A survey on Balancing the Load of Big Data for Preserving Privacy Access in Cloud

Mrs. Anooja Ali
Assistant Professor
Department of Computer Science
Reva University
Bangalore,India
anoojaali@gmail.com

Sumalatha D.P
Mtech in Data Engineering and Cloud Computing
Department of Computer Science
Reva University
Bangalore,India
Suma_devajana@yahoo.com

Abstract— **Data of users can be remotely stored on Cloud Storage and on- demand very good quality applications and services can be enjoyed without wasting any data storage space and maintenance cost locally. Documents which are sensitive should be stored in encrypted format for security reasons. Hence it would be difficult for document searching which is in encrypted format. There are many protocols which have been proposed to search keyword over encrypted data to overcome this issue. A method called Oblivious RAM can be used in hiding access patterns of the data. But such protocols are not vigorous and do not scale well. In this paper, we explain a different kind of attack which makes full use of data access pattern leakage which ensures only those important delicate information can be disclosed. A latest privacy preserving access control system is proposed for ensuring data to be safe in cloud. We have shown the simplified model of search over remote encrypted data from Mallory's point of view.**

Keywords:* *access patterns, authenticate,encryption,documents, oblivious RAM,storage space

1. INTRODUCTION

Cloud Computing is becoming interesting in the fields of science, engineering etc. Users can outsource their storage,computation to cloud thus reducing their storage space and maintenance cost.The user just need to pay only for the services he has used.Cloud provides many applications like Microsoft Online,Google apps etc ,Infrastructures like Amazon's EC2,Eucalyptus, Platforms which helps developers in writing applications like Amazon's S3, Windows Azure[1].

Security and privacy are the major concerns in Cloud as sensitive data are moved on cloud .If the user is unable to access or retrieve their data ,it is not worth moving to the cloud. Consider a set of clients and a server which is not trustable. A client(Tanya) has set of documents which are sensitive wherein she has to store in remote server which is owned by Ken. Alice doesnot want Ken to learn her the content of her documents as it is sensitive. Hence those documents are encrypted and stored using certain encryption techniques. An encrypted index structure is stored in the server along with the encrypted data would be easier to search on encrypted data. In this method, an authorized user will be having access to trapdoor generation function. Trapdoor is used in searching the intended keyword in the server. Assumption is that server will not have access to trapdoor generation function. However it is crucial to hide the corresponding keyword of given query from an opponent. Otherwise chances are likely that opponent learns the document that contains the given keyword and those documents which does not contain the keyword as well. A different kind of attack is introduced in this paper, where data access pattern leakage is exploited. This ensures only significant amount of sensitive information can be disclosed [2].

Databases which are accessible publicly are an absolutely necessary resource for receiving information till date. They also constitute noteworthy risks to isolation of the user, since a unusual database operator can track the user's queries and conclude what the user is planning to. Users are always careful to access the database in situation where user's plan has to be maintained in secrecy. When a single database is accessed to assure the privacy of the user, the whole database should be downloaded, namely there should be exchange of n bits.

HAIL(High Availability and Integrity Layer) is a system with distributed cryptography where a number of servers ensure a client that the files which are stored are not harmed and are recoverable[6]. Proofs in HAIL are computable efficiently and are highly compact whatever the file size is. HAIL proves and actively allots file shares[6]. The Proofs of Retrievability(POR) implemented on individual servers is made more efficient and secured by HAIL.

A new method for privacy protecting access control scheme is assessed for data security. The real and genuine identity of a user is verified by the cloud without knowing user's identity [8]. A aspect of access control is allowed only to users with privilege are able to decrypt the stored information. This authentication scheme is more efficient in performance than the other access control methods for clouds.

2. LITERATURE SURVEY

The invent of cloud computing has lead to huge amount of data storage in remote servers[2]. Storagethe data remotely reduces the overhead of data management for data owners which is cost effective. All the delicate documents are needed to be stored in encrypted format to give security for the data. But this would make it difficult to search the stored documents.

