RESEARCH ARTICLE OPEN ACCESS

Highly secured and effective keyword search over encrypted cloud data

Michael Shabi Nathan. S
PG Scholar
Department of Computer Science and Engineering,
University College of Engineering
Nagercoil, India
msn.s009@gmail.com

Dr.V.Kavitha, M.E., Ph.D.,
Dean
University College of Engineering
Nagercoil, India

Abstract—
Cloud computing economically enables the paradigm of data service outsourcing. However, to protect data privacy, sensitive cloud data have to be encrypted before outsourced to the commercial public cloud, which makes effective data utilization service a very challenging task. Although traditional searchable encryption techniques allow users to securely search over encrypted data through keywords, they support only Boolean search and are not yet sufficient to meet the effective data utilization need that is inherently demanded by large number of users and huge amount of data files in cloud. In this paper, we define and solve the problem of secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy. Specifically, we explore the statistical measure approach, i.e., relevance score, from information retrieval to build a secure searchable index, and develop a one-to-many order-preserving mapping technique to properly protect those sensitive score information. The resulting design is able to facilitate efficient server-side ranking without losing keyword privacy. Thorough analysis shows that our proposed solution enjoys “as strong-as-possible” security guarantee compared to previous searchable encryption schemes, while correctly realizing the goal of ranked keyword search. Extensive experimental results demonstrate the efficiency of the proposed solution.

Index Terms—cloud computing, confidential data, searchable encryption, ranked search

I. INTRODUCTION

Cloud Computing is a used to describe a new class of network based computing that takes place over the Internet. It is basically a step on from Utility Computing. It ia a general computing and it has a collection of integrated and networked hardware, software and Internet infrastructure. It uses the Internet for communication and transport that provides hardware, software and networking services to clients. These platforms hide the complexity and details of the underlying infrastructure from users and applications by providing very simple graphical interface or API. This kind of platform provides on demand services that are always on, anywhere, anytime and anyplace.

Cloud Computing has been envisioned as the next- generation architecture of IT Enterprise. Normally it moves the application software and databases to the centralized large data centers. The management of the data and services may not be fully trustworthy. This paradigm brings about many new security challenges which have not been well understood. Cloud storage enables users to remotely store their data and enjoy the on demand high quality cloud applications without the burden of local hardware and software management. Cloud computing provides convenient on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. The concept of cloud computing fills a perpetual need of IT. A way to increase capacity or capabilities on the fly without investing in new infrastructure, training, new personnel or licensing new software. Cloud computing encompasses any subscription-based or pay-per-use service that, in real time over the internet, extends its existing capabilities. Several trends are opening up the era of Cloud Computing which is an internet-based development and use of computer technology.

Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search[1], without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing cloud environment, they may suffer from the following problems. For each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved file in order to find ones most matching their interest, which demands possibly large amount of post processing over-head .Next, invariably sending back all files solely based on presence or absence of the keyword further incurs large unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm. In short, lacking of effective mechanisms to ensure the file retrieval accuracy is a significant drawback of existing searchable encryption schemes in the context of Cloud Computing.

In this paper we propose a ranked search which greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer toward practical deployment of privacy-preserving data hosting services in the context of Cloud Computing. To achieve the design goals on both system security and usability, we propose to bring together the advance of both crypto and IR community to design the ranked searchable symmetric encryption (RSSE) scheme, in the spirit of "as-strong-as-possible" security guarantee.

Specifically, we explore the statistical measure approach from IR and text mining to embed weight information (i.e., relevance score) of each file during the establishment of searchable index before outsourcing the encrypted file collection. As directly outsourcing relevance scores will leak lots of sensitive frequency information against the keyword privacy, we then integrate a recent crypto primitive order-preserving symmetric encryption (OPSE) and properly modify it to develop a one-to-many order-preserving mapping technique for our purpose to protect those sensitive weight information, while providing efficient ranked search functionalities.

