International Journal for Modern Trends in Science and Technology
Volume: 03, Special Issue No: 01, February 2017
ISSN: 2455-3778
http://www.ijmtst.com

An Outsourced Cloud Data with High Secured Ranked Keyword Search

Shabbir Hussain Shaik[1] | Prasad U[2] | Shanmukha Sai Y[3]

[1,2,3]Assistant Professor, Department of CSE, Nalanda Institute of Engineering & Technology, Guntur, India

To Cite this Article
Shabbir Hussain Shaik, Prasad U and Shanmukha Sai Y, "An Outsourced Cloud Data with High Secured Ranked Keyword Search", *International Journal for Modern Trends in Science and Technology*, Vol. 03, Special Issue 01, 2017, pp. 40-45.

ABSTRACT

Distributed computing financially empowers the worldview of information administration outsourcing. To ensure information protection, touchy cloud information must be encoded before outsourced to the business open cloud, which makes powerful information usage benefit an extremely difficult assignment. Albeit customary searchable encryption procedures permit clients to safely seek over scrambled information through watchwords, they bolster just Boolean hunt and are not yet adequate to meet the successful information use require that is innately requested by extensive number of clients and gigantic measure of information records in cloud. In this paper, secure positioned watchword look over scrambled cloud information is characterized and settled. Positioned look extraordinarily improves framework ease of use by empowering query item importance positioning as opposed to sending undifferentiated outcomes, and further guarantees the document recovery exactness. In particular, we investigate the factual measure approach, i.e., pertinence score, from data recovery to manufacture a safe searchable record, and build up a one-to-many request safeguarding mapping method to appropriately secure those delicate score data. The subsequent plan can encourage productive server-side positioning without losing watchword protection. Intensive investigation demonstrates that our proposed arrangement appreciates "as solid as could be allowed" security ensure contrasted with past searchable encryption plans, while effectively understanding the objective of positioned catchphrase look.

KEYWORDS: *Ranked keyword, search, High secured, confidential data, Outsourced, Cloud Data*

I. INTRODUCTION

1.1 History

The concept of cloud computing is not new. The power and scale of the cloud has changed greatly from what it was in the beginning. As the technology and business environments had progressed the status of cloud computing has changed. What was known as cloud computing long ago was the same in principle, but the uses in information today have changed by an immense degree.[11]

The beginning of cloud computing can be traced back to the mainframe days of the 1960s when the idea of "utility computing"[1] was coined by computer scientist and Turing award winner John McCarthy. Utility computing ended up becoming something of a big business for companies IBM. The concept was : that computing power could be broken down as a service for businesses much like how the power and telephone companies operated for their customers. Indeed, it was an article "The Computers of Tomorrow" for the Atlantic Monthly in May of 1964 where author Martin Greenberger pointed out the

concept that "advanced arithmetical machines of the future" were now being used not only institutionally for scientific calculation and research but for business functions such as accounting and inventory. The potential for huge profit to be made in this type of invested had for terminal machines that would cost less than $300. These ideas were indeed profound, but they never really took off as consumers were looking for more complete personal computer solutions that had, for example, some storage capacity available.

The rise of the Internet beginning in the mid-90s changed how computers could be used and how information could be disseminated. With the idea of utility computing long gone, companies such as Amazon began to harness the power of server farms to offer a gaggle of products to would-be buyers.

1.2 Open source in cloud

The cloud offering an open-source package called Cloud Foundry, a Platform-as-a-Service[2] that should strike fear in the hearts of its competitors, especially the likes of Salesforce.com, Microsoft and Rack space. The platform will offer developers the tools to build out applications[13] on public clouds, private clouds and anyplace else, whether the underlying server runs.

Developers and enterprise IT shops will soon have another option for PaaS (platform as a service) in the form of Cloud Swing, an upcoming offering from Open Logic that builds on its core business of providing technical support for open source software. Cloud Swing customers can use the platform to assemble software stacks of both open source and commercial products for use on cloud infrastructure services such as Amazon EC2 (Elastic Compute Cloud).

1.3 Cloud computing and its security

The security in cloud is to integrate seamlessly with the IT security in your own data centre. However, the cloud service provider implements its own IT security procedures.

- To protect customers from external threats.
- To ensure that individual customer environments are isolated from one another.
- For every type of cloud service, the provider delivers a good deal of the IT security.
- IT security software and hardware (firewalls, intrusion detection systems, virtual private networks (PNs), and secure connections) that the cloud provider has in place.
- Know how the cloud providers are protecting the overall computing environment.

