National Conference on 'Advances in Engineering, Technology and Applied Sciences' - NCAETAS-2019
© 2019 IJSRSET | Volume 5 | Issue 6 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099

A Review on Fuzzy Keyword Matching for Search Through Encrypted Data in Cloud Computing

Anushree Charde, Revati Khodke, Utkarsha Thakare, Tejas Gandole, Saurabh Dhale
Computer engineering, Bapurao Deshmukh College of Engineering, Wardha, Maharashtra, India

ABSTRACT

Cloud computing is now days booming data store for users. Users intend to store data on the cloud so as to reduce the burden on the system and saving their capital investment towards the management of data. As there is millions of users' uses cloud storage data has to be stored in secured or an encrypted form before outsourcing to the cloud. Because cloud computing allows a user to store and retrieve the data from anywhere and on any type of the device according to the users need. This transformation or manipulation of data at cloud gives rise to several security issues as data is going to be accessed by the user on the internet. In general sense the user may often misspell the keywords while searching for a particular file. In our solution to this, we are using N-gram algorithm to quantify keywords similarity and develop two advanced techniques such as based-trie traverse searching scheme. The paper reviews the concept of fuzzy keyword search over encrypted data using cloud computing and the algorithms and mechanisms utilized for it.

Keywords: Cloud Computing, Fuzzy Keyword Search, Encryption Technique

I. INTRODUCTION

Day by day we are utilizing the digital documents and stepping up into the big data era, terabytes of data is getting produced worldwide per day. Users, business analysts and other entities that use systems are getting attracted by cloud computing for its vast and scalable storage or computing resources, low cost and the payment policy which charges as per your storage goes.

Variety of users can access and share information stored in the cloud independent of locations. The outsourced data may contain very sensitive information such as e-mails, company financial data, government documents, Personal Health Care records, facebook photos and business documents [9]. The cloud storage facility have various privacy issues so, cloud clients tends to encrypt the data before outsourcing it to the cloud. This encryption of data obviates the searching based data utility [4] and hence the need of various mechanisms and algorithms to search over encrypted data appears.

Moreover, in Cloud Computing, data owners may share their outsourced data with a large number of users. The individual users might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways is to selectively retrieve files through keyword-based search instead of retrieving all the encrypted files back which is completely impractical in cloud computing scenarios. Such keyword-based search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios, such as Google search [11].

By integrating the trapdoors of keywords within the index information, effective keyword search can be realized while both file content and keyword privacy are well-preserved. Although allowing for performing searches effectively and securely, the existing searchable encryption methods do not suit for cloud computing scenario since they support only *exact* keyword search.

For any effective practical searching utilization of the key idea, a user should be allowed to search using the fuzzy keywords i.e. the user supposed to type the misspelled words as keywords to search for the particular file on stored on the cloud.

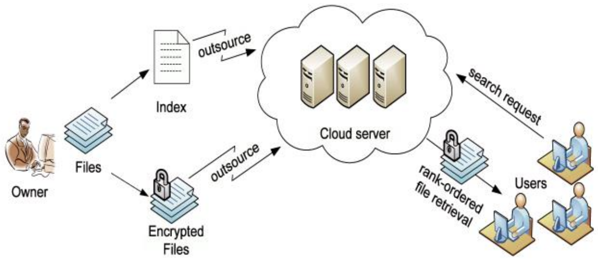

Figure 1. Architecture of fuzzy keyword search [7]

Although a client has mistaken to type the keyword but system should present it with the correctly spelled name of the files which are closest to match with the keywords that has been typed by the user in search. In this paper we focus on fuzzy keyword a matching technique which is supposed to be a non-trivial part in searching through encrypted cloud data.

The rest of paper is organized as follows: Section II presents Literature review that describes some necessary background for the techniques used in this paper. Section III introduces the proposed system and the algorithms that are going to be the part of our proposed system. Finally Section IV concludes the paper.

II. LITERATURE REVIEW

we have gone through the various mechanisms and techniques imposed by some authors. We come up with several techniques that we can apply to do our work better. Here is some related information that can be applied to the system described above,

Wen-Chung Kuo et al. [1] proposed that cloud computing is a trending data storage house in the networking industry. Storing data in the cloud can raise the storage overhead and reduce the costing of both hardware and software for the users. However, the clients cannot trust the cloud storage in the sense of security and privacy. So, there is a need of encryption of data before outsourcing to the cloud, There are some encryption techniques that supports exact keyword search. This paper provides the fuzzy keyword searching techniques for encrypted data.

Fang Zheng Li et al. [2] this paper elaborates the various techniques for searching the fuzzy words encountered by the users. The top-k algorithm has been employed in this paper for searching the results of the precision queries. Experiment results show that average precision and average recall are good enough to correspond to actual situation when query sets are come from normal school fuzzy sets.

