ISSN No. 0976-5697

Volume 4, No. 2, Jan-Feb 2013

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Security and Privacy Ensured Data Search Model for Encrypted Storage in Cloud Environment

Treesa Maria Vincent*
Dept. of Computer Science
Dept. of Computer Science
Thodupuzha,India
mariatreesa15@gmail.com

Mrs.J.Sakunthala
Dept. of Computer Science
Dept. of Computer Science
Erode,India
Sakunthalasubburam@gmail.com

Abstract: Cloud computing economically paradigm of data service outsourcing. Data owner share the data under the cloud servers. The cloud server may leak data information due to unauthorised entities or even be hacked. Encrypted storage and retrieval model is used. Searchable encryption technique supports only Boolean search process. Large amount of users and files are not handled by this search process. The privacy enabled data searching scheme provides solution for secure ranked key word search over encrypted cloud data. Ranked search enhances system usability by enabling search result relevance ranking. Relevance score is statistical measure approach is used in information retrieval. Relevance score is used in secure searchable index preparation process. One to many order preserving mapping technique is used to properly protect sensitive score information. Ranked searchable symmetric encryption is used to perform secured data retrieval process.

Keywords: Relevance score, Ranked Search, Encryption, Ranked searchable symmetric Encryption, One-to many order preserving mapping

I. INTRODUCTION

In information retrieval, inverted index is a widely used indexing structure that stores a list of mappings from keywords to the corresponding set of files that contain this keyword , allowing full text search. For ranked search purpose , the task of determining which files are most relevant is typically done by assigning a numerical score, which can be precompiled, to each file based on some ranking function. We will use this inverted index structure to give our basic ranked searchable encryption construction.

In information retrieval, a ranked function is used to calculate relevance score of matching files to given search request. The most widely used statistical measurement for evaluating relevance score in the information retrieval community uses the TF*IDF rule, where Term Frequency(TF) is simply the number of times a given term or keyword appears within a file and Inverse Document Frequency(IDF) is obtained by dividing the number of files in the collection by the number of files containing the term. Among several

II. OVERVIEW

Cloud Computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources [2]. The benefits brought by this new computing model include but are not limited to: relief of the burden for storage management, universal data access with independent geographical locations, and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc., [3].

As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud, such as e-mails, personal health records, company finance data, and government documents, etc. The fact that data owners and cloud server are no longer in the same trusted domain may put the outsourced unencrypted data at risk [4] the cloud server may leak data information to unauthorized entities [5] or even be hacked [6]. It follows that sensitive data have to be encrypted prior to outsourcing for data privacy and combating unsolicited accesses. However, data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. 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.

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 two main drawbacks. On the one hand, 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 overhead, On the other hand, invariably sending back all files solely based on presence/ 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. Nonetheless, the state of the art in information retrieval (IR) community has already been

utilizing various scoring mechanisms quantify and rank order the relevance of files in response to any given search query. Although the importance of ranked search has received attention for a long history in the context of plaintext searching by IR community, surprisingly, it is still being overlooked and remains to be addressed in the context of encrypted data search.

Having a correct intuition on the security guarantee of existing SSE literature is very important for us to define our ranked searchable symmetric encryption problem. As later, we will show that following the exactly same security guarantee of existing SSE scheme, it would be very inefficient to achieve ranked keyword search, which motivates us to further weaken the security guarantee of existing SSE appropriately and realize an "as-strong-as-possible" ranked searchable symmetric encryption. Actually, this notion has been employed by cryptographers in much recent work [7] where efficiency is preferred over security.

Therefore, how to enable a searchable encryption system with support of secure ranked search is the problem tackled in this paper. Our work is among the first few ones to explore ranked search over encrypted data in Cloud Computing. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria, thus making one step closer toward practical deployment of privacy-preserving data hosting services in the context of Cloud Computing. To achieve our 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 of each file during the establishment of searchable index before outsourcing the encrypted file collection [12].

