Multiple Keyword Search Over Encrypted Data On Cloud

P. Poorvaja[1], K. Priyadharshini[2], P. Rama[3], S. Vaishnavi[4]
[1, 2, 3, 4] Dept of Information Technology
[1, 2, 3, 4] Saranathan college of engineering, Tiruchirapalli-620012, Tamil Nadu, India

Abstract- *A trouble-free, colossal, and expandablestorage is provided by Cloud Storage at low cost, but the major concern is data privacy. It prevents users from storing files on the cloud trustingly. The way to enhance privacy from data owner point of view is by encrypting the files before outsourcing them onto the cloud and decrypt the files after downloading them. However, for mobile devices, there is a heavy overhead for data encryption and data retrieval process which incurs a complicated communication between data user and cloud. Generally, with limited bandwidth capacity and limited battery life, these issues introduce heavy overhead for computing and communication as well as a higher power consumption for mobile device users, which makes the encrypted search very challenging on mobile cloud. In this paper, we propose TEES, a bandwidth and energy efficient encrypted search architecture over mobile cloud. Here, the proposed architecture offloads the computation from mobile devices to the cloud, and further optimization has been done on the communication between the mobile clients and the cloud. It is illustrated that the data privacy does not degrade when the performance enhancement methods are applied.*

Keywords- TEES (Traffic and Energy saving EncryptedSearch), multiple keyword search, privacy preserving.

I. INTRODUCTION

Cloud storage system is a service model in which data are maintained, managed and backed up remotely on the cloud side, and meanwhile data keeps available to the users over a network. Mobile Cloud Storage (MCS)denotes a family of increasingly popular on-line services, and even acts as the primary file storage for the mobile devices [3]. MCS enables the mobile device users to store and retrieve files or data on the cloud through wireless communication, which improves the data availability and facilitates the file sharing process without draining the local mobile device resources [4]. The data privacy issue is paramount in cloud storage system, so the sensitive data is encrypted by the owner before outsourcing onto the cloud, and data users retrieve the interested data by encrypted search scheme. In MCS, the modern mobile devices are confronted with many of the same security threats as PCs, and various traditional data encryption methods are imported in MCS [5], [6]. However, mobile cloud storage system incurs new challenges over the traditional encrypted search schemes, in consideration of the limited computing and battery capacities of mobile device, as well as data sharing and accessing approaches through wireless communication. Hence, a suitable and efficient encrypted search scheme is necessary for MCS. Generally speaking, the mobile cloud storage is in great need of the bandwidth and energy efficiency for data encrypted search scheme, due to the limited battery life and payable traffic fee. Therefore, we focus on the design of a mobile cloud scheme that is efficient in terms of both energy consumption and the network traffic, while keep meeting the data security requirements through wireless communication channels. To this end, we introduce TEES (Traffic and Energy Saving Encrypted Search) architecture for mobile cloud storage applications. TEES achieves the efficiencies through employing and modifying the ranked keyword search as the encrypted search platform basis, which has been widely employed in cloud storage systems. Traditionally, two categories of encrypted search methods exist, that can enable the cloud server to perform the search over the encrypted data: ranked keyword search and Boolean keyword search. The ranked keyword search adopts the relevance scoresto represent the relevance of a file to the searched keyword and sends the top-k relevant files to the client. It is more suitable for cloud storage than the boolean keyword search approaches, since boolean keyword search approaches need to send all the matching files to the clients, and therefore incur a larger amount of network traffic and a heavier post-processing overhead for the mobile devices. By careful redesign of ranked keyword search procedure, TEES offloads the security calculation to the cloud side to save the energy consumption of mobile devices, and TEES also simplify the encrypted search procedure to reduce the traffic amount for retrieving data from encrypted cloud storage. Besides the energy and traffic efficiencies, TEES is implemented with security enhancement in consideration of the modified encrypted search procedure in order to mitigate statisticsinformation leak and keywords-files association leak [2] [3] for MCS, by adding noise in Term Frequency distribution function and keeping the Order Preserving Encryption attributes.

