European Journal of Applied Sciences 8 (3): 156-163, 2016
ISSN 2079-2077
© IDOSI Publications, 2016
DOI: 10.5829/idosi.ejas.2016.8.3.22993

Privacy and Auditing Bigdata Stored in Cloud with Verify Update

K. Gokulkrishnan and Ramya G. Franklin

Department of MCA, Faculty of Computing,
Sathyabama University, Chennai-600119,Tamilnadu, India

Abstract: Cloud storage service provides an effective way to users to outsource their data in remote location and use the on-demand application in high quality way without worry of maintenance and load of local storage from public pool of configurable computing resources. In this process users of cloud services can not able to have physical connection with outsourced data. It creates the issue of data integrity in cloud computing particularly for users with controlled computing resources. In addition, users can be worry free about data integrity of storage data in cloud if cloud storage is local. Therefore, maintain the data integrity in cloud storage most important to users. So enabling the public auditability for outsourced data by third party auditor provides better solution to users and users can be worry free. In existing works data integrity process has done without control of real data file. Normally trusted third party is done the verification process of the outsourced data, this operation is called data inspection and this believed third party is called auditor. We say this updates is coarse grained updates. In small data updates this type of update process will cause updating the entire file block and re-computing the all the data, which results the communication overheads and higher storage. In this paper we propose fine grained data updates this provides request of fine-grained update and support authorized auditing. In our scheme we also provide some enhancement for verifying small updates that can considerable minimize the overhead of communication. Ranked Merkly Hash Tree (RMHT) is the technique that is used in this scheme support to split the data into small units, it mainly supports data integrity verification and theoretical investigation. Our experimental results show that this system not only supports flexibility and improved security, but for big data application it considerable reduce the overhead when more no of regular small updates.

Key words: Cloud computing · Flexibility · Improved security · Worry free · Re-computing

INTRODUCTION

Now a day cloud computing is a rapidly increasing technology. This technology is stronger and cheaper processors, jointly with software service process are moving data from one place to on large scale data centers [1]. The improving networks and network feasibility helps to user to outsource their data on remote location with more quality services. Cloud storage provides remote location to user to store their data that enables users need not worry about the hardware component. Cloud computing is moving to the next age group information technology(IT) for many organization and individual users, because of its many extraordinary advantages in IT history: everywhere network access, independent location resource pooling, fast resource elasticity, self-service in on-demand way, risk transference and usage based costing [2]. Cloud computing is fully support to organization with very nature of business environment in distributive manner. One important feature of this paradigm shifting is offers data are outsourced or being collected to cloud. In users view, including IT organization and individuals, outsourcing data to cloud storage with better on- demand way provides attractive advantages: relief to user for hardware component for storage administration and users can be worry free about data. While Cloud technology has more benefits than always, it also contains some new and critical security

Corresponding Author: K. Gokulkrishnan, Department of MCA, Faculty of Computing,
Sathyabama University, Chennai-600119,Tamilnadu, India.

issues over the outsourced data. In our proposed system, we will examine the integrity verification problem for big data storage in cloud and security of the data. Although cloud offers many advantage to users, the data correctness is important issues in cloud because of the following reasons. In first, cloud computing having the strong and good infrastructure than personal devices, but it still having more external and internal warnings in data integrity [3,4]. When the data integrity is verified by the trusted third party that is called data auditing [5, 6]. Examples of Noteworthy security breaches and outages shows from time to time [7-9]. Second there may reason for Cloud Service Provider may act unfaithfully to data users for their outsourced rank of data. For example, Cloud Service Provider might hide the data loss issues to user for maintain a name and even recover storage for financial reason by removing data that are rarely accessed. At the same time one of the main problem in cloud computing is that this new technology comes with some security challenges. The security problem of the cloud is that the stored data may not have the control from where it is placed and also unauthorized user can hack the data. So if user wants to provide security of the data, user must provide security technology on the data. Normally data exchange with security is important for any network. Here we use RSA algorithms for providing security on the user details when moving to TPA. RSA is a Public-Key cryptography algorithm. RSA Algorithm uses two keys public and private and which are asymmetric because one is used for encryption and another is used for decryption. In this scheme we mainly concentrate on better solution for dynamic updates of small data, which profit the efficiency and scalability of Cloud Storage. To overcome the above problem, our scheme use ranked Merkle hash tree technique and data segmentation with flexible strategy. At the same time we address the security difficulties in public verifiability to create this scheme with more robust and secure, which accomplish by provide an addition authorization process between three parties of Data Owner (DW), Cloud Storage Service (CSS) and Third party auditor (TPA).