Searchable encryption. Traditional searchable encryption has been widely studied as a cryptographic primitive, with a focus on security definition formalizations and efficiency improvements. Song et al. first introduced the notion of searchable encryption. They proposed a scheme in the symmetric key setting, where each word in the file is encrypted independently under a special two-layered encryption construction. Thus, a searching overhead is linear to the whole file collection length. Goh developed a Bloom filter-based per-file index, reducing the workload for each search request proportional to the number of files in the collection. Chang and Mitzenmacher also developed a similar per-file index scheme. To further enhance search efficiency, Curtmola et al. proposed a per-keyword-based approach, where a single encrypted hash table index is built for the entire file collection, with each entry consisting of the trapdoor of a keyword and an encrypted set of related file identifiers. Searchable encryption has also been considered in the public-key setting. Boneh et al. presented the first public-key-based searchable encryption scheme, with an analogous scenario. In their construction, anyone with the public key can write to the data stored on the server but only authorized users with the private key can search. As an attempt to enrich query predicates, conjunctive keyword search over encrypted data have also been proposed. Aiming at tolerance of both minor typos and format inconsistencies in the user search input, fuzzy keyword search over encrypted cloud data has been proposed by Li et al. in [9]. Very recently, a privacy-assured similarity search mechanism over outsourced cloud data has been explored by Wang et al. in [11]. Note that all these schemes support only Boolean keyword search, and none of them support the ranked search problem which we are focusing on in this paper.

Following our research on secure ranked search over encrypted data, very recently, Cao et al. [10] propose a privacy-preserving multikeyword ranked search scheme, which extends our previous work in [1] with support of multikeyword query. They choose the principle of "coordinate matching," i.e., as many matches as possible, to capture the similarity between a multikeyword search query and data documents, and later quantitatively formalize the principle by a secure inner product computation mechanism. One disadvantage of the scheme is that cloud server has to linearly traverse the whole index of all the documents for each search request, while ours is as efficient as existing SSE schemes with only constant search cost on cloud server.

At the first glance, by changing the relevance score encryption from the standard indistinguishable symmetric encryption scheme to this OPSE, it seems to follow directly that efficient relevance score ranking can be achieved just like in the plaintext domain. However, as pointed out earlier, the OPSE is a deterministic encryption scheme. This inherent deterministic property, if not treated appropriately, will still leak a lot of information as any deterministic encryption scheme will do. One such information leakage is the plaintext distribution. For example, which shows a skewed relevance score distribution of keyword "network," sampled from 1,000 files of our test collection. For easy exposition, we encode the actual score into 128 levels in domain from 1 to 128. Due to the deterministic property, if we use OPSE directly over these sampled relevance scores, the resulting ciphertext shall share exactly the same distribution as the relevance score. Specifically, the authors have shown that the TF distribution of certain keywords from the Enron e-mail corpus3 can be very peaky, and thus result in significant information leak for the corresponding keyword. In [8], the authors further point out that the TF distribution of the keyword in a given file collection usually follows a power law distribution, regardless of the popularity of the keyword. Their results on a few test file collections show that not only different keywords can be differentiated by the slope and value range of their TF distribution, but even the normalized TF distributions, i.e., the original score distributions can be keyword specific. Thus, with certain background information on the file collection, such as knowing it contains only technical research papers, the adversary may be able to reverse engineer the keyword "network" directly from the encrypted score distribution without actually breaking the trapdoor construction, nor does the adversary need to break the OPSE.

We consider an encrypted cloud data hosting service involving three different entities as illustrated in Figure 1, data owner, data user, and cloud server. Data owner has a collection of n data files $C = \{F_1, F_2, \ldots, F_n\}$ that he wants to outsource on the cloud server in encrypted form while still keeping the capability to search through them for effective data utilization reasons. To do so, before outsourcing, data owner will first build a secure searchable index I from a set of m distinct keywords $W = \{w_1, w_2, \ldots, w_m\}$ extracted2 from the file collection C, and store both the index I and the encrypted file collection C on the cloud server.

We assume the authorization between the data owner and users is appropriately done. To search the file collection for a given keyword w, an authorized user generates and submits a search request in a secret form—a trapdoor T_w of the keyword w—to the cloud server. Upon receiving the search request T_w, the cloud server is responsible to search the index I and return the corresponding set of files to the user. We consider the secure ranked keyword search problem as follows: the search result should be returned according to certain ranked relevance criteria (e.g., keyword frequency-based scores, as will be introduced shortly), to improve file retrieval accuracy for users without prior knowledge on the file collection C. However, cloud server should learn nothing or little about the relevance criteria as they exhibit significant sensitive information against keyword privacy. To reduce bandwidth, the user may send an optional value k along with the trapdoor T_w and cloud server only sends back the top-k most relevant files to the user's interested keyword w.

