Document Retrieve On Cloud With Multikeyword Searching

Searching document on Cloud

Shreya Thite¹, Monali Yenpure², Surabhi Mohite³

¹Computer Engineering, NESGOI Of Naigaon Pune
²Computer Engineering, NESGOI Of Naigaon Pune
³Computer Engineering, NESGOI Of Naigaon Pune

Abstract — Cloud storage contain large amount of outsourced data. For large data require to provide security to store the data on cloud storage, large data cannot be handled by the user for searching document on cloud. The cloud provides data storage and sharing services for users, and has ample storage space. The third party auditor is able to verify the integrity of shared data based on requests from users, without downloading the entire data. To find the specific data on cloud is critical task. Finding data in large data is the big challenge in this case. So in this paper, we introduce to 'Multikeyword Searching' strategy. In multikeyword searching we can find data from cloud using date, Location or keyword if mention in the document. It is easiest technique of finding data on cloud. Also this technique increases the overall performance of the system. It require less time to find document on cloud. We provide integrity and confidentiality to the data while it is stored on cloud.

Keywords—Cloud computing, Multikeyword searching, Privacy preserving, Public auditing, Third party auditor etc.

I. INTRODUCTION

The cloud can be formidable and costly for the users. The overhead of using cloud storage should be minimized as much as possible such that a user does not need to perform so many operations to their outsourced data. Cloud stores large size of data. So user cannot directly retrieve data. Therefore it have chances to reduce performance of the system. So we have come up with a solution that this processing will be done on cloud. Despite of significant improvement in cloud storage computing capabilities, still computing requirements of cloud users, especially enterprise users, is not achieved.

Users no longer have physical possession of the possibly large size of outsourced data makes the data integrity protection in Cloud Computing a very challenging and potentially formidable task, especially for users with constrained computing resources and capabilities. Thus, enabling public auditability for cloud data storage security is of critical importance so that users can resort to an external audit party to check the integrity of outsourced data when needed. Instead of directly applying the old public auditing scheme to the multi-server setting. On cloud storage, it is also difficult to search the data as per user require. So in this paper we introduce the ‘multikeyword searching’ technique for searching various type of files. The files can be search by the keyword, location or date, which are mention in data. This technique is helpful to increase the system performance and efficiency.

II. LITERATURE SURVEY

A literature review or study forms an essential part for studing and searching on previouse topic:

In this paper, authors propose a public auditing scheme for the regenerating-code-based cloud storage. To solve the regeneration problem of failed authenticators in the absence of data owners, they introduce a proxy, which is privileged to regenerate the authenticators, into the traditional public auditing system model.

Cloud Computing is emerging technology and gaining remarkable popularity in the recent years for its benefits in terms of flexibility, scalability, reliability, efficiency and cost effectiveness. Despite of all these benefits, Cloud Computing has one problem: Security, they analyze the problems of data security in cloud storage, which is a distributed storage system. An effective and flexible scheme is proposed in our paper to ensure the integrity and correctness of the data stored in cloud server.

In this paper, they propose a privacy-preserving public auditing system for data storage and security in Cloud Computing. They utilize the homomorphism authenticator and random masking to assure that TPA would not learn any information about the data content stored on the cloud server during the efficient auditing process.


In this paper, authors propose a privacy-preserving public auditing system for data storage security in Cloud Computing. They utilize the homomorphic linear authenticator and random masking to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process, which not only eliminates the burden of cloud user from the tedious and possibly expensive auditing task, but also alleviates the users fear of their outsourced data leakage.

III. PROPOSED SYSTEM

Every coin has two sides, and cloud computing is no exception. There is criticism about privacy in cloud model, because of the fact that administrator have access to data stored in the cloud. They can unintentionally or intentionally access the client data. Traditional security or protection techniques need a reconsideration for cloud. Except for private cloud where organization does not have control over the equipment, the progress of cloud is seems little slow, because organizations think instead of compromising on the security of the data, they are still willing to invest in buying private equipment to setup their own infrastructure. Security issues which are of concern to the client can be classified into sensitive data access, data segregation, bug exploitation, recovery, accountability, malicious insiders and account control issues. Like different disease have different medicines, different cloud security issues have different solutions, like cryptography, use of more than one cloud provider, strong service level agreement between client and cloud service provider. Heavy investment is needed to secure the compromising data in cloud. Cloud can grow only if it is possible to build a trust in client, and which can be built only if security concerns are being addressed. Generally the cloud services are browser based, therefore any browser enabled device such as for instance laptop, desktop, smartphone, tablets can used to gain access to these services, the services at providers end may be hosted on any platform, from windows, Linux, etc. which are accessible via internet.

IV. SYSTEM ARCHITECTURE

In this paper, we consider data storage and sharing services in the cloud with three entities: the cloud, the third party auditor (TPA), and users who participate as a group (as shown in Fig. 1). Users in a group include one original user and a number of group users. The original user is the original owner of data, and shares data in the cloud with other users. Based on access control policies, other users in the group are able to access, download and modify shared data. The cloud provides data storage and sharing services for users, and has ample storage space. The third party auditor is able to verify the integrity of shared data based on requests from users, without downloading the entire data. When a user (either the original user or a group user) wishes to check the integrity of shared data, she first sends an auditing request to the TPA. After receiving the auditing request, the TPA generates an auditing message to the cloud, and retrieves an auditing proof shared data from the cloud. Then the TPA verifies the correctness of the auditing proof. Finally, the TPA sends an auditing report to the user based on the result of the verification.