Data access patterns are leaked by most of the protocols due to efficiency reasons. Protocols such as Oblivious RAM are used to hide data access patterns, but such protocols do not perform as expected for datasets used in real world [7]. A novel attack is introduced in this paper to exploit pattern leakage of data access where sufficient amount of sensitive information is disclosed using prior knowledge. Our analysis with a real world dataset shows that the proposed attack is able to disclose sensitive information with a very high accuracy. In addition to this, we analyze a simple technique to reduce the risk against attack with less increase in computing resource and communication expenses. Further, our proposed reduction technique is collective enough to be used in concurrence with any searchable encryption scheme that discloses data access pattern.

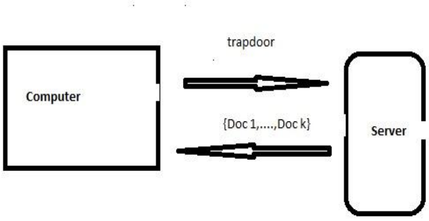

Figure. 1 Simplified model of searching remote encrypted data from Mallory's point of view.

To explain the attack,in this paper we present a simple model to simulate 'search over remote encrypted data'. This model does not rely on encryption method that is hidden and searchable. This model works as long as the search over encryption scheme discloses data access patterns to the attacker.

In our model, Tanya has keywords set which she wishes to search over the document set. Initially, a modified Inverted matrix over the set of documents is built by Tanya. That is Tanya builds a binary matrix in which rows are registered by the keyword set and columns by document set. The binary matrix is referred for an particular instance of 'search over encrypted data ' as *Index Map(ID)*.If the *ith* keyword appears in the *jth* document,the *(i,j)th* entry in the Index Map will be 1, otherwise 0. In this model, using Index Map the server Ken searches for various keywords.Tanya encrypts the rows of the matrix independently as sending the Index Map in plain text discloses document contents. Also she applies Trapdoor function on the set of keywords. Finally, Along with the trapdoor value of the related keyword Tanya sends each of the encrypted rows.

When Tanya wants to search for a particular keyword, the keyword is applied to the trapdoor function and sends the request to Ken. Ken does a simple text matching and returns Tanya the encrypted row of the Index Map. Tanya then decrypts and asks for the set of documents. Bob then sends back the required documents and the thus the search concludes.

Data exchanging among researchers and professionals in network and distributed system security is supported by NDSS[9]. The main goal is to support and facilitate the customers using internet to try and use available security mechanisms.

Databases which are accessible publicly are an absolutely necessary resource for retrieving up-to-date Information. They also constitute significant risks to the privacy of the user, because a curious database operator can follow the user's queries and can find out what the user is after. Perhaps, in cases where the users' aim is to keep secret, users are often careful about accessing the database. It can be shown that when trying to access single database, the complete database should be downloaded to fully guarantee the user privacy, namely n bits should be communicated .

In the multi server space, Gertner proposed conversion from any PIR scheme to SPIR scheme which was at the expense of increased number of non-cooperating servers[3]. Transformation from PIR to SPIR for the single server has been proved by Naor and Pinkas[5]. The transformation uses 1 out of n Oblivious Transfer(OT)protocol[9]. The idea is to generate l=log n key pairs,$(k^0_1,k^1_1),..,(k^0_1,k^0_1)$ and mask its i-th data item, i =1..n, with a subset of keys, which corresponds t the representation of i in binary. If the server's values are $x_1,..x_n$ and the binary representation of i is $_{i1}.._{i1}$,then the server disguises value xi using keys $k^{i1}_1,..,k^{i1}_1$.

The PIR to SPIR conversion uses any PIR scheme where only masked item can be transferred to the client. This conversion adds to the communication problem of the PIR scheme . The per query computation problem will increase with n log n pseudo function evaluations.

2.1 HAIL: A High-Availability and Integrity Layer for Cloud Storage

It is a system implemented in distributed manner, wherein a set of servers proves to a client that stored file is not damaged [6]. HAIL's proof are highly brief and are systematically computable by servers. HAIL verifies and reallocates file shares cryptographically.

2.1.1 Replication system

The first idea of HAIL is to replicate F on each of the n servers. Cross-server redundancy can be used to check fairness. The client simply chooses a random file-block position j and retrieves the corresponding F_j of F from each server. Since all the blocks which are returned are identical ,the client concludes that F is intact in that position. If any inconsistencies are found,it reconstructs F and either removes or replaces faulty servers. The client can increase its capability of detecting corruptions by sampling multiple file-block positions.

