Keywords Classification Technique Supporting Sub-Dictionaries over Encrypted Cloud Data

Tejaswini S[1], Vartika Sharma[2], Syed Thouheed Ahmed[3]

[1]*IV Semester M.Tech Student, Computer Science & Engineering, GSSSIETW, Mysuru*
[2]*Assistant Professor, Dept. of CSE, GSSSIETW, Mysuru*
[3]*Sr. Research Engineer, ThinkSoft Research and Information Technologies, B'lore*

Corresponding Email: **thejaswi.naik@gmail.com**

ABSTRACT

Cloud computing is considered as a utility which allows the users to store their data on to the cloud servers and therefore allow data access to the users through the servers. This uploaded data to the cloud contains privacy information and hence this leads to the necessity of encrypting the data before they are outsourced to the servers. This leads to difficulty of searching since the data is encrypted and to solve this issue searchable encryption is introduced. The objectives of this paper are: First, to introduce the preference and relevance factors in order to enable best keyword search and to improve user experience. Second, to apply logic operations of the keywords which are AND, OR and NO operations. Third, to achieve better search efficiency by building index using sub-dictionary technique. The document or the data retrieved using the keywords can later be downloaded using the key which will be provided by the data owner through email.

Keywords: cloud computing, encryption, logic operations, relevance factor

INTRODUCTION

The cloud computing is considered as a utility where the users can store their data on to remote cloud servers namely data outsourcing and allow data access to public users. This provides a low-cost, stable and more scalable way of the data access because the cloud servers are highly scalable and efficient, thus favoring small enterprises. The outsourced data are likely to contain privacy information and this leads to the necessity of encrypting the data documents. This limits the usability of the data due to difficulty of search over the encrypted data and this issue can be solved by searchable encryption [1], [2], [3].

In searchable encryption, the data owner generates keywords according to the outsourced data and these keyword are encrypted before they are outsourced to the server. When a data user needs access to some outsourced data, the user selects some relevant keywords and the ciphertext of the selected keywords will be sent to the cloud server. The cloud server retrieves the result by matching the outsourced encrypted keywords with the encrypted keywords of documents already present in the server.

RELATED WORK

Prof. C R Barde, et al., [4], proposes a secured search scheme for multi keyword search which uses the computation of the product to improve the privacy requirements. According to this paper, when any un-authorized user is trying to access the cloud data an alert can be sent in the form of message or email. The methodology adapted is a secure nearest neighbor technique which works according to the principle of matching of co-ordinates in order to find the similarity among the query and the documents.

Hongwei Li, et al., [5], proposes a searchable encryption scheme for multi-keyword ranked search over the storage data. Specifically, by considering the large

number of outsourced documents in the cloud, the relevance score and *k*-nearest neighbor techniques are used to develop an efficient multi-keyword search scheme that can return the ranked search results based on the accuracy. Within this framework, to further improve the search efficiency, the blind storage system is adopted to conceal access pattern of the search user.

Jiadi Yu, et al., [6], focuses on addressing the data privacy issues using Searchable Symmetric Encryption. The server-side ranking based on order-preserving encryption (OPE) inevitably leaks data privacy and to eliminate the leakage, two-round searchable encryption (TRSE) scheme is proposed that supports top-k multi keyword retrieval. In TRSE, vector space model and homomorphic encryption are applied. The vector space model helps to provide sufficient search accuracy, and the homomorphic encryption enables users to involve in the ranking while the majority of computing work is done on the server side by operations only on ciphertext.

PROPOSED SYSTEM

The system consists of three entities namely data owner, cloud server and data user/customer. The architecture diagram for the proposed system is given in figure 1.

Dataowner: The data owner is responsible for outsourcing the data to the cloud in order to provide convenient data access to search users. To protect the data encryption will be done using symmetric encryption. The keywords from the outsourced document will be generated to improve search efficiency. The index will be created using keywords and the secret key. After that, encrypted documents and the indexes will be sent to the cloud by the data owner. The secret key and symmetric key will be sent to the search user.