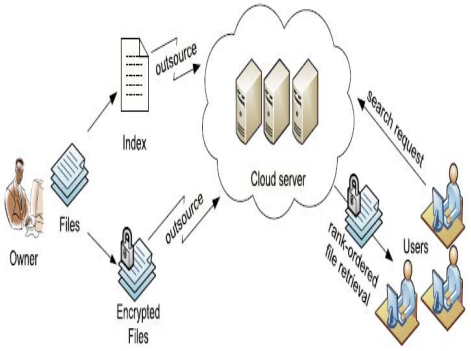

Figure 1: Architecture for search over encrypted cloud data

We primarily consider an "honest-but-curious" server in our model, which is consistent with most of the previous searchable encryption schemes. We assume the cloud server acts in an "honest" fashion and correctly follows the designated protocol specification, but is "curious" to infer and analyze the message flow received during the protocol so as to learn additional information. In other words, the cloud server has no intention to actively modify the message flow or disrupt any other kind of services. However, in some unexpected events, the cloud server may behave beyond the "honest-but-curious" model.

III. OBJECTIVES

To enable ranked searchable symmetric encryption for effective utilization of outsourced and encrypted cloud data under the aforementioned model, our system design should achieve the following security and performance guarantee. Specifically, we have the following goals: 1) Ranked keyword search: to explore different mechanisms for designing effective ranked search schemes based on the existing searchable encryption framework, 2) Security guarantee: to prevent cloud server from learning the plaintext of either the data files or the searched keywords, and achieve the "as-strong-as-possible" security strength compared to existing searchable encryption schemes, 3) Efficiency: above goals should be achieved with minimum communication and computation overhead.

A. Secure Ranked Keyword Search over Encrypted Cloud Data

Cloud Computing enables cloud customers to remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. The benefits brought by this new computing model include but are not limited to: relief of the burden for storage management, universal data access with independent geographical locations, and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc.

With the prevalence of cloud services, more and more sensitive information are being centralized into the cloud servers, such as emails, personal health records, private videos and photos, company finance data, government documents, etc. 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. 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.

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 search1, 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 two main drawbacks. On the one hand, 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 overhead; On the other hand, invariably sending back all files solely based on presence/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. Nonetheless, the state-of-the-art in information retrieval (IR) community has already been utilizing various scoring mechanisms quantify and rank-order the relevance of files in response to any given search query. Although the importance of ranked search has received attention for a long history in the context of plaintext searching by IR community, surprisingly, it is still being overlooked and remains to be addressed in the context of encrypted data search.

B. Searchable Symmetric Encryption

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. The area of searchable encryption has been identified by DARPA as one of the technical advances that can be used to balance the need for both privacy and national security in information aggregation systems. In addition, it can allow services such as Google Desktop offer valuable features without sacrificing the client's privacy.

Searchable encryption can be achieved securely in its full generality using the work of Ostrovsky and Goldreich on software protection based on oblivious RAMs. While oblivious RAMs hide all information about the RAM use from a remote and potentially malicious server with a polylogarithmic overhead in all parameters, this comes at the cost of a logarithmic number of rounds of interaction for each read and write. Therefore, the previously mentioned work on searchable encryption achieves more efficient solutions by weakening the privacy guarantees.

In the setting of **searching on private-key**-encrypted data, the user himself encrypts the data, so he can organize it in an arbitrary way and include additional data structures to allow for efficient access of relevant data. The data and the additional data structures can then be encrypted and stored on the server so that only someone with the private key can access it. In this setting, the initial work for the user is at least as large as the data, but subsequent work is very small relative to the size of the data for both the user and the server. Furthermore, everything about the user's access pattern can be hidden.

C. Order-Preserving Symmetric Encryption

Order-preserving symmetric encryption (OPSE) is a deterministic encryption scheme whose encryption function preserves numerical ordering of the plaintexts. OPSE has a long history in the form of one-part codes, which are lists of plaintexts and the corresponding cipher texts, both arranged in alphabetical or numerical order so only a single copy is required for efficient encryption and decryption. One-part codes were used, for example, during World War I. A more formal treatment of the concept of order-preserving symmetric encryption (OPE) was proposed in the database community by Agrawal et al. The reason for new interest in such schemes is that they allow efficient range queries on encrypted data. That is, a remote untrusted database server is able to index the data it receives, in encrypted form, in a data structure that permits efficient range queries. By "efficient" we mean in time logarithmic in the size of the database, as performing linear work on each query is prohibitively slow in practice for large databases.