II. LITERATURE SURVEY

AnkathaSamuyelu Raja Vasanthi ," Secured Multi keyword Ranked Search over Encrypted Cloud Data", 2012

In [1], main aim is to find the solution of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. A variety of multi- keyword semantics are available, an efficient similarity measure of "coordinate matching" (as many matches as possible), to capture the data documents' relevancy to the search query is used. Specifically, "inner product similarity", i.e., the number of query keywords appearing in a document, to quantitatively evaluate such similarity measure of that document to the search query is used in MRSE algorithm. The main limitation of this paper was, the user's identity (ID) is not kept hidden. Due to this, whoever puts the data on Cloud Service Provider was known. This may be risky in some situations where confidentiality of data need to be maintained. Hence, this drawback is overcome in the proposed system.

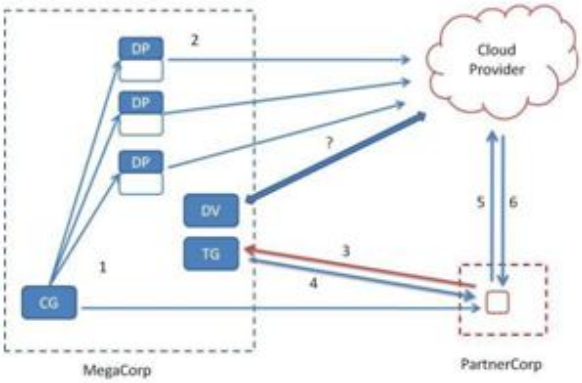
Y.-C. Chang and M. Mitzenmacher, "PrivacyPreserving Keyword Searches on Remote EncryptedData," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.

Consider the problem: a user U wants to store his files in an encrypted form on a remote file server S. Later the user U wants to efficiently retrieve some of the encrypted files containing specific keywords, keeping the keywords themselves secret and not to endanger the security of the remotely stored files. For example, a user may want to store old e-mail messages encrypted on a server managed by Yahoo or another large vendor, and later retrieve certain messages while travelling with a mobile device. In [2], solutions for this problem under well-defined security requirements are offered. The schemes are efficient as no public-key cryptosystem is involved. Indeed, the approach is independent of the encryption method chosen for the remote files. They are incremental too. In that, user U can submit new files which are secure against previous queries but still searchable against future queries. From this, the main theme taken is of storing data remotely on other server and retrieving that data from anywhere via mobile, laptop etc

S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. FinancialCryptography and Data Security, Jan. 2010.

Many organizations are outsourcing their storage and computing needs with the advances in networking technology and an increase in need for computing resources. When the benefits of using a public cloud infrastructure are clear, it introduces significant security and privacy risks. In fact, it seems that the biggest obstacle to the adoption of cloud storage (and cloud computing in general) is concern over the confidentiality and integrity of data. In [3], an overview of the benefits of a cryptographic storage service, for example, reducing the legal exposure of both customers and cloud providers, and achieving regulatory compliance is provided. Besides this, cloud services that could be built on top of a cryptographic storage service such as secure backups, archival, health record systems, secure data exchange and e-discovery is stated briefly.

Architecture of a Cryptographic Storage Service:



This architecture consists of three main components, they are a data processor (DP), that processes data before it is sent to the cloud; a data verifier (DV), that checks whether the data in the cloud has been tampered with; and a token generator (TG), that generates tokens that enable the cloud storage provider to retrieve segments of customer data; and a credential generator that implements an access control policy by issuing credentials to the various parties in the system

Y. Prasanna, Ramesh," Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data",2012.

On one hand, users who do not necessarily have prior knowledge of the encrypted cloud data, have to post process every retrieved file in order to find ones most matching their interest; On the other hand, invariably retrieving all files containing the queried keyword further incurs unnecessary network traffic, which is absolutelyundesirable in today's pay-as-you-use cloud paradigm. This paper has defined and solved the problem of effective yet secure ranked keyword search over encrypted cloud data [4]. 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. For the first time, the paper has defined and solved the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. The proposed ranking method proves to be efficient to return highly relevant documents corresponding to submitted search terms. The idea of proposed ranking method is used in our proposed system in order to enhance the security of data on Cloud Service Provider

Jain Wang, Yan Zhao, Shuo Jaing, and Jaijin Le,"ProvidingPrivacyPreservinginCloudComputing",2010.