Related Works: In cloud computing cloud user (CU) is responsible for outsourcing the data files in Cloud. The Cloud server (CS) is responsible for providing storage space and services to data storage. Normally CSP (Cloud service provider) manages the CS (Cloud Server). The trusted TPA (Third Party Auditor), is responsible for data auditing on cloud storage which is stored by cloud users. The auditing is conduct by TPA based upon user request. In common users depend on cloud server (CS) for storing the data on cloud storage and data maintenance. In this type of storage users interact with cloud server dynamically to update and access their data for different purpose. Users can not able to physical connection with CS every time. So users may choose TPA (Third party auditor) for protecting the integrity of outsourced data, at the same to keeping their data secure from Third Party Auditor (TPA). In business Third party Auditor is responsible for auditing, is independent and TPA has no connectivity with users and Cloud Storage during auditing process. At the same time data leakage of user data by TPA during the auditing should be banned[10-12]. User granting permission to the TPA for auditing data by provide audit rights to TPA's Public key via the signed certificate and TPA auditing is authenticated by that certificate.

Data privacy and security related problems has been extensively studied in past works [13,14,15]. In this work, the main aim is frequent and small data updates, it is important issues in cloud application because these updates process is exists in various cloud application like online social networks and some business transaction. More cloud users likes to cut their huge datasets into small datasets and keep that splitter dataset in various physical servers for data privacy-preserving, better processing and reliability. In cloud computing privacy and security is the important issues and these are most frequently occurred affect. In existing many technological approaches were used to increase the privacy and safety of cloud computing. In existing there was more work trying to increase the data privacy and security with Technological methods on CSP side. Now a days some extensive research is conducted on outsourced data. In [16] Ateniese etal are first proposed scheme of public auditability in their "provable data possession (PDP" representation to securing files control for outsourcing data on unfaithful storage places. This model achieved the public auditability by use of Homomorphic tags based with RSA. Using this way the auditability of public is achieved. In this system this author do not take the problem of dynamic storage of data and their model may struggle security and design problem when extending this model from static to dynamic data storage. However, this system support dynamic operation with priori bound of the queries and it does not co-operate for fully dynamic operation. i.e, it support only fundamental data chunk operations with minimum functionality and placing of data block cannot be assisted. In [17,18] Wang *et al.* proposed the protocol of response-challengel to identify the errors and correctness of data. The main focus of this scheme is to store the data in dynamic storage at distributed scenario. Alike to [17], in this they only provide limited support for dynamic operation. In [19] Jules *et al.* define a "proof of irretrievability (PoR)" system. In this system error-correcting and position verifing codes are proposed to protect both "retrievability" and "possession" of data on archive systems. Particularly to detecting the data file F some extra blocks called "sentinels" are added with data file F. Further the F is encrypted to secure the extra blocks position. In [17], a user can perform the no of queries is also conformed priori and precomputed "sentinels" block are used in dynamic data updates development. In [20] Erway *et al.,* first explore construction scheme for dynamic provable data possession. In this they used PDP model to support provable data updates to stored data using with rank-based authenticated skip lists. This method is important for dynamic version of PDP solution. To keep up the data privacy and integrity shan *et al.,* [21] proposed the TPA concept. This scheme reduces the users online burden and maintain the data privacy preserve. In [22] Chen *et al.,* introduces scheme for auditing the data correctness in multiple server. In existing schemes they were trying to provide data integrity and privacy for various storage systems. But the main problem in existing schemes, data dynamics and public auditability has not been addressed fully. In this scheme we will propose the public auditing process for maintain the data integrity and security on cloud. It benefits efficiency and scalability of CSS (Cloud Storage Server).

