Volume : 4 | Issue : 6 | Special Issue June 2015 • ISSN No 2277 - 8179

ENCRYPTED CLOUD DATA USAGE FOR SEARCHING OF RANKED MULTI-KEYWORD

Engineering

KEYWORDS: Cloud computing, keyword search, privacy-preserving, ranked search searchable encryption

ELSON KURIAN,

M.Tech Student, Dept. of Computer Science and Engineering, Shree Devi Institute of Technology, Mangalore, Karnataka (India).

ABSTRACT

With the onset of cloud computing, data holders are inspired to deploy their miscellaneous data management systems to the commercial public cloud for extreme flexibility and utility. Sensitive data have to be encrypted before outsourcing, in-order to assure data privacy. This obscures traditional data usage based on plain-text keyword search. This makes it essential to provide a service to search encrypted cloud data. Plaintext key word is replaced by multiple keywords; consider the huge amount of data holders and documents in the cloud. In this paper, the asserting problem of searching encrypted cloud data using ranked multi-keyword (MRSE) is defined and solved. Out of distinct multi-keyword semantics, the adequate similarity measuring of "coordinates matching" and "inner product similarity, i.e., possibilities of many matches for capture the documents from query search perceptible evaluations for similarity measures. Adopting the basic idea for the MRSE based on secure inner product computation and archive privacy requirements in two distant thread models. To assure the searching reality of the data searching service, this proposal supports more and more search semantics. Experiments based on the real-world data further showing an indeed advent of low overhead on computation and communication.

1 Introduction

Cloud computing is the deep vision of computing as a great service, where the cloud users can casually store their data in the cloud, providing good quality applications and services. Protection of privacy and unauthorised access in the cloud and farther perceptive data eg: tax documents, emails, financial transactions, personal documents etc... The data users have to be encrypted before deployed to pubic cloud [2]. Ranked search structure permits quicker significant information .An adequate search utility over encrypted cloud data is covered instead of downloading and decrypting the whole data from cloud [5]. Multiple key words search needs instead of single keyword for ranking system yields too rude results eg: Google search [8]. Problems that arise in cryptographic schemes are search over an encrypted data's and its security proofs. They provide provable mystery on encryption and query segregation for searches [6]. The peculiarity of cloud data storage is allows various keywords in search query and returns the documents in systematic order. The searching methods focus on feature of similarity and inner product matching [1]. The "Latent Semantic Analysis" explained the relation between terms and documents. It implied higher-order structure in the terms with documents and follows reduced-dimension vector space for representing words and documents.[7] System usability increases on ranked searching by returning the identical files in ranked order in certain criteria, thus practical formation of privacy-preserving over a data hosting services in Cloud Computing increased [3]. While preserving data sharing in multi owner method and privacy indentation imposing issue owed the change of membership of users. The only method for privacy of data is encrypting the data files before deploying to the cloud [4]. In this paper, data holders are inspired to deploy their miscellaneous data management systems to the commercial public cloud for extreme flexibility and utility. Sensitive data have to be encrypted before outsourcing, in order to ensure data privacy. This obscures traditional data usage based on plain-text keyword search. This makes it essential to provide a service to search encrypted cloud data [8].The number of query keywords emerge in a document, possibilities of many matches for capture the documents from query search perceptible evaluations for similarity measures Meanwhile the index creation, each document is combine with a binary vector as a sub-index where each bit shows comparable keywords is enclose in the document. The search query is also expressed as a binary vector. So the similarity could be truly measured by the inner product of the query vector with the data vector [9].

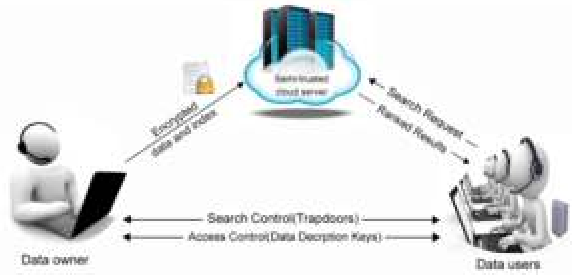
Semi-trusted cloud server
Encrypted data and index
Search Request
Ranked Results
Search Control(Trapdoors)
Access Control(Data Decription Keys)
Data owner
Data users

Fig. 1: Architecture of the search over encrypted clou data.

It can be summarized as follows:

1. Exploring the problem of multi keyword ranked search over encrypted cloud data and endowing privacy concern secure cloud data.

