VOL 1 (2017) NO 2

e-ISSN : 2549-9904
ISSN : 2549-9610

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

# Encrypted Arabic Documents Search Model using Fuzzy Keywords in Cloud Computing

Nidal Hassan Hussein [#], Ahmed Khalid [*], Khalid Khanfar [**]

[#] *PhD. Program in Computer Science, Sudan University of Science and Technology, Sudan*
[*] *Department of Computer Science, Community College, Najran University, Najran, KSA*
[**] *Head of Information Security Department at Naif Arab University for Security, Saudi Arabia*

*E-mail: nidal276@hotmail,com*

***Abstract*— Cloud computing is a burgeoning and revolutionary technology that has changed how data are stored and computed in the cloud. This technology incorporates many elements into an innovative architecture. Among them are autonomic computing, grid computing, and utility computing. Moreover, the rapid storage of data in the clouds has an impact on the security level of organizations. The chief challenge of cloud computing is how to build a secured cloud storage. The reason for this difficulty is that before data transfer, data are usually encrypted in order to achieve a high utilization. Another real challenging task of cloud computing is how to apply a search over encrypted data. As many techniques support only exact keyword matches, we propose a model to search over encrypted data that are written in Arabic. If an exact keyword match fails, our model will approximate the file as a secondary result. Our model will also use a fuzzy keyword search to enhance system usability by obtaining matching result whenever users input exact matches or the closest possible matches based on keywords. To the best of our knowledge, our model is considered to be the first research work that applies fuzzy search over Arabic encrypted data.***

***Keywords*— fuzzy keyword, cloud computing, encryption, edit distance, multi cloud, wildcard, Arabic Fuzzy Search Scheme(AFSS)**

## I. INTRODUCTION

Cloud computing is a new method of technology that supports the vision of computing as a utility. [12] It helps provide a fast, efficient, convenient and reliable aggregation of computing resources to a centralized network. [13]

A summary of cloud computing advantage includes fewer risks of cyber hacking, on-demand self-service, the ubiquity of network access, location independent resource pooling, rapid resource elasticity, and affordability based on usage pricing.[13,14] Moreover, cloud computing can assist users to avoid large capital outlays in deployment and management of both hardware and software. As cloud computing can be protected with the right tools and expertise, sensitive information is centralized in the cloud. Such information includes email, government documents, personal health records, private photos and videos, and company finance information. With this technology, users can store their data in the cloud as far as data owners and cloud servers are in different trusted domain. However, if they are on the same network, there may be such a security breach that cloud servers can be hacked, thus leaking classified data to unauthorized entities. Also, sensitive data must be encrypted prior to outsourcing to ensure privacy and combating access to unauthorized information. [15]

Searching over encrypted data is the most popular and interesting technique in the cloud computing system. The main idea about this concept is that before transferring data to cloud servers, they need to be encrypted to ensuring maximum protection of important information. Besides, techniques which make use of multiple domains are required to design an effective search system over encrypted data [16]. In this paper, we propose an Arabic Fuzzy Search Scheme(AFSS) model. This model provides privacy, thus preserving keywords over encrypted Arabic data file. The main objective of AFSS model is to allow users to use Arabic fuzzy search over encrypted data and to enhance system usability. The rest of this paper is organized as follows: Section II focuses on the related work concerning cloud computing. Section III explores the problem statement. Section IV highlights the proposed model. Section V indicates the mathematical model design. The last section, Section V1, concludes the paper.

## II. Related Work

The word search over encrypted data was first proposed by Song et al. [3]. It helps users to obtain practical solutions on how to search problems and protect untrusted servers. Public key encryption with keyword search (PEKS) [5] was first proposed by Boneh in 2004. Since then, a keyword search problem has been divided into two parts: public model and private model. Many researchers have focused on cloud cryptography, especially on efficiency improvements and security definition [1,2,6].

One of the academic scholars is Goh [4]. He suggested how to improve the search of information on data files using indexes. Like Goh [4], Change et al. [7] and Curtmola et al. [8] proposed how to use a single encrypted hash table index to search for data. Their approach employs a codified and unique identifier for each data file containing the corresponding keywords. However, this method is not beneficial in cloud computing as the encrypted set of file identifiers only recognizes keywords. A better method is fuzzy search, and it helps users to find information effectively using string matching. Fuzzy search works using a formulated approach where n is the number of encoded data files (c= {f1, f2, f3,….fn}) transferred to and stored in the cloud server, w is the set of a particular keyword (W={w1, w2, w3,…wn}), d is the predefined edit distance, and (w, k) is the searching input. For instance, by assuming k<d, the system searches data files and display the keywords with the word w.