II. ENCRYPTION

Encryption is a process of converting data into code which is called as ciphertext. It is used to help the server to encrypt the document using any encryption algorithm and to convert the encrypted document to the File with activation code and then activation code send to the user for download. To protect data privacy and combat unsolicited accesses, sensitive data has to be encrypted before outsourcing so as to provide end-to-end data confidentiality assurance in the cloud and beyond. However, data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files.

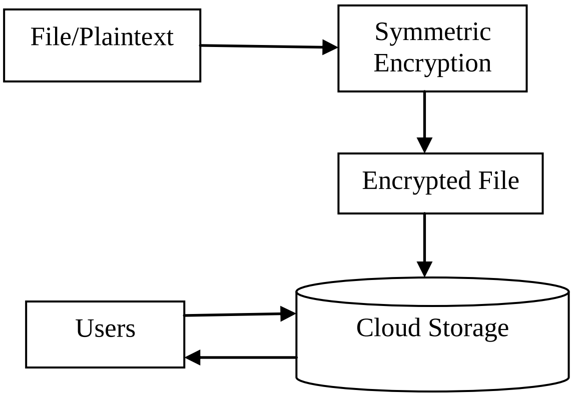

Fig 1: Process of encrypting and storing cloud data

Besides, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. Unfortunately, data encryption, which restricts user's ability to perform keyword search and further demands the protection of keyword privacy, makes the traditional plaintext search methods fail for encrypted cloud data.

III. SEARCHING ON ENCRYPTED DATA

Cloud computing is very popular. It is dangerous to upload plain data. The solution to the problem is to encrypt these data before outsourcing. The main goal is to make the cloud be able to do keyword search but learn nothing about the data and search. Encryption hides all partial information about data. Client must download all data, decrypt and perform operations locally.

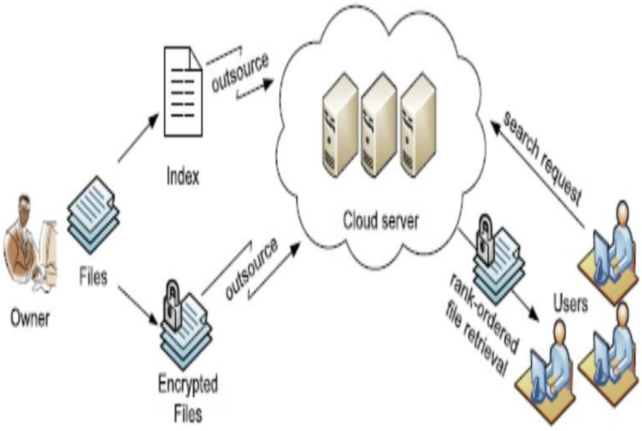

Fig 2: Architecture for search over encrypted cloud data

Here the owner has encrypted the file using symmetric encryption then it is stored in the cloud storage. Once the encrypted data are stored in the cloud storage then the users can use that data using single keyword search. Searchable encryption schemes allow users to perform keyword based searches on an encrypted database. Almost all existing such schemes only consider the scenario where a single user acts as both the data owner and the query. However, most databases in practice do not just serve one user; instead, they support search and write operations by multiple users. In this paper, we systematically study searchable encryption in a single user setting and to build a secure searchable index, and develop a one-to-many order-preserving mapping technique to properly protect those sensitive score information.

IV. SEARCHABLE SYMMETRIC ENCRYPTION

Symmetric Encryption is a process of encrypting and decrypting a message or file using a single same key. Searchable symmetric encryption (SSE) [5] allows a party to outsource the storage of his data to another party in a private manner, while maintaining the ability to selectively search over it. This problem has been the focus of active research and several security definitions and constructions have been proposed. Then present two constructions that show secure under our new definitions. Interestingly, in addition to satisfying stronger security guarantees, our constructions are more efficient than all previous constructions. Further, prior work on SSE only considered the setting where only the owner of the data is capable of submitting search queries. Private-key storage outsourcing allows clients with either limited resources or limited expertise to store and distribute large amounts of symmetrically encrypted data at low cost. Since regular private-key encryption prevents one from searching over encrypted data, clients also lose the ability to selectively retrieve segments of their data. To address this, several techniques have been proposed for provisioning symmetric encryption with search capabilities; the resulting construct is typically called searchable encryption. An index is a data structure that stores document collections while supporting efficient keyword search, given a keyword, the index returns a pointer to the documents that contain it.

