ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Efficient Central Keyword Based Search Method over Encrypted Data in Cloud

[1] A.Ramya,[2] D.udhaya chandrika,[3] S.Sureka, [4]R.Vinothini
[1] Assistant Professor, [2][3][4]UG Student
[1][2][3][4] Department of CSE, VSB Engineering College, Karur

***Abstract* – Keyword based search is a important aspect, when searching the data in cloud. Keywords may have a certain grammatical relationship among them which reflect the importance of keywords from the user's perspective intuitively. Proposed system has relationship among query keywords into consideration and designs a keyword weighting algorithm to show the importance of distinction of the keywords. Key word weighting algorithm accurately and efficiently localizes the central keyword that the user is interested in. We can choose the central keyword (not all keywords) of the query to extend. When a user inputs some query keywords, our scheme can effectively and accurately locate and extend the semantics of the central keyword. The returned results should be relevant to both the multiple keywords that the user inputs and the extension keyword. To calculate the relevance scores between keywords and files, we use the widely used TF-IDF rule, where TF (term frequency) denotes the frequency of a given keyword in a document and IDF (inverse document frequency) represents the importance of a keyword. When the data owner updates the dataset, the TFIDF values are also changed. To enable updating, we make a few changes in the trapdoor and index generation by inserting the IDF values into the query vector and the TF values into the index vector, respectively.**

Index Terms— Trapdoor, TF-IDF rule, MRSE, weighting algorithm, central keyword.

INTRODUCTION

Cloud computing is a popular platform because of more people are inclined to outsource their data to the cloud, due to its flexibility and unlimited resources. Enable effective searches over encrypted data, the data owner first builds an encrypted index based on the extracted keywords from data files and the corresponding index-based keyword matching algorithm, and then outsources both the encrypted data and the index structure to the cloud server. To search over the encrypted files, the cloud server integrates the trapdoors of keywords with the index information and finally returns the target files to the data users. Proposed system takes the relationship among query keywords into consideration and designs a keyword weighting algorithm to show the importance of distinction of the keywords. Using the keyword weights, we can accurately and efficiently localize the central keyword that the user is interested in. TF-IDF rule is introduced in our design to analyze relevance between the query and files. Choosing the central keyword (not all keywords) of the query to extend, our scheme can greatly reduce the trapdoor generation time.

EXISTING SYSTEM

Existing systems take the semantic relationship among query keywords into consideration. They regard the input keywords as independent and irrelevant. In fact, the importance of a keyword is quite different from that of the others, and this can be clearly shown by the semantic relations among the query keywords.

Another problem is that the existing keyword-based search techniques can only return files that contain the exact query keyword. Although some mechanisms like semantic or extension searches have been designed to solve this problem, they all require additional computations at the client side and thus introduce extra overhead.

If the user queries terms semantic relevant to the index terms, it is possible that none or only a few matched results will be returned. As a result, the user must perform more query operations to obtain the desired files. To address this issue, a common technique is query extension, which can extend original query terms according to certain rules before submitting a search request. Also provide multi-keyword search scheme by using the MDB-tree to construct the index structure which greatly reduces the search complexity, where the cloud server only needs to search over a part of the tree. Multi-keyword ranked search scheme (MRSE) which used coordinate matching to realize the ranked search.

Disadvantages

• Unable to hit the files which contain semantic-relevant keywords.

• More computation overhead at the client side.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

PROPOSED SYSTEM

Proposed system has following entities data owner, the data user and the cloud server. The data owner has a set of data files and wants to outsource it to the cloud. As these data files may contain sensitive information, the data owner encrypts the data before outsourcing due to privacy concerns.

