BST-MRPAS: Binary Search Tree based Multi Replica for Public Cloud Auditing System

Savita Vishwakarma
Department of Computer Science and Engineering
Medicaps University, Indore, India

Rudresh Shah
Department of Computer Science and Engineering
Medicaps University, Indore, India

ABSTRACT

Data integrity and storage efficiency are two essential needs for cloud storage. Cloud computing has emerged as a long-dreamt vision of the utility computing paradigm that provides reliable and resilient infrastructure for users to remotely store data and use on-demand applications and services. Currently, many individuals and organizations mitigate the burden of local data storage and reduce the maintenance cost by outsourcing data to the cloud. However, the outsourced data is not always trustworthy due to the loss of physical control and possession over the data. With the growing awareness of data privacy, more and more cloud users choose to encrypt their sensitive data before outsourcing them to the cloud. Therefore, in order to resolve the issues in cryptographic security a new methodology is required to develop. That provides the data owner and data management with efficient. Thus, the proposed solution incorporates the TF-IDF for calculating frequency of each word of text data and performing indexing on selected words.

In this paper, the investigation is done about the outsourced data and their sensitivity and security issues. In this, a mechanism is proposed for public cloud data security by means of *BST-MRPAS* i.e. Binary Search Tree based Multi Replica for Public Cloud Auditing System for the end user applications. Additionally for providing end user trust and security management the upload, download and Update services are provided.

Keywords

Cloud Storage, Public Auditing, Multi-Replica, TF-IDF, Data integrity, Cloud Server.

1. INTRODUCTION

While cloud computing is gaining popularity, diverse security and privacy issues are emerging that hinder the rapid adoption of this new computing paradigm. And the development of defensive solutions is lagging behind. To ensure a secure and trustworthy cloud environment it is essential to identify the limitations of existing solutions and envision directions in current trends [1]. Cloud storage provides scalable and Quality of service guaranteed resources for storage, users can store and compute their data from any location at any time by a device which can be connected with Internet to visit that cloud. Besides these powerful advantages of cloud Storage, however, many people and companies is still feel hesitant to store their data in cloud. The reason behind this hesitancy is the fear of people and companies regarding loss of control on their data because there are some incidents of data loss and data leakage which make people to think about it [2].

1.1 Cloud Data Storage

In the last decade, the demand of outsourcing data is greatly increased. Data storage and high performance computation are the main needs, which have to be fulfilled. These services are provided by many cloud computing service providers like Drop box, Google App Engine, Amazon Simple Storage Service (AmazonS3), etc. The advantage of storing data in cloud servers is that the data owners can reduce the overhead of buying extra strong servers and also avoid hiring of server management engineers. The technology used for internet based development is nothing but cloud computing. Cloud provider offers one of the most fundamental services that are data storage. Data encryption is a basic solution to maintain security of data and the encrypted data is uploaded into the cloud. Depending on the possibility to identify privacy and security users cannot join the cloud computing systems [3] [4] [5].

1.2 Secure Cloud Storage Auditing

Cloud storage is nowadays very popular storage system. Cloud storage is storing of data off-site to the physical storage, which is maintained by third party. Cloud storage is saving of digital data in logical pool and physical storage spans multiple servers, which are manage by third party. Third party is responsible for keeping data available and accessible and physical environment should be protected and running at all time. Instead of storing data to the hard drive or any other local storage, we save data to remote storage, which is accessible from anywhere and anytime.