In real-life scenarios, the value of d can be different from a particular keyword, and if the matching is not successful, {FIDwi} will be returned and ed (w,wi)⩽min{k,d} will be satisfied.

To determine the level of firmness of the tow strings, a reliable method is to use the edit distance [10]. For instance, to complete the edit distance against a large dictionary, two words w1 and w2 are assumed. The number of processes required to change them from one form to another is the edit distance between w1 and w2, and the three primitive operations are substitution, deletion, and insertion.

## III. Problem Statement

Figure 1 shows an encrypted cloud data hosting service involving three base units. The data owner has a collection of data file written in Arabic F= {f1, f2, .....fn}. The user can also store the information on a cloud server in an encrypted form using standard symmetric algorithms such as AES. Another requirement is to ensure that it can search through the server. To obtain effective data utilization, the data owner needs to first build the searchable index before outsourcing. The particular keywords w={w1, w2,.....,wn}are identified from the collection of files F and stored in an encrypted form on the cloud sever. We assume that full authorization between the user and the cloud sever has been done if the authorized person wants to retrieve any vital information on the system. This authorization includes the encoded keywords or search words of the information on the server. The cloud server utilizes the requested keywords to search and return the corresponding set of file to the authorized user.

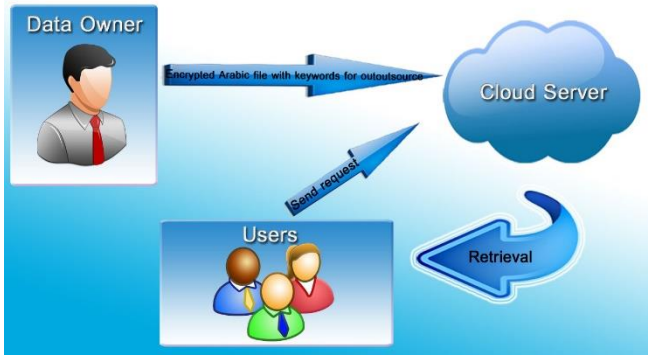

Figure1.Ranked keyword search in cloud model

## IV. The Proposed Model

Our proposed model AFSS consists of three units, which consist of many modules including data owner, data user, and cloud storage provider (CSP) (see Fig. 2).

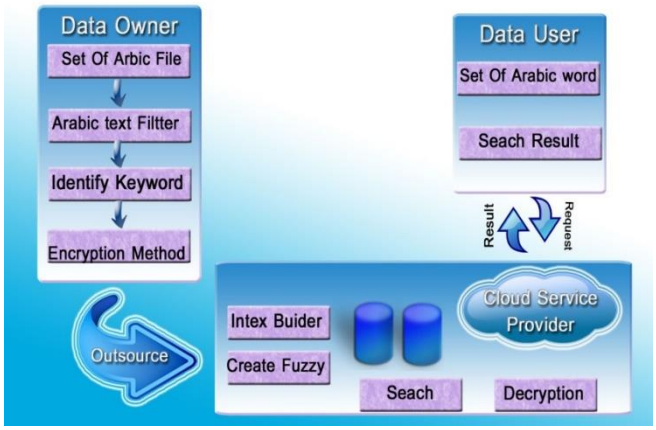

Figure 2: Architecture Arabic Fuzzy Search Scheme(AFSS)

The data owner has a collection of files written in Arabic and wishes to outsource them to the cloud server in an encrypted form with the support of secure ranked search. After the user has inputted a particular keyword on the system, the cloud server searches files for any words that match the search words using the calculated relevance score. The cloud server displays the result based on this calculated relevance score. The file data are stored on the server after the extracted keywords are encrypted with the CSP

## V. Mathematical Model Design: Notation

f denotes an Arabic file
I(fi) denotes an index for a file f whose order is i

$$W(f_i) = w_1, w_2, \dots, w_m) \quad (1)$$

denotes the keywords of the file fi.
the mathematical model for Arabic Fuzzy Search Scheme is as follows

$$AFSS = (F, KE(F), IB(F), F_z(w), Search(w)) \quad (2)$$

Where

$$F = < f_1, f_2, \dots, f_n > \quad (3)$$

denotes a collection of Arabic files

$$KE(F) = \sum_i^n W(f_i) \quad (4)$$

It's the total keywords for the collection of Arabic files.

$$IB(F) = \sum_i^n (W(f_i), I(f_i)) \quad (5)$$

Is the database for the keywords and indexfor the Arabic collection files,where I(fi) is the index of the file fi and the w(fi) is the keywords for the file fi.