Proposed Work

Overview: In cloud data integrity and security is important problem for data users. To achieve the data security and integrity, we propose user verification process for security and public auditing process for integrity of data. To achieve the small dynamic updates our scheme uses the adaptable data segmentation scheme and a RMHT (ranked Merkle hash tree). At the same time we will address the data security problem by adding the RSA algorithm for transferring user data from CSS to TPA. Here Data integrity and user personal details are maintained by TPA. Third Party Auditor (TPA) is responsible for auditing process. In public auditing TPA can check the correctness of data in Cloud Storage Server (CSS) without recover copy of whole data. This public auditing gives advantages for reduce the online burden and save the computation resources. In this scheme we use Ranked Merkle Hash Tree (RMHT) for split the Data Owner data whenever they upload the data on CSS for store. This Ranked Merkle Hash Tree (RMHT) technique helps for TPA for auditing process and big data storage on cloud.

Data Owner Registration: In this proposed system for providing the security of data from unauthorized users we create a user Id and verify Id by TPA for Data Owner to outsource the data for storing on cloud with securely. In registration process when ever Data Owner registers on Cloud Storage Server (CSS) for upload data, CSS will generate the private and public key for encrypt the user details and forward to Third party auditor (TPA). Then TPA will create user id and verify id for security purpose. The user id will be send to Data Owner by TPA. Mean While verify id send to CSS. That User id will be checked by CSS with verify id when the Data Owner upload the data to CSS. Using this Registration process unauthorized users access is controlled on CSS.

Secure Data Moving and Storing: The common public key algorithm is RSA, This algorithm is named based on the name of creator (Rivest, Shamir and Adleman-RSA). Basically RSA algorithm is an asymmetric type. Asymmetric means it uses two keys first is the public key issued to all, using this public key one can encrypt the data and second is the private key is used for decryption the data and it is kept secret and should not disclose to everyone. Here working procedure of RSA in our cloud computing is described as: Mainly RSA algorithm is used for provide the security of data. In our project we use RSA algorithm for encrypting the user registration details for securely moving to the TPA. The main aim of the data securing is that authorized user can only access the data.

TPA Auditing: In proposed system, it fully supports the fine grained update requests and authorized auditing. To protect the integrity of data and cloud user online burden also save the computation resources, it is important to allow public auditing process for Cloud Storage. So achieve the public auditing users may use trusted TPA (Third Party Auditor) to examine the data when ever user needed auditing. The TPA, who has knowledge and ability for auditing can check the data integrity on the