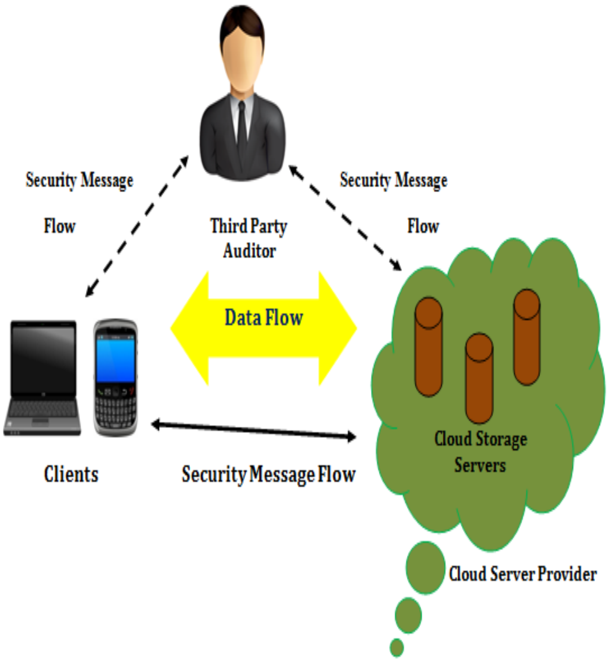

Figure 1: Cloud Data Storage Auditing

It reduces efforts of carrying physical storage to everywhere. By using cloud storage, we can access information from any computer through internet, which omitted limitation of accessing information from same computer where it is stored.

While considering data privacy, we cannot rely on traditional technique of authentication, because unexpected privilege escalation will expose all data. Solution is to encrypt data before uploading to the server with user's own key. Data sharing is again important functionality of cloud storage, because user can share data from anywhere and anytime to anyone. For example, organization may grant permission to access part of sensitive data to their employees. But challenging task is that how to share encrypted data. Traditional way is user can download the encrypted data from storage, decrypt that data and send it to share with others, but it loses the importance of cloud storage [6].

2. LITERATURE SURVEY

The given section provides the understanding about the Relevant previous technique of cloud storage secure auditing system that are recently contributing in cloud environment therefore a number of research articles and research papers are included in this section.

Kai, He, et al. [7] proposes a public batch auditing protocol for multi-cloud storage for data integrity. In this, a TPA can verify the multiple auditing requests simultaneously from different users on multiple servers. To achieve this system makes use of homomorphic cipher text verification and recoverable coding approach. This system provides quick detection of corrupted data with minimum communication cost.

A.L. Ferrara et al. [8] introduce batch auditors including regular, identity-based, group, ring and aggregate signature schemes. This paper has implemented all this algorithms and compares batching algorithms. This paper identifies that whether the batch auditing is practical or not.

To achieve efficient data dynamics, ***Q. Wang et al [9]*** enhance the proof of storage models. Block tag authentication system have made some changes in classic Merkle Hash Tree construction structure. System also includes the bilinear aggregate signature for simultaneous auditing of multiple user requests with TPA. This system is secure and more efficient.

Some systems provide data confidentiality using private key encryption techniques during TPA audit process. But some system does not consider data freshness property during development of new system. ***Priya, K et al. [10]*** show the disadvantage is overcome by implementing HMAC mechanism. This scheme is beneficial for metadata secrecy, integrity checking. This system randomly verifies the data instead of whole data checking. This is the disadvantage of this system.

Chakraborty et al. [11] implements the homomorphic encryption scheme for providing the data confidentiality. This scheme is based on the Elliptic curve cryptography technique. System also implements a data possession scheme. It has provided dynamic operations on data. The third party auditor verifies and modifies the data on behalf of the client. Merkle hash tree (MHT) is used to fast access on data stored in cloud. This system checks the correctness of data and also identifies the misbehavior activities in server.

Yan et al. [12] have implemented a novel remote integrality checking approach in cloud computing. It has efficient data integrity checking, dynamic update and data confidentiality. This system has checked the mass of files remotely with constant storage and resources of communication.