Dr. Nanendra Shekokar et al. [3] proposed that With the increased rate of growth and adaptation of cloud computing, daily, more and more sensitive information is being centralized onto the cloud. For the protection of valuable proprietary information, the data using keywords but these techniques account for only exact keyword search. This makes paper we focus on secure storage using Advance Encryption Standards (AES) and information retrieval by performing fuzzy keyword search on this encrypted data. The AES is used to encrypt the files that are going to be outsourced .In the proposed solution, we exploit edit distance to quantify keywords similarity and develop an efficient technique for constructing fuzzy keyword sets, which focus on reducing the storage and representation overheads.

M A Manazir Ahsan et al. [4] proposed that, user like business entities, researchers are getting attracted towards the cloud storage. Though, cloud provides scalable data storage utility most users are not comfortable to store data on the cloud for security purpose. However, existing schemes suffer either from efficient handling of multi-letter errors or cannot distinguish anagrams. In this paper, we propose a scheme for fuzzy keyword search on encrypted data concentrating on fuzzy word matching among dictionary words. Our proposed scheme construct a transformed keyword (monogram set) based on each letter and its position in the word, which enables to find out original word from its typo secured with maximum similarity metric. As a similarity metric, we have chosen a modified version of Jaccard similarity which ensures maximum similarity for the closest word possibly the original one. Our experiment also suggests the applicability of our scheme.

Jing Chen et al. [5] in this paper that motivated by privacy preservation requirements for outsourced data, keyword searches over encrypted cloud data have become a hot topic. Compared to single-keyword exact searches, multi-keyword fuzzy search schemes attract more attention because of their improvements in search accuracy, typing mistakes, and user experience in general. However, existing multi-keyword fuzzy search solutions are not sufficiently efficient when the file set in the cloud is large.. In this framework, a novel two-stage index structure is exploited to ensure that search time is independent of file set size. The multi-keyword fuzzy search function is achieved through a delicate design based on the Gram Counting Order, the Bloom filter, and the Locality-Sensitive Hashing index.

III. PROPOSED WORK

Implementation of fuzzy keyword search :

consider a scenario where a private enterprise would like to centralize its data. Files are encrypted using AES and outsourced to cloud storage. At the same time the information is stored in a FILE INDEX: 1.File ID. 2. Keywords 3. File Name. We derive out fuzzy keyword sets from this FILE INDEX using N-GRAM algorithm and JACCARD coefficient to calculate the keywords similarity. To avoid enumerating all fuzzy keywords, and thus reducing the search space, we eliminate keywords with JACCARD coefficient smaller than our threshold value.. Search results are ranked according to our proposed weighted ranking function. These fuzzy keywords sets are associated with their respective file identification. On the fuzzy keyword sets generated, the trap-door function is applied. The keyword trapdoors and file ID's are now outsourced to cloud storage.

GRAM – BASED TECHNIQUE:

Another efficient technique for constructing fuzzy set is based on grams. The gram of a string is a **substring** that can be used as a signature for efficient approximate search. While gram has been widely used for constructing inverted list for approximate string search, we use gram for the matching purpose. We propose to utilize the fact that any primitive edit operation will affect at most one specific character of the keyword, leaving all the remaining characters untouched. In other words, the relative order of the remaining characters after the primitive operations is always kept the same as it is before the operations.

For example, the gram-based fuzzy set SCASTLE, 1 for keyword CASTLE can be constructed as
{CASTLE, CSTLE, CATLE, CASLE, CASTE, CASTL, ASTLE}.

ADVANCE ENCRYPTION STANDARD

AES is an encryption algorithm used for encryption of files that are going to be outsourced . The 4x4 matrix of bytes made from 128-bit input block is referred to as the sate array. Before any round-based processing for encryption can begin, input state is XORed with the first four words of the schedule. The importance of fuzzy search has received attention in the realisation of plaintext searching for information retrieval. This problem was addressed by allowing user to search relevant information based on approximate string matching. It seems possible for one to directly apply these string matching algorithms to the context of searchable encryption by computing the trapdoors on a character base within alphabets. However, this simple construction suffers from dictionary and statistical attacks due to lack of privacy-preserving encryption methods. Among the searchable encryption techniques, most of those works are focused on efficiency improvement and formalization of security definitions. Searchable encryption's first construction was proposed by Song et al, in which each word in the document is encrypted independently under a special two-layer encryption construction. Goh proposed views Bloom filters for constructing the indexes for data files to achieve more efficient search Chang et al And Curtmola et al both proposed similar approaches, in which a single encrypted hash table index is built for the whole of file collection. In this, each entry consists of the trap-door of a keyword and an encrypted set of file identifiers whose corresponding data files contain the keyword. A complementary approach was presented by Bonehet alas public-key based searchable encryption scheme