2. Two MRSE proposals can be modulated based on similarity measure on "coordinate matching" while met with two different threat models.

3. Further enhancements of our ranked search mechanism to support more search semantics and dynamic data operations are investigated.

4. Thorough analysis inspecting privacy and efficiency assurances of the proposed schemes is given. Experiments based on the real-world data further showing an indeed advent of low overhead on computation and communication

This version supports the data/index dynamics in the mechanism designing. Moreover, we enhance the experimental works by including the analysis and evaluation of new schemes. By adding more analysis on secure inner product and the privacy part for improvements

2 Problem Formulations

2.1 System Model

Cloud data hosting service holding different entities listed in Fig.1: the data owner, the data user, and the cloud server. The data owner outsourced the data documents F in the encrypted form C to the cloud server. Before outsourcing, build an encrypted searchable index I for effective data utilization. Then the collection of encrypted document and the index I are outsourced to cloud server. An authorized user earn corresponding trapdoor T over search control mechanisms, for searching the document collections using t given keywords. Upon accepting T from a data user, the cloud server is culpable to search the index I and return the analogues set of encrypted documents. According to some ranking touchstone, the cloud server ranked the documents accurately. The data user sends an alternative number k forward with trapdoor T so that the cloud server sends back top-k documents to scale down the communication cost. The access control mechanism is occupied to manage decryption efficiency given to users and the data collection can be amended in terms of inserting, updating, and deleting existing documents.

2.2 Threat Model

In this model, "honest-but-curious" cloud server is persistent with related works on cloud security. The cloud server performs in an "honest" fashion and properly following the nominated protocol specification. Its "curious" to assume and evaluate data in its storage and message outflows received during the protocol providing additional information. According to the information known by the cloud server, considering two threat models listed below.

Known Ciphertext model: In this model cloud server reveals only the encrypted data set C and searchable index I which are deployed from the data owner.

Known background model: In this model, the cloud servers possess additional knowledge than the known Ciphertext model. The correlation relationship of trapdoors and the statistical information of data sets enrich the models.

2.3 Design Goals

To implement the ranked search for useful application of outsourced cloud data under the preceding model, system design should be attain security and performance guarantees as follows.

- Multi-keyword ranked search. To design search strategy which grant multi-keyword query and grant result similarity ranking for data retrieval, instead of return the identical results.

- Privacy-preserving. To avoid the cloud server from learning extra information from the index and the data set.

- Efficiency. Above target on purpose and privacy should carry out with minor communication and computation overhead.

2.4 Notations

- F—the plaintext document collection, a set of m data documents F= (F1, F2,.... Fm).

- C –– The encrypted documents collections in the cloud server, denoted as C =(C1,C2, . . . ,Cm).

- W—the dictionary, the keyword set having of n keyword, denoted as W=(W1,W2, . . .,Wn).

- I—the searchable index combined with C, denoted as (I1,I2, . . . , Im) where each subindex Ii is create for Fi.

- Ŵ––the subset of W , the keywords in a search request, Ŵ= (Wj1, Wj2, ...,Wjt).

- TŴ–– the trapdoor for the searching request Ŵ.

- FŴ—the ranked id list of all documents in order to Ŵ.

2.5 Preliminary on Coordinate Matching

Coordinate matching is transitional similarity measure between conjunctive search and disjunctive search which uses the number of query keywords emerging in the document to specify the importance of query. The large amount of outsourced data in cloud computing malleable for users to mention a list of keywords reveals their interest and retrieve the most valuable documents with a ranked order.

3 Frameworks and Privacy Requirements for MRSE

In this part, describing the framework of multi-keyword ranked search over encrypted cloud data (MRSE) and providing privacy requirements for a secure cloud data utilization system.

3.1 MRSE Framework

Data documentations are not shown in framework since the data owner employing symmetric key cryptography to encrypt and then outsource data. With the index and query, the four algorithms as follows.

- Setup (l^l). Taking a security parameter ^{l} as input, the owner outputs a Symmetric Key as SK.

- BuildIndex (F, SK). Based on the dataset F, the owner creates a searchable index I which are encrypted by the key SK and then deployed to the cloud. After the index construction, the document collection can be encrypted and outsourced.

- Trapdoor (Ŵ). With t keywords of interest in W as input, this algorithm generates a corresponding trapdoor TŴ.

- Query (TŴ,k,I). When the cloud server get the query request as(TŴ,k), it perform the ranked search on the index I with support of trapdoor TŴ and then return FŴ, the ranked id lists of documents in top-k sorted order.