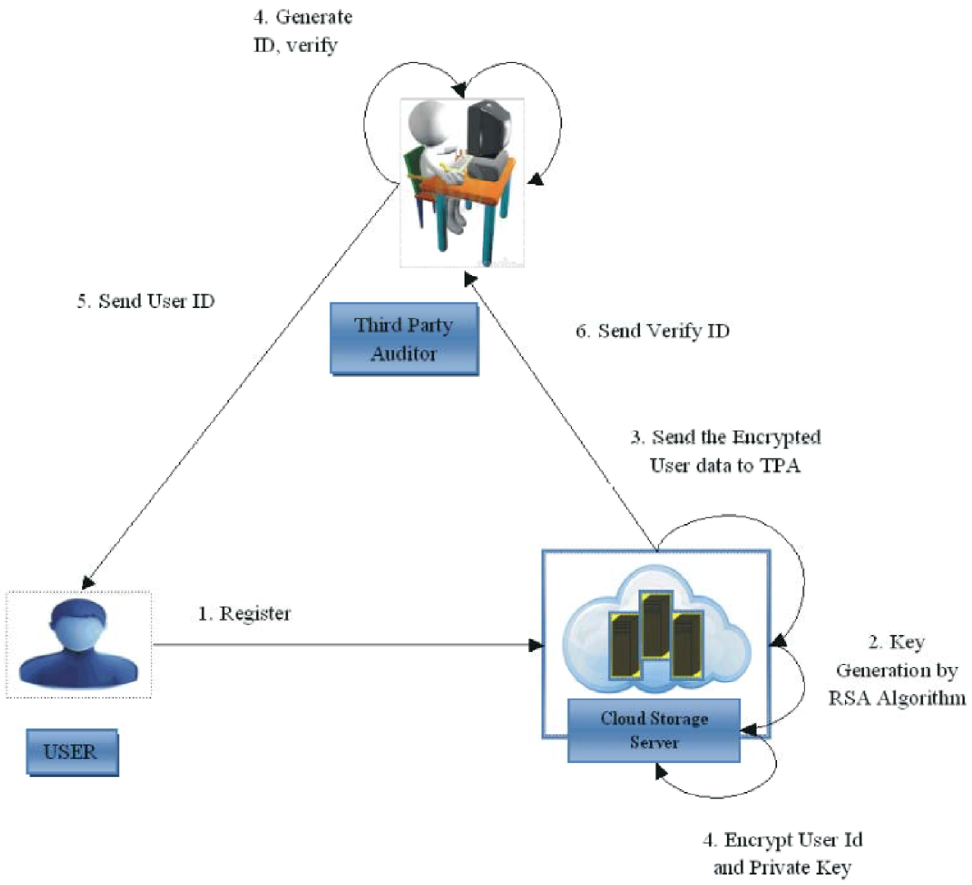


Fig. 1: Data Owner Registration and Generate User ID

cloud in periodic way on behalf of user request. It provides easiest and low-cost way for cloud user to protect their correctness of data on the cloud. In existing system TPA auditing is achieved by sending the message to cloud server and cloud server auditing it based on the request by the TPA. In our proposed system our TPA auditing is done on directly on the cloud servers. Here TPA provides the advantage for auditing the data on based on the user request. By this type of TPA auditing we provide more integrity for user data. Normally public auditing schemes contain four algorithms. KeyGen, SigGen, GenProof, Verify Proof. KeyGen algorithm is used to generate the key, which is executes by the users to set up system. In SigGen the metadata of verification is created by user. In which it contain digital signature. GenProof algorithm is executes by CS to create data storage evidence. TPA executes the Verify Proof algorithm for verify and audit the proof.

In public auditing scheme, it contain two process describes below.

Setup: By running the algorithm of KeyGen, user can assign the secret and system public parameters. Using the algorithm of SigGen it creates metadata for proof by data file F preprocessing. The data file F and proof metadata F is stored on cloud by user and remove its copy in local. User can change the data file by enlarging it.

Audit: Third Party Auditor send message or challenge the audit to CS to ensure that the CS has keep data file F when auditing process is going. When running the GenProof Algorithm, CS receive reply message by file F and input as metadata. Finally TPA checks the response message provided by the CS using VerifyProof alogorothm.

Verifiable Fine-Grained Dynamic Data Operations: Users in cloud computing may update their data for different application needed in cloud server. Public auditing is a process to support the data dynamics, using this user can do changing on outsourced data. Some existing auditing

