IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 01-05
www.iosrjournals.org

Survey on Efficient Keyword Search Scheme over Encrypted Data on Mobile Cloud

Anju Chandran, Asso. Prof. P Krishna kumaran Thampi

Final Year MTech Degree Dept. of Computer Science and Engineering Narayana Gurukulam College of Engineering Kerala, India
Dept. of Computer Science and Engineering Sree Narayana Gurukulam College of Engineering Sree Kerala, India

Abstract: *Mobile cloud storage is a service model in which data are maintained, managed and backup remotely on the cloud side and meanwhile data keeps available to the user over network. To store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risk. Encryption and decryption are overhead for mobile devices because it is very small and have low computation power. TEES (Traffic and Energy Efficient Storage) offloads computation from mobile devices to cloud to reduce the power consumption and bandwidth. Data privacy does not degrade when the performance enhancement methods are applied. The network traffic also significantly reduced during file retrieval. This paper presents a study on Traffic and Energy Efficient Search over encrypted data on mobile cloud.*

Keywords:*TEES, Mobile Cloud Storage, Energy Efficient, Traffic Efficient*

I. Introduction

Clouds are a large pool of easily usable and accessible virtualized resources and services. Cloud storage is a model of online data storage and access data from multiple distributed and connected resources that comprise a cloud over network. Mobile Cloud Storage (MCS) provides services to users to store and retrieve data or files on the cloud through wireless communication. MCS is increasingly popular online service which facilitates the file sharing process without draining the local mobile device resources and high data availability.

The data privacy issue is the utmost important in cloud storage system, so data is encrypted by the owner prior to outsourcing on to cloud, and retrieve data by encrypted search scheme. MCS also imported many traditional data encryption methods. But the traditional encryption methods incurs new challenges, in consideration of limited computing power, battery capacities, data sharing and accessing approaches through wireless communication. So the design of mobile cloud storage scheme should be efficient in both energy consumption and the network traffic with security requirements through wireless communication.

TEES (Traffic and Energy Efficient Search) introduces architecture to achieve efficiencies of mobile cloud storage application. TEES employs and modifies ranked keyword search scheme over encrypted data on mobile cloud storage system. Traditionally, two categories exist for encrypted keyword search: ranked keyword search and Boolean keyword search. The ranked keyword search sends top-k relevant files to client. The Boolean keyword search sends all the matching files to the client, which causes a larger amount of network traffic. So ranked keyword search is most suitable for the mobile cloud storage.

The TEES architecture with ranked keyword search offloads the security calculation to the cloud to save energy consumption of mobile devices and simplify the encrypted search procedure to reduce the traffic for retrieving the data from encrypted cloud storage.

To mitigate the statistics information leak and keyword-files association leak TEES implements the advanced security scheme for modified search procedure by adding noise in Term Frequency distribution function and keeping Order Preserving Encryption attributes. The advantages of TEES in comparison with the traditional complex encrypted search procedure

1) TEES reduces the energy consumption by offloading the computation of the relevance scored to the cloud search.
2) TEES reduces the network traffic for the communication of the selected index and reduces the file retrieval time.
3) TEES redistributes the encrypted index to avoid statistics information leak and wraps keywords adding noise in order to render them indistinguishable to the attackers. TEES guarantees enhanced security level for MCS wireless communication.

II. Study Of File Retrieval In Mobile Cloud Storage

Traditional cloud storage system includes file/ index encryption by data owner then outsource the data to the cloud storage and encrypted search/retrieval procedure of the data users in cloud computing.

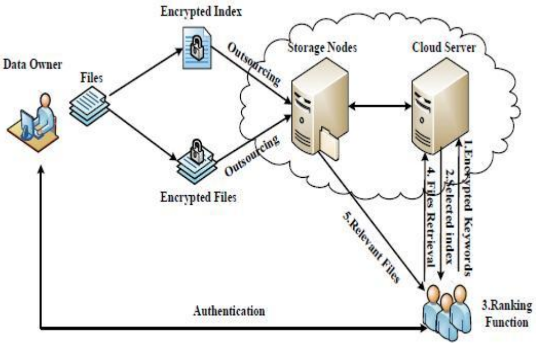

Fig. 1. Traditional encrypted search Architecture

The data owner executes the preprocessing and indexing work for that he should invert files to store on the cloud for search engines as shown in figure 2. Every word in the file undergoes stemming to retain the word stem. After that data owner encrypt and hashes every word stem. Then data owner creates the index and store the encrypted index in to the cloud with encrypted file set.

