

# Fuzzy Keyword Search Over Data in Cloud Computing

Saharsh Gera<sup>1</sup> Shreya Kapoor<sup>2</sup>

<sup>2</sup>Assistant Professor

<sup>1,2</sup>Department of Information Technology

<sup>1,2</sup>MAIT, GGSIPU, Delhi, India

*Abstract*— Cloud computing is on-demanding delivery of computing power, databases storage, application, & other IT resources by the use of cloud services platforms via the internet with pay as you go pricing. Cloud computing provided a simple and easy way to accessing servers, storages, database & a broad set of application services over the Internet / network. A fuzzy search is the process of locating Web pages that're likely to be relevant to a searching arguments even when the arguments doesn't exactly similar to the desired information. A fuzzy keyword search is done by the means of a fuzzy matching algorithms, which return a list of result based on similar words even though search argument words and spellings may not exactly similar. Exactly and highly relevant matches are appears on the top of the list. A fuzzy keyword search matching program can operates like a spelling checker and spelling-error correctors. For example, if the user types "Misissippi" into Google, a list of hit is returns along with the questions like, "Did you mean Mississippi?" Alternative spelling, & words that sounds the same but spelled differently, are given. A fuzzy keyword matching programs can compensate for a common input typing errors, as well as errors introduce by optical characters recognition scans of printed document. This program can returns hits with contents that contains a specified based words along with the prefix and suffix. For example, if "planet" is enter as a searching word, hit occur for site contains words such as "protoplanet" or "planetary." This program can also used to find synonyms & related terms, it working as an on-line thesaurus or encyclopaedic cross-references tools. In the Ask Jeevs searching engines, if the word 'galaxy' is entered, hit are returned such as "Galaxy Photography", "Milky Way" and "The 9 Planet Solar System". Fuzzy searching can also be used to locate individuals based on incomplete or partially inaccurate identifying information.[1].

**Key words:** Fuzzy Keyword Search, Cloud Computing, AES, Wildcard, String Matching Algorithm

## I. INTRODUCTION

Cloud computing means that instead of all the computer hardware and software you're using sitting on your desktop, or somewhere inside your company's network, it's provided for you as a service by another company and accessed over the Internet, usually in a completely seamless way. Exactly where the hardware and software is located and how it all works doesn't matter to you, the user—it's just somewhere up in the nebulous "cloud" that the Internet represents.

Cloud computing is a buzzword that means different things to different people. For some, it's just another way of describing IT (information technology) "outsourcing"; some other people use it as computing services provided over the network; and some other defined it as bought in computer services that people uses that sites outside their firewall.[2]

Data security is consistently being a major issued in information technology. In cloud computing, it becomes more attention able because the data's are stored in different servers. Privacy protections & Data securities are the main factors of user's concern of cloud computed technologies. There are various of techniques on the topic in cloud-computing is being investigated in academic and industry, data security & privacy protections are becomes more and more importants for the further developments of cloud technologies in industries and other business. Data security & privacy protectioning issues are related to both software & hardware in the cloud architecture[3]

In this paper, I generally focouses on Fuzzy matching. Fuzzy matching searches a translation memory database for a query's phrases or words, finding derivatives by suggesting words with approximate matching in meanings as well as spellings. The fuzzy match technique applies a match percentage. The DB return possible match for the query phrase among certain. Till now, fuzzy matches is not compatible of replaces human in language translation process, but with more and more researches and artificial intelligence techniques application, it may be capables of replacing human in the future with the nearly 100 % accuracy. Fuzzy matching-based translation software is available and is used to generate a basic translated form, which can be manually processed to reach a final, readable form. In this reasearch paper I also develop another technique, that is, a wildcard based techniques, this is used for the constructing of fuzzy word sets. With helps of security analysis I find that propose system is secure.

## II. ALGORITHM USED

The algorithm that is used in this is entitled as String Matching Algorithm.

### A. Algorithm's Description

In computer science & technology string matching is technique of searching strings that matches pattern approximately (rather than exactly). The problem of approximating string matches is typically divides into 2 sub-problem: 1. finding approximately substring insides a given phrase and 2. finding dictionary string that matches the patterns approximately.

The nearness of a matches is measured in the terms of the numbers of primitive operation necessarily to convertes the string into an exact matches. These numbers are called as the edit distance between the strings and the patterns. The usually primitive operation are[4]

- insertion: cot → coat
- deletion: coat → cot
- substitution: coat → cost

### III. SYSTEM ARCHITECTURE

The architecture of fuzzy keyword search over data in cloud computing is given in fig 1[7]

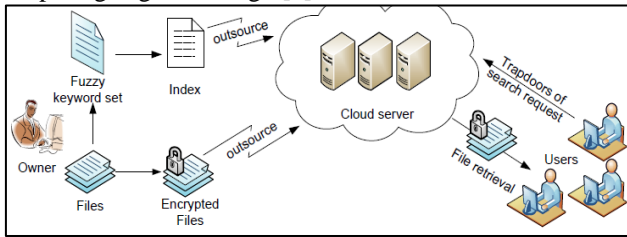


Fig. 1: Architecture of Fuzzy Keyword

#### A. Existing System

Previous System explain about the straightforward approach of searching file over encrypt file by provides fuzzy keyword, this is to be done to achieve searching privacy using the techniques of secure trapdoor. There are many prones of using this type of approach. Larger storage complexities were introduce by simple enumerated method by construction of fuzzy key sets, the usability of this type will greatly affects.

