ISSN: 2321-7782 (Online)
Impact Factor: 6.047

Volume 4, Issue 11, November 2016

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Nearest Ranked Keyword Search for Effective Utilization of the Outsourced Data in the Cloud Computing Applications

P. Sai Bharathi[1]
M.Tech Scholar, Dept of CSE
St.Marys Womens Engineering College
Budampadu,Guntur(dt)
Andhra Pradesh – India

V. Radha Kumari[2]
Assistant Professor
St.Marys Womens Engineering College
Budampadu,Guntur(dt).
Andhra Pradesh – India

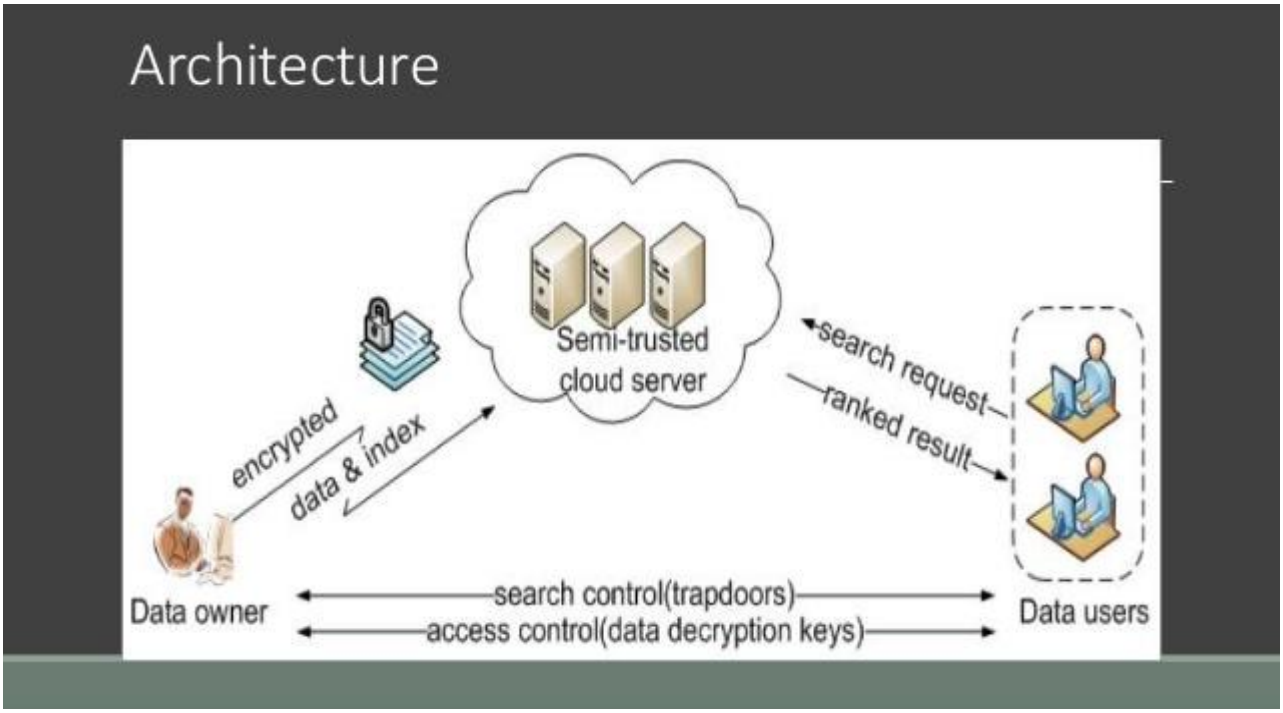
Abstract: ***Cloud computing is a subscription-based service where the networked storage space and computer resources can be obtained. Cloud computing economically enables the paradigm of data service outsourcing. However, to protect data privacy, sensitive cloud data have to be encrypted before outsourced to the commercial public cloud, which makes effective data utilization. In the proposed system, the problem of effective secure ranked keyword search over encrypted cloud data is done. Ranked keyword search greatly enhances sthe system usability by returning the matching files in a ranked order. The existing technique resolves the optimization complexities in ranked keyword search and its effective utilization of remotely stored encrypted cloud data. But it limits the further optimizations of the search results by preventing cloud server to interact with cloud users to maintain the integrity of actual owner's keyword and the data associated with it. The aim is to define a framework which enhances the accuracy of the ranked keyword search by secured machine learning, which does not affect the data integrity. Introducing new and interactive access permissions allows only specific group of people to guide the search engine. This technique lists the exact or necessary search results for any encrypted keyword. Due to this learning the privacy of the keyword does not get to be violated because, the owner of the encrypted keyword has some lists of users to whom only the machine should learn for secured and improved search results.***

