Maintaining Data Storage Integrity in Cloud Computing

UPENDER TALLURI,
Associate Professor,
Department of Computer Science and Engineering,

SHIRISHA MUNASA D,
Assistant Professor,
Department of Electronics and Communications Engineering,

Siddhartha Institute of Technology and Sciences,
Narapally, Hyderabad, Telangana – 500 088.

Abstract

Cloud computing has been envisioned as the IT enterprise's next-generation architecture. It shifts application software and databases to centralised massive data centres, where data and service management may not be completely reliable. This novel paradigm introduces a slew of new security issues that have yet to be fully grasped. Because of the enormous number of data users and documents in the cloud, it is critical for the search service to support multi-keyword queries and result similarity ranking in order to satisfy the effective data retrieval need. Similar efforts on searchable encryption concentrate on single keyword or Boolean keyword searches, seldom distinguishing between the two. This paper investigates the topic of maintaining data storage integrity in Cloud Computing, as well as Multi Keyword Ranked Search over Encrypted Cloud Data. While previous efforts to ensure distant data integrity have generally lacked either public auditability or dynamic data operations, this project does both. We augment current proof of storage models by altering the basic Merkle Hash Tree design for block tag authentication to achieve efficient data dynamics. We study the concept of bilinear aggregate signature to expand our primary result into a multiuser situation, where TPA can execute many auditing jobs simultaneously, to facilitate efficient management of numerous auditing activities. The suggested systems are very efficient and provably secure, according to extensive security and performance study.

1. Introduction

Cloud computing is a concept for providing on-demand network access to a shared pool of customizable computing resources (e.g., networks, servers, storage, applications, and services) that may be swiftly supplied and released with minimum administration effort or service provider contact. Cloud computing refers to a distant server that can be accessed through the internet and is used for corporate applications and functionality as well as computer software consumption.

Users save money on annual or monthly subscriptions by using cloud computing. More and more sensitive data, such as emails, personal health records, private films and images, firm financial data, government papers, and so on, is being consolidated into cloud servers as a result of the benefits of cloud services. In order to offer end-to-end data confidentiality assurance in the cloud, secret data must be encrypted prior to outsourcing.

Because there may be a significant number of outsourced data files, data encryption makes optimal data use a difficult process. Furthermore, with Cloud Computing, data owners may share their outsourced data with a large number of users, who may only want to download particular data files during a given session. Keyword-based search is one of the most popular techniques to do so.

This keyword search strategy, which has been frequently used in plaintext search contexts, allows users to choose retrieve files of interest. Unfortunately, standard plaintext search methods for encrypted cloud data fail due to data encryption, which inhibits users' capacity to execute keyword searches and further necessitates the protection of keyword privacy. By matching files in a ranked order based on specified relevancy criteria, ranked search dramatically increases system usability (e.g., keyword frequency).

2. Literature survey

Cloud Computing doesn't really modify these reasons, but it really does give so much application suppliers the option of dispatching their product as SaaS without needing to create or requirement a datacentre: just as semiconductor foundries enabled chip companies to design and sell chips without owning a fab, Cloud Computing enables SaaS deployment and scaling on demand without having to build or provision a datacentre. The SaaS provider may now transfer some of his difficulties to the Cloud Computing provider, similar to how SaaS allows users to unload some of their problems to the SaaS provider.

The ability to verify the integrity of F without having explicit knowledge of the entire file is a more difficult challenge. Blum et al. [8] were the first to discuss the topic in broad terms, focusing on the challenge of effectively testing the validity of a memory-management algorithm. Follow-up research has looked at the issue of dynamic memory-checking in a variety of contexts. While many Byzantine-failure storage systems rely on storage duplication, a new line of study has focused on using information dissemination and error-coding to decrease the amount of file redundancy necessary to satisfy robustness guarantees, such as in.

This isn't simply a terrifying prospect. Many internet services enable end users and corporations to retain data for as long as they wish, and some even charge for it. Even the most famous websites can lose data, according to press reports, and users have no reasonable foundation for evaluating the danger of data loss or choosing between storage options. Although the examples in this article focus on online storage services, the issue is applicable to the fledgling online service oriented economy (OSOE), in which enterprises and end users buy IT services from a range of online service providers (OSPs).