Privacy is an important issue for cloud computing, both in terms of legal compliance and user trust and needs to be considered at every phase of design. The [5] paper tells the importance of protecting individual's privacy in cloud computing and provides some privacy preserving technologies used in cloud computing services. Paper tells that it is very important to take privacy into account while designing cloud services, if these involve the collection, processing or sharing of personal data. From this paper, main theme taken is of preserving privacy of data. This paper only describes privacy of data but doesn't allow indexed search as well as doesn't hide user's identity. Thus, these two drawbacks are overcome in our proposed system.

AnuradhaMeharwade,G.A.Patil, "Efficient Keyword Search over Encrypted Cloud Data"

Cloud computing provides a secured data outsourcing and high-quality data services. Therefore, many sensitive information like personal health records, company finance data, and government documents are being outsourced to cloud. Since the sensitive data may be leaked to unauthorized entities, data has to be encrypted before outsourcing. This protects data to some extent, but at the cost of compromised efficiency. In cloud computing, data owners share their outsourced data with large number of data users, who will retrieve the specific file of their interest. This can be done through keyword-based search; this allows users to selectively retrieve files of interest. There may be files that contain more keyword related information with less frequency of occurrence of akeyword in the document. Therefore, here we are using concept of ranked searchable symmetric encryption model with some advancement and presented a system which will efficiently retrieve files containing information related to specified keyword in rank order from an encrypted file collection. As a result the application developed support the secure and efficient data retrieval.

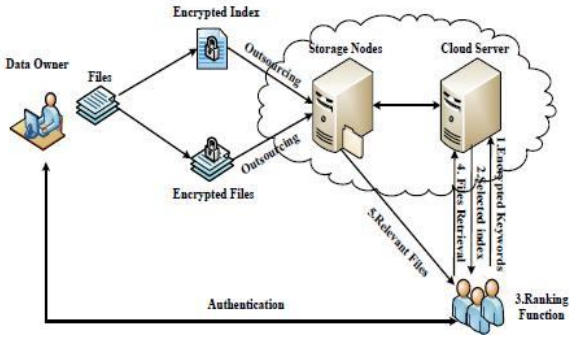
III. RELATED WORKS

Encrypted Search Schemes:

Over the past recent years, encrypted search has evolved towards the ability of data sharing with protection of user's privacies. Up to now, encrypted search includes Boolean keyword search and ranked keyword search. In Boolean keyword search [6] the server sends back files only based on the existence or absence of the keywords, without looking at their relevance.

Traditional Encrypted Search over Cloud Data:

Traditional cloud storage system architecture and general procedures are shown in Figure 1, which include: file/index encryption by the data owner, outsourcing the data to the cloud storage, and encrypted data search/retrieval procedure of the data users in cloud computing, intersection of artificial intelligence, machine learning, statistics, and systems.



The overall goal of the data mining process is to extract information from a dataset and transform it into an understandable structure. It involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post processing of discover structures, visualization and online updating.