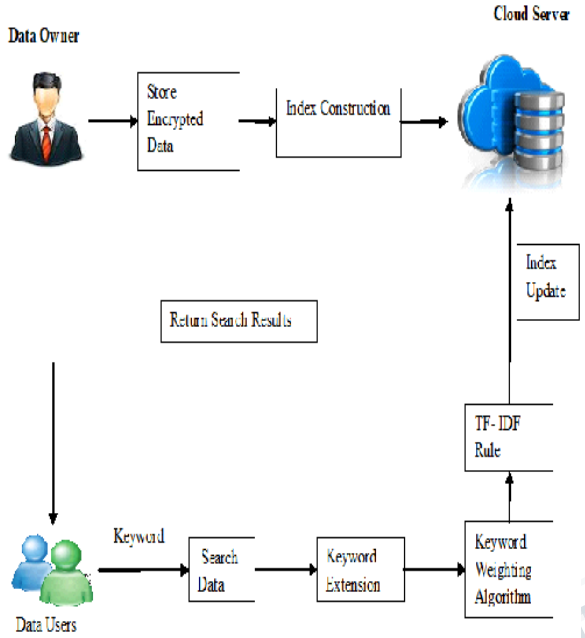
To facilitate the efficient use of these encrypted data files, the data owner needs to build a searchable encrypted index I based on the keyword set W extracted from F. The index will then be outsourced to the cloud server along with the corresponding encrypted files. The data user is authorized by the data owner and searches the outsourced data files stored on the cloud via some input keywords. Based on these keywords, the user chooses the central keyword to extend, computes its trapdoor T and sends it to the cloud server. The cloud server stores the encrypted data files and index, and also handles search requests from the data user. Upon receiving the trapdoor T generated by an authorized user, the cloud server then searches over the index I and returns the top-K relevant files as the search results to the user.

Proposed system takes the relationship among query keywords into consideration and designs a keyword weighting algorithm to show the importance of distinction of the keywords. Using the keyword weights, we can accurately and efficiently localize the central keyword that the user is interested in. TF-IDF rule is introduced in our design to analyze relevance between the query and files. Choosing the central keyword (not all keywords) of the query to extend, our scheme can greatly reduce the trapdoor generation time. Proposed framework has following substances information proprietor, the information client and the cloud server. The information proprietor has an arrangement of information records and needs to outsource it to the cloud. The index will then be outsourced to the cloud server along with the corresponding encrypted files. The data user is authorized by the data owner and searches the outsourced data files stored on the cloud via some input keywords

Advantages

•Accurately locate and extend the semantics of the central keyword.
•Return search result based on multiple keywords.
•TF-IDF used to resolve the index problem.

Proposed System Architecture



Index Construction

Build Index(K, C)
1. Initialization: i) scan C and extract the distinct words W = (w1, w2, ..., wm) from C. For each wi ∈ W, build F(wi);
2. Build posting list:
i) for each wi ∈ W
• for $1 \le j \le |F(wi)|$:
a) calculate the score for file Fij according to equation 2, denoted as Sij ;
b) compute Ez(Sij), and store it with Fij 's identifier (id(Fij)||Ez(Sij)) in the posting list I(wi);
3. Secure the index I:
i) for each I(wi) where $1 \le i \le m$:
• encrypt all Ni entries with padding 0 s, (0 ||id(Fij)||Ez(Sij)), with key fy(wi), where $1 \le j \le v$.
• set outstanding $v - Ni$ passages, assuming any, to irregular estimations of an indistinguishable size from the current Ni sections of I(wi) and replace wi with πx(wi);
4. Output I

CONCLUSION

In proposed work, efficient search technique is constructed to improve the search results. The relationship among the query keywords into consideration and designed a keyword weighting algorithm based on the

relations. We also designed a central keyword semantic extension scheme according to the keyword weights. By choosing the central keyword instead of not all the keywords to extend, our scheme achieves a tradeoff between functionality and efficiency. To express the relevance between the query and the files better, we introduced the TF-IDF rule when building the trapdoor and index. By storing the IDF value in the dictionary, our scheme can support updates for adding new files.

REFERENCES

[1] S. Hu, Q. Wang, J. Wang, Z. Qin, and K. Ren, "Securing sift: Privacypreserving outsourcing computation of feature extractions over encrypted image data," IEEE Trans. on Image Processing, vol. 25, no. 7, pp. 3411–3425, 2016.

[2] Z. Fu, J. Shu, X. Sun, and N. Linge, "Smart cloud search services: verifiable keyword-based semantic search over encrypted cloud data," IEEE Transactions on Consumer Electronics, vol. 60, no. 4, pp. 762– 770, 2014.

