NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 3, ISSUE2, Feb.-2016

MHT BASED INTEGRITY AUTHENTICATION FRAMEWORK: A DIFFERENT APPROACH FOR CLOUD DATA SECURITY

Ms. Sajjan R.S.
Department of Computer Science and Engineering
VVP Institute of Engineering and Technology Solapur (Soregaon), India

Mr. Vijay Ghorpade
Ph.D. SRTM University, Nanded

Mr. Arkas B.D.
Department of Computer Science and Engineering
VVP Institute of Engineering and Technology Solapur (Soregaon), India

ABSTRACT

Cloud computing that concern the latest trend in application development for Internet services, relying on clouds of servers to manage multiple tasks that used by individual machines. The developers take important services, such as email, calendars, and word processing, and host them entirely online, proposed by the cloud of special servers with cloud computing. Cloud storage has become more popular for storing and sharing data across multiple users. Cloud computing is a recent trend in IT where data is stored in data centers rather than on personal portable PCs. But cloud do not ensures data security parameters such as data integrity, data authentication, data auditing and control to its users. Also in order to improve data reliability and availability, cloud service providers use common strategies like storing multiple replicas along with original datasets.
Public data auditing schemes allow users to verify their outsourced data storage without having to retrieve the whole dataset. But existing data auditing techniques suffers from efficiency and security problems such as communication overhead (bandwidth) to maintain data integrity, and public auditing. To address these problems, the proposed system presents novel public auditing schemes like Data Integrity Verification and Data Modification & Data Insertion based on the *Merkle Hash Tree* (MHT). Also the proposed system assures users to securely store and share their data efficiently and effectively on the cloud.

KEYWORDS—Cloud; Merkle Hash Tree; Authentication; Integrity; Security;

INTRODUCTION

Cloud is most intensive research topic that incorporates the Internet services. For the management of huge upcoming data and to provide infinite computing resources on demand, so the cloud is highly scalable. The cloud computing has acquired popularity in recent IT technologies. Cloud computing is combination of multiple existing technologies such as web computing, parallel and distributed computing, grid computing, utility computing, virtualization etc. Cloud storage provides special features for customers with benefits, ranging from cost saving and simplified convenience, to mobility opportunities and scalable

service. These great features attract much more customers to utilize and manage the storage of their personal data to the cloud storage. The cloud servers mostly utilized to relieve clients from the intensity of storage management and maintenance. Cloud computing constructs and allows us to access the applications that actually reside at remote location. Based on the three special different definitions of cloud computing that conclude services such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Cloud computing confirms advantages for organizations parsing to centralize the management of software and data storage, with guarantees on reliability and security for their users. Mostly, many efforts of the commercialization of the cloud such as Amazon's

RELATED WORK

In quick time people are expecting for utilizing data explosion, reading, and writing. One of the specialties utilizes cloud storage with an algorithm engine that find the exact way for data storage. The system provides a secure and performs data storage on the public cloud for use of number of users. Cloud storage has become popular by business users due to its vigorous benefits, proceeding lower cost and better resource utilization. Cloud data storage has provided significant benefit by allowing users to store massage amount of data on demand in cost effective manner. Cloud system performs data storage on public cloud for use of number of users. Hence data security is important functional method for cloud data storage. Data integrity and authentication enhances the completeness, correctness & freshness of data.

MERKLE HASH TREE

Merkle Hash tree is tree in which every non leaf node is labeled with the hash of the labels of its children nodes. Hash trees based on binary trees concept which are useful because they allow efficient and secure verification of the contents of larger data structures, those encrypted form of data would split into batches and those batch files are stored in cloud. The root node has the top hash key stored in local database of the owner. Authenticated Data Structures is a technique in which some kind of authentication data is stored on the DSP. On the client's query, a DSP returns the queried data along with some extra authentication data that is then used by the client to verify the authenticity of returned data. In this scheme the main MHT is divided into smaller MHTs and the root hashes of these sub-trees are signed.
In Merkle proposed the use of binary trees to authenticate a large number of public keys with a single value, namely the root of the tree. That is how the definition of a Merkle tree comes into use. It is a complete binary tree with a *k*- bit value associated to each node such that the interior node value is a hash function of the main hash tree.