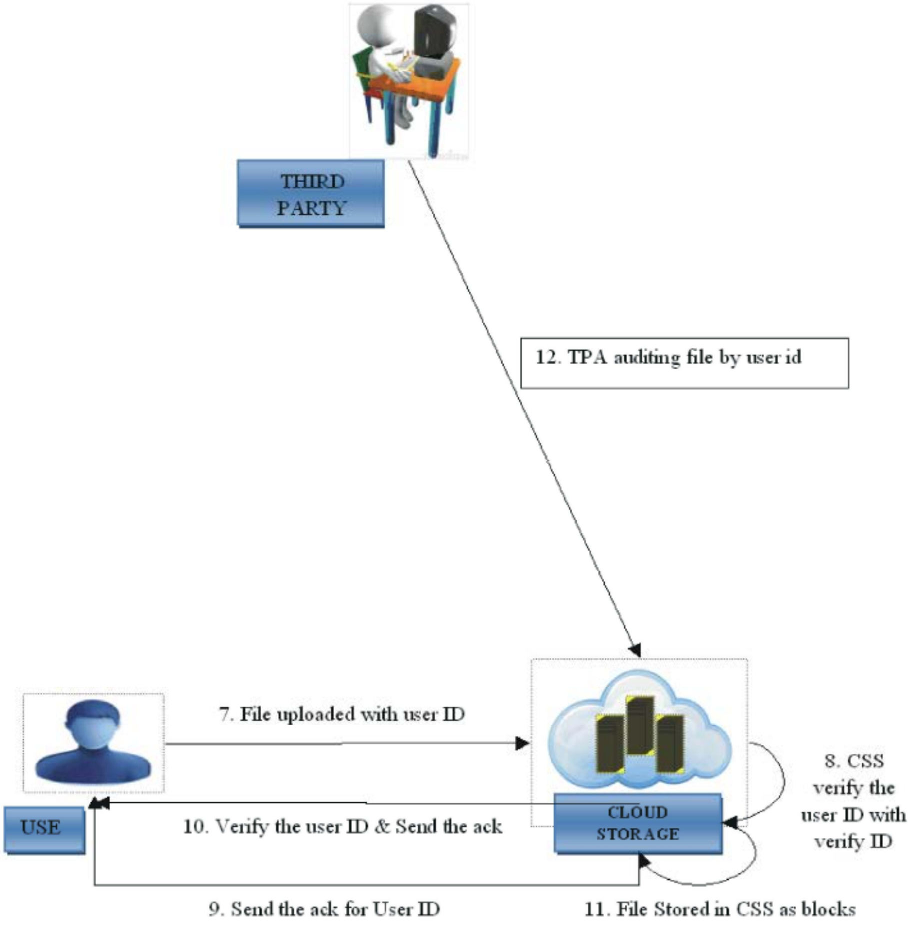

THIRD PARTY
12. TPA auditing file by user id
7. File uploaded with user ID
USE
10. Verify the user ID & Send the ack
CLOUD STORAGE
8. CSS verify the user ID with verify ID
9. Send the ack for User ID
11. File Stored in CSS as blocks

Fig. 2: Data Splitting and Storing on CSS

schemes only support the full data dynamics. In that models, only modification, insertion and deletions has been done only with fixed sized blocks.

In past Merkly Hash Tree has been studied intensively. But in this scheme we use an improved Merkly Hash Tree (MHT) with ranks called RMHT. Like the binary tree, every N node will have utmost two child nodes. Depend on the algorithm of update, non-leaf nodes will always contain two child nodes. In RMHT information holded Tree T is stand as {H, rn} where rN represent rank of this node and H represent hash value. Tree T is created as follows, For a LN (leaf node) based on a mi (message), we have=h (mi), rln=is a parent node N1 = {H1; rN1} and N2= {H2; rN2} is builded as Np = {hH1kH2; rN1, rN2} where || is a sequence operator.

Perform Update Algorithm: Cloud Storage Server (CSS) executes the Update algorithm when updating the user data files F. Cloud Storage Server(CSS) CSS parses Update req and receive {PM,I,o,m new}. Cloud Storage Server update T and mi accordingly when type PM, then output updates and updated file F. After ending this algorithm, Cloud Storage Server will forward P for update to user.

Verify Update Algorithm: Once the user get H (mi), it will first calculate R by H(mi));Ωi and validate sig, then using the newg it will start to calculate mi and calculate Rnew with m;Wig and evaluate Rnew. User will return signature to Cloud Storage Server for it to update accordingly if Rnew R.

