Improving the Efficiency of Data Retrieval in Secure Cloud

Nisha T.M. and Lijo V.P.

MES College of Engineering, Kuttippuram
Kerala, India
nishunni@gmail.com

Abstract. Cloud computing allows much more efficient computing by centralizing storage, memory, procession and bandwidth. The data is stored in off-premises and accessing this data through keyword search. Traditional keyword search was based on plaintext keyword search. But for protecting data privacy the sensitive data should be encrypted before outsourcing. One of the most popular ways is selectively retrieve files through keyword-based search instead of retrieving all the encrypted files back. Present methods are focusing on the fuzzy keyword search and which efficiently search and retrieve the data in most secure and privacy preserved manner. The existing system uses single fuzzy keyword searching mechanism. A conjunctive/sequence of keyword search mechanism will retrieve most efficient and relevant data files. The conjunctive/sequence of keyword search automatically generates ranked results so that the searching flexibility and efficiency will be improved.

Keywords*:* fuzzy keyword, conjunction keyword, sequence keyword, edit distance, wildcard method.

1 Introduction

Cloud computing is so named because the information being accessed from a centralized storage, and does not need any user to be in a specific place to access it. This is a method in which information is delivered and resources are retrieved by web-based tools and its applications, rather than a direct connection to a server.

There are so many issues in storing the data securely in the cloud because most of the sensitive information is centralized in to clouds. But the most important aspect arises in data retrieval part. The data owner stores their data in the cloud and any authorized person can access those files. The cloud server provides the authorization, otherwise on retrieval they can do modification, insertion and deletion in the original files and can store back in clouds. So the original data can be mishandled, which may cause security problems. So here encryption plays an important role. That is these sensitive data are encrypted before outsourcing.

One of the most popular ways or techniques is to selectively retrieve files through keyword based search instead of retrieving all the encrypted files back. The data encryption also demands the protection of keyword privacy since keywords usually

V.V. Das and J. Stephen (Eds.): CNC 2012, LNICST 108, pp. 238–241, 2012.

contain important information related to the data files. The existing searchable encryption techniques do not suit for cloud computing scenario because they support only exact keyword search. This significant drawback of existing schemes signifies the important need for new methods that support searching flexibility, tolerating both minor types and format inconsistencies. The main problem is how efficiently searching the data and retrieves the results in most secure and privacy preserving manner. The existing system is mainly focusing on the 'fuzzy keyword search' method. The data that is outsourced is encrypted, constructs fuzzy sets based on both wild card technique and gram based technique, and also introduced a symbol-based trie-traverse search scheme [2, 4], where a multi-way tree was constructed for storing the fuzzy keyword set and finally retrieving the data.

2 Related Works in Keyword Search Mechanism

The main challenges are security in data storage, searching, data retrieval etc. Here it is mainly focusing on the data searching and retrieval part. Traditional encryption techniques support only exact keyword search. This technique is insufficient, because it will retrieve the data, only if the given keyword matches for the predefined keyword set. So for increasing the flexibility in searching so many new searching techniques were introduced.

S. Ji, proposed a new computing paradigm, called *interactive, fuzzy search* [3]. This uses 'Straight forward method' for keyword construction. It gives the idea about the queries with a single keyword. Here the keyword set construction needs more space for storing the keywords. So in order to reduce the space complexity J. Li, [4, 6] proposed another fuzzy keyword search method which includes 'Wild-card' [2, 4, 6] based method and 'Gram based' [2, 4, 6] method for constructing fuzzy keyword sets, a 'symbol-based trie-traverse search scheme' [2,6] for data retrieval. Philippe Golle, proposed a Secure Conjunctive Keyword Search [1], which gives a clear idea about the conjunction of the keyword. By introducing the conjunction of keywords the relevancy will be increase. That is the efficiency will be increasing and generate ranking automatically. Xin Zhou, proposed a Wild card search method for Digital dictionary based on mobile platform [2]. It gives a good idea about the Wild card method, and trie tree approach which reduces the search range largely, which includes the fuzzy pointer field, and also gives the idea for the process of inserting a word in the tree.

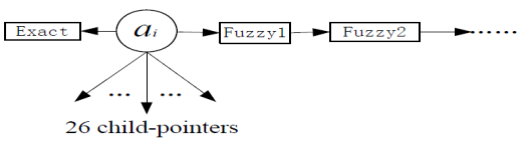

Fig. 1. A Node of Advanced Trie-Tree

Figure1. shows the node of an advanced trie-tree [2]. It consists of nodes, in which each node has 26 child pointers. In the proposed work the existing trie-tree method is replaced with the advanced trie-tree method. Here two fields namely exact index and fuzzy pointers are added in to the node. The exact index's value contains the offset of the word's entry in the dictionary, which contain all the information about the word. The fuzzy pointer is a pointer to record a list of keyword's offsets in the dictionary.

3 Fuzzy Keyword Search

The working of fuzzy keyword search is described here. Initially given *'n'* encrypted data, which is stored in the cloud server along with that an encrypted predefined set of distinct keywords are also stored in the cloud server. Cloud server provides the authorization to the users who want to access the encrypted files. Only the authorized person can access the stored data. When a user search for a request the cloud server maps the request to the data files, which is indexed by a file ID and is linked to a set of keywords. Fuzzy keyword search will return the results by keeping the following two rules.

