Volume No:4, Issue No:2 (July-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

Privacy-Preserving Multi-Keyword Top-K Similarity Search through Encrypted Data in Cloud Computing

Atthar Mabjan
Department of Computer Science & Engineering, MJR College of Engineering & Technology, Piler, AP- 517214, India.

Reddi Durga Sree
Department of Computer Science & Engineering, MJR College of Engineering & Technology, Piler, AP- 517214, India.

Karamala Suresh
Department of Computer Science & Engineering, MJR College of Engineering & Technology, Piler, AP- 517214, India.

ABSTRACT:
Cloud computing provides the more facility to store and manage data remotely. The volume of data is increasing per day. The owners choose to store the sensitive data on the cloud storage. To protect the data from unauthorized accesses, the data must be uploaded in encrypted form. Due to huge amount of data is stored on the cloud storage; the association between the documents is hiding when the documents are encrypted. It is necessary to design a search technique which gives the results on the basis of the similarity values of the encrypted documents.

In this paper a cosine similarity clustering method is proposed to make the clusters of similar documents based on the cosine values of the document vectors. We also proposed a MRSE-CSI model used to search the documents which are in encrypted form. The proposed search technique only finds the cluster of documents with the highest similarity value instead of searching on the whole dataset. Processing the dataset on two algorithms shows that the time needed to form the clusters in the proposed method is less. When the documents in the dataset increases, the time needed to form clusters also increases. The result of the search shows that increasing the documents also increases the search time of the proposed method.

Keywords:
Cloud computing, multi-Keyword search, cosine similarity clustering, encrypted data.

I. INTRODUCTION:
Cloud computing becomes popular as it provides large amount of storage space and high and good quality services. The huge amount of data is created per day. It is a difficult task for the owner of the data to store and manage this large amount of data. To overcome this difficulty, the data owners can store their data on the cloud server to use the on demand applications and services from shared resources [1]. The cloud server providers agreed that their cloud service is armed with strong security constraints though security and privacy are major drawback which avoid the use of cloud computing services [2]. To protect the sensitive data on the cloud server from hackers, the data holder may encrypt the documents and uploads to cloud server [3]. In the earlier various strong cryptography methods were used to design the search techniques on the cipher text [4], [5], [6]. These techniques needs many operations and need large amount of time. So these techniques are not suitable for big data where information volume is huge. The property of a document depends on its association with other documents. Therefore maintaining the relationship between documents is important to fully express a document. The results of search returned to the customers may have damaged information due to hardware failure or storage corruption.

Cite this article as: Atthar Mabjan, Reddi Durga Sree & Karamala Suresh, "Privacy-Preserving Multi-Keyword Top-K Similarity Search through Encrypted Data in Cloud Computing", International Journal of Research in Advanced Computer Science Engineering, Volume 4, Issue 2, 2018, Page 1-4.

Volume No:4, Issue No:2 (July-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

Thus a mechanism should be given to customers to check the accuracy of the search results. The proposed architecture of search technique is based on the cosine similarity clustering which maintain the association between plain text and encrypted text to upgrade the efficiency of search.

II. EXISTING SYSTEM:

Searchable encryption (SE) is a hot research field, especially with the development of cloud computing. In this section, we review and analyze the existing searchable encryption schemes. SE can be divided into public key searchable encryption [4], [9] and symmetric searchable encryption (SSE) [3] [7], [8] according to different cryptography primitives. In this paper, we focus on the symmetric searchable encryption because public key searchable encryption usually is computationally expensive. Cao et al. proposed the multi-keyword ranked search over encrypted data for the first time and built a searchable index based on the vector space model, and chosen ”coordinate matching” to measure the similarity between queries and documents.

However, in their schemes, the time complexity of search is O(nm) (n is the number of keywords in dictionary, m is the size of the documents that stored in the cloud server), and the time complexity of trapdoor construction is also high. Sun et al. proposed a tree-based index structure which is based on the vector space model and the TF_IDF model. This structure achieves sub-linear time complexity, but it is vulnerable in protecting data privacy. One step further, Xia et al. [16] proposed a Greedy Depth-first Search tree-based searchable encryption scheme EDMRS, which achieved more efficiency than early works, but the cost of search remains high and the time complexity of creating trapdoor is high O(n2). The works add random numbers j in indexes or queries to disturb the importance scores between queries and documents, and they claimed that the value of Σj can be adjusted to control the level of query unlink ability.

But they cannot secure the query unlink ability thoroughly, because in order to guarantee the accuracy of queries, the level of query unlink ability is usually get limited. Actually, the cloud server can easily link two identical queries by Analyzing and comparing the results.

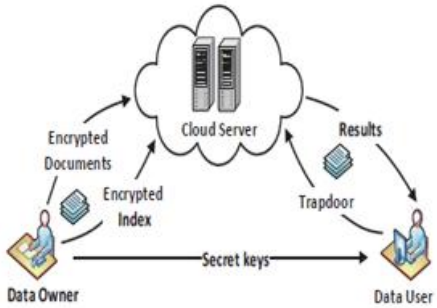

Fig: 1.Searching over outsourced encrypted data.