RESULT AND DISCUSSION

For ensure the security and data integrity of the cloud storage data, various techniques have been used to achieve that issues. In this proposed system we use user ID and verify ID for security of data on the cloud and

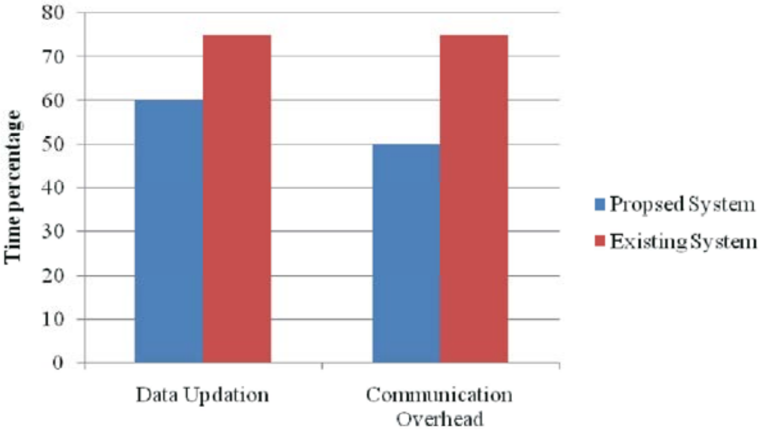
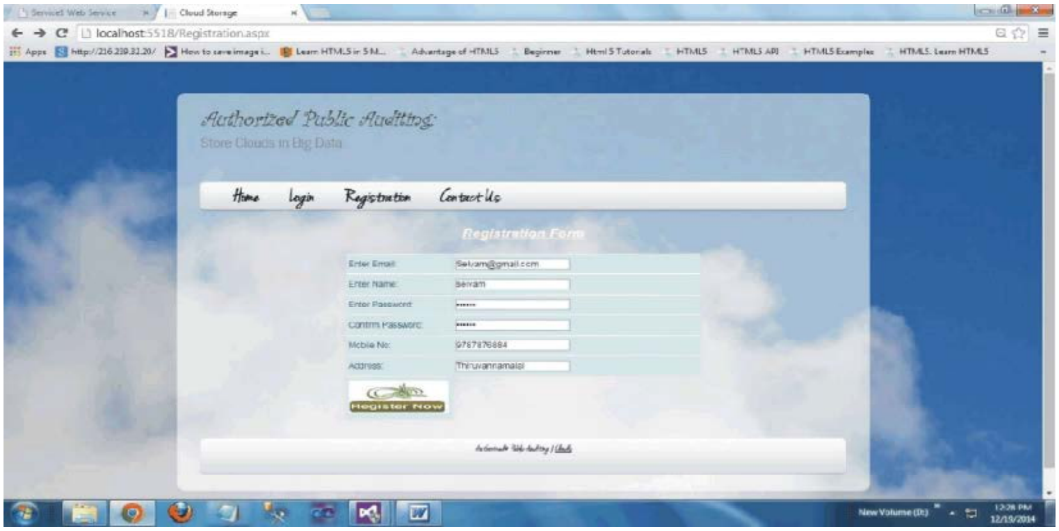


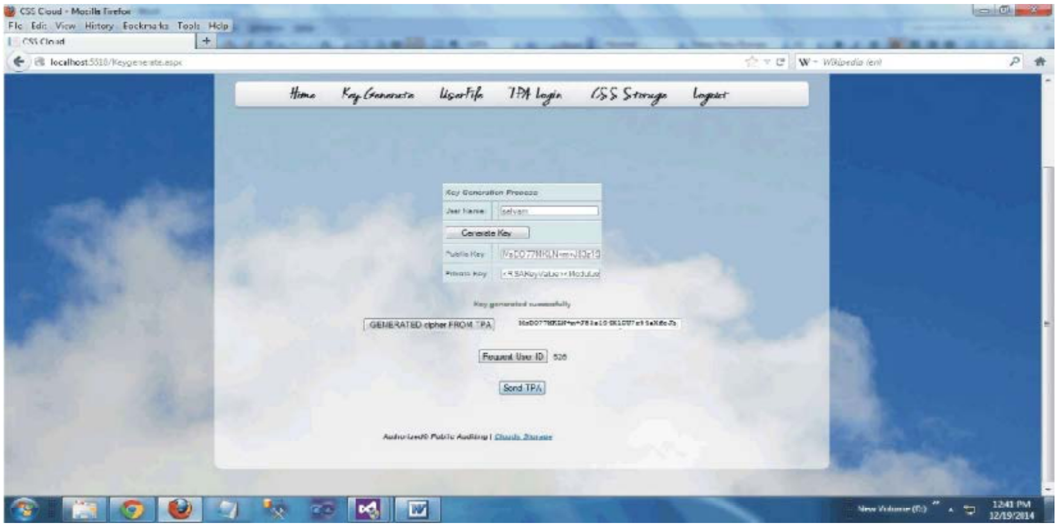
Fig. 3: Data updation and Communication overhead comparison with existing system