Cloudserver: The cloud server stores the encrypted documents and indexes that it received from owner of the data. The cloud server provides services and data access to data user. When the data user sends keywords to the cloud server, the matching documents will be retrieved based on operations.

Datauser: A search user sends queries to the cloud server with the following steps. First, the symmetric key and secret key is received from the owner of the data. Second, the search is performed using the keywords and trapdoor generated from those keywords will be sent to the cloud server. Last, the matching documents from the collection of documents will be received from the server and it can be decrypted using symmetric key to obtain readable text.

EXPERIMENTS AND RESULTS

Figure 2 gives the screenshot for the home page and figure 3 shows the new user screen where the new user can register by providing the information and data owner screenshot is shown in figure 4 where the data owner logs in to outsource the data file to the cloud server.

REFERENCES

[1] D. X. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in *Proceedings of S&P*. IEEE, 2000, pp. 44–55.

[2] R. Li, Z. Xu, W. Kang, K. C. Yow, and C.-Z. Xu, “Efficient multi keyword ranked query over encrypted data in cloud computing,” *Future Generation Computer Systems*, vol. 30, pp. 179–190, 2014

[3] H. Li, D. Liu, Y. Dai, T. H. Luan, and X. Shen, “Enabling efficient multi-keyword ranked search over encrypted cloud data through blind storage,” *IEEE Transactions on Emerging Topics in Computing*, 2014.

[4] Prof. C. R. Barde, Pooja Katkade, DeepaliShewale, RohitKhatale “Secured Multiple-keyword Search over Encrypted Cloud Data” *International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 2, February 2014).*

[5] Hongwei Li, Dongxiao Liu, Yuanshun Dai1, Tom H. Luan2, and Xuemin (Sherman) Shen “Enabling Efficient Multi-Keyword Ranked Search over Encrypted Mobile Cloud Data through Blind Storage” *IEEE transactions on emerging topics in computing, 6 march 2015.*

[6] Jiadi Yu, Peng Lu, Yanmin Zhu, GuangtaoXue, Member, and Minglu Li

"Toward Secure Multikeyword Top-k Retrieval over Encrypted Cloud Data" IEEE *Transactions On Dependable And Secure Computing, Vol. 10, No. 4, July/August 2013.*