$$F_z(w) = \{sub(*, ch(i).w)\} \quad (6)$$

where Sub(*,ch[(i).w)∈KE(F) is function that gives the fuzzy words for the keyword w by substituting the character its position i of the wordw by * whereI = 0 ..length[w]. For example, for the user keyword "شعاع" with preset edit distance 1, its fuzzy key words construct is as follows:

Sub(*,ch(i).شعاع)={*شعاع, ....., ش*اع, *عاع, *شعاع, شع*ع, شعاع } ∈KE

$$search(w) = \{f_1, f_2, \dots, f_m\} \quad (7)$$

Where w ∈ W($f_i$) ∩ KE(F) i=1, …, m

Considering the above mathematical model, the data owner has a collection of files written in Arabic and wishes to outsource these collections to the cloud server in encrypted form before outsourcing the se collections that are responsible for creating the keywords w(fi) for each file and encrypting the collections and the keywords. The user writes the search words and sends these encrypted words in the search tool to the cloud server before waiting for the appropriate results. The role of CPS is to receive the encrypted Arabic collection files and the keywords sent by the data owner. Then, for each Arabic file, the user creates the corresponding index I(fi) using the keywords W(fi). Afterward, the collection, keywords and the index are saved in the cloud. Whenever data users send their search request using keywords w, CPS receives the request and search for the w in the KE(F)if there is an exact match. After that, the corresponding files are retrieved and sent to the user using the index of these files; otherwise, CPS generates the set of fuzzy keywords using the function Fz(w)that is used to retrieve the closest possible matching files based on keyword similarity. Finally, the CSP decrypts files and sends these files to the user upon request.

## VI. CONCLUSION

In this paper, we introduce Arabic Fuzzy Keywords Search Model which deal with a collection of Arabic files outsourced in a cloud server. The model consists seven equations which will help to solve the problems of privacy for Arabic files and support the efficient fuzzy search of remotely stored encrypted data in cloud computing. The future work is concern on applying and evaluating the model

## REFERENCES


[1] Bao F, Deng R, Ding X, Yang Y. Private query on encrypted data in multi-user settings. Proc.Of ISPEC. 2008.

[2] Bellare M, Boldyreva A, O'Neill A. Deterministic and efficiently searchable encryption. Proceedings of Crypto 2007; 4622 of LNCS.Springer-Verlag.

[3] Song D, Wagner D, Perrig A. Practical techniques for searches on encrypted data. Proc. of IEEE Symposium on Security and Privacy. 2000.

[4] Goh EJ. Secure indexes, Cryptology ePrint Archive. http://eprint.iacr. org/. 2003/216.

[5] Boneh D, Crescenzo GD, Ostrovsky R, Persiano G. Public key encryption with keyword search. Proc. of EUROCRYP. 2004.

[6] Waters B, Balfanz D, Durfee G, Smetters D. Building an encrypted and searchable audit log. Proc. Of 11th Annual Network and Distributed System. 2004.

[7] Chang YC, Mitzenmacher M. Privacy preserving keyword searches on remote encrypted data. Proc. of ACNS. 2005.

[8] Curtmola R, Garay JA, Kamara S, Ostrovsky R. Searchable symmetric encryption: improved definitions and efficient constructions. Proc. of ACM CCS. 2006.

[9] Li J, Wang Q, Wang C, Cao N, Ren K, Lou W. Fuzzy keyword search over encrypted data in cloud computing. Proc. of IEEE INFOCOM. 2010

[10] INF 3800, Edit Distance, 2011.02.21

[11] M. Hellmann, Fuzzy Logic Introduction, LaboratoireAntennes Radar Telecom, F.R.E CNRS 2272

[12] D. Parkhill, "The challenge of the computer utility," Addison-Wesley Educational Publishers Inc., US,1966.

[13] P. Mell and T. Grance, "Draft nist working definition of cloud computing," Referenced on June. 3rd, 2009 Online at http://csrc.nist.gov/groups/SNS/cloud-computing/index.html, 2009.

[14] M. Armbrust and et.al, "Above the clouds: A berkeley view of cloud computing," Tech. Rep., Feb 2009. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html

[15] Jin Li1, Qian Wang1, Cong Wang1, Ning Cao2, Kui Ren1, and Wenjing Lou2, " Enabling Efficient Fuzzy Keyword Search over Encrypted Data in Cloud Computing",

[16] Jospin Jeya, J. and E. Kannan," EFFICIENT RANKED AND SECURE FILE RETRIEVAL IN CLOUD COMPUTING", American Journal of Applied Sciences 11 (6): 906-911, 2014 ISSN: 1546-9239