3. PROPOSED WORK

3.1 Problem Statement

The cloud environment provides support for efficient computing and enables to provide the efficient computing and storage solutions at the remote end. In this presented work the main aim to address the problem of data integrity to between two parties when our data is too outsourced on cloud storage. The security and privacy of the data is one of the major issues in adoption of cloud computing. In comparison with conventional system, the cloud users will be completely isolated with having direct control over their data. This concept will investigate the problem of integrity verification big data storage in cloud. This problem is known as data auditing, when verification is conducted by third party. From cloud users point of view it is known as 'Auditing as a service'. In case of remote verification scheme, the cloud storage server (CSS) cannot certify the integrity proof of a given fraction of data to a verifier until whole data is intact. To ensure stored data integrity this support is no less important than any data protection mechanism deployed by cloud service provider (CSP), how much ever secure they pretend to be in that it will provide the verifier a piece of direct, trust worthy and real-timed intelligence of the integrity of cloud user's data through a challenge request. It is required that an improved and efficient data auditing scheme is to be conducted on regular intervals of time for users who ensure higher security demand for their data.

3.2 Proposed Methodology

A problem associated with the MHTs based data integrity schemes is the efficient storage and retrieval of MHTs in our MySQL Database. Numerous approaches exist to store hierarchical or tree like data in a database, e.g., adjacency list, nested set, nested interval and closure table etc. Each approach has its pros and cons. For this specific problem of storing MHTs, a new technique named Binary Search Tree Based Multi Replica Public Auditing System has been proposed in this research work.

For the advancement of the current security scenario here it present a proposed BST-MRPAS working flow for ensures security and privacy enhancement of the system:

Description: The figure 2 shows the working flow model of the proposed system. The proposed system proves the authentication and integrity with improving the data auditing and replication for cloud data storage. Users are firstly uploading the data with delegation means that cover required or initiated data. The security and authentication system uses the Binary Search Tree based algorithm that improves the efficiency, bandwidth in terms effectiveness of cloud storage data. In this, firstly is done process of cloud setup where two modules are establish one is user panel and second one is server side panel. User panel is to responsible managed all the activities e.g. upload, update, download etc. Furthermore, server side panel is responsible for storing user data and provide authenticity and legitimacy. Also main task is that it prevents from data integrity which is based on internal process. So that for accessing the function of user panel users need to get registered with the system through user registration process. While registering, users need to submit their personal details for completion of registration process. User registered with their details such as identity (user name, password and email-id).