OPSE not only allows efficient range queries, but allows indexing and query processing to be done exactly and as efficiently as for unencrypted data, since a query just consists of the encryptions of a and b and the server can locate the desired cipher texts in logarithmic-time via standard tree-based data structures. Indeed, subsequent to its publication, has been referenced widely in the database community, and OPSE has also been suggested for use in in-network aggregation on encrypted data in sensor networks and as a tool for applying signal processing techniques to multimedia content protection. Yet a cryptographic study of OPE in the provable-security tradition never appeared. Our work aims to begin to remedy this situation.

D. Privacy-Preserving Public Auditing for Secure Cloud Storage

Cloud Computing has been envisioned as the next-generation information technology (IT) architecture for enterprises, due to its long list of unprecedented advantages in the IT history: on-demand self-service, ubiquitous network access, location independent resource pooling, rapid resource elasticity, usage-based pricing and transference of risk. As a disruptive technology with profound implications, Cloud Computing is transforming the very nature of how businesses use information technology. One fundamental aspect of this paradigm shifting is that data is being centralized or outsourced to the Cloud.

While Cloud Computing makes these advantages more appealing than ever, it also brings new and challenging security threats towards users' outsourced data. Since cloud service providers (CSP) are separate administrative entities, data outsourcing is actually relinquishing user's ultimate control over the fate of their data. As a result, the correctness of the data in the cloud is being put at risk due to the following reasons. First of all, although the infrastructures under the cloud are much more powerful and reliable than personal computing devices, they are still facing the broad range of both internal and external threats for data integrity. Examples of outages and security breaches of noteworthy cloud services appear from time to time. Secondly, there do exist various motivations for CSP to behave unfaithfully towards the cloud users regarding their outsourced data status. For examples, CSP might reclaim storage for monetary reasons by discarding data that has not been or is rarely accessed, or even hide data loss incidents to maintain a reputation. In short, although outsourcing data to the cloud is economically attractive for long-term large-scale storage, it does not immediately offer any guarantee on data integrity and availability. This problem, if not properly addressed, may impede the success of cloud architecture.

To fully ensure the data integrity and save the cloud users' computation resources as well as online burden, it is of critical importance to enable public auditing service for cloud data storage, so that users may resort to an independent third party auditor (TPA) to audit the outsourced data when needed. The TPA, who has expertise and capabilities that users do not, can periodically check the integrity of all the data stored in the cloud on behalf of the users, which provides a much more easier and affordable way for the users to ensure their storage correctness in the cloud. Moreover, in addition to help users to evaluate the risk of their subscribed cloud data services, the audit result from TPA would also be beneficial for the cloud service providers to improve their cloud based service platform, and even serve for independent arbitration purposes. In a word, enabling public auditing services will play an important role for this nascent cloud economy to become fully established; where users will need ways to assess risk and gain trust in the cloud.

E. Confidentiality-Preserving Rank-Ordered Search

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/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/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 document(s) 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, we focus in 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.

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.

IV. EXISTING SYSTEM

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, 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 two main drawbacks. On the one hand, 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 overhead, On the other hand, invariably sending back all files solely based on presence/ 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. Nonetheless, the state of the art in information retrieval (IR) community has already been utilizing various scoring mechanisms quantify and rank order the relevance of files in response to any given search query. Although the importance of ranked search has received attention for a long history in the context of plaintext searching by IR community, surprisingly, it is still being overlooked and remains to be addressed in the context of encrypted data search.

Therefore, how to enable a searchable encryption system with support of secure ranked search is the problem tackled in this paper. Our work is among the first few ones to explore ranked search over encrypted data in Cloud Computing. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria, thus making one step closer toward practical deployment of privacy-preserving data hosting services in the context of Cloud Computing. To achieve our 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 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.

A. Drawbacks of existing system

The privacy enabled data searching scheme provides solution for 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. The statistical measure approach, i.e., relevance score, from information retrieval is explored to build a secure searchable index. One-to-many order-preserving mapping technique is developed to properly protect those sensitive score information.

- Static relevance score model

- Complex reversible operation under order preserving scheme
- Result authentication is not provided
- Retrieval latency is high

V. PROPOSED SYSTEM

The cloud data center manages the transactional data values. The data values are maintained in encrypted format. The data values are queried using the encrypted query values. The system is designed to provide data security and privacy for the transactional data over the cloud environment. The order preserving mapping model is used for the encryption process. The score functions are used to fetch the data values in a ranked manner. The dynamic scoring mechanism is used in the system.

A. Description of the system

The system is divided into two applications. They are data source and client application. The data source manages the transactional data values. The client application issues the query value and collects the data from the data source. The data values are updated in the data source in an encrypted format. The data retrieval and ranking operations are carried out on the encrypted data format only. The system secures the data under the storage and query transmission process.