2.2. Load Balancing

A secured storage for multi owner is identified in cloud which is used for sharing of data.Initially, data is uploaded by the Data Owner in encrypted format in cloud server. Once the data is uploaded, the cloud server returns the public key and private key.Later, the trusted Third Party Auditor audits the data using Hash Tree Algorithm and send to Cloud Service Provider who stores it. When a user wants to see or download the data, he has to present the public key given, wherin the Owner verifies it. If the key is genuine, the key for decryption will be sent to the user. The Load balancing technique is also used here. Initially, request from users are sent to the cloud service Provider's data center. The request will be lined up through communication channel. Later the work will be handed over to sub server.

2.2.1 Advantages

- Since the public and private keys are to be used , only the valid users can access the data.
- The Auditor being trusted, there exists trust among users and Cloud Service Providers.
- The data is audited with batch auditing process in multi-level.
- With high security and trusted auditing, Customers also increase for organization.

2.2.2 Algorithm

Merkle Hash Tree

Step A: A given file is split into 'n' data blocks.
Step B: Hash value of each data block is evaluated and these form the leaves in a hash tree.
Step C: Nodes are the hashes of their concerned child.
Step D: The final hash value obtained in a single node will be the hash value at the top.

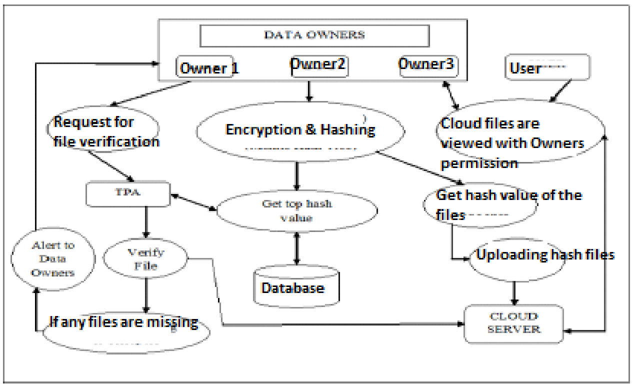

Fig 2. Overall Process Design

3. COMPARISON

Approach	Advantages	Disadvantages
Privacy preserving access control with authentication for data security	Balancing speed is faster. Highly efficient and Less communication overhead	Layout of the cell cannot be maintained.
Access pattern declaration on searchable encryption	Adding noise to the index matrix is quite effective in preventing Inference attacks.	Additional computational overhead can be incurred.

4.Conclusion

We presented a simple technique to reduce the hazards against the proposed attack at the cost of a slight increase in the computational resources and cost of communication. High Availability and Integrity Layer(HAIL) for cloud storage is a distributed cryptographic system wherein a set of servers proves to a client that stored file is not damaged and are recoverable. We also show that data is secure when top hash value is kept in local database and hash code files in Cloud Server.

5. REFERENCES

[1] Sushmita Ruj, Milos Stojmenovic, Amiya Nayak "Privacy Preserving Access Control with Authentication for Securing Data in Clouds",2012.

[2] M. Islam, M. Kuzu, and M. Kantarcioglu, "Access pattern disclosure on searchable encryption: Ramification, attack and mitigation," in *Network and Distributed System Security Symposium*, 2012.

[3]B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval," *Journal of the ACM (JACM)*, vol. 45, no. 6, pp. 965–981,1998.

[4] R. Sion and B. Carbunar, "On the computational practicality of private information retrieval," in *Proceedings of the Network and Distributed Systems Security Symposium*, 2007, pp. 2006–06

[5] O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious rams," *Journal of the ACM (JACM)*, vol. 43, no. 3, pp.431–473, 1996..

[6] K. D. Bowers, A. Juels, and A. Oprea, "Hail: A high-availability and integrity layer for cloud storage," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, 2009, pp. 187– 198.

[7] R. Ostrovsky, "Efficient computation on oblivious rams," in *Proceedings of ACM STOC*, 1990, pp. 514–523.

[8] O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious rams," *Journal of the ACM (JACM)*, vol. 43, no. 3, pp. 431–473, 1996.

[9] F. Dabek, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica, "Widearea cooperative storage with cfs," in *ACM Symposium on Operating Systems Principles*, 2001, pp. 202–215.