Multi Keyword Ranked Search over Encrypted Cloud Data

AMBIKA GANNU,
Assistant Professor,
Department of Electronics and Communications Engineering,

VENKATESH THOTA,
Assistant Professor,
Department of Electronics and Communications Engineering,

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

Abstract

Because of the massive amount of data users and documents in the cloud, it is necessary for the search service to handle multi-keyword searches and result similarity ranking in order to meet the effective data retrieval need. Similarly, work on searchable encryption focus on single keyword or Boolean keyword searches, with no distinction between the two. This study looks at the subject of data storage integrity in cloud computing, as well as Multi Keyword Ranked Search over Encrypted Cloud Data. Unlike prior efforts to maintain distant data integrity, this project includes both public auditability and dynamic data operations. To achieve efficient data dynamics, we extend current proof of storage models by modifying the fundamental Blockchain Merkle Tree architecture for block tag authentication. We investigate the idea of bilinear aggregate signature in order to extend our primary result into a multiuser situation in which TPA can do many auditing jobs at the same time, allowing for more efficient administration of multiple auditing operations. According to comprehensive security and performance testing, the proposed solutions are exceedingly efficient and provably secure. Cloud computing has been envisioned as the next-generation architecture for the IT company. It moves application software and databases to centralized huge data centers, where data and service management may not be totally trustworthy. This unique paradigm poses a plethora of new security challenges that have yet to be completely understood.

1. Introduction

Furthermore, with Cloud Computing, data owners may share their outsourced data with a large number of users, each of whom may only want to retrieve particular data files they are interested in during a given session. One of the most frequent methods is to conduct a keyword-based search. This keyword search strategy, which allows users to obtain files of interest selectively, is extensively used in plaintext search contexts.

Unfortunately, typical plaintext search methods for encrypted cloud data fail due to data encryption, which limits the user's capacity to execute keyword searches and also necessitates the protection of keyword privacy. Normal matching files in a ranked order based on specified relevancy criteria considerably increases system usability with ranked search.

Cloud computing is a concept for giving on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provided and released with minimal management effort or service provider interaction. Cloud computing refers to a remote server that can be accessed via the internet and is utilised for both business applications and software consumption.

Cloud computing allows users to save money on annual or monthly subscriptions. As a consequence of the benefits of cloud services, more and more sensitive data, such as emails, personal health records, private films and photographs, company financial data, government papers, and so on, is being pooled onto cloud servers. Secret data must be encrypted prior to outsourcing in order to provide end-to-end data confidentiality guarantee in the cloud.

2. Literature survey

Focuses on concerns relating to future SaaS Providers (Cloud Users) and Cloud Providers, which have gotten less attention in the past. The opportunity, which is still a research problem, is to create a storage system that not only meets these needs but also combines them with the cloud advantages of scalability, data durability, and high availability, as well as meeting programmer expectations in terms of resource management for scalability, data durability, and high availability. Many established enterprises, including newspapers like the Washington Post, movie studios like Pixar, and universities, use the elasticity of Cloud Computing on a regular basis.

Similarly, in our primary POR architecture, we use error-correction to bind the effects of errors in a storage archive. Our primary contribution may be divided into three categories. First, provide a formal, tangible security definition for PORs that is thought to be of wide interest and practical use. Second, a sentinel-based POR method was devised, which has several noteworthy properties: The data in sentinels, and therefore the consequent PORs, may be made independent of the stored file, which has theoretical implications; a strong proof can be very small.

Unless there is a compelling need, providers will not give auditing interfaces. When designing system interfaces that facilitate auditing, bear in mind that the mechanisms that give such incentive are more likely to be social than technological. These behavior-altering systems often employ either fines or incentives, or a mix of both. Penalty-based processes include rules, legislation (and the possibility of jail), and loss of reputation (which can force a supplier out of business).

Clients' perceptions of overall system performance are limited by the rate at which they can create data to outsource. There are, however, a number of mitigating circumstances. (1) Unlike challenging outsourced data, which is done on a regular basis, data outsourcing is a one-time activity. (2) The procedure can be run in parallel. Each file can be processed at a separate processor in turn. If processors share important material, a single file may be parallelized easily. This model fits our PDP solutions: They have a minimal (or even constant) server overhead and only need a tiny bit of communication every task.