This variant's transmission cost is now somewhat greater than a file block's size. Leave as an open problem the development of publicly verifiable PDP methods with challenges and replies smaller than a single file block, with a focus on determining if an untrusted server keeps a client's data. Introduced a methodology for proving data ownership in which file block visits, server processing, and client-server communication should all be minimised. Pre-processing is the limiting performance element for E-PDP, given the efficiency of computing problems.

There are several research issues that have yet to be uncovered. Future study in this field might go in a number of different directions, as I've imagined. An approach that enforces public verifiability appears to be the most promising. TPA can audit cloud data storage without requiring users' time, feasibility, or resources because to public verifiability, which is provided in. An intriguing topic under this paradigm is whether it is possible to build a strategy that ensures both public verifiability and storage accuracy of dynamic data. In addition, we intend to examine the topic of fine-grained data error localisation in conjunction with our research on dynamic cloud data storage.

3. Methodology

According to our hypothesis, the cloud server is "honest but inquisitive." The cloud server, in particular, both respects the protocol definition and analyses data in its storage and message flows received via the protocol to gain more information.

We want to find out what the latent semantic link between words and documents is. We assess the latent semantic structure and remove the obscuring "noise" using statistical approaches. The suggested approach aims to group items that are related in some way in order to deliver data to the user. The files include phrases that are latently semantically connected with the query keyword.

Multi-keyword Ranked Search: It enables multi-keyword queries as well as rating of results.

Our solution is built to fulfil the privacy criterion by preventing the cloud server from learning more information from the index and trapdoor.

The following modules are used in cloud computing to provide auditability and data dynamics.

- Network architecture for cloud data Storage.
- Verification protocol using RSA signature for data Integrity.
- Merkle Hash Tree for block Tag Authentication.
- Batch Auditing of Multi client Data

i. Network architecture for cloud data Storage

Multiple cloud components commonly communicate with each other using application programming interfaces, mainly web services, in cloud architecture, which is the systems architecture of the software systems involved in the delivery of cloud computing. This is similar to the UNIX principle of having several programmes that each do one thing well and communicate with one another through universal interfaces. The resultant systems are more controllable than monolithic versions because the complexity is managed.

ii. Verification protocol using RSA signature for data Integrity

The use of a PKC-based homomorphic authenticator (e.g., a BLS signature or an RSA signature-based authenticator) to provide public auditability to the verification protocol is proposed. The BLS-based strategy to show our approach with data dynamics assistance is offered in the following description.

iii. Merkle Hash Tree for block Tag Authentication

Merkle hash trees are a widespread type of hash tree, thus the name. The root hash, as well as the overall size of the file set and the piece size, are now the only pieces of data in the system that must come from a reliable source. Any piece can be checked by a client who only has the root hash of a file set. It begins by computing the hash of the chunk it has received.

iv. Batch Auditing of Multi client Data

Given K signatures on K separate data files from K clients, it is more beneficial to combine all of these signatures into a single short one and verify it all at once, because cloud servers can manage numerous verification sessions from various clients at the same time. Provable data updates and verification in a multi-client system are permitted to attain this purpose.

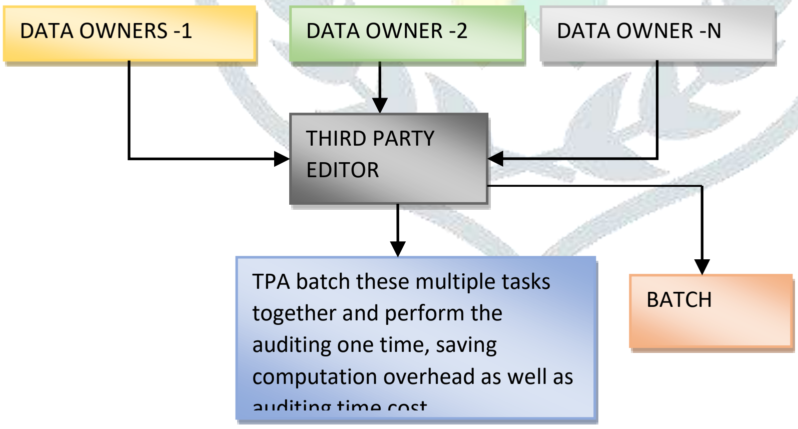

DATA OWNERS -1
DATA OWNER -2
DATA OWNER -N
THIRD PARTY EDITOR
TPA batch these multiple tasks together and perform the auditing one time, saving computation overhead as well as auditing time cost
BATCH