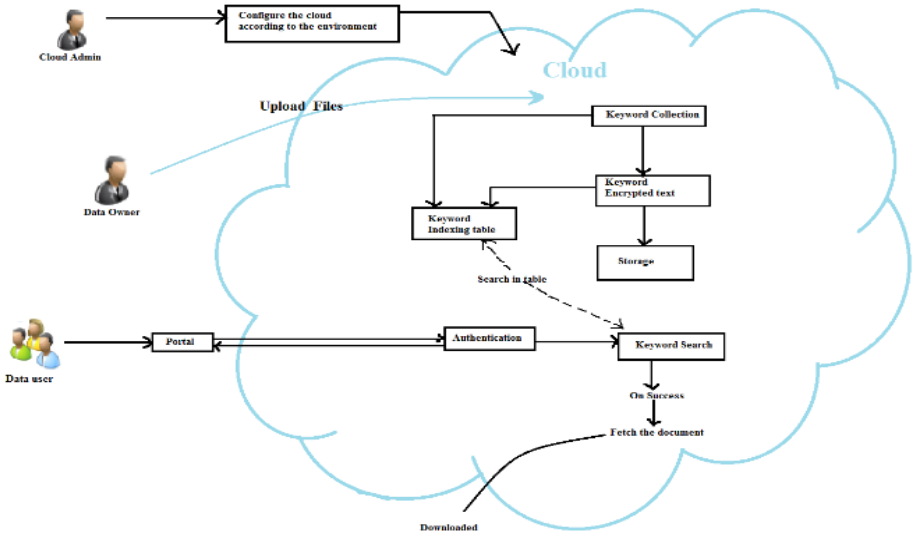

Figure 1: Architecture Diagram

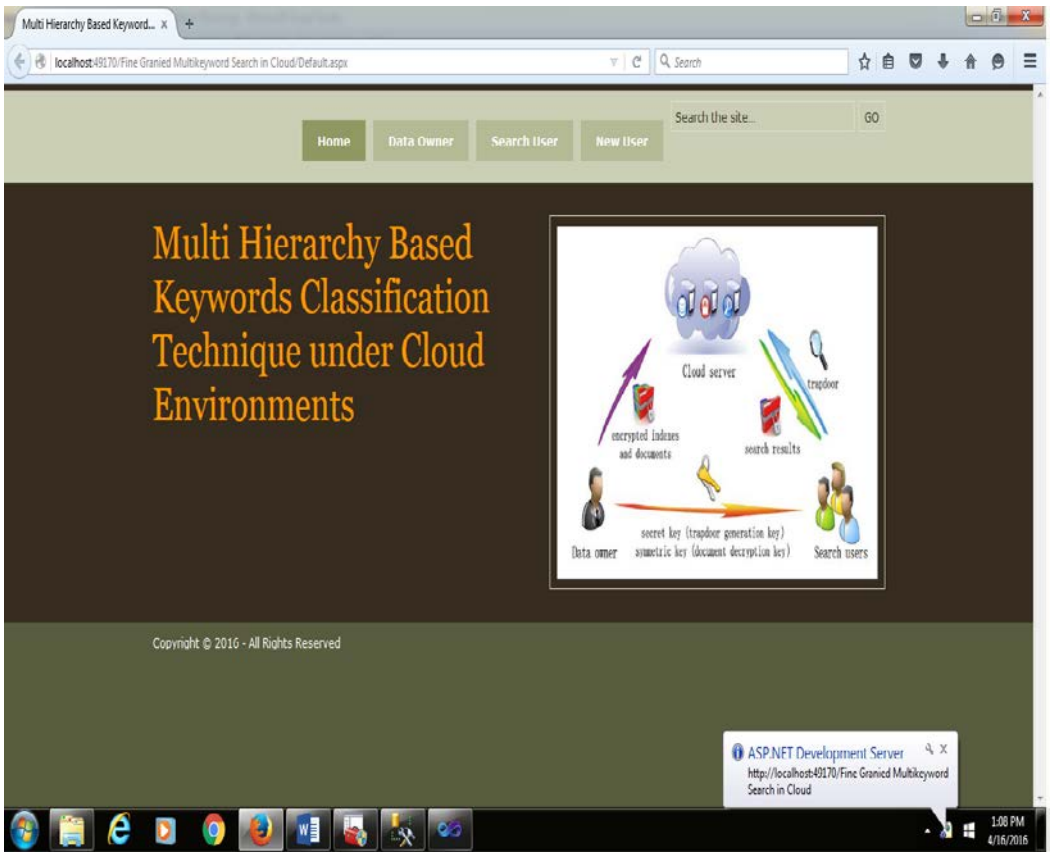

Figure 2: Home screen