General Terms: ***Efficient Ranked Keyword Search, Search engine in Cloud, Security in Search engine, confidential data, searchable encryption.***

Keywords: ***Search in cloud, secured search engines, Inter cloud communication in cloud search engines, dynamic secret key.***

I. INTRODUCTION

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. However, cloud computing technology challenges many traditional approaches to datacenter and enterprise application design and management. The effectiveness and efficiency of traditional protection mechanisms are being reconsidered as the characteristics of this innovative deployment model can differ widely from those of traditional architectures. An alternative perspective on the topic of cloud security is that this is but another, although quite broad, case of "applied security" and that similar security principles that apply in shared multi-user mainframe security models apply with cloud security. Secure search over encrypted data has recently attracted the interest of many researchers. Song et al.[3] first define and solve the problem of secure search over encrypted data.



They propose the conception of searchable encryption, which is a cryptographic primitive that enables users to perform a keyword-based search on an encrypted dataset, just as on a plaintext dataset. Searchable encryption is further developed by [4], [5], [6], [7], [8]. However, these schemes are concerned mostly with single or Boolean keyword search. Extending these techniques for ranked multi keyword search will incur heavy computation and storage costs. Secure search over encrypted cloud data is first defined by Wang et al. [9] and further developed by [10], [11], [12]. These researches not only reduce the computation and storage cost for secure keyword search over encrypted cloud data. To protect data privacy, sensitive data has to be encrypted before outsourcing so as to provide end-to-end data confidentiality assurance in the cloud and beyond. Thus, exploring privacy-preserving and effective search service over encrypted cloud data is of paramount importance. Data owners may share their data with large number of on-demand data users and huge amount of outsourced data documents in cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability and scalability. On the one hand, to meet the effective data retrieval need, large amount of documents demand cloud server to perform result relevance ranking, instead of returning undifferentiated result. Such ranked search system enables data users to find the most relevant information quickly. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data. The other hand, to improve search result accuracy as well as enhance user searching experience, it is also crucial for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse result. As a common practice indicated by today's web search engines (e.g., Google search), data users may tend to provide a set of keywords instead of only one as the indicator of their search interest to retrieve the most relevant data.

II. Related Work

The evolution of cloud computing is one of the major advances in the technologies which represent cloud computing: Platform-as-a-service (PaaS), Software-as-a-Service (SaaS) and Infrastructure-as-a-Service (IaaS). Public key encryption [8] deals with the privacy of database data. There are two different scenarios: public databases and private databases. Private databases: A user wishes to upload its private data to a remote database and wishes to keep the data private from the remote database administrator. An additional privacy requirement is to hide any information from the database administrator regarding the access pattern, i.e. if some item was retrieved more than once, some item was not retrieved. Public Databases: The database data is public (such as stock quotes) but the user is unaware of it and wishes to retrieve some data-item or search for some data-item, without revealing to the database administrator which item it is. All keyword searches are based on this index; hence our