The substitution operation on the 1st character of word are as follows:

SAHARSH: {AAHARSH, BAHARSH, CAHARSH, DAHARSH..... YAHARSH, ZAHARSH}.

#### B. Proposed System

The proposed system is developed with the help of these three algorithms. These three algorithms are:[6]

- 1) WILDCARD&BASED TECHNIQUE
- 2) GRAM&BASED TECHNIQUE
- 3) SYMBOL&BASED TRIE-TRAVERSE SEARCH SCHEME

##### 1) Wildcard Based Technique

Form the keyword SAHARSH with the preset edit distance 1, its wild based fuzzy key set can be expressed as:

SSAHARSH, 1 = {SAHARSH, \*SAHARSH, \*AHARSH, S\*AHARSH, S\*HARSH..... SAHARS\*H, SAHARS\*, SAHARSH\*}.

Edit Distance

- a) Substitution
  - b) Deletion
  - c) Insertion
- a) Substitution: changes 1 alphabet to other alphabet in a word.
  - b) Deletion: deletion of 1 alphabet from word.
  - c) Insertion: insertion another alphabet into word.

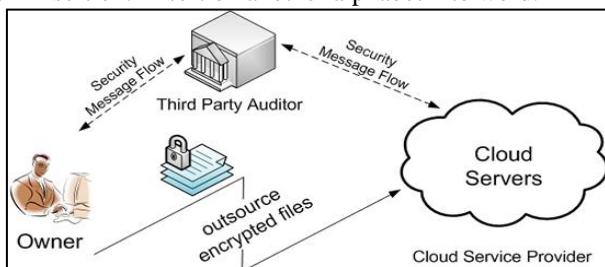


Fig. 2: Wildcard Based-Technique

##### 2) Gram & Based Technique

The gram based fuzzy keyword set SSAHARSH, 1 for key SAHARSH can be expressed as:

{SAHARSH, SHARSH, SAARSH, SAHASH, SAHARH, SAHARS, AHARSH}

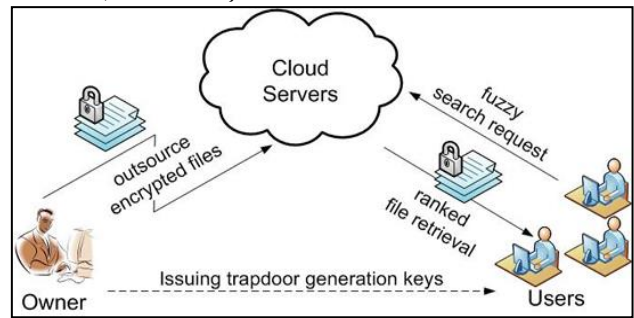


Fig. 3: Gram Based-Technique

##### 3) Symbolbased-Trietraverse Search

This is a combination of both wildcard based and gram based.

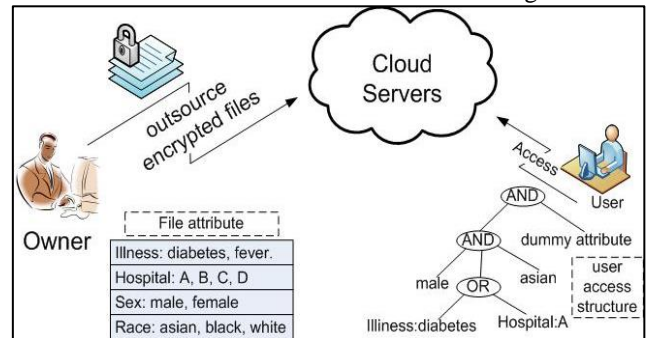


Fig. 4: Symbol Based-Trie Traverse Search

### IV. CONCLUSIONS

- 1) This paper achieves effectively utilization of remotely store data in Cloud.
- 2) I designed two advance techniques (wildcard based & grambased technique) to constructs the storage-efficient fuzzy keywords sets.
- 3) Based on first two technique I formalized new symbol based trie traverse searching scheme.
- 4) With securities analysis, I show that the solution is very much secure.
- 5) Search semantics that takes into consideration conjunction of keywords, sequence of keywords, and even the complex natural language semantics to produce highly relevant search results.
- 6) Search ranking that sorts the searching results according to the relevance criteria.

### REFERENCES

- [1] <http://whatis.techtarget.com/definition/fuzzy-search>.
- [2] <http://www.explainthatstuff.com/cloud-computing-introduction.html>.
- [3] <http://journals.sagepub.com/doi/full/10.1155/2014/190903>
- [4] Boytsov, Leonid (2011). "Indexing methods for approximate dictionary searching: Comparative analysis". Jea Acm. 16 (1): 1-91. doi:10.1145/1963190.1963191
- [5] B. Waters, D. Balfanz, G. Durfee, and D. Smetters, "Building an encrypted and searchable audit log," in Proc.

- [6] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc.
- [7] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc.
- [8] D. Boneh, G.D.Crescenzo, R. Ostrovsky, and G.Persiano, "Public key encryption with keyword search," in Proc
- [9] A review paper on fuzzy search over encrypted data in cloud computing by neel gala issn:2393-9842.

