International Journal of Research in Advent Technology, Vol.2, No.12, December2014
E-ISSN: 2321-9637

Enhancing Security over Encrypted Cloud using Ranked Multi keyword Search

Runali S. Kumbhare
Department of Computer Engineering, Lokmanya Tilak College of engineering, Mumbai University, Navi Mumbai, India.

Abstract- In recent years due to captivating features of cloud computing most of the data owners are driven to outsourced their large amount of data from local site to commercial public cloud. Cloud provide great flexibility and economic saving but privacy and security of the sensitive data is a big concern to mitigate security, data have to encrypted before outsourcing in cloud computing. Hence, to enhance security searching data becomes a mandatory feature. To achieve search it is necessary to allow multiple keywords in the search request and produce search results in ranked order to retrieve the result accordingly. In this paper, we propose an efficient scenario for searching is "Edit distance algorithm". Also we choose a similarity matching called interrelate matching to identify the relevance of the data documents to the search query. Here we extend these schemas to support search semantics and verify the results with observation in search dataset in cloud computing.

Keywords- Cloud computing, enhancing security, keyword search, ranked result.

1. INTRODUCTION

Cloud computing is imagined as a dreamed vision of information technology in coming years where the user can store their large amount of information, data records and also enjoy easy access of data in the cloud. But in order to share this data in cloud pool, we must need to protect all data and enhance the security before outsourcing this complex data to the end users. Its great flexibility, scalability, robustness and high economic savings have driven many enterprises and individuals to store their data into cloud. Cloud contain all datasets, for example: e-mails, personal health records, photo albums, tax related files, financial transactions, secret data and many more which needs to be highly preserved in cloud.

It is difficult to search everything on cloud or to find the desired data. There is a possibility that data stored by data owners might possess threat by other owner in order to harm the stored data. For this reason it is mandatory to use certain mechanism to enhance privacy and security of data items.

On one hand, to ensure effectiveness and reliability, the search mechanism provide the schema of Multi keyword search in cloud which drives the cloud server to produce results in ranked order instead to providing undifferentiated mass of relevant data items. This ranking of documents will not only make the data clear but also avoid confusion among data users. As cloud is "pay-as-you-go" service, one need to make sure that whatever data one is searching must be relevant and useful to him. This Multi keyword scheme will show all the files related to the keyword by using a technique called as " edit distance" based on similarity and interrelate matching of words in cloud computing.

In single keyword search, the results produced were too very limited and not clear but Multi keyword search provides us withoptions and make the searching process easy and efficient. Here we define "interrelate matching" which is an efficient similarity measure which interrelates between the word entered by user and words present ineach datasets which is stored by the data owner into the cloud. It is more flexible for users to specify a ranked list of relevant documents.