It requires an amount of work from the server that is linear in the number of documents that contain the keyword requires constant storage on the client, and linear storage on the server. While the construction also performs searches in one round, it can induce false positives, which is not the case for our construction. Additionally, all the constructions require the server to perform an amount of work that is linear in the total number of documents in the collection. The ranked keyword search [2] over encrypted data is used to achieve economies of scale for Cloud Computing.

Cryptographic encryption protects data from compromise due to theft or intrusion. In addition to outsider attacks, security measures should also be taken against potential insider attacks. For example, when information storage is outsourced to a third-party data center, system administrators and other personnel involved may not be trusted to have decryption keys and access the content of the data collections. When an authorized user remotely accesses the data collection to search and retrieve desired documents, the large size of the collections often makes it infeasible to ship all encrypted data to the user’s side, and then perform decryption and search on the user’s trusted computers. Therefore, new techniques are needed to encrypt and organize the data collections in such a way as to allow the data center to perform efficient search in encrypted domain. There are a number of scenarios where the content owner may want to grant a user limited access to search a confidential collection. For example, the searcher could be a scholar or a low-level analyst who wants to identify relevant documents from a private or classified collection, and may need clearance only for the top-ranked documents; the searcher could also be the opposing side during document discovery phase of a litigation, who would request relevant documents from the content owner’s digital collection be turned over.

The requirements of balancing privacy and confidentiality with efficiency and accuracy pose significant challenges to the design of search schemes for a number of search scenarios. This problem has attracted interests from the cryptography community in recent years to investigate theories and techniques for “searchable encryption.” However, existing work only supports Boolean searches to identify the presence or absence of terms of interests in encrypted documents. Advances in information retrieval have gone well beyond Boolean searches; scoring schemes have been widely employed to quantify and rank-order the relevance of a document to a set of query terms. The goals of this paper are to explore a framework to securely rank-order documents in response to a query, and develop techniques to extract the most relevant documents from a large encrypted data collection. To our best knowledge, this is the first attempt in the research community to explore secure rank-ordered search. As an initial step, this paper on modeling common scenarios of secure rank-ordered search and exploring indexing and search techniques built upon existing established cryptographic primitives. The understandings obtained from this exploration will pave ways to bring together researchers from information retrieval and applied cryptography to establish a bridge between these areas. To accomplish our goals, we collect term frequency information for each document in the collection to build indices, as in traditional retrieval systems for plaintext. We further secure these indices that would otherwise reveal important statistical information about the collection to protect against statistical attacks. During the search process, the query terms are encrypted to prevent the exposure of information to the data center and other intruders, and to confine the searching entity to only make queries within an authorized scope. Utilizing term frequencies and other document information, we apply cryptographic techniques such as order-preserving encryption to develop schemes that can securely compute relevance scores for each document, identify the most relevant documents, and reserve the right to screen and release the full content of relevant documents. The proposed framework has comparable performance to conventional searching systems designed for non-encrypted data in terms of search accuracy.

V. RANKED KEYWORD SEARCH

The review of existing searchable symmetric encryption (SSE)[5] schemes and provides the definitions and framework for our proposed ranked searchable symmetric encryption (RSSE). It would be very inefficient to support ranked search[2] functionality over encrypted data, as demonstrated in our basic scheme. In this section, we develop a framework to perform ranked search securely and efficiently with minimum disclosure of the indexing information. We assume that the data center can only be trusted with data storage and should not be allowed to obtain information about the stored data. This module is used to help the user to get the accurate result based on the single keyword concepts. The users can enter the multiple words query, the server is going to split that query into a single word after search that word file in our database. Finally, display the matched word list from the database and the user gets the file from that list.