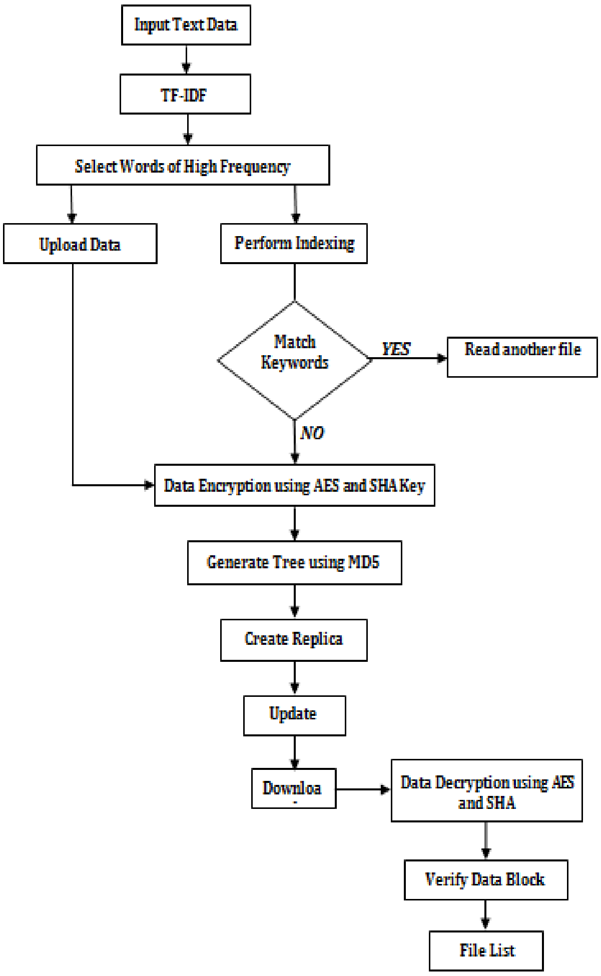

Figure 2: Proposed BST-MRPAS Working Flow

The user can login successfully only if user id and password are entered correctly. The login is a failure if the incorrect user id or wrong password is entered by the user. This helps in preventing unauthorized access. In next, it takes an input text data from the user panel. When browse the data for uploading it creates 3 folders on cloud server of that file. The folder name should be similar to the file name. Example, if the file name is B then folder name should beB_1, B_2, B_3. Each folder contains exact one replica of the text file. Text data is splits into 3 replicas i.e. b_1, b_2, b_3. Hence, B_1, B_2, B_3 contains b_1, b_2, b_3 respectively. We find frequency on the basis of a particular word that how many times encounter in each file. Keep top of the words, which have high TF-IDF. Select top 50 words, which have high frequency of overall text data. In this flow diagram, shown base upload data and proposed upload. In base approach, upload the data on cloud server, and store data. In proposed BST-MRPAS, it performed indexing using inverted index technique. Inverted index used when we upload data file to match the data if file is already exist on server. If the response is return true that means file is exist and upload another text file on second time. If indexing return false means that data keyword does not match by existing file, so that perform data encryption using AES algorithm on selected words for base and proposed. After encryption process, generate hash key using SHA. The SHA calculate using byte value of file. And it required hash key value for data encryption and decryption. The key length is 256 bits of SHA for AES where it encrypts and decrypts data using SHA key. After that server generate tree on basis of MHT and Binary Search Tree. In propose tree generating using MD5 algorithm. This is main concept of the system. In next, user can update its data that have been uploaded. Finally, original file can be downloaded by verifying data block file. For downloading file, data decryption is performed using AES algorithm and verifies data using SHA key to ensure the data integrity.

Verifying the authenticity of data has emerged as a critical issue in storing data on untrusted servers. The Proposed system analyzes various types of update requests on dynamic data of varied size of files. Additional authorized authorization process is incorporated in this scheme to eliminate malfunctioning of unauthorized challenges from unknown users during auditing process. This scheme also investigates how the efficiency can be improved while updating and verifying frequent minor updates which prevails in many cloud applications.

4. RESULT ANALYSIS

4.1 Time Consumption

The amount of total time required to perform execution of the selected algorithm is termed as the time complexity of the proposed method. The total required time is measured in terms of the milliseconds of the proposed and traditional system and is demonstrated using figure 3.

$$\begin{aligned} \text{Time consumption} \\ &= \text{Algorithm End Time} \\ &- \text{Algorithm Start Time} \end{aligned}$$

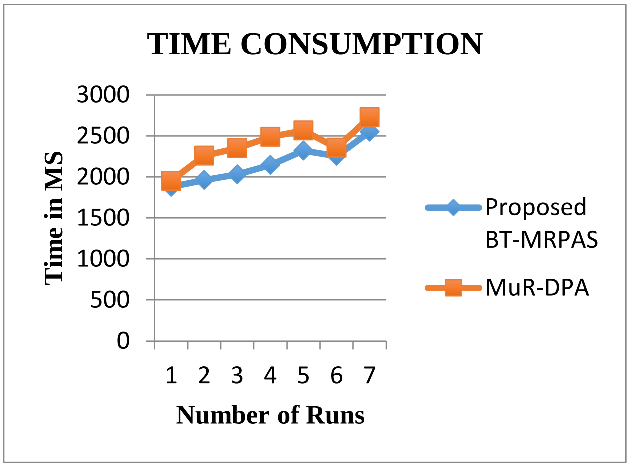

Figure 3: Time Consumption

In order to show the performance of implemented approaches the time consumption is reported in figure 3. In this diagram the X axis shows the different experiments on which different values generated and the Y axis shows the amount of time consumed for processing the data file. Additionally the performance of proposed BST-MRPAS is given using blue line and traditional system MuR-DPA depicts using orange line. According to the given results, the proposed system consumes less time as compared to other traditional algorithm. Additionally the results shows the amount of time consumed is depends on the amount of data provided for execution. But the respective performance of the system shows their effectiveness over the traditional algorithm. Moreover, while implementing BST-MRPAS, is enhancing the security for cloud data storage.