3. Methodology

Confidentiality is a term used to describe the degree to which information is kept private. The index contains the TF values for keywords. As a result, the cloud server's index must be encrypted.

Unlikability of the Trapdoor Over the search result, the cloud server might perform some statistical analysis. In the meanwhile, if you search the same question repeatedly, you should get different trapdoors. The cloud server shouldn't be able to figure out how trapdoors are connected.

Privacy is a key term. The cloud server was unable to identify the keyword in the query, so it created an index based on statistical data such as phrase frequency. For the cloud server, the data user provides t keywords. Through search control techniques, a similar trapdoor T w is created. We presume that the authorization between the data owner and the data user is nearly complete in this work.

Encryption/Decryption Module

For file E/D we using Jenkins hash function.

Jenkins Hash Function

1. Choose one of the following user profiles.

Username, password, phone number, name of the document, and size of document.

2. Convert from bits to bytes

3. Fill in the 64*64 cells with the byte conversion solution.

4. There is no such thing as a unique insertion order.

5. Rearrange the elements and move them outside the cell.

6. Finally, to rearrange an element, add the mobile number or document size.

Block Modification operations

Our system can manage dynamic data operations for cloud data storage openly and efficiently:

Data Modification: One of the most commonly utilized processes in cloud data storage is data modification. The replacement of specified blocks with new ones is referred to as a fundamental data alteration procedure.

Data Insertion: In contrast to data modification, which does not alter the logic structure of the client's data file, data insertion refers to adding new blocks after certain predetermined points in the data file F.

Data Deletion: The operation of data deletion is the polar opposite of data addition. The term "single block deletion" refers to removing the given block and pushing all subsequent blocks ahead one block. The protocol processes are comparable to data alteration and insertion procedures, hence they are removed here.

Performance Evaluation

In Cloud Computing, the difficulty of delivering simultaneous public auditability and data dynamics for distant data integrity checks is identified. The building was purposefully constructed to achieve these two crucial purposes while keeping efficiency in mind. The current proof of storage models for block tag authentication are upgraded to achieve efficient data dynamics by altering the standard Merkle Hash Tree architecture.

The concept of bilinear aggregate signature is utilized to expand our primary result into a multi-user situation, where TPA can execute numerous auditing jobs concurrently, to facilitate efficient management of many auditing activities. The suggested approach is very efficient and provably secure, according to extensive security and performance study.

4. Result and discussion

The hardware or software environment in which a program operates is referred to as a platform. Some of the most popular systems, such as Windows 2000, Linux, Solaris, and MacOS, have previously been discussed. The majority of platforms are made up of a mix of operating system and hardware. In contrast to most other platforms, Java is a software-only platform that works on top of other hardware-based systems.

Microsoft Open Database Connectivity (ODBC) is a programming interface that is used by application developers and database system providers. Prior to ODBC becoming the de facto standard for Windows program to communicate with database systems, programmers had to utilize proprietary languages for each database to which they wished to connect. From a coding standpoint, ODBC has rendered the database system selection practically unimportant, which is exactly what it should be. When business demands unexpectedly change, application developers have much more essential things to worry about than the syntax required to transfer their software from one database to another.