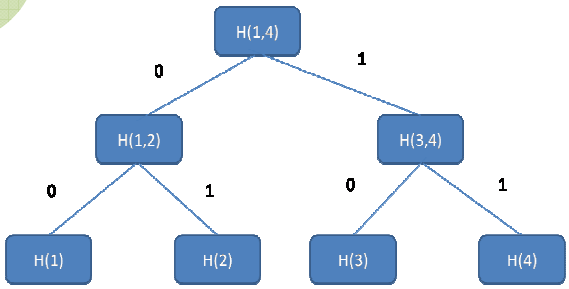

Fig1. Merkle Hash Tree structure

A. Merkle Hash Tree (MHT)

It is same as the binary tree each node having two child nodes, according to this algorithm every non leaf node having two child nodes. Information contained in one node N in an MHT T is constructed as follows. For a leaf node based on a file block m_i, node value is computed as $h_i = H(m_i)$. A parent node of N_1 and N_2 is constructed as $N_P = \{H(h1||h_2)\}$. The main lacuna of Lamport - Diffie scheme is the size of the public key that to be constructed for user. To verify the validity of any signature all the verifiers need an authenticated copy of this public key. Merkle proposed the use of binary trees to authenticate a large number of public keys with a single value, namely the root of the tree. It is a complete binary tree with a *k*- bit value associated to each node such that the interior node value is a hash function of the node values of its children (Figure 1):

User may choose the leaf value, but usually it is a cryptographic hash function of the values that need to be authenticated. In this scheme that values are called *leaf - preimages.* The leaf node can be verified with respect to a publicly concerned root value and its *authentication path.*

B. Ranked Merkle Hash Tree (RMHT)

A Merkle Hash Tree is a novel authenticated data structure designed for efficient verification of data updates by ranks, known as RMHT. According to the update algorithm, every non-leaf node will constantly have 2 child nodes. Information contained in one node N in an RMHT T is represented as $\{H, r_N\}$ where H is a hash value and r_N is the rank of this node. T is constructed as follows. For a leaf node LN based on a message m_i, we have $H = h(m_i)$, $r_{LN} = si_;$ A parent node of $N1 = \{H1, r_{N1}\}$ and $N2 = \{H2, r_{N2}\}$ is constructed as $NP = \{h(H1||H2), (r_{N1} + r_{N2})\}$ where || is a concatenation operator.

C. Verifiable Data Updating

Actually in the verification, the client will be able to detect any fault caused by accident or dishonest behaviours in the update. In the verifiable update process in both our basic scheme and the modification, *Cloud Storage Server (CSS)* cannot provide the client with the satisfactory result, i.e., R0 cannot match the Rnew computed by the client

D. Hash functions

The term hash apparently comes by way of analogy with its special meaning in the physical world, to "chop and mix". The first use of the concept was in a memo from 1953, some ten years later the term hash came into use. Mathematically, a hash function (or hash algorithm) is a method of turning data into a number suitable to be handled by a computer. It provides a small digital “fingerprint” from any kind of data. The function substitutes or transposes the data to create that “fingerprint”, usually called hash value. This value is represented as a short string of random – looking letters and numbers (for example binary data written in hexadecimal notation).

PRAPOSED SYSTEM

The cloud server provider maintains required storage space for outsourced data. The clients are responsible to store and retrieve data as and when required while the *Third Party Auditor (TPA)* is responsible to verify the integrity of data which is being flown between data owner and service provider. But existing data auditing techniques suffers from efficiency and

security problems such as communication overhead (bandwidth) to maintain data integrity, and public auditing.

To address these problems, the proposed system presents novel public auditing schemes like Data Integrity Verification and Data Modification & Data Insertion based on the MHT. The proposed system supports both data dynamics and public auditing. Also the proposed system assures users to securely store and share their data efficiently and effectively on the cloud. In the proposed system implementation of secure cloud storage system is done.