JACCARD ALGORITHM

this paper proposed the similarity measurement method between words by deploying Jaccard Coefficient. Technically, we developed a measure of similarity Jaccard with Prolog programming language to compare similarity between sets of data. Furthermore, the performance of this proposed similarity measurement method was accomplished by employing precision, recall, and F-measure. Precisely, the test results demonstrated the awareness of advantage and disadvantages of the measurement which were adapted and applied to a search for meaning by using Jaccard similarity coefficient. A measure of similarity of the search words. The determination of the association between two words with Jaccard coefficient. Jaccard index is a name often used for comparing similarity, dissimilarity, and distance of the data set.

N-GRAMS ALGORITHM

N-Grams is a word prediction algorithm using probabilistic methods to predict next word after observing N-1 words. Therefore, computing the probability of the next word is closely related to computing the probability of a sequence of words.

Simple (Unsmoothed) N-grams

The simplest probabilistic model for word prediction can be assigning equal probability to each word. So suppose that there are N words in a language, then the probability of any word following another word would be 1/N. However, this approach ignores the fact that some words are more frequent than the others in languages.

An improvement to the model above would be assigning the probability of a word wi following the word wi-1 is the probability of the word wi itself. For example, word "the" occurs 7% in Brown corpus, and "rabbit" occurs at a frequency of 1/10.000. Then, for any word, the probability of the next word being "the" is %7. However, this ignores that in some contexts, occurrence of the "rabbit" after a word is much more probable than occurrence of "the". For instance, "rabbit" following the word "white" seems much more logical than the word "the" following "white".

IV. CONCLUSION

The paper studies the various searching techniques for the encrypted cloud data. The paper gives information about the limitations of previously available searching techniques that ere suitable for searching plaintext data and not to the cloud scenarios. The techniques that we have studied like N-gram, Jaccard, AES are efficient and yet privacy preserving as well as easy to implement.

We have presented the various security and data utilization issues in cloud storage related to available searching techniques.

The various issues that has been overcome are keyword privacy, index privacy, query privacy, fine search, scalability, index Security, efficiency.

V. REFERENCES

[1]. Wen-Chung kuo, I-Hsien Liu, Yu-Hsin Lin, Jung-ShianLi,Chuan-Gang Li,"a secure multi-keyword fuzzy search with polynomial function for encrypted data in cloud computation", in 3rd international conference on computer and communication system 2018.

[2]. FangZheng Li, Da Yong Luo,DongXie, "Fuzzy queries of numerical attributes for keyword-based search over relational databases", in IEEE 978-1-4244-4738-1/9/$25.00 in 2009.

[3]. Dr. NarendraShekokar, KunjitaSampat, ChandaniChandawall, Jahnavi Shah, "Implementation of Fuzzy Keyword Search Over Encrypted Data in Cloud Computing", International Conference on Advance Computing Technologies and Application 2015.

[4]. MA Manazir Ahsan, MusaratSabilah, Ainuddin Wahid Bin Abdul Wahab, Mohad Yamani Idna Bin Idris,"An Efficient Fuzzy Keyword Matching Technique for searching through Encrypted Cloud Data", in IEEE 978-1-5090-6255-3/17/$31.00 in 2017

[5]. Jing Chen,Kun He, Lan Deng, QuanYuan,Ruiying Du, Jie Wu, "ElimFS: A Chieving Efficient, Leakage-Resilient And Multi-Keyword Fuzzy Search On Encrypted Cloud Data",in IEEE 2017.

[6]. Seo-Young Noh, AmolJaikar, "NOVEL:NO-VOWEL Technique To Search Fuzzy Keyword",in 2017 IEEE 18TH International Conference On Mobile Data Management.

[7]. QianWang,JunLi,Cong Wang, NingCao, KuiRen and Wenjing Lou, "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing", in Mini-Conference IEEE INFOCOM 2010.

[8]. Sonal Rahul Jandade,Jyoti N. Nandimath, "Survey on Keyword Search in Encrypted Data With Privacy Presentation", in International Journal Of Science and Research in 2016.

[9]. KalyaniSonawane, Rahul Dagade,"A Survey On Multi-Keyword Ranked Search Over Encrypted Cloud Data Owners", in International Journal Of Computer Application(0975-8887)Volume 162-No 11 March 2017.

[10]. QanqunXu, Hong Shen, YingPengSang, HuiGian, " Privacy-Preserving Ranked Fuzzy Keyword Search Over Encrypted Cloud Data", in 2013 International Conference On Parallel And Distributed Computing Application And Technologies.

[11]. Google, "Briteny Spears Spelling Correction,"Referenced online at http://www.google.com/jobs/briteney.html,june 2009