scheme does not order full pattern-matching generality with the actual text. In practice, this should be sufficient for most users. It is worth noting that this framework can have complete control over what words are keywords that can be useful for many applications. In this section, main research areas related to keyword search are presented. When we talk about cloud service the work is more specific and the parametric. Many researchers performed a lot of work in the same direction. In year 2007, Byron Y.L. Kuof presented a tag based summarization approach for the web search. The presented work is suggested on the public cloud. In which the integration of the web architecture and the database extraction is integrated. The work includes the refinement of the user query based on the cloud tags. The words extracted from the query are been summarized and this summarized query is passed to the public cloud. The cloud interface enabled the extraction of new and required information [1].In year 2009, Hang Guo presented personalization architecture for the cloud services. The work includes the individual access to the cloud to perform the user query. The author work is presented in two main parts one is client side and other is cloud side. The client side basically fetches the periodic information from the system where as the cloud data search engine presents the data for the modeling [2].In year 2011, Ju-Chiang Wang presented a content oriented tag based search for the database search. In his work the music database is selected for the query analysis. The query performed by the user is analyzed and divided to different colors or the levels to perform the effective content based retrieval. The probabilistic fusion model was defined based on Gaussian mixture model and the multinomial mixture model. The author evaluated the proposed system for the Effectiveness of the user query and the related results [3].In Year 2012;Cengiz Orencik presented a rank based keyword search on the data cloud. In this work the document retrieval is performed on the cloud server based on the keyword analysis and the information search is performed relative to the defined information. The presented work is performed on the encrypted data that has improved the security and the reliability of the retrieval. On this basis a secure protocol is suggested called Private Information Retrieval. The system will performed the query and present the final results on the basis of parametric ranking. The presented work is the efficient computation and communication of the requirement analysis [4]. Another cloud search is suggested by Daniel E. Rose in 2012 which is based on the information retrieval. The author presented his work on Amazon cloud service. The work is tested under different criteria such as scalability, configuration etc. The presented search reduce the barrier to allow a person or the organization to perform the content oriented search and the search is tested under the enterprises environment as well as on web search[5].

III. PROBLEM STATEMENT

The existing technique resolves the optimization complexities in ranked keyword search and its effective utilization of remotely stored encrypted cloud data. But it limits the further optimizations of the search results by preventing search engine to interact with cloud users to maintain the integrity of actual owner's keyword and the data associated with it. Consider an encrypted cloud data hosting service involving three different entities, as illustrated in Fig 1data owner, data user, and cloud server. Data owner has a collection of n data files that he wants to outsource on the cloud server in encrypted form while still keeping the capability to search through them for effective data utilization reasons.

The problem with the techniques available for implementing search engine in an environment consists of sensitive outsourced cloud data can be summarized as:

a) Lacking of effective mechanisms to ensure the file retrieval accuracy is very difficult.
b) Security is not addressed fully and limits search engine's accuracy. The ranked keyword search over encrypted data is to achieve economies of scale for Cloud Computing. This process start from the review of existing searchable symmetric encryption schemes and provides the definitions and framework for this proposed ranked searchable symmetric encryption.

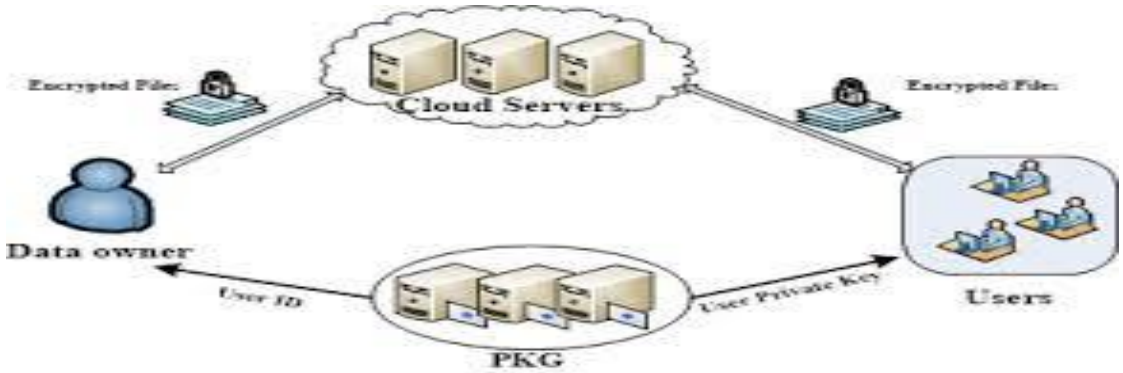


Fig 1: Ranked keyword search in cloud model

Searchable encryption allows data owner to outsource his data in an encrypted manner while maintaining the selectively search capability over the encrypted data. In order to achieve more efficient solutions, almost all the existing works on searchable encryption literature resort to the weakened security guarantee, i.e., revealing the access pattern and search pattern but nothing else. Result, i.e., which files have been retrieved. The search pattern includes the equality pattern among the two search requests (whether two searches were performed for the same keyword), and any information derived thereafter from this statement.