In this system, first the users will register with the cloud service providers, during the registration phase the Private Key will be generated for the users of multiple groups. Using MHT algorithm the cloud server split the data stored in the files into batches and those data parts will be encrypted. The cloud server will allow the TPA to audit the data that was stored in the cloud server as requested by the user.

The user is allowed to access the files by providing the private key, it is verified by the cloud server and if the key values are correct then the data user can download the file with approval of authentication belonging to multi groups.

E. Setup

The setup phase is used to generate security keys like private key and public key by invoking KeyGen() function. In the pre-processing, it makes use of homomorphic authenticators and Meta data. The methods needs two arguments namely file and key that responsible to audit data being flown for verification of integrity.

Content of file is divided into multiple blocks, Hash code is computed for each block. The actual hash code of two blocks is merged then this merged key is merged with other key made up of two merged keys. This process continues until all leaf nodes are found in the Merkle hash tree. After the processing of path the root element sent to cloud server.

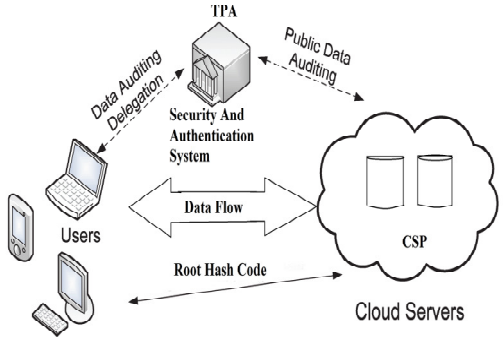

Fig2: A MHT framework for cloud data security.

F. Data Integrity Verification

The data integrity verification is done by third party auditor. The TPA challenges server for block level data verification at regular intervals by sending file name and block randomly. On challenge, the root hash code is computed by server. In cloud computing public auditing ability and data dynamics for remote data integrity are checked. The system construction is conceptually designed to compute these two important goals while efficiency being also

measured. To achieve efficient data dynamics, the system improves the existing proofs of storage models by manipulating the classic Merkle Hash Tree construction for block tag authentication and integrity.

- The *Third Party Auditor* (TPA) is trusted who does not create any security problems.

- There might be latent storage inconsistencies that are not disclosed by cloud service providers.
- Cloud service providers may delete some data of data owner for monetary gains or other reasons.

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 MHT based algorithm that improves the efficiency, bandwidth in terms effectiveness of cloud data.

G. Data Modification and Data Insertion

Cloud users need to perform modifications to their data online frequently these are known as data dynamics. Unlike some of the existing systems, the proposed system supports data dynamics. Data modification and data insertion are the two important operations required by the system. The former modifies existing data while the latter inserts new data.

H. Algorithm

1. Data Integrity Verification

Step1: Start.
Step2: TPA generates random set.
Step3: CSS computes root hash code based on the block input.
Srep4: CSS computes the originally stored value.
Step5: TPA decrypts the given content and compares with generated root hash.
Step6: After verification, the TPA can determine whether the integrity is fetched.
Step7: Stop

2. Data Modification and Data Insertion

Step1: Start.
Step2: Client generates new Hash for tree then sends it to CSS.
Step3: CSS updates F and computes new R'.
Srep4: Client computes R.
Step5: Client verifies signature. If it fails output is False.
Step6: Compute new R and verify the update.
Step7: Stop

PERFORMANCE EVALUATIONS

TABLE 1
Comparisons of External authentication and Integrity verification scheme

	MR-PDP [2]	DPDP [3]	SIR-DPA [4]	FU-DPA [5]	Proposed Scheme
Block less Verification	Yes	Yes	Yes	Yes	Yes
Stateless Verification	Yes	Yes	Yes	Yes	Yes
Infinite Verifications	Yes	Yes	Yes	Yes	Yes
Public Auditability	**No**	**No**	**Yes**	**No**	**Yes**
Coarse-grained Verifiable Data Updating	No	Yes	Yes	Yes	Yes
Variable-sized Data Blocks	Yes	Yes	No	Yes	Yes
Authorized Auditing	No	No	No	Yes	Yes
One Interaction for Updating All Replicas	**No**	**No**	**No**	**No**	**Yes**