4.2 Memory Used

The amount of main memory required to execute the algorithm with the input amount of data is known as the memory consumption. The total memory consumption of the algorithm is computed using the following formula.

$$\text{consumed memory} = \text{total memory} - \text{free memory}$$

The figure 4 shows the total memory consumption of the system. In this diagram, the amount of main memory consumed is given in Y axis and different experiments performed is reported in X-axis. According to the obtained results, the proposed algorithm consumes fewer resources as compared to the traditional security technique.

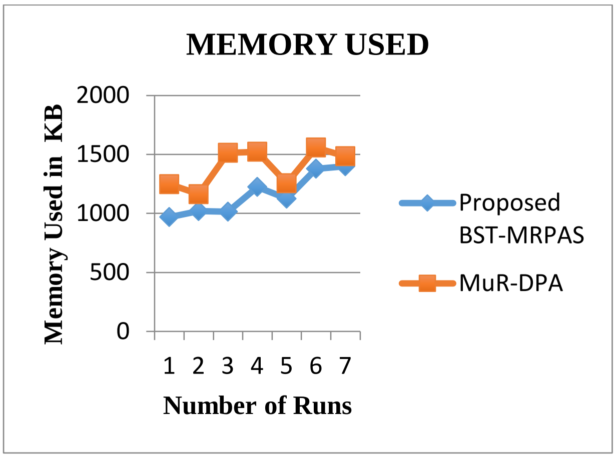

Figure 4: Memories Used

4.3 Tree Levels

A binary search tree is one where each node has, at most, 2 children, creatively named left and right. In a tree data structure, the total number of edges from leaf node to a particular node in the longest path is called as Level of the tree. Simply we can define; maximum number of height of the tree is showing that number of level of the tree. Following are the formula to calculate level of tree:

$$\text{Level of Tree (n)} = 2^{(\text{Height}+1)} - 1$$

In this, Figure 5 shows that clear demonstration of the levels between Proposed and traditional system i.e. BST-MRPAS and MuR-DPA respectively. In this figure, X-axis Shows different runs and Y-axis shows that number of levels of both implemented method. By analyzing of both where minimum tree levels shows, that it minimizes time complexity and space complexity of the approach. Therefore, we can say that levels of the tree are directly affected to the time and memory of the BST-MRPAS and MuR-DPA. Tree levels are constructed using binary search tree and merkele hash tree (MHT). In this, propose BST-MRPAS have less number of tree levels and traditional method generating high number of levels. So that proposed approach is more efficient to deliver high level of security.

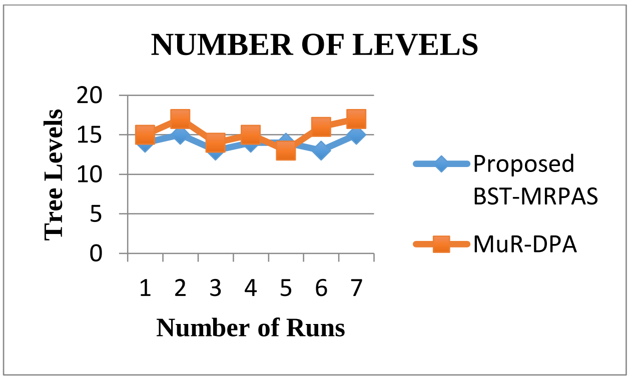

Figure 5: Levels of Tree

5. CONCLUSION

Public clouds are popular nowadays, where they are generally used in the storage and retrieval of the user's information. With the advent of cloud computing, more and more sensitive data are outsourced to the cloud server to reduce the management cost and enjoy the ubiquitous access. However, this novel computing paradigm introduces serious privacy challenges in that users' data are no longer locally possessed but stored on the remote server, which belongs to a different trust domain compared with the data users'. In this chapter, it focuses on the privacy and security concerns of public auditing based data transmission between server and end user, function performed over encrypted cloud data.