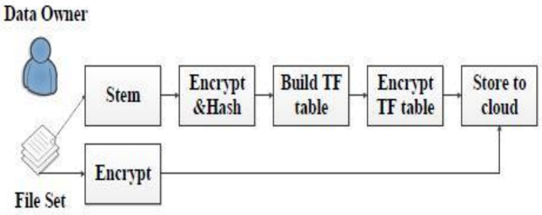

Fig. 2. Process of Preprocessing and Indexing

The data user authenticated by the data owner for accessing files. The data owner checks the identity provided by the data user and sends the encrypted key back, if the user is a legal user.

The search and retrieval process are illustrated in the figure 1. An authenticated user stems the keyword, encrypt with the key and hashes to get the index and send the encrypted keyword to the cloud. The cloud server searches for the keyword and send back the keyword related index to the user. The user calculates relevant score with selected index to find the top-k relevant files to follow up request to the cloud in order to retrieve files. The position of these files is selected and sent back to the data user from cloud server. After that the data user decrypts the file and recovers the original data.

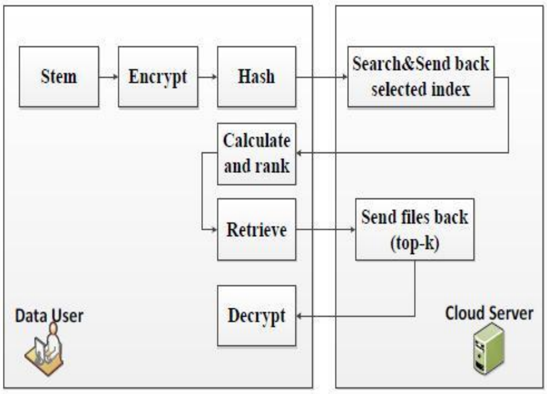

Fig. 3. TRS Two Round Trip Encrypted search

The above process illustrated in the figure 3. Traditional two round trip scheme is implemented for the authenticated user. This scheme is very complicated and requires more computation power. It is more complex than the simple Plain text search scheme, where searching and retrieving of file is done in one round trip time without security service.

The TRS scheme is suitable for the users with personal computer, but in the case of mobile, the computation in the user side incur heavy burden. To solve these problems a novel idea is introduced Traffic and Energy Efficient Search.

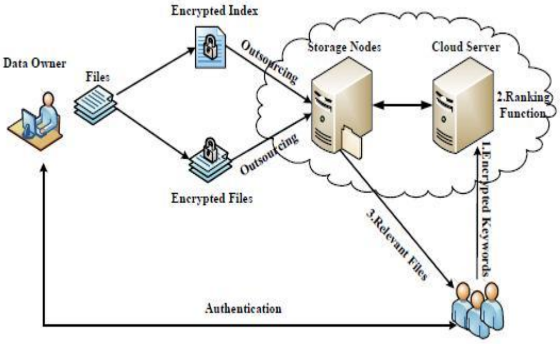

Fig. 4. Encrypted Search Architecture of TEES

The figure 4 shows the architecture of TEES. Compared to the traditional architecture the computation in the user side, to find the top-k files in the index table is offloaded to the cloud server. This helps to reduce the energy consumption, file search and retrieval time and traffic overhead.

During the preprocessing and indexing stages data owner gets TF table as index and uses Order Preserving Encryption. The cloud server calculates the relevance score and ranks them without decrypting the index. The data user wants to retrieve the top-k relevant files based on the encrypted keyword. The data user authenticated by the data owner using the user provided details.