IV. PROPOSED FRAME WORK

4.1 Framework Definition

This process defines and solves the problem of effective yet secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria e.g., keyword frequency), thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. It first gives a straightforward yet ideal construction of ranked keyword search searchable symmetric encryption (RSSE)[6] security definition, and demonstrates its inefficiency. To achieve more practical performance, the process then propose a definition for ranked searchable symmetric encryption, and give an efficient design by properly utilizing the existing cryptographic primitive, order-preserving symmetric encryption (OPSE).

4.2 Design Goals:

To enable ranked keyword search for effective utilization of outsourced cloud data under the a fore mentioned model, our system design should achieve the following security and performance guarantee.

a) Ranked key word search: For efficient searching process the process use the mechanism of Topic detection and tracking 2004. The search time includes fetching the posting list in the index, decrypting, and rank ordering each entry.
b) Security guarantee: For providing the security in the cloud server, this process uses the privilege method.

4.3 Mechanisms for Implementation:

Topic detection and tracking:

TDT refers to automatic techniques for finding topically related material in streams of data techniques that could be quite valuable in a wide variety of applications where efficient and timely information access is important. For example, a lot of useful information could be gleaned from a multitude of news sources, but no one has the time to watch, listen to, or read carefully each of the many news sources available. Tasks can vary in focus and size from hypothetical applications to enabling technologies. In brief, the goal of each of the tasks is:

Topic Tracking–detect stories that discuss a target topic, Link Detection–detect whether a pair of stories discuss the same topic, Topic Detection–detect clusters of stories that discuss the same topic, First Story Detection–detect the first story that discusses a topic, and Story Segmentation–detect story boundaries.

4.4 Privilege:

A privilege is a special entitlement to immunity granted by the state or another authority to a restricted group, either by birth or on a conditional basis. It can be revoked in certain circumstances. For Example, a system administrator or, in the case of network resources such as access to a particular device, a network administrator assigns privileges to users. System software then automatically enforces privilege to the files.

V. BASIC CONCEPT OF RANKED SEARCH ALGORITHM

5.1 Ranking Algorithm:

1. Define the list of available clouds on any public cloud server called Cloud(1),Cloud(2).....Cloud(n)
2. For i=1 to n{Identify parameters for Cloud (i) called Availability (i), Response Time (i), Security(i)}
3. Accept the User Query called Request under the specification Request Keyword, Request Security, Request Dead Line,
4. Activate the Middle layer to provide the best service selection
5. Accept the user query and filter it to retrieve the keywords under the following step
    - Remove the stop list words from the query list
    - Rank the different keywords respective to category
    - Find the frequency of keywords
    - Keep the most occurring keywords and present as relevancy measure
6. As the keywords retrieve perform query on each public cloud and perform the content and tag based match.
7. Find the list of M clouds that satisfy the relevancy criteria as well as identify the other cloud parameters like response time, security measure.
8. For i=1 to M [Perform the Content based similarity measure as]
9. Relevency Vector = 0For j=1 to Length (User Keywords) {Relevency Vector =Relevency Vector + Keyword Occurrence (Clound(i) ,Keyword(j)) /Total Keywords (Clound(i),Keyword(j));}
10. Security Vector=0;If (User Security Req=Security (Cloud(i)) Security Vector=1;
11. Response Time Vector=0 If (User Deadline> Response Time (Cloud(i))

    {Response Time Vector=User Deadline-Response Time(Cloud(i) }
12. Rank (Clound(i))= Relevency Vector*w1 +Security Vector *w2 +Response Time Vector*w3;
13. As user get Ranked list of clouds, selection can be performed for best cloud service provider respective to user interest.