In this paper, it presents privacy-preserving public auditing system for data storage security in Cloud Computing. The proposed Binary search tree based Multi Replica for Public Auditing System based Secure Data Transmission i.e. *"BST-MRPAS"* is initiated in a dispersed environment where the single storage is secure way of authenticity and storage or hosting services. In addition of that for preventing the unauthorized access to the system a strong user authentication technique using the normal credential and multi replica of different uploaded data files. Data files are stored on cloud server and have been updating data files at the same place. Multi replica based data blocks are ensure that data integrity between two parties are secure in such a way that public cloud data are accessible throughout the session. Furthermore, for securing the data in storage and untrusted network an AES used to performing encryption and decryption and MD5 used for generating tree on the basis of data split is implemented. It has also demonstrated that it provides enhanced security and flexibility and significantly lower overhead for big data applications with large number of frequent small updates such as application in social media and business transactions.

6. REFERENCES

[1] Liu, Yuhong, "A survey of security and privacy challenges in cloud computing: solutions and future directions", Journal of Computing Science and Engineering 9.3 (2015): 119-133.

[2] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, "Above the clouds: A berkeley view of cloud computing", EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28, 2009.

[3] J. F. Yang and Z. B. Chen, "Cloud Computing Research and Security Issues", 2010 IEEE International Conference on Computational Intelligence and Software Engineering (CiSE), Wuhan pp. 1-3, Dec. 2010.

[4] M.A. Shah, M. Baker, J.C. Mogul, and R. Swaminathan, Auditing to Keep Online Storage Services Honest, Proc. 11th USENIX Workshop Hot Topics in Operating Systems (HotOS 07), pp. 1-6, 2007.

[5] N. Praveen Kumarga and D. Sireesha, "Ensuring Data Integrity in Cloud Computing", IJCSNS International Journal of Computer Science and Network Security, Volume 14 No.9, September 2014.

[6] G. Janani and C. Kavitha, "Public Auditing of Dynamic Big Data Storage with Efficient High Memory Utilization and ECC Algorithm", International Journal of Innovative Research in Computer and Communication Engineering, Volume 3, Issue 3, March 2015.

[7] Kai, He, "An efficient public batch auditing protocol for data security in multi-cloud storage", 2013 8th ChinaGrid Annual Conference (ChinaGrid), IEEE, 2013.

[8] A.L. Ferrara, M. Green, S. Hohenberger, and M. Pedersen, "Practical Short Signature Batch Verification", Proc. Cryptographers' Track at the RSA Conf. 2009 on Topics in Cryptology (CT-RSA), pp. 309-324, 2009.

[9] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing", IEEE Trans. Parallel and Distributed Systems, vol. 22, no. 5, pp. 847-859, May 2011.

[10] Priya, K., and I. Gunavathi. "Ensure cloud storage correctness based on public auditing mechanism", Communications and Signal Processing (ICCSP), 2015 International Conference on, IEEE, 2015.

[11] Chakraborty, TamalKanti, "Enhanced public auditability & secure data storage in cloud computing" 3rd International IEEE, 2013, Advance Computing Conference (IACC), 2013 IEEE.

[12] Yan, Xiangtao, and Yifa Li. "A wew remote data integrity checking Scheme for cloud storage with privacy preserving", 14th International conference Communication Technology (ICCT), 2012 IEEE.