CONCLUSION AND FUTUREWORK

In this paper, the proposed system is presented by secure public auditing schemes named as Data Integrity Verification and Data Modification & Data Insertion based on the *Merkle Hash Tree* (MHT). The One Iteration for Updating All Replicas scheme incorporated a novel authenticated data structure based on the Merkle hash tree. The level values of nodes in MHT are generated in a top-down order, and all replica blocks for each data block are organized into a same replica sub-tree. In this paper Merkle Hash Tree is more flexible for storing and sharing files across multiple users. The development of Merkle Hash Tree based system algorithm proves efficiency, effectiveness and bandwidth. An exhaustive analysis of efficiency of our algorithm is function that gives advantage over the existing system.

REFERENCES

[1] William stalling, A Handbook on "*Cryptography & Network Security, Principles & Practices" by Pearson Education, Fifth edition published by Pearson 2011*

[2] R. Curtmola, O. Khan, R. C. Burns, and G. Ateniese, *"MR-PDP: Multiple-replica provable data possession," in Proc. 28th IEEE Int.Conf. Distrib. Comput. Syst., Beijing, China, 2008, pp. 411–420.*

[3] C. Erway, A. C. Papamanthou, and R. Tamassia, "*Dynamic provable data possession" in Proc.16th ACM Conf.Comput.Commun. Security, Chicago, USA, 2009, pp. 213–222.*

[4] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "*Enabling public audit ability and data dynamics for storage security in cloud computing," IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5,pp. 847–859, May 2011.*

[5] C. Liu, J. Chen, L. T. Yang, X. Zhang, C. Yang, R. Ranjan, and K. Ramamohanarao, *"Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates," IEEE Trans. Parallel Distrib. Syst. 2014.*

[6] Chang Liu, Rajiv Ranjan, Chi Yang, Xuyun Zhang, Lizhe Wang, Senior Member " *MuR-DPA: Top-Down Levelled Multi-Replica Merkle Hash Tree Based Secure Public Auditing for Dynamic Big Data Storage on Cloud " IEEE Transactions On Computers, Vol. 64, No. 9, September 2015*

[7] Yubin Xia, Yutao Liu, Haibing Guan, Yunji Chen, Tianshi Chen, Binyu Zang, Haibo Chen, "*Secure Outsourcing of Virtual Appliance" IEEE Transactions on Cloud Computing DOI. 2015*

[8] Christina Delimitrou and Christos Kozyrakis *Security Implications of Data Mining in Cloud Scheduling", IEEE Computer Architecture Letters 2015*

[9] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, ''*Cooperative Provable Data Possession for Integrity Verification in Multi-Cloud Storage'' IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 12, pp. 2231-2244, Dec. 2012.*

[10] X. Zhang, C. Liu, S. Nepal, S. Panley, and J. Chen, ''*A Privacy Leakage Upper-Bound Constraint Based Approach for Cost- Effective Privacy Preserving of Intermediate Datasets in Cloud,'' IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 6, pp. 1192-1202, June 2013.*

[11] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang *Senior Member,* IEEE and Mohammad Mehedi Hassan *Member, IEEE* and Abdulhameed Alelaiwi *Member,* IEEE *"Secure Distributed Deduplication Systems with Improved Reliability" IEEE Transactions on Computers. 2015*

[12] IEEE Explore *Abstract; [Online].Available: http://ieeexplore.ieee.org*

[13] Amazon Web Services *[Online].Available: https://aws.amazon.com/*

[14] *Videos of Data security in cloud computing. [Online].Available: https://www.youtube.com/results?search_query=Videos+of+Data+security+in+cloud+computing*

[15] *Open Stack Open Source Cloud Software [Online].Available: http://openstack.org/*