2. PROBLEM FORMULATION

2.1. System model

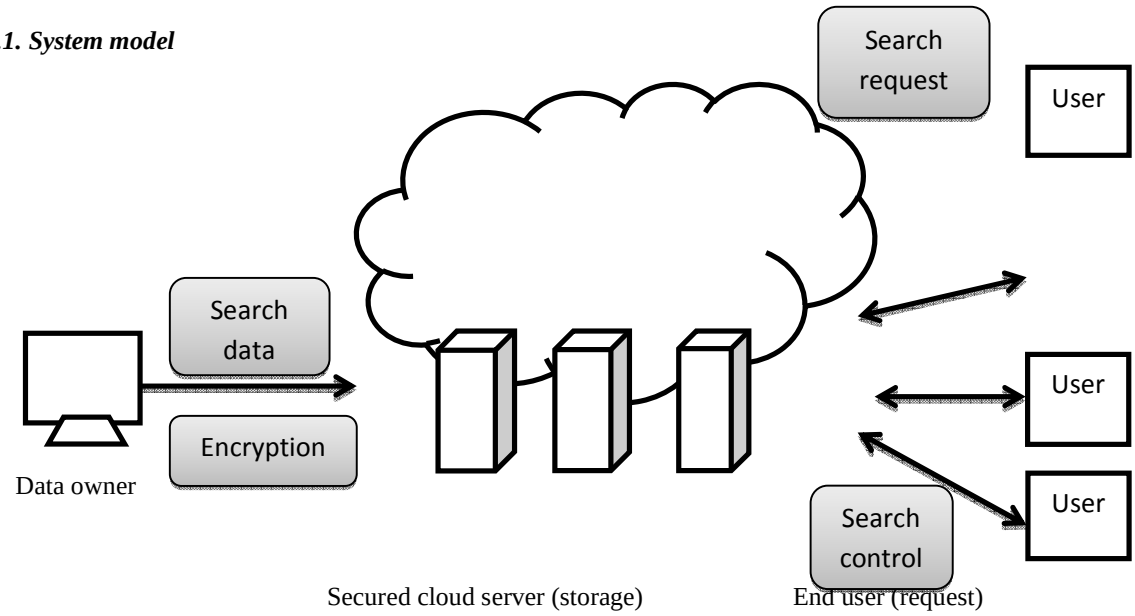

Fig 1: System architecture of search mechanism

Consider a system architecture which consists of three different entities as illustrated in fig.1. The data owner, the end user and the encrypted cloud server. The data owner contains set of documents which he wants to outsource in cloud server, the data owner needs to login himself in cloud to store his files and information in cloud server. Cloud sever: Once the storing of data is done in cloud by data owner, the data is highly preserved and encrypted so that no one can change, update, or retrieve the data. Even cloud server is not allowed to review the data, its only job is to store in it and provide security. End user: The data user will search for certain files in cloud server. For that the user will first enter the keyword of interest and submit it to cloud server. After requesting for files related to the keyword, the cloud server will search all those documents or files which contain the entered keyword and display them as a result to the user. The displayed list is ranked according its priority i.e. Files containing more number of words is displayed first.

2.2. Design goals

To enable search over encrypted cloud data and its effective utilization, the system should achieve high security and privacy as follows:

- Multi keyword search- It define the search which allows the end user to find results by entering the relevant keyword and receive a list of files as results in ranked order.
- Enhance security- The data stored in cloud must be highly secured so that no one other than owner can change, modify or retrieve it.
- Efficiency- The above goals of security and searching should be clear and achieved with low cost.

2.3. Interrelate Matching

It is a transitional similarity technique which uses number of query keywords present in the data stored in cloud to quantify the relevance of that document to the query. When users are aware about the exact dataset to be retrieved, queries perform well with search requirement as specified by the user. It is more convenient for the user to specify a list of keywords indicating of their interest and retrieve the most relevant documents with the rank order.

3. PROPOSED SYSTEM

3.1. Edit- distance Mechanism

In order to design a search mechanism, we propose a “edit-distance” mechanism which will find out similarity over search keywords. It is the way of quantifying how dissimilar two strings are to one another by counting minimum number of operations

required to transform one string to another It is a string metric for measuring the difference between two sequences. Informally, this distance between two words is the minimum number of single-character edits (i.e. insertions, deletions or substitutions) required to change one word into the other.

3.2. Applications

In approximate string matching, the objective is to find matches for short strings in many longer texts, in situations where a small number of differences are to be expected. The short strings could come from a dictionary, for instance. Here, one of the strings is typically short, while the other is arbitrarily long. This has many uses forinstance: spell checkers, correction systems for optical character recognition, and software to assist natural language translation based on translation memory.

It can also be computed between two longer strings, but the cost to compute it, which is roughly proportional to the product of the two string lengths, makes this impractical. Thus, when used to aid in fuzzy string searching in applications such as record linkage, the compared strings are usually short to help improve speed of comparisons.

3.3. Formal definitions

Given two strings a and b on alphabet $\sum$, the edit distance d(a,d) is series of edit operations that transforms a into b. Insertion: insertions of single symbol. If u=uv, then inserting the symbol x produces uxv. Deletion: deleting a single changes uxv to uv. Substitution: substituting a single symbol x for a symbol $y \neq x$ changes uxv to uyv ($x \rightarrow y$). Here each of the operation has unit cost hence efficiency is maintained.

4 PERFORMANCE ANALYSIS

In this section, we analyze the search mechanism using edit distance technique over Multi keyword search scenario. Performance is evaluated in order to ensure high security and retain privacy of data in cloud.

5. RELATED WORK

5.1. Fuzzy keyword search over- In this we greatly enhance system usability by returning results that exactly match to user's keyword or input. Here the closest possible match is also considered but undifferentiated results are produced to the user.

5.2. Single keyword search- it supports single keyword search where user enter a single keyword and exact keyword related data isdisplayed in ranked manner. Here, multiple and partially matched words are not considered.

6. CONCLUSION

Here, we have enhanced the security feature of Multi keyword searching over encrypted cloud domain. We use interrelate matching due to which as many match as possible can be determine. We proposed efficient similarity searchable scheme which enable a highly secure environment and preservers privacy during search mechanism in encrypted cloud.

REFRENCES

[1]N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. 829-837, Apr, 2011.

[2] 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, Mar. 2010.

[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] Efficient Similarity Search over Encrypted Data Mehmet Kuzu, Mohammad Saiful Islam, Murat Kantarcioglu Department of Computer Science, The University of Texas at DallasRichardson, TX75080,USA