1.4 Servers Performance Management

Performance management is all about how your software services run effectively inside your own environment and through the cloud. If you start to connect software that runs in your own data centre directly to software that runs in the cloud, you create a potential bottleneck at the point of connection.

Services connected between the cloud and your computing environment can impact performance if they aren't well planned. This is especially likely to be the case if there are data translations or specific protocols to adhere to at the cloud gateway. As a customer, your ability to directly control the resources will be much lower in the cloud. Therefore,

- The connection points between various services must be monitored in real time. A breakdown may impact your ability to provide a business process to your customers.
- There must be expanded bandwidth at connection points.

1.5 Scope

As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud, such as e-mails, personal health records, company finance data and government documents, etc. The fact that data owners and cloud server are no longer in the same trusted domain may put the outsourced unencrypted data. Cloud computing is the long dreamed vision of computing[16] as a utility, where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. The benefits brought by this new computing model include but are not limited to relief of the burden for storage management, universal data access with independent geographical locations and avoidance of capital expenditure on hardware, software and personnel maintenances etc. It follows that sensitive data has to be encrypted prior to outsourcing for data privacy and combating unsolicited accesses. However, data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files[11].

In Cloud Computing, cloud may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files. One of the most popular ways to do so is

through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. This existing searchable scheme will support only Boolean keyword search, which will combine words and phrases using the words AND, OR, NOT operators. This leads to following drawbacks.

- Non relevant data search result
- Large unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm.
- Decrease the efficiency and File retrieval accuracy.

So this kind of plaintext search method fails for cloud data. In order to improve the efficiency of ranked keyword search concept based searching techniques is used for file retrieval in which search words are conceptually related to the topic[13].

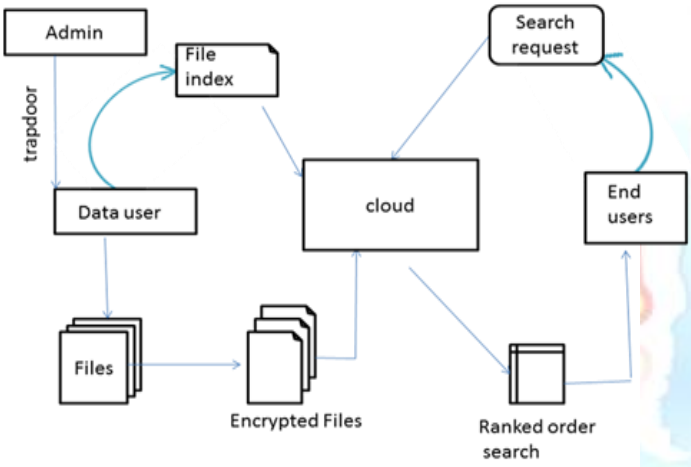

Fig. 1: Basic model of system

1.6 Efficient Ranked Keyword Framework

1.6.1 Setup

The data user collects the data files and encrypts the files using DES encryption and generates a secret key. Then data user generates the searchable index terms from the unique words which was extracted from file collection. The below table 1 contains the sample words and index terms which was extracted from file collection. Then the index terms are published on cloud.

Table 1: Index for words

Word	File Index
Soap	ravi.txt(1,1000)
Software	file3.txt(7,1002)
Protocol	ravi.txt(4,1001)
Dot net	file3.txt(2,1001)

1.6.2 Rank calculation

Once file indexing over, next rank is calculated. For calculating the rank for each file, term frequency, document frequency, the length of files and the number of documents that the data user has in his collection needs to be know. The term frequency calculated based on how many times the keywords occurs in the same document , and for each file , and for each term this needs to be calculated. The document frequency calculated based on how many times a particular keyword exists in the different documents. Given below Table 2 shows the ranks obtained for different keywords in different files, for example soap keyword is repeated in file Ravi 3 times and in sha it is repeated 2 times so rank 1 is given to ravi and second rank is given to sha file.

