PRACTICAL TECHNIQUES FOR SEARCHES ON ENCRYPTED DATA IN CLOUD SERVICES

[#1]MD.KHALEEFA, M.Tech Student,
[#2]MALSOOR, Associate Professor,
[#3]Dr.M.SUJATHA, Associate Professor,
Department Of CSE,
JYOTHISHMATHI INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR,T.S.INDIA.

ABSTRACT: Cloud computing is a technology, which provides low cost, scalable computational capacity. The storage and access of document have been major problem in this area. While, many schemes have been proposed to perform conjunctive keyword search ,less attention has been noted. The storage and access of confidential documents are identified together of the central issues within the space. especially, several researchers investigated solutions to go looking over encrypted documents keep on remote cloud servers. whereas several schemes are planned to perform conjunctive keyword search, less attention has been noted on additional specialized looking out techniques. During this paper, we have a tendency to gift a phrase search technique supported Bloom filters that's significantly quicker than existing solutions, with similar or higher storage and communication price. In this project, at the time of file uploading on cloud we check file deduplication. We store only unique files on cloud. Using MD5 Algorithm We check file deduplication. File deduplication checking is used for cloud storage management. Our technique uses a series of n-gram filters to support the practicality. The theme exhibits a trade-off between storage and false positive rate, and is filmable to defend against inclusion-relation attacks. A style approach supported Associate in Nursing application's target false positive rate is additionally represented. Secure data deduplication can significantly reduce the communication and storage overheads in cloud storage services, and has potential applications in our big data-driven society.

Keywords:—Conjunctive keyword search, file deduplication, Phrase search, Privacy, Security, Encryption.

I. INTRODUCTION

CLOUD computing has emerged as a disruptive trend in both IT industries and research communities recently, its salient characteristics like high scalability and pay-as you-go fashion have enabled cloud consumers to purchase the powerful computing resources as services according to their actual requirements, such that cloud users have no longer need to worry about the wasting on computing resources and the complexity on hardware platform management. Nowadays, more and more companies and individuals from a large number of big data application shave outsource their data and deploy their services into cloud servers for easy data management, efficient data mining and query processing tasks. Data encryption has been widely used for data privacy preservation in data sharing scenarios, it refers to mathematical calculation and algorithmic scheme that transform plaintext into cipher text, which is a non-readable form to Secure data deduplication can significantly reduce the communication and storage overheads in cloud storage services, And has potential applications in our big data-driven society. Unauthorized parties. A variety of data encryption models have been proposed and they are used to encrypt the data before outsourcing to the cloud servers. However, applying these approaches for data encryption usually cause tremendous cost in terms of data utility, which makes traditional data processing methods that are designed for plain text data no longer work well over encrypted data . Secure data deduplication can significantly reduce the communication and storage overheads in cloud storage services, and has potential applications in our big data-driven society. Data encryption has been wide used for knowledge privacy preservation in knowledge sharing situations, it refers to mathematical calculation and algorithmic theme that remodel plaintext into cipher text, that may be a non-readable type to unauthorized parties. a spread of knowledge secret writing models have been planned and that they square measure accustomed write in code the data before outsourcing to the cloud servers. However, applying these approaches for encryption typically cause tremendous price in terms of knowledge utility, that makes traditional processing strategies that square measure designed for plaintext knowledge now not work spill encrypted knowledge. Data encryption has been wide used for knowledge privacy preservation in knowledge sharing situations, it refers to mathematical calculation and algorithmic theme that remodel plaintext into cipher text, that may be a non-readable type to unauthorized parties. a spread of knowledge secret writing models have been planned and that they square measure accustomed write in code the data before outsourcing to the cloud servers. However, applying these approaches for encryption typically cause tremendous price in terms of knowledge utility, that makes traditional processing strategies that square measure designed for plaintext knowledge now not work spill encrypted knowledge.

II. BACKGROUND

Boneh et al. proposed one of the earliest works on keyword searching. Their scheme uses public key encryption to allow keywords to be searchable without revealing data content. Waters et al. investigated the problem for searching over encrypted audit logs. Many of the early works focused on single keyword searches. Recently, researchers have proposed solutions on conjunctive keyword search, which involves multiple keywords. Other interesting problems, such as the ranking of search results and searching with keywords that might contain errors termed fuzzy keyword search, have also been considered. The ability to search for phrases was also recently investigated. The ranking of search results was looked at by Wang. The authors described a solution based on the commonly used TFIDF (Term Frequency X Inverse Document Frequency) rule and the use of order preserving symmetric encryption. Liuet considered the search for potentially erroneous keywords termed fuzzy keyword search. The index-based solution makes use of fuzzy dictionaries containing various misspelling of keywords including wildcards. Some of the existing system has examined the security of the proposed solutions and, where flaws were found, solutions were proposed.

