements have to be met: 1) TPA should be able to efficiently audit the cloud data storage without demanding the local copy of data, and introduce no additional on line burden to the cloud user; 2) The third party auditing process should bring in no new vulnerabilities towards user data privacy.

Limitation of Existing System are as follows:

- User's files are not encrypted on some open source cloud storage systems. So, privacy is not preserve.
- The storage service provider can easily access the user's files. This brings a big concern about user's privacy.
- The user has no supreme control over the software applications including secret data. User has to depend on the provider‘s action, maintenance and admin it.

V. PROPOSED SYSTEM

There is a need to develop an effective public auditing protocol which overcomes the limitation of the existing

auditing scheme. The proposed system is developed to verify the correctness of cloud data by 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 using Cipher algorithm, generating key for each, concatenating the keys and generates a signature on it. The cloud server is used to store the encrypted blocks of files. 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 key for each block of encrypted files. It uses the same signature generation algorithm which was used by client. It later concatenate those keys and generates a 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 they 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.

A. Advantages of proposed system

- Without knowing the local copy of data the TPA will audit the data.
- It reduces the overhead of communication as well as computation as compared to common approaches of data auditing.
- Online burden is reduced for data owners
- Our strategy is apparently safe against adversaries under random oracle model.

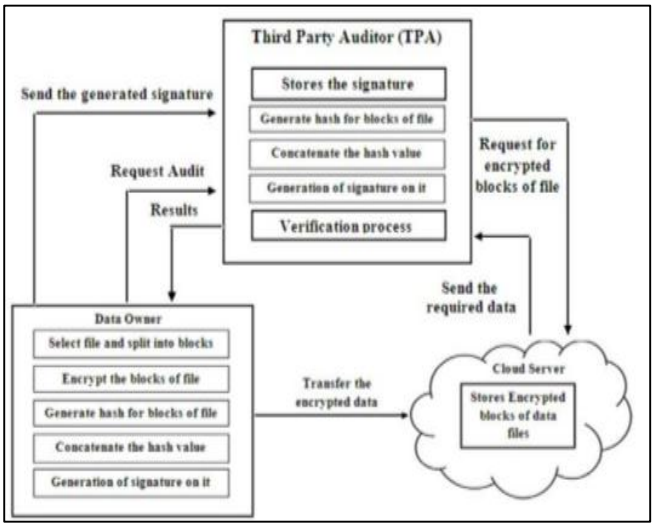

Fig. 2: Proposed Auditing Scheme Architecture

Data owner is an important part of our 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 him will be split into number of blocks. The blocks which are split are now encrypted using Cipher algorithm by the data owner. Each blocks of file will be encrypted and stored on the client. A copy of the encrypted file will be transferred to cloud server for storage purpose. After encrypting the blocks, now a key for the blocks are generated separately. After keys are generated, the keys for each blocks are concatenated and digital signature is performed on it. Digital signatures are used to authenticate the source of messages. 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.

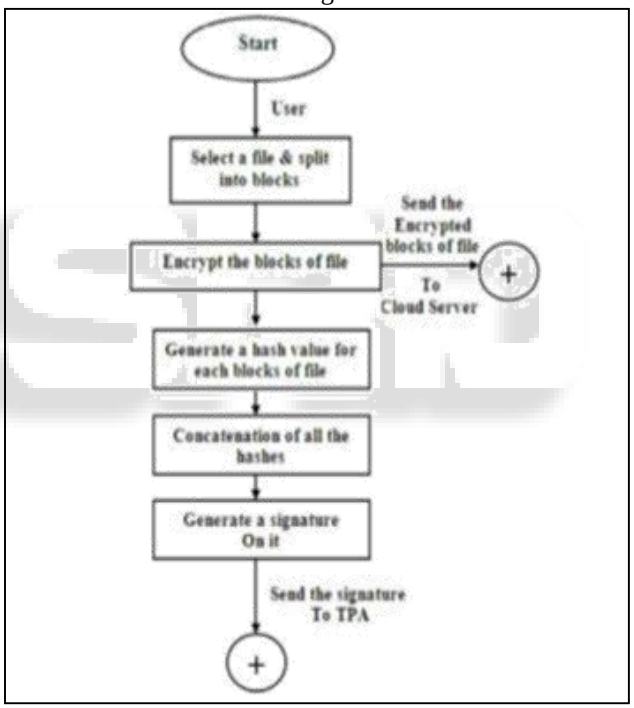

Fig. 3: Flowchart for working of Data Owner

The data owner make use of cloud storage to store the encrypted form of data. 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 our proposed auditing scheme.