Table 2: Rank calculation

Keywords	File1	File 2
Soap	ravi.txt(3,1000)	sha.txt(2,1001)
Software	myfile1.txt(5,1000)	lucky.txt(2,1001)
Protocol	sha.txt(4,1000)	Ravi.txt(2,1001)
Programming	Lucky.txt(7,1000)	sha.txt(2,1001)

II. Literature Survey

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

Boneh.D, Crescenzo G. D.,Ostrovsky.R and Persiano.G[1] describe the concept of public key encryption with keyword search .Consider user Bob who sends email to user Alice encrypted under Alice's public key. An email gateway wants to test whether the email contains the keyword "urgent" so that it could route the email accordingly. Alice, on the other hand does not wish to give the gateway the ability to decrypt all her messages. A mechanism that enables Alice to provide a key to the gateway that enables the gateway to test whether the word "urgent" is a keyword in the email without learning anything else about the email is shown. This mechanism is represented as Public Key Encryption with keyword Search. As

another example, consider a mail server that stores various messages publicly encrypted for Alice by others. Using this Alice can send the mail server a key that will enable the server to identify all messages containing some specific keyword, but learn nothing else.

In fact, the serious problems with previous notion of security for SSE are observed, and show how to design constructions which avoid these pitfalls. Further, second solution also achieves what we call adaptive SSE security, where queries to the server can be chosen adaptively (by the adversary) during the execution of the search, this notion is both important in practice and has not been previously considered. Surprisingly, despite being more secure and more efficient, SSE schemes are remarkably simple.

As an additional contribution, multiuser SSE is considered. All prior work on SSE studied the setting where only the owner of the data is capable of submitting search queries. the natural extension where an arbitrary group of parties other than the owner can submit search queries is also taken into considerations. SSE in the multi-user setting, and present an efficient construction that achieves better performance than simply using access control mechanisms is defined.

Singhal.A[3] defines how to assign a similarity measure to each document that indicates how closely it matches a query. Boolean queries are not the only method of searching for information .If some exact subset of the document being sought is known, then they are certainly appropriate, which is why they have been so successful in areas such as commercial databases and bibliographic retrieval systems .Often, however, the information requirement is less precisely known. For this reason, it is sometimes useful to be able to specify a list of terms that give a good indication of which documents are relevant, though they will not necessarily all be present in the documents sought. The system should rank the entire collection with respect to the query, so that the top 100, say, ranked documents can be examined for relevance and those that constitute the answer set extracted.

Song.D, wagner.D, and perrig.A [4] propose cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting such crypto systems. It is desirable to store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. But this usually implies that one has to sacrifice functionality for security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query without loss of data confidentiality.. The techniques have a number of crucial advantages.

III. Proposed Approach

3.1 Module description

3.1.1 Providing secured data transfer between owner and cloud

As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud, such as emails, personal health records, company finance data, and government documents, etc. The fact that data owners and cloud server are no longer in the same trusted domain may put the outsourced unencrypted data at risk[16]: the cloud server may leak data information to unauthorized entities or even be hacked. It follows that sensitive data has to be encrypted prior to outsourcing for data privacy and combating unsolicited accesses. This module is used to help the user to upload data in a secured way using encrypt the document by DES Algorithm and to convert the encrypted document with some keys and then keys are send to the user for to retrieve the results. This module provides secured data transfer into cloud with series of encryption methods.

3.1.2 Maintenance of index files with relevant keywords

Index structure

In information retrieval, indexing structure is used to stores a list of mappings from keywords to the corresponding set of files that contain this keyword, allowing full text search. For ranked search purposes, the task of determining which files are most relevant is typically done by assigning a numerical ranks, which can be pre computed, to each file based on some ranking method introduced below.

Ranking method

Collect the words which are greater than 4 letters, then search those words in file which is uploaded by user. Count those repetitions and save keyword in index along with file name, then again search the same word in other files and make count how many times it is repeated. Now rank one is given to the highest counted file for that particular keyword like that rank is calculated for remaining keywords. For example take keyword soap two files contains it so ranking is done like this. Soap – ravi.txt(3,1001) sha.txt(2,1001) here, Ravi file soap is repeated 3 times and in sha file it is

repeated 2 times so rank 1 is given to Ravi and rank 2 is given to sha.

3.1.3 Updating the index file at every updated data in the cloud

For every new file updation, the index has to be modified and it should be updated with new relevancy scores and with new ranking. Two types of updation are present. They are as follows:

• Index is updated based on the previous indexing results by taking the old scores we will calculate the new scores and add to the index

• Rehashing of index is done based on users criteria .When the users choose the file based on their interest then ranking order will change.