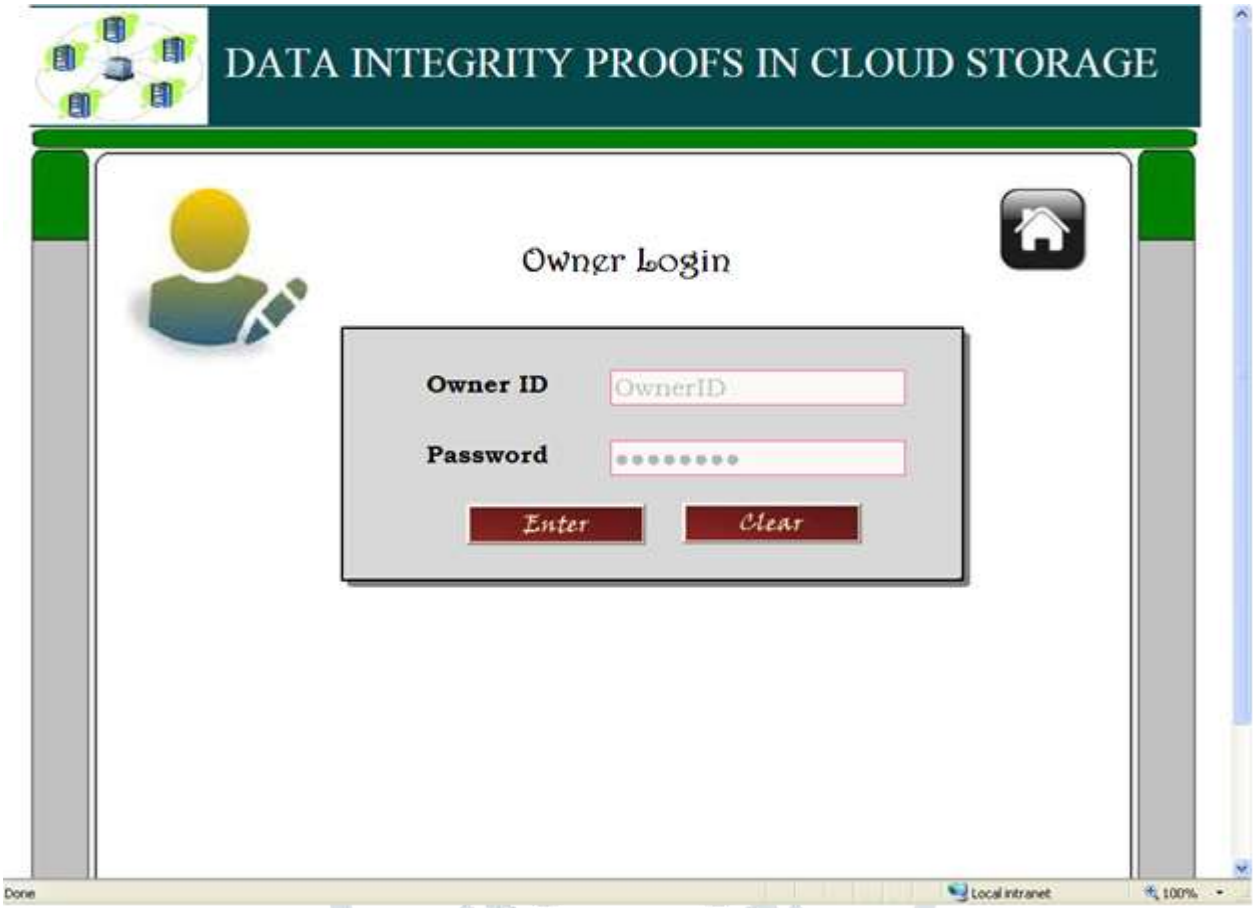

Fig.1 Owner login

This option will be used by the owner to log in. The authenticated user will be granted access.