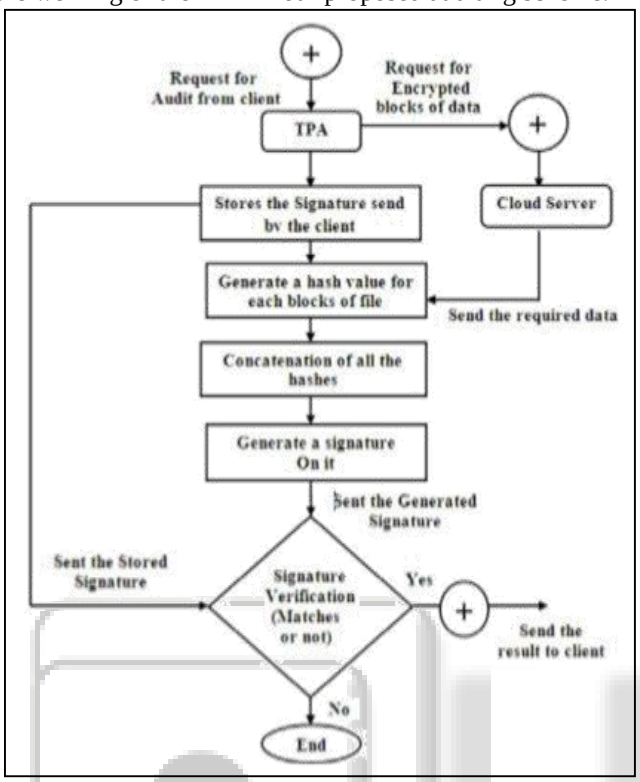

Fig. 4: Flowchart for working of TPA

VI. SECURITY ISSUES IN CLOUD COMPUTING

Data security must be taken into account as shifting data into the cloud could result in many security breaches such as data leakage, data loss, data alteration, and unauthorised data access. Data security here refers to data confidentiality, integrity, and availability. Data confidentiality means data must not be accessed except by authorised users [20]. Data confidentiality can be achieved by the adoption of encryption techniques. Data availability means data must be always accessed even in the case of unpleasant disasters such as power outages. Data integrity means data must not be intentionally or accidentally tampered. Data integrity can be achieved by the adoption of hash function and digital signature techniques. However, those techniques cannot be used directly to detect data integrity drifts for the following reason. When users shift their data into the cloud, they delete their local copy of the data files. Thus, the challenge is how to ensure the integrity of the data without having a local copy of the data. Downloading data files for the purpose of checking their integrity is not an efficient solution due to the high network bandwidth required [5].

VII. METHODS FOR DETECTING DATA INTEGRITY DRIFTS

There are various methods that can be used to detect data integrity drifts. These methods can be used to build an efficient system that requires less communication and computation overhead. There are three methods that will be discussed in this section, which are digital signature, Message Authentication Code (MAC) and hash function.

1) Digital Signature

Digital signature is basically an electronic signature that authenticates the sender of the messages to the recipient. So, the recipient will ensure that the message has been sent by a known sender . In addition, digital signature can be used to assure that the data has not been tampered while it is in transit. Thus, the integrity of the data is checked.

The digital signature method consists of three algorithms, namely, keyGen, Signature, and Validation:

- KeyGen: this algorithm produces a private key and its corresponding public key.
- Signature: this algorithm takes the message and the private key, and produces the digital signature.
- Validation: this algorithm takes the signature and the public key to recover the message.

If the recovered message matches the original one, the signature is valid (the message is not tampered).

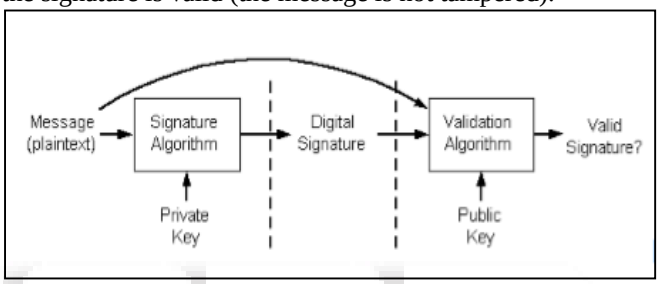

Fig. 5: Digital Signature Method

However, to prevent the verifier from accessing the content of the data during the verification process, the data can be encrypted or hashed prior to signing it. Therefore, the verifier will only access the data in an encrypted format.

- Advantages:
  - This method can detect data integrity drifts.
  - This method can preserve the privacy of the data if the data is encrypted prior to generating the signature.
- Drawbacks:
  - This method is based on asymmetric cryptography as it requires two keys, which are public and private keys. Thus, it is not efficient for encrypting large messages.

2) Message Authentication Code (MAC)

Message Authentication Code (MAC) is a portion of information that can be used to provide an assurance regarding the authenticity and integrity of the data [24]. MAC method with the help of the secret key can be used to detect data alteration.