File/Index Encryption:

The data owner first executes the preprocessing and indexing work as shown on Figure 2. He should invert files, that are selected to store on the cloud, for text search engines [6]. Every word in these files undergoes stemming to retain the word stem. After this step, the data ownerencrypts and hashes

every term (word stem) to fix its entry in the index. The index is then created by the data owner. Finally, the data owner encrypts the index and stores it into the cloud server, together with the encrypted file set. Most of the previous schemes under this architecture use Order Preserving Encryption (OPE) [4] to encrypt the file index. This file index is often a TF (Term Frequency) table composed of TF values [19]. The TF-IDF table could be used to determine word relevance in documents

IV. SYSTEM ARCHITECTURE

This system shows that the data owner encrypts and uploads the file to the cloud server while encrypted index will be created. The data user searches for the corresponding file he/she wants for, using a keyword or multiple keyword. The corresponding files will be retrieved based on the ranked search algorithm and top-k relevant files will be retrieved. The user can select his/her preferred file and send request to the data owner for the password to download that file.The data owner now authenticates the user and sends the one-time password through registered mail id to the data user for downloading that file. Using the one-time password provided by the data owner, the data user downloads the file.

Searches incur following steps,

1. An authenticated user stems the keyword to be queried, encrypts it with the keys and hashes it to get its entry in the index. Then the encrypted keyword is sent to the cloud server.
2. On receiving the encrypted keyword, the cloudserver first searches for it in the index. Then the index related to this keyword is sent back to the data user.
3. The data user calculates the relevance scores with the selected index to find the top-k relevant files and sends a follow-up request to the cloud server in order to retrieve the files.
4. The position of these files is selected and they are sent back to the data user from the cloud server.
5. The data user decrypts the files and recovers the original data.

V. PROPOSED SYSTEM

TEES (Traffic and Energy saving Encrypted Search) architecture for mobile cloud storage applications is proposed. TEES achieves the efficiencies through employing and modifying the ranked keyword search as the encrypted search platform basis, which has been widely employed in cloud storage systems.Traditionally, two categories of encrypted search methods exist, that can enable the cloud server to perform the search over the encrypted data: ranked keyword search and boolean keyword search. The ranked keyword search adopts the relevance scores to represent the relevance of a file to the searched keyword and sends the top-k relevant files to the client. It is more suitable for cloud storage than the boolean keyword search approaches since boolean keyword search approaches need to send all the matching files to the clients, and therefore incur a larger amount of network traffic and a heavier post-processing overhead for the mobile devices.

TEES employs the architecture redesign over traditional encrypted search procedure, and our comprehensive experiments prove the TEES has following advantages in comparison with the traditional complex encrypted search procedure:

1. TEES reduces the energy consumption by35%_55% by offloading the computation of the relevance scores to the cloud server. This reduces the computing workload on the mobile device side while at the same time significantly speeding up the mobile file access speed (e.g. it doubles the speed for accessing a 100KB file).
2. With a simplified search and retrieval process, TEES reduces the network traffic for the communication of the selected index, and reduces the file retrieval time by 23%_46% in our experiments.
3. In implementing the redesigned encrypted search procedure, TEES redistributes the encrypted index to avoid statistics information leak, and wraps keywords adding noise in order to render them indistinguishable to the attackers.

VI. CONCLUSION AND FUTURE WORK

In this paper we proposed a efficient method to authenticate to moving KNN query. Result proved that the method is VO-optimal ,i.e.., the verification object has the minimal size with respect to the given tree. Developed optimization techniques that can further reduce the computation cost communication frequency and cost between a moving client and the LBS further more extended the

solution to handle moving KNN queries that involve multiple data sets. The experimental results show that the authentication method achieves low communication cost and CPU overhead.

REFERENCES

[1] AnkathaSamuyelu Raja Vasanthi," Secured Multi keyword Ranked Search over Encrypted Cloud Data", 2012
[2] Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.
[3] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptography and Data Security, Jan. 2010.
[4] Y. Prasanna, Ramesh. " Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data", 2012.
[5] Jain Wang, Yan Zhao, ShuoJaing, and Jaijin Le," Providing Privacy Preserving in Cloud Computing",2010.
[6] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. S&P 2000.Proceedings.2000 IEEE Symposium on. IEEE, 2000, pp. 44–55. Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 4, Issue 5, May 2017.
[7] Jiawei Han, MichelineKamber, Jian Pei, "Data Mining Concepts and Techniques" Third Edition.