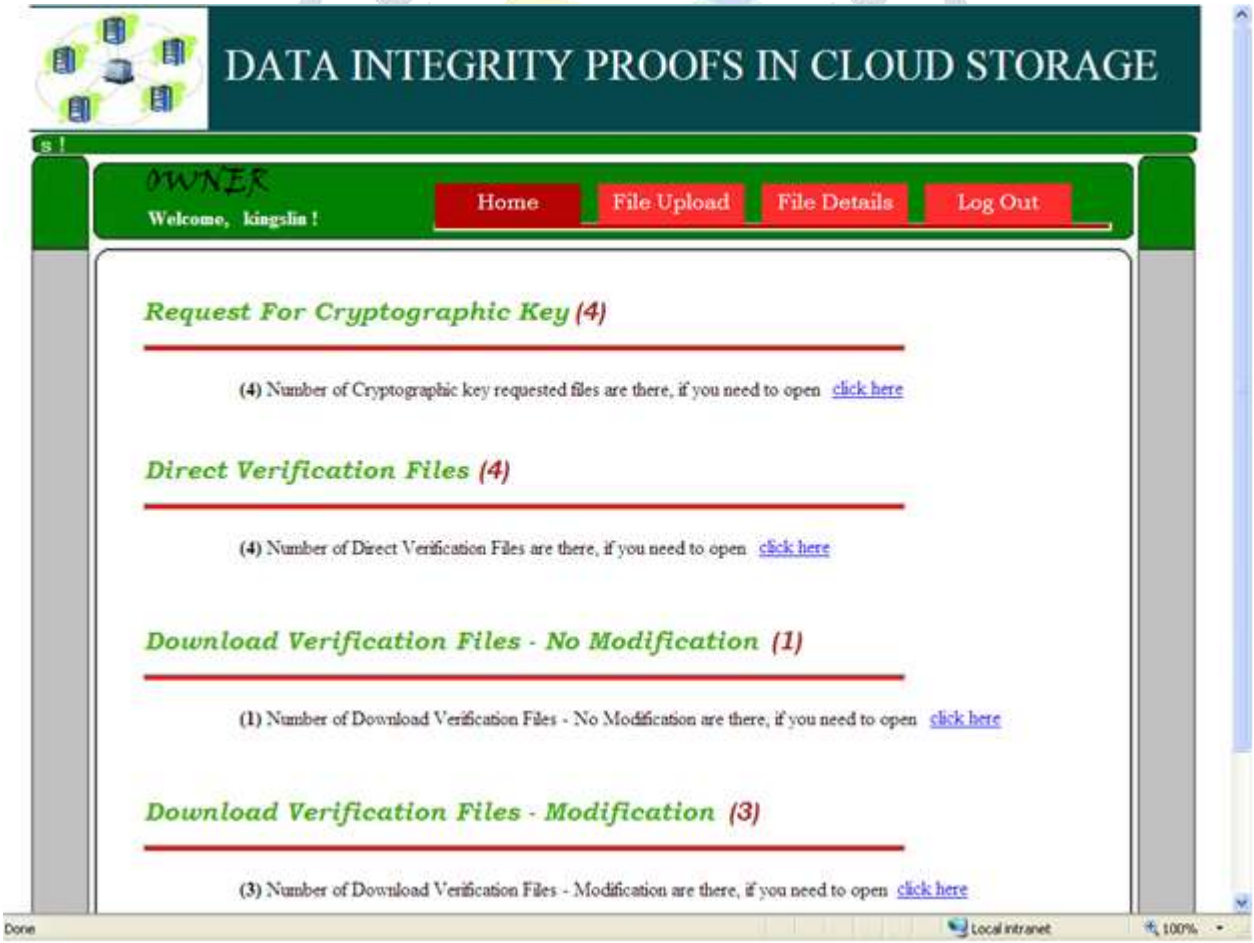

Fig.2 File verification

When you enter the owner login, you will be able to validate files. Output: Here you may do download and direct file verification.

5. Conclusion

This unique paradigm raises a plethora of new security concerns that have yet to be completely understood. Because of the vast number of data users and documents in the cloud, it is necessary for the search service to handle multi-keyword searches and result similarity ranking in order to meet the effective data retrieval need. Similar work on searchable encryption focus on single term or Boolean keyword searches, with no distinction made between the two.

References

1. F. Zhang, R. Safavi-Naini, and W. Susilo, *An Efficient Signature Scheme from Bilinear Pairings and its Applications*, Springer, Berlin, Germany, 2004.
2. A. Juels and B. S. Kaliski, "PORs: proofs pf retrievability for large files," in *Proceedings of ACM CCS*, pp. 584–597, Alexandria, VA, USA, October 2007.
3. G. Ateniese, R. Burns, R. Curtmola et al., "Provable data possession at untrusted stores," in *Proceedings of ACM CCS 2007*, pp. 598–610, Alexandria, VA, USA, October 2007.
4. Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 5, pp. 847–859, 2011.
5. K. Yang and X. Jia, "An efficient and secure dynamic auditing protocol for data storage in cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 19, pp. 1717–1726, 2013.
6. B. Wang, B. Li, and H. Li, "Panda: public auditing for shared data with efficient user revocation in the cloud," *IEEE Transactions on Services Computing*, vol. 8, no. 1, pp. 92–106, 2015.
7. J. Yuan and S. Yu, "Efficient public integrity checking for cloud data sharing with multi-user modification," in *Proceedings of IEEE INFOCOM*, pp. 2121–2129, Hong Kong, China, March 2014.
8. C. H. Chen and P. C. Lee, "Enabling data integrity protection in regenerating-coding-based cloud storage: theory and implementation," *IEEE Transactions on Reliability*, vol. 64, no. 3, pp. 840–851, 2015.
9. C. Liu, J. Chen, L. T. Yang et al., "Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 9, pp. 2234–2244, 2014.
10. S. Ghimire, J. Y. Choi, and B. Lee, "Using blockchain for improved video integrity verification," *IEEE Transactions on Multimedia*, vol. 22, no. 1, pp. 108–121, 2020.