Depending upon the two results the re-index is calculated

3.1.4 Searching and Retrieval of files

Searchable and Retrieval of files through encryption has been widely studied in the context of cryptography. Among those works, most are focused on efficiency improvements and security definition formalizations. The first construction of searchable encryption was proposed by Song et al in which each word in the document is encrypted independently. To achieve more efficient search similar "index" approaches is used, where a single encrypted hash table index is built for the entire file collection. In the index table, each entry consists of the trapdoor of a keyword and an encrypted set of file identifiers whose corresponding data files contain the keyword. As a complementary approach presented a public-key based searchable encryption scheme, with an analogous scenario. [10]shows the search and retrieval process, where the users can send the search keyword request to the cloud but the cloud will give the ranking order of the files retrieval response to only authorized users.

3.1.5 Ranking of search results

At the point when the client scan for any magic word, the cloud server will send the top –n documents which are put away in the cloud focused around positioning in the list record. The client can choose any of the document in the top –n records focused around their advantage. The positioning request of the records will be changed focused around the picked document. Case when he chose the main 3 record in the rundown then the rank of the third document turns into one, next time when client look the decisive word the positioning of the document request is changed. This positioning of documents is spared and further utilized for the planning re-index.

3.2 Algorithm for Ranking

Description

step 1: Read word by word from the file which is to be uploaded

step 2: count number of time the word repeated in that file

step 3: check for that word whether present in index file or not

step 4: *if* found, re arrange the order of previously saved file according to their word counts.

else

append the word with its file name and count at the end of index file

step 5: upload the file to cloud

step 6: close

3.3 Algorithm for Search

step 1: Enter keyword to search

step 2: verify whether keyword is present in index file or not

step 3: if not present , display error message " word not found" . Else display the order of filenames , as placed in the index file(ranked order)

step 4: If updated by user, re arrange the order, placing the selected file 1st , and add to the index file in place of previous order

step 5: save index file, step 6: close

IV. Conclusion & Future Scope

Ranked keyword search on remotely stored data is done by saving files in cloud and retrieve the files by searching through the keywords. Retrieved files are presented in ranked order which is done by using ranking algorithm in the index page. Security for data stored in cloud is done trough saving encrypted files and privacy of data is maintained by providing different trapdoors to different users. Ranked analysis is done by score dynamics i.e. taking the user choices into consideration and giving highest rank to user chosen file so that user can get more efficient results.

As for the future work, the effectiveness of ranked keyword search is increased by concepts public-key systems that support comparison queries on encrypted data as well as more general queries such as subset queries and has to support arbitrary conjunctive queries (P1 and ... and Pn) without leaking information on individual conjuncts.

REFERENCES

[1] P. Mell and T. Grance, "Draft nist working definition of cloud computing," Referenced on Jan. 23rd, 2010 Online at http://csrc. nist.gov/groups/SNS/cloud-computing/index.html, 2010.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A berkeley view of Utiliy computing," University of California, Berkeley, Tech. Rep. UCBEECS-2009-28, Feb 2009.

[3] Cloud Security Alliance, "Security guidance for critical areas of focus in cloud computing," 2009, http://www. cloudsecurityalliance.org.

[4] Z. Slocum, "Your google docs: Soon in search results?" http:// news.cnet.com/8301-17939 109-10357137-2.html, 2009.

[5] B. Krebs, "Payment Processor Breach May Be Largest Ever," http://voices.washingtonpost.com/securityfix/200 9/01/ payment processor breach may b.html, Jan. 2009.

[6] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images," Morgan Kaufmann Publishing, San Francisco, May 1999.

[7] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of IEEE Symposium on Security and Privacy'00, 2000.

[8] E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, Report 2003/216, 2003, http://eprint.iacr.org/.

[9] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. of EUROCRYP'04, volume 3027 of LNCS. Springer, 2004.

[10] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS'05, 2005.

[11] Shabbir Hussain Shaik, Pranathi.K and Kranthi.S "Outsourced Cloud Data with Ranked Keyword Search" " in IJARCSMS ISSN: 2321 7782 (Online).

[12] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS'06, 2006.

[13] G Bharath Kumar, E Sai Kumar "Prism: Portion of Resources in Phase-Level Using MapReduce In Hadoop" in International Journal of Research Volume 03 Issue 10 June 2016.

[14] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchablesymmetric encryption: improved definitions and efficientconstructions," in Proc. of ACM CCS'06, 2006.

[15] A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Engineering Bulletin, vol. 24, no. 4, pp. 35–43, 2001.

[16] Poojitha Koneru, Dr. S.Prabakaran "A Secured and High Octane Rank Based Analysis in Cloud Computing Environment" in International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1973-1976.