The system is divided into five major modules. They are data source, storage management, score assignment, client and query process. The data source module is designed to manage the data values. The storage management module is designed to perform the data encryption and update operations. The score assignment module is used to assign the relevance score the for the transactional data values. The client application is used to fetch the data value from the data source. The query process module is designed to submit and collect the data values.

B. Data Source

The data source application is designed to manage the transactional and user information. The user information are updated with their access information. All the query history is maintained under the data source application. The transactional data values are maintained for different domains. The data values are updated in encrypted format. The data retrieval is performed under the data source application.

C. Storage Management

The storage management is designed to handle data encryption and update operations. The order preserving mapping technique is used to encrypt the data values. The system includes the reversible order preserving map model for the encryption process. The data update operation can be dynamically performed on the system. The data values are updated and stored in the encrypted format. The transactional data and its encryption process are carried out under the data source environment.

D. Score Assignement

The score assignment module is designed to assign the score values for the transactions. The similarity value is estimated to assign the score values. The relevance score is used to rank the transaction data values. The data retrieval is carried out with the score functions. The incremental data update initiates the dynamic score assignment process. The dynamic score assignment process updates the score values based on the new transaction data values.

E. Client

The client application is designed to perform the data retrieval operations. The data values are collected from the server and updated into the client interface. Each client is authenticated with unique identification value. The client collects the data values with query keywords.

F. Query Process

The query process module is designed to fetch the transactional data values. Query keyword is collected from the client. The query keyword is encrypted and transferred to the data source. The data source performs the searching process. The transactional data values are compared and similarity values are estimated. The results are prepared using the similarity value and threshold levels. The client application decrypts the transactional data values and produces the results in a ranked way.

VI. CONCLUTION

Cloud customers can remotely store their data on a shared pool of configurable computing resources in cloud. Searchable Symmetric Encryption scheme is used to provide storage and retrieval security. Order Preserving Symmetric Encryption scheme is enhanced in reversible mechanism. The system is improved with result authentication and similarity based ranking model. The data storage and search process is carried out with encrypted query model. The system performs index operations on encrypted data values. The system also secures the search results. The system supports incremental data update scheme.

VII. ACKNOWLEDGMENT

I express my sincere thanks to Mrs.J.Sakunthala ME, Assistant professor, Department of Computer Science and Engineering (PG), who has always been with me throughout the process of design and construction of this paper. I am very much thankful for her guidance, constant encouragement, support and valuable suggestions to successfully carryout this paper.

VIII. REFERENCES

[1] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Secure Ranked Keyword Search over Encrypted Cloud Data,” Proc. IEEE 30th Int’l Conf. Distributed Computing Systems (ICDCS ’10), 2010.

[2] P. Mell and T. Grance, “Draft Nist Working Definition of CloudComputing,”http://csrc.nist.gov/groups/SNS/cloudcomputing/ index.html, Jan. 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,” Technical Report UCB-EECS-2009-28, Univ. of California, Berkeley, Feb. 2009.

[4] Cloud Security Alliance “Security Guidance for Critical Areas of Focus in Cloud Computing,” http://www.cloudsecurityalliance.org, 2009.

[5] Cloud Security Alliance “Security Guidance for Critical Areas of Focus in Cloud Computing,” http://www.cloudsecurityalliance.org, 2009.

[6] Cloud Security Alliance "Security Guidance for Critical Areas of Focus in Cloud Computing," http://www.cloudsecurityalliance.org, 2009.

[7] A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill, "Order-Preserving Symmetric Encryption," Proc. Int'l Conf. Advances in Cryptology, 2009.

[8] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k Retrieval from a Confidential Index," Proc. Int'l Conf. Extending Database Technology: Advances in Database Technology (EDBT '09), 2009.

[9] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy Keyword Search over Encrypted Data in Cloud Computing," Proc. IEEE Infocom '10, 2010.

[10] N. Cao, C. Wang, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE Infocom '11, 2011.

[11] C. Wang, K. Ren, S. Yu, K. Mahendra, and R. Urs, "Achieving Usable and Privacy-Assured Similarity Search over Outsourced Cloud Data," Proc. IEEE INFOCOM, 2012.

[12] Cong Wang, Ning Cao, Kui Ren and Wenjing Lou, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data" IEEE Transactions On Parallel And Distributed Systems, Vol. 23, No. 8, August 2012.