A Secure and Dynamic Multi Keyword Ranked Search Scheme over encrypted.

The major aim of this paper is to resolve the matter of multi-keyword hierarchical search over encrypted cloud knowledge (MRSE) at the time of protective actual technique wise privacy within the cloud computing construct. knowledge holders area unit inspired to source their tough knowledge management systems from native sites to the business public cloud for big flexibility and monetary savings. but for protecting knowledge privacy, sensitive knowledge got to be encrypted before outsourcing, which performs ancient knowledge utilization supported plaintext keyword search. As a result, permitting Associate in Nursing encrypted cloud knowledge search service is of supreme significance. visible of the massive range of information users and documents within the cloud, it's essential to allow many keywords within the search demand and come back documents within the order of their acceptable to those keywords. Similar mechanism on searchable cryptography makes centre on single keyword search or Boolean keyword search, and infrequently type the search results. within the middle of various multi-keyword linguistics, deciding the well-organized similarity live of coordinate matching, it means as several matches as doable, to capture the suitable knowledge documents to the search question. notably, we consider dot product similarity i.e., the number of question keywords shows in a document, to quantitatively estimate such match live that document to the search question. Through the index construction, each document is connected with a binary vector as a sub index wherever every bit characterize whether or not matching keyword is contained within the document.

III. COMMUNICATION FRAMEWORK

We'll describe our keyword search framework using two parties: The data owner and an untreated cloud server. Our algorithms can easily be adapted to the scenario of an organization wishing to setup a cloud server for its employees by implementing a proxy server in place of the data owner and having the employees/users authenticate to the proxy server. A standard keyword search protocol is shown in figure. During setup, the data owner generates the required encryption keys for hashing and encryption operations. Then, all documents in the database are parsed for keywords. Bloom filters tied to hashed keywords and grams are attached. The documents are then symmetrically encrypted and uploaded to the cloud server. To add files to the database, the data owner parses the files as in setup and uploads them with Bloom filters attached to the cloud server. To remove a file from the data , the data owner simply sends the request to the cloud server, who removes the file along with the attached Bloom filters. Security In terms of security, we assume a semi-honest cloud server, which is interested in learning about stored data but will follow our keyword search protocol as described and will not modify or misrepresent any data in order to gain an advantage. Two of the main security issues regarding keyword searches are the privacy of the document sets and the privacy of the queried keywords. Briefly, a secure keyword search protocol should prevent the cloud server from obtaining non-negligible amount of information on the stored documents or the keywords in the query requests. Note that, in our target application, users are employees of the data owner's organization and are authorized to search for any documents in the data set. Should an application requires that users be restricted from accessing certain files, an access control system such as would be required to verify the matched results and returned only those which the user has the required credential to access. Our basic scheme in section achieves these goals under the assumption that the cloud has no prior knowledge on the stored data. Should the cloud provider has significant statistical knowledge on the stored data, such as the distribution of the keywords, it may be able to infer partial knowledge on its content. Under the security model where the cloud provider has some knowledge over the distribution of keywords or queries on the stored data, we describe modifications to the basic scheme which would offer protection against statistical attacks in section 4.6 and inclusion-relation.

IV. PROPOSED METHOD

In this paper, we present a phrase search scheme which achieves a much faster response time than existing solutions. The scheme is also scalable, where documents can easily be removed and added to the corpus. We also describe modifications to the scheme to lower storage cost at a small cost in response time and to defend against cloud providers with statistical knowledge on stored data.Although phrase searches are processed independently using our technique, they are typically a specialized function in a keyword search scheme, where the primary function is to provide conjunctive keyword searches. Therefore, we describe both the basic conjunctive keyword search algorithm and the basic phrase search algorithm along with design techniques.

SCOPE:

• The scheme is also scalable, where documents can easily be removed and added to the corpus.

• We also describe modifications to the scheme to lower storage cost at a small cost in response time and to defend against cloud providers with statistical knowledge on stored data.

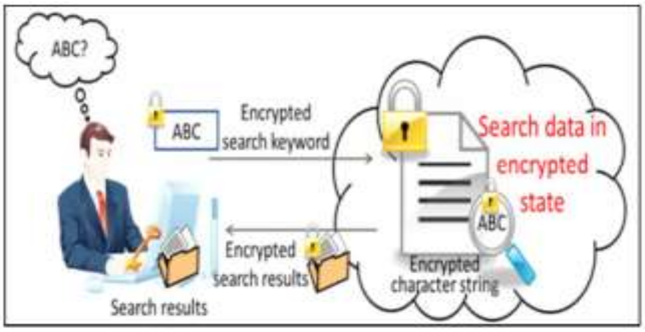

Fig.1: Proposed system architecture.