1. If the user's searching input exactly matches the predefined keyword, then the server is expected to return the files containing that keyword.
2. If there is no exact match or some inconstancies in the searching input, the server will return the closest possible results based on pre-defined similarity semantics.

Here it uses 'Wild-card' based method and 'Gram based' method for constructing fuzzy keyword sets, a 'symbol-based trie-traverse search scheme' where a multi-way tree was constructed for storing the fuzzy keyword set and finally retrieving the data. This greatly reduces the storage and representation overheads. It also exploits 'Edit distance' to quantify keywords similarity, to build storage- efficient fuzzy keyword sets to facilitate the searching process.

In this method first we implemented the single fuzzy keyword search mechanism. Wild card method and Gram based method are used for fuzzy keyword set construction, Edit distance for similarity measure and normal trie-tree for data retrieval. Then introduced conjunction of keyword or sequence of keywords (AND, OR, NOT, BOTH) in the existing method, so that we can get an idea of the difference between single fuzzy keyword search and sequence of fuzzy keyword search. Here the normal trie-tree is replaced by advanced trie-tree, which contains two fields called exact index and fuzzy pointers. The exact index's value contains the offset of the word's entry in the dictionary, which contain all the information about the word. The fuzzy pointer is a pointer to record a list of keyword's offsets in the dictionary. The root node contain the sequence of words AND, OR, NOT, BOTH.

Performance: This will introduce low overhead and on computation and communication. Low computation means the construction of keyword index should be less compared to the existing system. Low overhead on communication means which will retrieve the data where all the conjunctive fuzzy keyword should match. So it will retrieve the exact files that are searched for. A ranked result will enhance the performance for data retrieval. Ranked search greatly enhances system usability

by returning the matching files in a ranked order given in the keyword sequence regarding to certain relevance criteria like keyword frequency. Here a ranked list is generated for the conjunctive/sequence of keywords, and retrieve the data according to that ranking. It is highly secure, privacy preserving and efficient.

In wild card method for a key word 'SIMPLE' with the pre-set edit distance 1, the total number of variants constructed is 13+1. In general for ***'ℓ'*** number of words the number of variants constructed are **2*ℓ +1+1.** So for conjunction of keywords having ***'n'*** number of keywords ***n*** **(2*ℓ +1+1)** variants are constructed. For gram based method for the key word 'SIMPLE' with the pre-set edit distance 1, the total number of variants constructed is **ℓ+1.** So for conjunction of keywords having ***'n'*** number of keywords ***n*** **(ℓ +1)** variants are constructed. The complexity of searching will reduce from **O(m*n*l)** in to **O(m*n),** where ***'m'*** is the length of hash value. The space complexity is reduced from **O(m*p*n*l)** in to **O(m*n*l).**

Acknowledgments. I take this opportunity to convey my deep and sincere thanks to our Principal Dr. V. H. Abdul Salam and Head of the Department Dr.P. P. Abdul Haleem. I also extend my deep gratitude to my guide Mr. Lijo V P who gave the complete guidance and support in presenting the project. I express my sincere gratitude to all the staff of Computer Science & Engineering Department and my beloved family members who helped me with their timely suggestions and support. I also express my sincere thanks to all my friends who helped me throughout the successful completion of the work .All glory and honor be to the Almighty God.

References

1. Golle, P., Staddon, J., Waters, B.: Secure Conjunctive Keyword Search over Encrypted Data. In: Jakobsson, M., Yung, M., Zhou, J. (eds.) ACNS 2004. LNCS, vol. 3089, pp. 31–45. Springer, Heidelberg (2004)
2. Zhou, X., Xu, Y., Chen, G., Pan, Z.: A New Wild- card Search Method for Digital Dictionary Based on Mobile Platform. In: Proceedings of the 16th International Conference on Artificial Reality and Telexistence Workshops, ICAT 2006. IEEE (2006)
3. Ji, S., Li, G., Li, C., Feng, J.: Efficient interactive fuzzy keyword Search. In: Proc. of WWW 2009 (2009)
4. Li, J., Wang, Q., Wang, C., Cao, N., Ren, K., Lou, W.: Enabling Efficient Fuzzy Keyword search over encrypted data in cloud computing. In: Proc. of IEEE INFOCOM 2010 Mini-Conference, San Diego, CA, USA (March 2009)
5. Kaufman, L.M.: Data Security in the World of Cloud Computing. IEEE Security and Privacy 7(4), 61–64 (2009)
6. Li, J., Wang, Q., Wang, C., Cao, N., Ren, K., Lou, W.: Fuzzy keyword search over encrypted data in cloud computing. In: Proc. of IEEE INFOCOM 2010 Mini-Conference, San Diego, CA, USA (March 2010)
7. Wang, C., Cao, N., Li, J., Ren, K., Lou, W.: Secure ranked key word search over encrypted cloud data. In: Proc. of ICDC 2010 (2010)
8. Cao, N., Wang, C., Li, M., Ren, K., Lou, W.: Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data. In: Proc. of IEEE INFOCOM 2011 (April 2011)
9. Boneh, D., Waters, B.: Conjunctive, Subset, and Range Queries on Encrypted Data. In: Vadhan, S.P. (ed.) TCC 2007. LNCS, vol. 4392, pp. 535–554. Springer, Heidelberg (2007)