[3] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009, pp. 139–152.

[4] http://nlp.stanford.edu/software/lex-parser.shtml.

[5] http://nlp.stanford.edu:8080/parser/.

[6] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to wordnet: An on-line lexical database," International journal of lexicography, vol. 3, no. 4, pp. 235–244, 1990.

[7] I. H. Witten, A. Moffat, and T. C. Bell, Managing gigabytes: compressing and indexing documents and images. Morgan Kaufmann, 1999.

[8] J. Zobel and A. Moffat, "Exploring the similarity space," in ACM SIGIR Forum, vol. 32, no. 1. ACM, 1998, pp. 18–34.

[9] H. Li,Y. Yang, T. H. Luan, X. Liang, L. Zhou, and X. S. Shen, "Enabling fine-grained multi-keyword search supporting classified sub-dictionaries over encrypted cloud data," IEEE Transactions on Dependable and Secure Computing, vol. 13, no. 3, pp. 312–325, 2016.

[10] Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, "Enabling personalized search over encrypted outsourced data with efficiency improvement," IEEE transactions on parallel and distributed systems, vol. 27, no. 9, pp. 2546–2559, 2016.

[11] D. Lin et al.," An information-theoretic definition of similarity." in Icml, vol. 98, no. 1998, 1998, pp. 296–304.

[12] T. Pedersen, S. Patwardhan, and J. Michelizzi, "Wordnet:: Similarity: measuring the relatedness of concepts," in Demonstration papers at HLTNAACL 2004. Association for Computational Linguistics, 2004, pp. 38–41.

[13] Z. Xu, W. Kang, R. Li, K. Yo w, and C.-Z. Xu, "Efficient multi-keyword ranked query on encrypted data in the cloud," in Parallel and Distributed Systems (ICPADS), 2012 IEEE 18th International Conference on. IEEE, 2012, pp. 244–251.

[14] Z. Fu, X. Sun, N. Linge, and L. Zhou, "Achieving effective cloud search services: multi-keyword ranked search over encrypted cloud data supporting synonym query," IEEE Transactions on Consumer Electronics, vol. 60, no. 1, pp. 164–172, 2014.

[15] J. Yu, P. Lu, Y. Zhu, G. Xue, and M. Li, "Toward secure multikeyword top-k retrieval over encrypted cloud data," IEEE transactions on dependable and secure computing, vol. 10, no. 4, pp. 239–250, 2013.

[16] Request For Comments Database, http://www.ietf.org/rfc.html.

[17] Z. Xia, X. Wang, X. Sun, and Q. Wang, " A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 2, pp. 340–352, 2016.

[18] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003.

[19] Z. Fu, X. Sun, Q. Liu, L. Zhou, and J. Shu, "Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing," IEICE Transactions on Communications, vol. 98, no. 1, pp. 190–200, 2015.

[20] B. Wang, W. Song, W. Lou, and Y. T. Hou, "Inverted index based multi-keyword public-key searchable encryption with strong privacy guarantee," in

Computer Communications (INFOCOM), 2015 IEEE Conference on. IEEE, 2015, pp. 2092–2100.

[21] F. Cheng, Q. Wang, Q. Zhang, and Z. Peng, "Highly efficient indexing for privacy-preserving multi-keyword query over encrypted cloud data," in International Conference on Web-Age Information Management. Springer, 2014, pp. 348–359.

[22] B. Samanthula, W. Jiang, and E. Bertino, "Privacy-preserving complex query evaluation over semantically secure encrypted data," Computer Security-ESORICS 2014, 2014.

[23] C. Liu, L. Zhu, L. Li, andY. Tan, "Fuzzy keyword search on encrypted cloud storage data with small index," in Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on. IEEE, 2011, pp. 269–273.

[24] M. Chuah and W. Hu, "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data," in Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference on. IEEE, 2011, pp. 273–281.

[25] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in Data Engineering (ICDE), 2012 IEEE 28th International Conference on. IEEE, 2012, pp. 1156–1167.