Fig.1 Batch Auditing

4. Result and discussion

The goal of testing is to find mistakes. Testing is the practise of attempting to find all possible flaws or weaknesses in a work product. It allows you to test the functionality of individual components, subassemblies, assemblies, and/or a final product. It is the process of testing software to ensure that it satisfies its requirements and meets user expectations, and that it does not fail in an undesirable way. There are many different sorts of tests. Each test type is designed to satisfy a distinct testing need.

A server is a type of programme that serves and supports clients on a network. Web servers, proxy servers, mail servers, and print servers are examples of servers. A servlet is another customised software. A servlet is similar to a server-side applet in that it operates on the server. Java Servlets are a popular alternative to CGI scripts for creating interactive web applications. Servlets are similar to applets in that they are application runtime extensions.

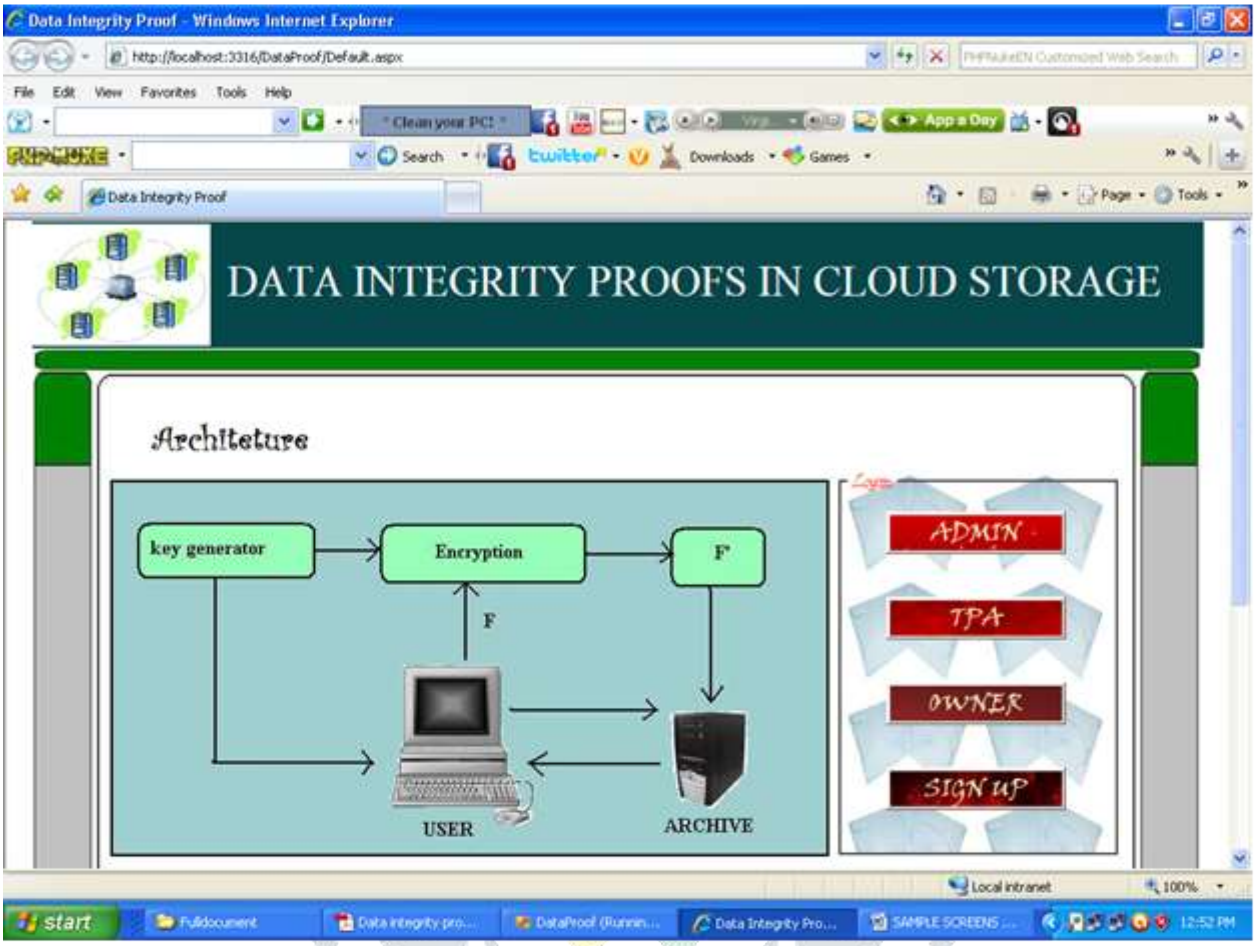
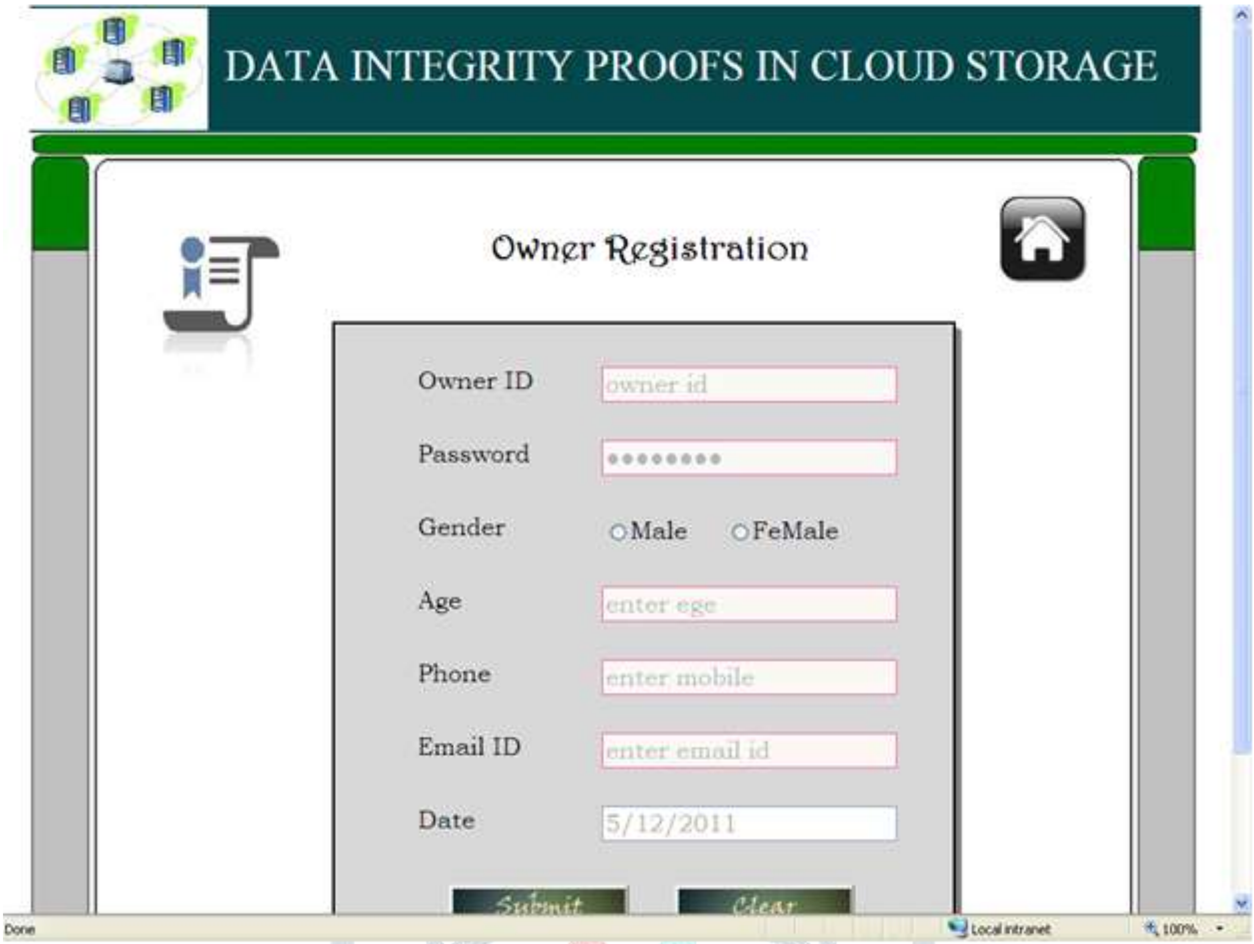