Fig 1: System Architecture
In this architecture, user can use multikeyword for searching file on cloud. User can use location name, date and keyword to find file on cloud. It increases the system performance and efficiency. User can register on cloud to get authority to preserve security on cloud. Once user can register on cloud then it can login in cloud using its own account for store or download data. Users have two type in this architecture. First is author which can upload the data on cloud to store it, and second is user's which can download the data from cloud using multikeyword strategy. Using this technique user require less time to get its result. In this paper, we used MS-SQL database for store the data. When user register on cloud then information about user is store on MS-SQL database.

IV. ALGORITHM

In this paper, we use binary vector generation algorithm for view the data without extraction of data. Algorithm is given below:

**Input**: Each client $C_m$ has an binary vector $b_m \in Z^n_2$, $1 \leq m \leq M$.

**Output**: $b = T_i(b_1, ..., b_M)$

1. Each $C_m$ selects $M$ random share vectors $b_{m,l} \in Z^6_{M+1}$, $1 \leq l \leq M$, such that $\sum_{l=1}^{M} b_{m,l} = b_m \mod(M+1)$.
2. Each $C_m$ sends $b_{m,l}$ for all $1 \leq l \neq m \leq M$.
3. Each $C_i$ computes $s_i = (s_i(1), ..., s_i(n)) = \sum_{m=1}^{M} b_{m,l} \mod(M+1)$.
4. Clients $C_i$, $2 \leq i \leq M - 1$, send $s_i$ to $C_1$.
5. $C_1$ computes $s = (s(1), ..., s(n)) = \sum_{i=1}^{M} s_i \mod(M+1)$.
6. for $i = 1, ..., n$ do
   7. If $(s(i) + s_M(i)) \mod(M+1) < t$ set $b(i) = 0$ Otherwise set $b(i) = 1$.
8. end for
9. Output $b = (b(1), ..., b(n))$.

VI. PERFORMANCE ANALYSIS

We focus on evaluating performance of our document retrieve on cloud with multikeyword searching setup, audit and repair procedure in our experiment all codes are written in .net in Visual Studio 2.10 platform on windows. All entities in our prototype in intel core i3 2450 2.5 GHz, 4 GB RAM, 7200 RPM Hitachi 5000 SATA Drive.

![Security Analysis Performance](image)

**Fig 2**: Performance analysis

VII. RESULT

This paper is based on multikeyword searching and also we provide security to this system. We compare our system with previous system and calculate computational time. Our system require less computational time as compare to existing system. Proposed system uses keyword to retrieve data, therefore its require less time than the existing system which can shown in bellow:
VIII. ADVANTAGES

The advantages of the proposed system are given below:

- **Usability**: Allows users to drag and drop files between Cloud storage and their local storage.
- **Bandwidth**: You can avoid emailing files to individuals and instead send a web link to recipients through your email.
- **Accessibility**: Storage files can be accessed from anywhere via internet connections.
- **Computation Time**: It requires less computational time compared to previous systems. Because the proposed system uses a multikeyword searching technique.
- **Cost Reduce**: Business and organizations can often reduce annual operating costs by using cloud storage; cloud storage costs about 3 cents per gigabyte to store data internally. Users can see additional cost savings because it does not require power to store information remotely.

IX. DISADVANTAGES

- **Network**: If you have no internet connection you’ve no access to your data.
- **Bandwidth**: Several cloud storage have specific bandwidth allowances. If an organization surpasses the given allowance, the additional charges could be significant. However, some providers allow unlimited bandwidth. This is the factor that companies should consider when looking at cloud storage providers.

X. APPLICATIONS

1. Use in big data where daily millions of documents generated.
2. Use in hospitals.

XI. CONCLUSION

The homomorphic linear authenticator and random number generation utilize to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during efficient auditing process, which not only eliminates the burden of cloud user from the tedious and possibly expensive auditing task. Multikeyword searching technique is used in the system. It improves the overall performance of the system and reduces the computation time as compared to existing systems. It is highly efficient system. Also, this system provides security to the data to preserve from unauthorized persons.

XII. FUTURE SCOPE

To enhance the security more, a mechanism to secure the keys in security cloud can be an area of research. Also to reduce the overhead of network traffic can be another area of research.

XIII. ACKNOWLEDGEMENT

This acknowledgment is intended to be thanks giving gesture to all those people who have been involved directly or indirectly with our dissertation work. First and foremost, we express our special thanks with gratitude and great respect to our valuable Head of Department Prof. P. V. Mahadik, and our Project guide Prof. S. Autade for their keen interest, fruitful suggestions and valuable guidance with motivation and constant encouragement. We are also thankful to their great patience, constructive criticism & useful suggestions apart from valuable guidance. We owe a great debt to our principal prof. R. J. Patil, who guided us through the maze of details associated with the seminar. Lastly we are deeply grateful to everyone who has been associated with this seminar.
REFERENCES