The region of a data centre determines how safe and protect all your data will be. Keeping this in mind we propose both its data centres in very low cost and security in very high. Moreover, none of the centres is easily usable and each of these has been protected with proper security measures. The payment cost of a cloud storage service consists of the costs for Storage, information Gets/Puts and Transfers. Different data centres of a CSP or different CSPs offer different prices for Storage, information Gets/Puts and Transfers. , the second objection is introduced: how to allocate information to data centres belonging to different CSPs and make resource reservation to minimize the service payment cost? Since website visit frequency varies over time, unexpected events may introduce a burst of requests in a short time. It may affect the accuracy of forecast the visit frequency. Thus, we need to dynamically adjust the data centres Get serving rate to save the payment cost [7].

III PROPOSED SYSTEM:

We used the cosine similarity clustering algorithm to review the time required for the search results. We have developed a search technique which is based on cosine similarity clustering algorithm to upgrade the efficiency of search.

Volume No:4, Issue No:2 (July-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

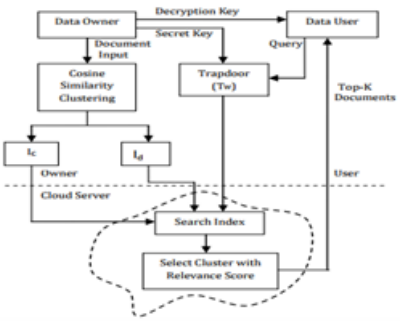

Fig.2: Building of Proposed System.

A. Methodology

Step 1: Key Generation:

The data owner of the data generates a key using the size of dictionary and a random generator. This secret key is used for encryption of the document index.

Step 2: Encrypted Index Generation:

The document vector is prepared and applies the cosine similarity clustering algorithm to form the clusters of proper documents. Each and Every cluster has its cluster index. These cluster indices are encrypted using matrix multiplication with the secret key. These encrypted cluster indices are uploaded to the cloud server.

Step 3: Document Upload:

The documents are encrypted using AES algorithm. These encrypted documents are transfer to the cloud server.

Step 4: Trapdoor Generation:

The data user send query to the data owner. After analyzing the keywords of query, data owner builds query vector. This query vector is encrypted using matrix multiplication with secret key and send to the user.

Step 5: Document Searching:

The user uploads this encrypted query vector (trapdoor TW) to the cloud server. The similarity value between trapdoor and encrypted cluster indices (IC) is calculated by the cloud server.

The most relevant cluster with the highest similarity value is selected. This cluster is extracted and again the similarity value is calculated between trapdoor and each document vector. The top-k documents will be returned to the user on the basis of similarity value.

Step 6: Decrypt Documents:

The search result contains the documents in encrypted form. The data user must send request for the key to the owner of data. Data owner sends the key to decrypt the documents.

IV CONCLUSION:

We have developed a MRSE-CSI model based on cosine similarity based clustering and word relevance technique. We analyze the search efficiency of the system using two clustering algorithms. The experimental result proves that the time needed to form the clusters of relevant documents is reduced by using cosine similarity clustering. The experimental result also proves that the speed of the search increases by using the cosine similarity clustering algorithm. The proposed architecture improves the search efficiency and rank security.

V REFERENCES:

[1] J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, and R. Buyya, “Ensuring security and privacy preservation for cloud data services,” ACM Computing Surveys, 2016.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I.Stoica, and M. Zaharia, “A view of cloud computing,” Communications of the ACM, vol. 53, no. 4, pp. 50–58, 2010.

[3] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: Improved definitions and efficient constructions,” in Proceedings of the 13th ACM Conference on Computer and Communications Security. ACM, 2006, pp. 79–88.

Volume No:4, Issue No:2 (July-2018)
ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

[4] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in Cryptology- Eurocrypt 2004. Springer, 2004, pp. 506–522.

[5] Z. Ying, H. Li, J. Ma, J. Zhang, and J. Cui, "Adaptively secure ciphertext-policy attribute-based encryption with dynamic policy updating," Sci China Inf Sci, vol. 59, no. 4, pp. 042 701:1–16, 2016.

[6] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. SP 2000. Proceedings. 2000 IEEE Symposium on, 2000, pp. 44–55.

[7] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003.

[8] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Applied Cryptography and Network Security. Springer, 2005, pp. 442–455.

[9] Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Pairing-Based Cryptography–Pairing. Springer, 2007, pp. 2–22.

[10] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Applied Cryptography and Network Security. Springer, 2004, pp. 31–45.

Author's Profile:

Mabjan Atthar
Pursuing M.Tech at MJR College of Engineering & Technology, Department of CSE, Piler, Chittoor Dist.

Reddi Durga Sree
Working as a Assistant Professor in MJR College of engineering & technology, Department Of CSE, Piler, Andhra Pradesh.

Karamala Suresh
Working as a Head of the Department in MJR College Of Engineering And Technology, Department Of Cse, Piler, Chittoor dist. He is having 14 years of teaching experience in engineering colleges, He received B.Tech(CSE) from JNTU Hyderabad in 2002,Received M.E(CSE) from satyabama university Chennai in 2006,and Received M.Tech(CSE) in 2015 from JNTU Anantapuramu, he published several research papers in various national and international journals. He is interested in Computer Networks.