Data Integrity Proof - Windows Internet Explorer
http://localhost:3316/DataProof/Default.aspx
File Edit View Favorites Tools Help
Data Integrity Proof
DATA INTEGRITY PROOFS IN CLOUD STORAGE
Architeture
key generator
Encryption
F'
F
USER
ARCHIVE
ADMIN
TPA
OWNER
SIGN UP
Local intranet
100%
start
12:52 PM

Fig.2 Creation of front page with options for Admin, TPA, Owner and Signup



DATA INTEGRITY PROOFS IN CLOUD STORAGE

Owner Registration

Owner ID	owner id
Password	●●●●●●●●●
Gender	○Male ○FeMale
Age	enter ege
Phone	enter mobile
Email ID	enter email id
Date	5/12/2011

Fig.3 Registration form

5. Conclusion

Keyword-based search is one of the most popular techniques to do so. This keyword search strategy, which has been frequently used in plaintext search contexts, allows users to choose retrieve files of interest. Unfortunately, standard plaintext search methods for encrypted cloud data fail due to data encryption, which inhibits users' capacity to execute keyword searches and further necessitates the protection of keyword privacy. By matching files in a ranked order based on specified relevancy criteria, ranked search dramatically increases system usability.

References

1. M. Li and P. P. C. Lee, "STAIR codes," *ACM Transactions on Storage*, vol. 10, no. 4, pp. 1–30, 2014.
2. M. Antonio, M. Antonio, and G. Javier, "Dynamic security properties monitoring architecture for cloud computing," *Security Engineering for Cloud Computing: Approaches and Tools*, IGI Gobal, PA, USA, 2012.
3. T. Jamal, M. Antonio, and S. Nesmachnow, "Evolution oriented monitoring oriented to security properties for cloud applications," in *Proceedings of the 13th International Conference on Availability, Reliability and Security*, Hamburg, Germany, August 2018.
4. C. Wang, S. S. M. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for secure cloud storage," *IEEE Transactions on Computers*, vol. 62, no. 2, pp. 362–375, 2013.

5. K. Yang and X. Jia, *TSAS: Third-Party Storage Auditing Service," Security for Cloud Storage Systems*, Springer, Berlin, Germany, 2014.

6. Y. Zhang, C. Xu, and X. Shen, "Blockchain-based public integrity verification for cloud storage against procrastinating auditors," *IEEE Transactions on Cloud Computing (Early Access)*, vol. 7, 2019.

7. A. Maña and A. Mana, "TPM-based protection for mobile agents," *Security and Communication Networks*, vol. 4, no. 1, pp. 45–60, 2011.

8. S. N. Bitcoin, *A Peer-To-Peer Electronic Cash System*, 2008, https://bitcoin.org/bitcoin.pdf.

9. D. Yue, R. Li, Y. Zhang, W. Tian, and C. Peng, "Blockchain based data integrity verification in P2P cloud storage," in *Proceedings of the IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 561–568, Singapore, December 2018.

10. H. Wang and D. He, "Blockchain-based private provable data possession," *IEEE Transactions on Dependable and Secure Computing (online)*, vol. 18, 2020.