VI. MODULES

- Cloud Service Provider: It stores all files uploaded by the authorized data owner, and allow accessing the data only to the authorized User. It accepts the query from ADL and sends the result according to the query send by ADL.
- Data Owner: The data owner is person who uploads the data. Data owner should login first to store their data on cloud. Only authorized owner can store their data on cloud and new data owner should register first.
- Cloud User: In this module, the cloud user should be authenticated so the user should register and login for cloud usage. The vendor which verifies that whether the user is authenticated or not, If he/she is unauthorized user then CSP cannot

continue further processing. Once login successfully then he/she can obtain the basic information from cloud storage for data integration. Here the user can request the data to cloud using certain keywords.

- Query Search: Based on ranking In this module, each user set the rank to their query and cloud searches the results on the basis of the rank. Cloud returns certain percentage of matched file to the ADL. The basic idea of this module is to protect the privacy and rank the user queries. In this module, Cloud sends the file along with the keywords.

VII. CONCLUSION

The problem of solving efficient ranked keyword search is to achieve the effective utilization of remotely stored encrypted data in Cloud Computing. A basic scheme shows that by following the same existing searchable encryption framework, it is very inefficient to achieve ranked search. This appropriately weaken the security guarantee, resort to the newly developed encrypted algorithms, which allows the efficiency in cloud. Investigations of privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset shows how our proposed schemes introduce low overhead on both computation an communication. It is a secure and privacy preserving, while correctly realizing the goal of ranked keyword search. In this paper, as an initial attempt, we motivate and solve the problem of supporting efficient ranked keyword search in a cloud. We proposed a framework that allows thinking of keyword searches more naturally. In this present work, the GUI interface will be created to pass the user query like a search engine. The first output will be drawn in terms of query filtration and extraction of keyword from query analysis. Once the keyword analysis is performed, keyword reduction will be done and finally the keywords will be drawn as output. Now, this extracted keyword will work as input to the cloud search architecture and based on algorithmic approach, it will return the effective URL list along with ranking. The proposed ranking method proves to be efficient in analyzing user query and returning highly relevant document corresponding to submitted search terms.

References

1. Byron Y-L. Kuo (2007),"Tag Clouds for Summarizing Web Search Results", WWW 2007, May 8–12, 2007, Banff, Alberta, Canada. pp. 1203-1204.
2. Hang Guo (2009),"Personalization as a Service: the Architecture and a Case Study",CloudDB'09, November 2, 2009, Hong Kong, China. Pp. 1-8
3. Ju-Chiang Wang (2011),"Colorizing Tags in Tag Cloud: A Novel Query-by-Tag Music search System", MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.ACM p 293-302
4. Cengiz Orencik (2012),"Efficient and Secure Ranked Multi-Keyword Search on Encrypted Cloud Data",PAIS 2012, March 30, 2012, Berlin, Germany. ACM, p 186-195
5. Daniel E. Rose (2012),"Cloud Search and the Democratization of Information Retrieval",SIGIR'12, August 12–16, 2012, Portland, Oregon, USA.PP. 1022-1023.
6. Y.-C. Chang and M. Mitzenmacher, Privacy preserving keyword searches on remote encrypted data, 2005.
7. S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, Top-k retrieval from a confidential index, 2009.
8. D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search, Springer, 2004.
9. Y. H. Hwang and P. J. Lee,Public Key Encryption with Conjunctive Keyword Search and Its Extension to a Multi-User System, 2007.
10. INFOCOM, 2011Proceedings IEEE April 2011.
11. P.Paillier,Public-Key Cryptosystems Based on Composite Degree Residuosity Classes, in Proc. EUROCRYPT, 1999, pp.223-238.V.Hristidis and Y.Papakonstantinou, DISCOVER: Keyword Search in Relational Databases,in Proceedings of the 29th International Conference onVery Large Data Bases, VLDB Endowment, August 2002,pp.670681.
12. Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou, Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data, IEEE Transactions On Parallel And Distributed Systems, Systems, VOL. 23,NO.8