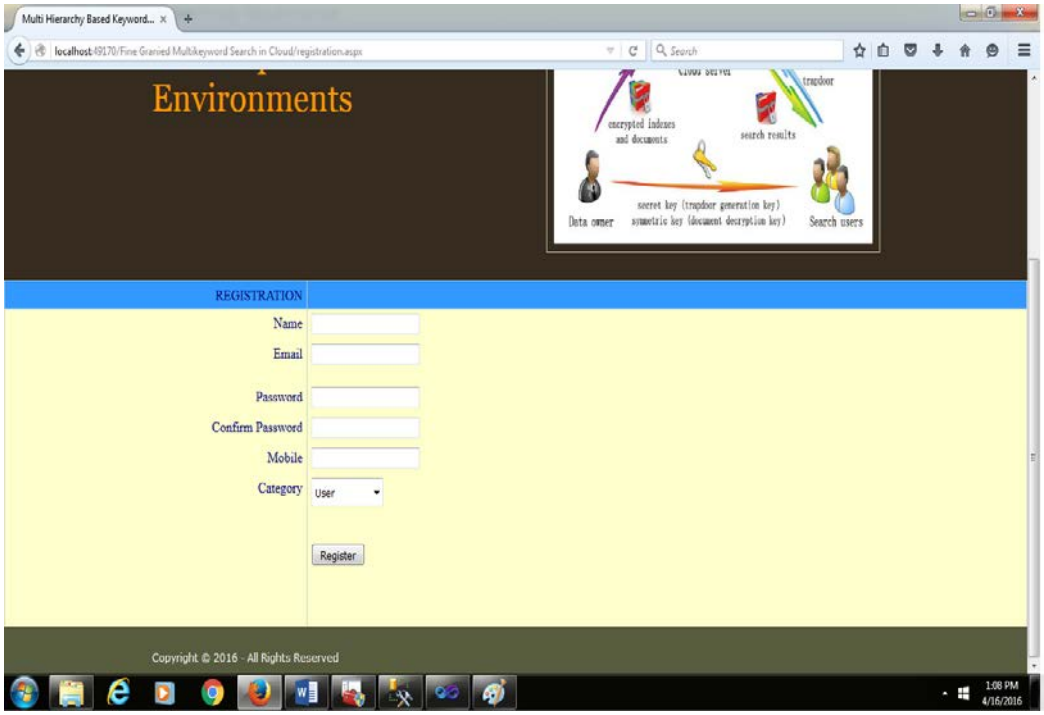

Figure 3: New user registration form

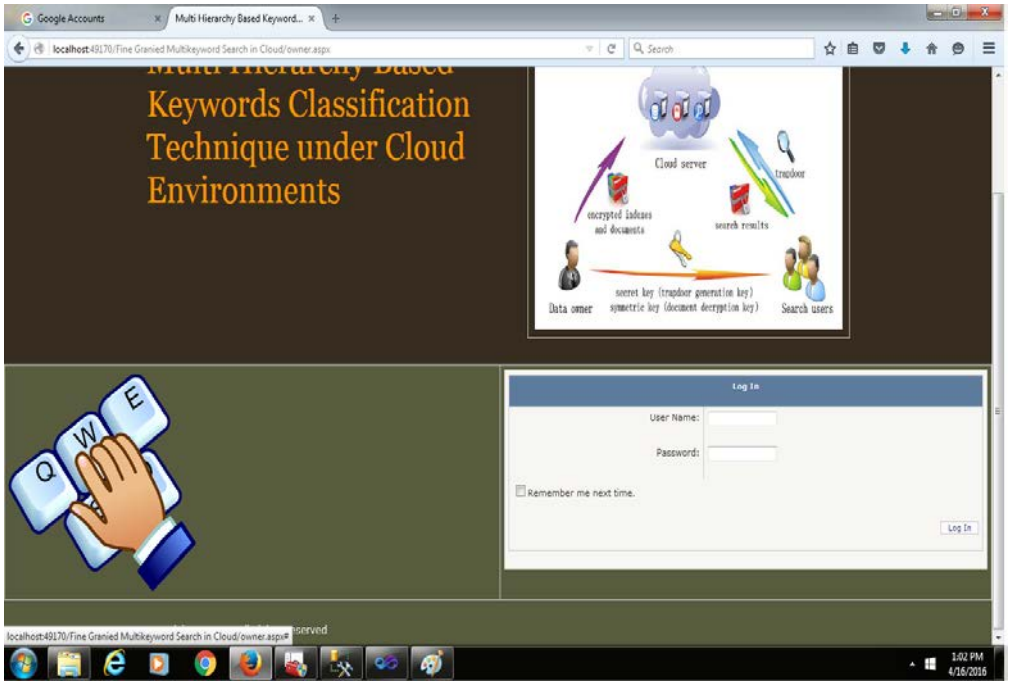

Figure 4: Data owner screen