Registration Form

Enter Email:	Selvam@gmail.com
Enter Name:	selvam
Enter Password:	•••••••
Confirm Password:	••••••
Mobile No:	9787876884
Address:	Thiruvannamalai

Fig. 4: Data Owner Registration


CSS Cloud - Mozilla Firefox
File Edit View History Bookmarks Tools Help
localhost:5518/Keygenerate.aspx
Home
Key Generate
UserFile
TPA login
CSS Storage
Logout
Generate Key
GENERATED cipher FROM TPA
Request User ID
Send TPA
Authorized© Public Auditing | Clouds Storage

Key Generation Process

User Name	selvam
Public Key	MsDO77MKLN+m+J83e19
Private Key	<RSAKeyValue><Modulus

Key generated successfully

MsDO77MKLN+m+J83e19 0K1D07e9 5aX6c7s

Request User ID: 526

Fig. 5: Key Generation on CSS

Ranked Merkle Hash Tree (RMHT) is used to split the data to blocks for storing on CSS. In existing scheme data updating on cloud storage fixed size blocks format. So for if every small updates on data it consume more computation time and provide overheads on network. In our proposed system we use fine grained updates for every data updates so it minimizes the size of data blocks and computation time.

The above Fig 3 explains the comparison of data updation and communication overhead with existing system. In existing system it when updating the data it takes more times and more space for even a small data updates. In proposed scheme it takes minimum and less space for small data updates. Then also compare to the existing scheme our proposed scheme reduce the communication overhead.

The above Fig 4 shows the screen of registration details of data owner for get the user id to upload the data on Cloud Storage Server. After successful finished the registration form, that details will forward to CSS. Using that detail CSS generate the key for encrypt the Data owner detail.

Fig 5 shows the process of key generation for send the details of data owner. Using the public key CSS will encrypt the details and forward to TPA for create ID. TPA will create user ID and verify ID using the details send by CSS. User ID will send to Data Owner and verify id will send to CSS from TPA.

CONCLUSION

In cloud computing data integrity and security is main problem for cloud user while using cloud services. To achieve the Data security in our system we use verification process with user id and verify id that are generated by TPA it later checked by TPA when user enter into the cloud server. Data integrity is achieved using public auditing by TPA. TPA can be used to ensure the integrity and security of data. In common trusted party can behave a Third Party Auditor (TPA) to fix the clash between user and CSP. Here in this proposed scheme, we are using public auditing process to supports the dynamic data updates. Public auditing scheme also supports hashing technique. Ranked Merkle Hash Tree (RMHT) is technique used for data dynamic operations on cloud storage. Using this RMHT user data can be splitter to blocks when uploading data to Cloud Storage. Our experimental and theoretical analysis results have indicate that our proposed scheme not only provide flexibility and security it also reduce the big data overheads process application with more number of compact update in frequent manner such as business transaction and social media application.