Advantage Of Proposed System

• Our framework differs from some of the earlier works, where keywords generally consist of meta-data rather than content of the files and where a trusted key escrow authority is used due to the use of Identity based encryption.

- When compared to recent works, where an organization wishes to outsource computing resources to a cloud storage provider and enable search for its employees, where the aim is to return properly ranked files.

- Most other recent works related to search over encrypted data have considered similar models such as, where the client acts as both data owner and user.

IV. Queries on Long Phrases

Long phrase queries are often used to locate known items rather than to locate resources for a general topic. In many cases, the goal is to identify a single document. Longer phrases also have a very low probability of occurrence and yield fewer matches. Therefore, even with a precision rate of 50%, we would rarely see more than a single false positive for a search query of longer phrases. In our experiment, we never encountered more than a single false positive in queries with phrases containing more than 4 keywords. The small number of false positives can also be easily identified and removed client-side. As a result, the effect of low precision rate in longer phrases should not have a noticeable detrimental effect in practice.

V. CONCLUSION

Our approach is also the first to effectively allow phrase search to run independently without first performing a conjunctive keyword search to identify candidate documents. The technique of constructing a Bloom filter index introduced in section 4.2 enables fast verification of Bloom filters in the same manner as indexing. According to our experiment, it also achieves a lower storage cost than all existing solutions except [13], where a higher computational cost was exchanged in favor of lower storage. While exhibiting similar communication cost to leading existing solutions, the proposed solution can also be adjusted to achieve maximum speed or high speed with a reasonable storage cost depending on the application. An approach is also described to adapt the scheme to defend against inclusion-relation attacks. Various issues on security and efficiency, such as the effect of long phrases and precision rate, were also discussed to support our design choices.

References

[1] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in In proceedings of Eurocrypt,2004, pp. 506–522.

[2] B. Waters, D. Balfanz, G. Durfee, and D. K. Smetters, “Building an encrypted and searchable audit log,” in Network and Distributed System Security Symposium, 2004.

[3] M. Ding, F. Gao, Z. Jin, and H. Zhang, “An efficient public key encryption with conjunctive keyword search scheme based on pairings,” in IEEE International Conference onNetwork Infrastructure and Digital Content, 2012, pp. 526–530.

[4] F. Kerschbaum, “Secure conjunctive keyword searches for unstructured text,” in International Conference on Network and System Security, 2011, pp. 285–289.

[5] F. Kerschbaum, “Secure conjunctive keyword searches for unstructured text,” in International Conference on Network and System Security, 2011, pp. 285–289.

[6] C. Hu and P. Liu, “Public key encryption with ranked multikeyword search,” in International Conference on Intelligent Networking and Collaborative Systems, 2013, pp. 109–113.

[7] Z. Fu, X. Sun, N. Linge, and L. Zhou, “Achieving effective cloud search services: multi-keyword ranked search over encrypted cloud data supporting synonym query,” IEEE Transactions on Consumer Electronics, vol. 60, pp. 164–172, 2014.

[8] C. L. A. Clarke, G. V. Cormack, and E. A. Tudhope, “Relevance ranking for one to three term queries,” Information Processing and Management: an International Journal, vol. 36, no. 2, pp. 291–311 , Jan. 2000.

[9] H. Tuo and M. Wenping, “An effective fuzzy keyword search scheme in cloud computing,” in International Conference on Intelligent Networking and Collaborative Systems, 2013, pp. 786–789.

[10] M. Zheng and H. Zhou, “An efficient attack on a fuzzy keyword search scheme over encrypted data,” in International Conference on High Performance Computing and Communications and Embedded and Ubiquitous Computing, 2013, pp. 1647–1651.

[11] S. Zittrower and C. C. Zou, “Encrypted phrase searching in the cloud,” in IEEE Global Communications Conference, 2012, pp. 764 – 770

[12] C. Liu, L. Zhu, L. Li, and Y. Tan, “Fuzzy keyword search on encrypted cloud storage data with small index,” in 2011 IEEE International Conference on Cloud Computing and Intelligence Systems, 2011, pp. 269–273.

[13] P. F. Brown, P. V. deSouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai, “Class-based n-gram models of natural language,” Computational Linguistics, vol. 18, no. 4, pp. 467–479, 1992.

[14] D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics. Prentice Hall, 2009