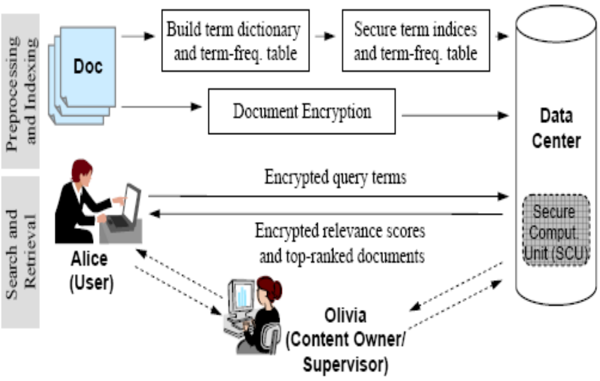
Preprocessing and Indexing
Doc
Build term dictionary and term-freq. table
Secure term indices and term-freq. table
Document Encryption
Data Center
Search and Retrieval
Encrypted query terms
Encrypted relevance scores and top-ranked documents
Secure Comput. Unit (SCU)
Alice (User)
Olivia (Content Owner/ Supervisor)

Fig 3: A framework for confidentiality preserving ranked search

The pre-processing is executed once by Olivia, when she stores the documents, all in encrypted form, in the data center. The major task of the pre-processing stage is to build a secure term frequency table and a secure inverse document frequency table, so as to facilitate efficient and accurate information retrieval. For an unprotected term frequency table, both the search term and its term frequency information are in plaintext. To protect the confidentiality of the search, we encrypt each of them in an appropriate way. Here, we create a new framework for confidentiality preserving rank-ordered search and retrieval over large document collections by using secure index, encrypted domain search & ranked retrieval. The use of this method is to provide good accuracy for a wide range of applications. It is designed only for non-encrypted data. Here the ccomplexity is high. It also focuses on protecting communication links and combating traffic analysis.

VI. ORDER-PRESERVING MAPPING TECHNIQUE

Normally Order preserving mapping is used to preserve the order. The order of mapped points on the two boundaries must be monotonically non-decreasing.

It is allowing different levels of detail like One-to-one, Many-to-one, One-to-many. Here the proposed method follows one-to-many order preserving mapping technique. The authorization between the data owner and users is appropriately done. To search the file collection for a given keyword, an authorized user generates and submits a search request in a secret form to the cloud server. Upon receiving the search request the cloud server is responsible to search the index and return the corresponding set of files to the user. This is considered as the secure ranked keyword search problem.

VII. CONCLUSION

In the proposed system, we motivate and solve the problem of supporting efficient keyword search for achieving effective utilization of remotely stored encrypted data in Cloud Computing. We first give a basic scheme and show that by following the same existing searchable encryption framework, it is very inefficient to achieve keyword search. We then appropriately weaken the security guarantee, resort to the newly developed crypto primitive OPSE, and derive an efficient one-to-many order-preserving mapping function, which allows the effective RSSE to be designed. Through security analysis, we show that our proposed solution is secure and privacy-preserving, while correctly realizing the goal of ranked keyword search. Extensive experimental results demonstrate the efficiency of our solution.

VIII. FUTURE ENHANCEMENT

In further enhancements of our search mechanism, we are going to search an Image or Video file in the cloud server which includes the efficient support of relevance score dynamics, the authentication of search results, and the reversibility of our proposed one-to-many order-preserving mapping technique.

REFERENCES

[1] C. Wang, N. Cao, K. Ren, and W. Lou, “Enabling Secure ranked keyword search over encrypted cloud data,” IEEE Transactions on Parallel & Distributed systems, Vol.23, Pg.no:8, Aug 2012.

[2] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Secure ranked keyword search over encrypted cloud data,” in *Proc. of ICDCS’10*, 2010.

[3] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the clouds: A berkeley view of cloud computing,” University of California, Berkeley, Tech. Rep. UCBEECS-2009-28, Feb 2009.

[4] Cloud Security Alliance, “Security guidance for critical areas of focus in cloud computing,” 2009, http://www. cloudsecurityalliance.org.

[5] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: improved definitions and efficient constructions,” in *Proc. of ACM CCS’06*, 2006.

[6] Y.-C. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data,” in *Proc. of ACNS’05*, 2005.

[7] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in *Proc. of EUROCRYP’04, volume 3027 of LNCS*. Springer, 2004.