The MAC method takes a secret key and a message as input, and generates a MAC. To verify the integrity of the data, the verifier has to use the secret key to generate a new MAC and then compare it with the received one. If the two MACs are the same, then the data is in its original state

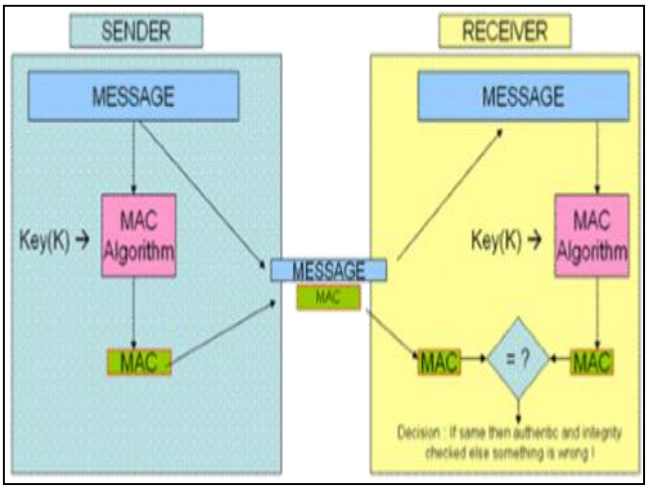

Fig. 6: Message Authentication Code (MAC) Method

- Advantages:
  - This method uses a single secret key to generate and verify MACs. Thus, this method is suitable for encrypting large files.
- Drawbacks:
  - The secret key needs to be shared between the sender and the receiver. Thus, the distribution ofthe key might impose a risk towards the data if the key is compromised

3) Hash Function

A cryptographic hash function can be used to detect data integrity drifts. It takes a message of a variable length as input, and produces a fixed size message digest . It does not require any key for generating the message digest unlike the MAC method.

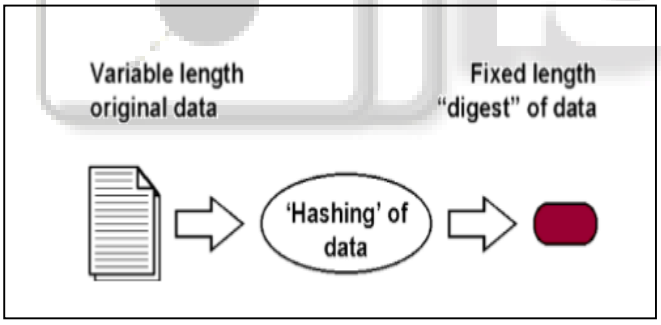

Fig. 7: Hash Method

To verifier the integrity of the message, the verifier has to compute a fresh hash value of the message and then compare it with the received one. If the two hash values are the same, the message is not tampered.

- Advantages:
  - This method does not require any key.

VIII. IMPLEMENTATION DETAILS

An efficient and secure protocol is proposed to ensure both data integrity and data confidentiality. To ensure data integrity, the protocol will make use of the Elliptic Curve Cryptography (ECC). For data confidentiality, data files will be encrypted using Pseudo Random Function. Thus, the content of users data cannot be leaked to the TPA or malicious parties. The protocol consists of six algorithms, namely, KeyGen, DataEnc, MetadataGen, ChallengeGen, ProofGen, and VerifyProof.

- KeyGen: This algorithm is run by the user to produce a pair of keys (public key (Upu) and private key (Upr)).
- DataEnc: This algorithm is also run by the user to encrypt the file (F) before uploading it to the CSP.
- MetadataGen: This algorithm is also run by the user to generate a verification metadata that will be sent to the TPA for auditing purposes.
- ChallengeGen: This algorithm is run by the TPA to send a verification challenge to the CSP.
- ProofGen: This algorithm is run by the CSP to build a proof of the correctness of the data and then send it to the TPA.
- VerifyProof: This algorithm is run by the TPA to determine whether the data has been tampered or not based on the proof received.

The protocol can be divided into two main phases namely, Initialisation and Verification.