Search control is to control how authorized users achieve trapdoors and access control is to regulate users' access to outsourced documents.

3.2 Privacy Requirements for MRSE

`The representative privacy related in searchable encryption guarantees search results.

Data privacy: The data owner employing symmetric key cryptography to encrypt and then outsource data, and well avoid the cloud server from forwarding into the outsourced data.

Index privacy: The cloud servers deduct association among keywords and encrypted documents from index; it can realise the major content of a document. The searchable index should be created to prevent the cloud server from association attacks.

Keyword privacy: The users usually prefer to hide what they are searching. i.e., the keywords represented by the corresponding trapdoor. Trapdoor can be formed in a cryptographic way to assure the query keywords

Trapdoor unlinkability: The deterministic trapdoor formation give the cloud server favour to accrue frequencies of various search requests regarding other keyword(s) can disobey the keyword privacy requirement. Introduce proper non-determinacy into the trapdoor generation procedure protection by trapdoor unlinkability.

Access pattern: Within the ranked search, the access pattern is the array of quest results where each results is a set of documents with rank order. A few searchable encryption entireties have been suggested to apply private information retrieval (PIR) technique, to with hold the access pattern.

4 Performance Analyses

In this section, a detailed evaluation of the proposed technique on a real-world data set is illustrated. The performance of the technique is evaluated according the efficiency of MRSE schemes.

4.1 Precision and Privacy

For this Section, some of the dummy keywords are inserting into data vector and some of them are elected in every query. Based on similarity scores of data vectors to query vector the server returns back top-k documents, some of real top-k document query may be ignored. Due to the impact of dummy keywords, similarity scores are decreased in the original documents and the similarity scores of others documents out of the real top-k are raised. This proposal provides a balance parameter for data users to assure their different desires on precision and rank privacy.

4.2 Efficiency

4.2.1 Index Construction

To construct a searchable sub index Ii for each document Fi in the data set F, initially we map the keyword set drive from the document Fi to a data vector Di, pursue by encrypting each data vector. The time worth of mapping or encrypting directly depends on the dimensionality of data vector which is resolved by the size of the dictionary, i.e., the amount of indexed keywords. And the time outlay of building the entire index is also associated to the number of subindex which is equivalent to the number of documents in the data set. The time cost of creating the entire index is almost linear with the size of data set given that the time cost of creating each subindex is set.

4.2.2 Trapdoor Generation

Like index building, every trapdoor creation incurs two multiplications of a split and matrix and query vector, where the dimensionality of query vector or matrix is different in two proposed schemes and becomes outsized with the increasing dictionary size. Additional, it shows that the no of query keywords has small influence on the trapdoor generation, which is a major advantage over associated works on multi-keyword searchable encryption.

5 Conclusion

In this paper, the asserting problem of searching encrypted cloud data using ranked multi-keyword (MRSE) is defined and solved. Out of distinct multi-keyword semantics, the adequate similarity measuring of "coordinates matching" and "inner product similarity, i.e., possibilities of many matches for capture the documents from query search perceptible evaluations for similarity measures. Adopting the basic idea for the MRSE based on secure inner product computation and archive privacy requirements in two distant thread models. Experiments based on the real-world data further showing an indeed advent of low overhead on computation and communication.

In future, the cloud server is treated as entrusted state, the integrity checking of the rank order in search results analysed.

REFERENCE

[1] AnkathaSamuyelu Raja, Vasanthi, "A Secured Multi-keyword Ranked Search over Encrypted Cloud Data" Proc . IJARCSSE October 2012 | [2] Ning Cao, Cong Wang, Ming Li, " Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data" Proc. Ieee Infocom January-2014 | [3] 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. | [4] Mr. Parjanya C.A, Mr. Prasanna Kumar M, "Advance Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud" Proc . IJARCSSE march, 2014. | [5] A. Singhal, "Modern Information Retrieval: A Brief Overview,"IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001. | [6] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000 | [7] Li Chen, Xingming Sun, Zhihua Xia and Qi Liu, "An Efficient and Privacy-Preserving Semantic Multi-Keyword Ranked Search over Encrypted Cloud Data",Jan 2013. | [8] I.H. Witten, A. Moffat, and T.C. Bell, Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kaufmann Publishing , May 1999. | [9] W.K. Wong, D.W. Cheung, B. Kao, and N. Mamoulis, "Secure kNN Computation on Encrypted Databases," Proc. 35th ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), pp. 139-152,2009