REFERENCES

1. Franz, M., P. Williams, B. Carbunar, S. Katzenbeisser and R. Sion, 2011. Oblivious Outsourced Storage with Delegation, in Proc. Finan-cial Cryptography and Data Security Conference (FC), pp: 127-140.
2. Wang, C., Q. Wang, K. Ren and W. Lou, 2010. Privacy-Preserving Public Auditing for Storage Security in Cloud Computing, Proc. IEEE INFOCOM', pp: 10.
3. Mell, P. and T. Grance, 2009. Draft NIST Working Definition of Cloud Computing, http://csrc. nist.gov/groups/SNS/cloudcomputing/ index.html.
4. Armbrust, M., A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Kaminski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica and M. Zaharia, 2009. Above the Clouds: A Berkeley View of Cloud Computing, Technical Report UCB-EECS-2009-28, Univ. of California, Berkeley.
5. Arrington, M., 2006. Gmail Disaster: Reports of Mass Email Deletions, http://www.techcrunch.com 2006/12/28/gmaildisasterreports of-mass-email-deletions/.
6. Wang, Q., C. Wang, K. Ren, W. Lou and J. Li, 2011. Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing, IEEE Trans. Parallel Distrib. Syst., 22(5): 847-859.
7. Wang, C., Q. Wang, K. Ren and W. Lou, 2010. Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing, in Proc. 30st IEEE Conf. on Comput. and Commun. (INFOCOM), pp: 1-9.
8. Kincaid, J., 2008. MediaMax/TheLinkup Closes Its Doors, http:// www.techcrunch.com /2008/07/10/ mediamaxthelinkupclosesits- doors.
9. Amazon.com, 2008. Amazon s3 Availability Event: July 20, 2008," http://status.aws.amazon. com/s3-20080720.html, July.
10. Wang, Q., C. Wang, K. Ren, W. Lou and J. Li, 2011. Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing, IEEE Trans. Parallel and Distributed Systems, 22(5): 847-859.
11. Ateniese, G., R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson and D. Song, 2007. Provable Data Possession at Untrusted Stores, Proc. 14th ACM Conf. Computer and Comm. Security (CCS '07), pp: 598-609.

12. Shah, M.A., R. Swaminathan and M. Baker, 2008. Privacy- Preserving Audit and Extraction of Digital Contents, Cryptology ePrint Archive, Report 2008/186.
13. Erway, C., A. Ku pcu, C. Papamanthou and R. Tamassia, 2009. Dynamic Provable Data Possession, in Proc. 16th ACM Conf. on Comput. and Commun. Security (CCS), pp: 213-222.
14. He, Y., S. Barman and J.F. Naughton, 2011. Preventing Equivalence Attacks in Updated, Anonymized Data, in Proc. 27th IEEE Int'l Conf. on Data Engineering (ICDE), pp: 529-540.
15. Ateniese, G., R.D. Pietro, L.V. Mancini and G. Tsudik, 2008. Scalable and Efficient Provable Data Possession, in Proc. 4th Int'l Conf. Security and Privacy in Commun. Netw. (SecureComm), pp: 1-10.
16. Ateniese, G., R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson and D. Song, 2007. Provable Data Possession at Untrusted Stores, Proc. 14th ACM Conf. Computer and Comm. Security (CCS ?07), pp: 598-609.
17. Ateniese, G., R.D. Pietro, L.V. Mancini and G. Tsudik, 2008. Scalable and Efficient Provable Data Possession, Proc. Intl Conf. Security and Privacy in Comm. Networks (SecureComm 08), pp: 1-10.
18. Wang, C., Q. Wang, K. Ren and W. Lou, 2012. Towards Secure and Dependable Storage Services in Cloud Computing, IEEE Trans. Service Computing, 5(2): 220-232.
19. Juels, A., J. Burton and S. Kaliski, 2007. PORs: Proofs of Irretrievability for Large Files, Proc. ACM Conf. Computer and Comm. Security (CCS 07), pp: 584-597.
20. Erway, C., A. Kupcu, C. Papamanthou and R. Tamassia, 2009. Dynamic Provable Data Possession, Proc. 16th ACM Conf.Computer and Comm. Security (CCS ?09).
21. Shah, M.A., M. Baker, J.C. Mogul and R. Swaminathan, 2007. Auditing to Keep Online Storage Services Honest, Proc. 11th USENIX Workshop Hot Topics in Operating Systems (HotOS 07), pp: 1-6.
22. Chen, B., R. Curtmola, G. Ateniese and R. Burns, 2010. Remote Data Checking for Network Coding-based Distributed Stroage Sys-tems, in Proc. ACM Cloud Computing Security Workshop (CCSW), pp: 31-42.