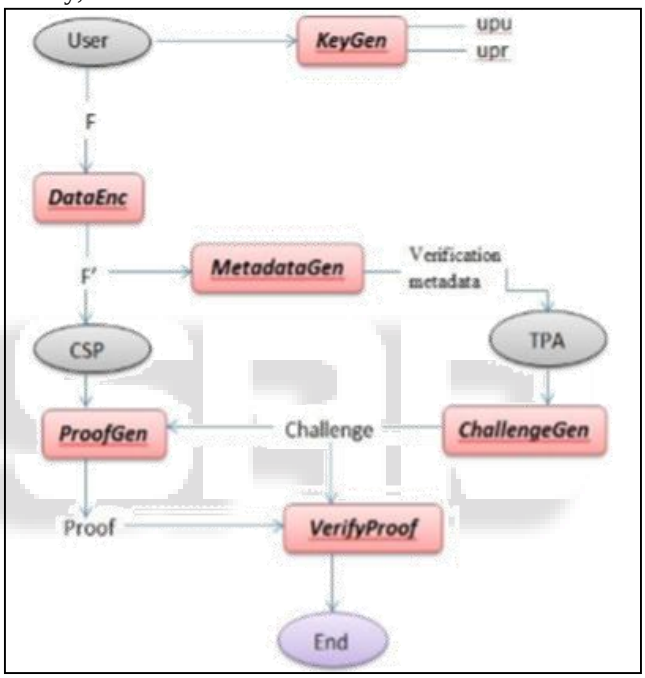

Fig. 8: Algorithm used in the Protocol

1) Initialization Phase

Initialization phase is for pre-processing the data file(F) before uploading it to the CSP.

- Step 1 : The user first needs to run the KeyGen algorithm that generate algorithm as input.
- Step 2 : After generating the keys, the file will be divided into a number of data blocks(n) as follows.
  F = m1, m2, m3,,mn, where mi is the ith block
  where F is the file and m1,m2,m3,mn are the blocks of file.
  SSBT's College of Engineering and Technology, Bambhori, Jalgaon (MS) 31
- Step 3 : The user then runs the DataEnc algorithm to encrypt the file blocks to maintain the confidentiality of the data. So, the encrypted file will be as follows:
  F = M1, M2, M3,, Mn
- Step 4 : After that, the user has to pre-process the F and generate its metadata verification (D) using the MetadataGen algorithm. This can be done as follows.

D = d1, d2, d3,, dn where di is the metadata for the ith block.

- Step 5 : The metadata for the ith block (di) will be computed using the keys, and the encrypted block, which is as follows. The encrypted file (F) will be sent to the CSP, while the verification metadata (D) will be sent to the TPA for auditing purposes.

2) Verification Phase

Verification phase is for auditing users data.

- Step 1 : The TPA runs the ChallengeGen algorithm to issue a verification challenge that will be sent to the CSP.
- Step 2 : Upon receiving the challenge, the CSP has to respond to the challenge by using the ProofGen algorithm. The CSP will generate a proof(R) and then send it to the TPA.
- Step 3: After the TPA has received the proof (R), the TPA will compute R' and then compare it with the received proof (R) by using the VerifyProof algorithm. After that, it uses the keys and the verification metadata to compute.
- Step 4: If R = R, the data is in its original state (not tampered). Finally, the verification result will be sent to the user.

IX. 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 checks whether the stored data is tampered or not and informs about it to the user. 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.

In future, we would like to perform data dynamic operations such as users can add a new block, modify an existing block, or delete an existing block and Batch auditing this feature allows the TPA to handle and perform numerous auditing tasks at the same time. As there are a great number of data users in the cloud, the TPA may receive a multitude of auditing requests from different users at once.

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] Boyang Wang, Baochun Li and Hui Li, "Public auditing for shared data with efficient user Revocation in the cloud", IEEE Xplore Digital Library, vol 8, Issue 1, Sep 2015.

[4] Kai He, Chuanhe Huang, Kan Yang and Jiaoli Shi, "Identity- preserving public auditing for shared cloud data," in the 23rd IEEE International Symposium on Quality of Service (IWQOS), 2015.

[5] P.Divya and B. Sivananthan, "A Privacy-preserving access control with robust data authenticity for cloud group," Journal of Scientific and Computational Intelligence, vol. 2, issue 1, Sep 2015.

[6] G. Shreedevi and K.G. Arunkumar, "Survey of public auditing of shared data with multiple third party auditor with efficient user revocation in cloud" Journal of Computer Technology and Applications, vol.6 (2), Mar-Apr 2015.

[7] Prof. Sawan Baghel and Prof. Gaurav Saboo, "Efficient Cryptographic algorithms for cloud storage security," Journal of Emerging Technologie in Engineering Research, vol.3, issue 2, Nov 2015

[8] Priyadharshini, B., and P. Parvathi. "Data integrity in cloud storage." Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on. IEEE, 2012.

[9] Goyal, Renuka, and Navjot Sidhu. "Third Party Auditor: An Integrity Checking Technique for Client Data Security in Cloud Computing." International Journal of Computer Science & Information Technologies 5.3 (2014).

[10] Jansma, Nicholas, and Brandon Arrendondo. "performance comparison of elliptic curve and digital signatures." nicj. net/files (2004).

[11] 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