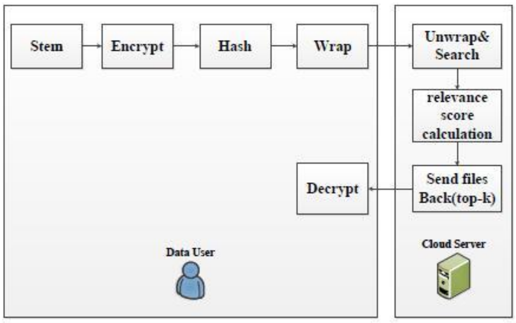

Fig. 5. ORS: Novel Process of Search and Retrieval

If an authenticated data user tries to retrieve top-k relevant files, the data owner send back key to encrypt the keyword. The data user encrypt the keyword with key, then wrap the encrypted keyword to a tuple to add noise to avoid statistical information leak; this tuple is used to perform the retrieval. On receiving the wrapped keyword, the cloud server first makes sure that it is accessed by a legal user. If the server is notified by the data owner that this user is to become invalid in a near future, the search is performed but a warning is also issued. If this is a legal user, the server unwrap the tuple to recover the entry of the keyword and searches for it in the index. After calculating the relevance scores, the position of the files corresponding to the keyword is picked and the top-k relevant files are sent back to the data user's mobile clients without performing any decryption on these files. The data user decrypts these files in the mobile client and recovers the original data.

III. Designing Of Tees System

The proposed system introduces modified algorithms to achieve security enhancement with energy and traffic efficiency. This system builds in the aspects of the data owner, data user and cloud server. The data owner builds the TF table as index, OPE algorithm for encryption. The cloud server implements both unwrap and rank function.

3.1 Data Owner

The data owner is responsible for building index table, encryption and authentication process. TF-IDF (Term Frequency –Inverse Document frequency) is a product of term frequency and inverse document frequency. The term frequency is the number of times that term occurs in the document. The inverse document frequency measures the term common or rare in the whole document.

To build index the data owner collects the files to store in to the cloud. Then extract the terms from the file, then encrypt and hash the term to store in to TF table. After that calculate the frequency of term and compute and store the index.

OPE (Order Preserving Encryption) method is used for encrypting the data. This approach uses one-to-many mapping to map every TF value to a random number in a certain range. This helps to prevent the attacker from collecting statistical information from the TF table. The OPE algorithm is very simple and consumes less energy.

The data owner maintains legal user set and invalid user set. At the time of authentication data owner checks the identity of the data user. If the user belongs to the legal set then data owner sends the keys and hash table to the user. The data owner checks the International Mobile Equipment Identity of user’s mobile and stores its encrypted version. The data owner updates the legal table periodically to ensure the security.

3.2 DATA USER

The data user module executes in the mobile client side. The data user wraps the keyword using some random number after hashing. The wrap function adds some noise to the keyword to control the keyword-files association leak. The wrapped keyword sends to the relevance score calculation to the cloud server. The user decrypts the files corresponding to the encryption done by the data owner. The authentication function used for the user authentication.

3.3 CLOUD SERVER

The cloud server unwrap the keyword and searches into the TF table. The cloud server calculates the relevance scores and returns top-k relevant files to the legal data user.

IV. Related Work

In 2000 D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data” proposed techniques for searches on encrypted data. This scheme uses primitives from classical symmetric-key cryptography to define security. The proposed scheme encrypts each word in the document separately. This encrypted search improves the confidentiality. But one who can able to use statistical technique to learn important information. To overcome the problem periodical key change is needed and re-encrypt the document. This will be a burden for mobile devices and this file encryption method is not compatible with the existing scheme and cannot deal with data compression.

In 2004 R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, “Order preserving encryption for numeric data” proposed one to one mapping OPE which will lead to Statistics information leak control. The proposed algorithm is very complex for the mobile devices.

In 2005 Y. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data” proposes single keyword search remotely. This scheme proposes pattern matching technique instead of keyword matching. It ensures security for newly submitted files and cannot ensure security of previously submitted files.

In 2009 A. Boldyreva, N. Chenette, Y. Lee, and A.

O´ neill, “Order preserving symmetric encryption” describes deterministic encryption scheme. The encryption function preserves numerical ordering of plaintext. This scheme allows efficient range queries as well as indexing query processing as efficiently as for unencrypted data.

In 2010 K. Kumar and Y. Lu, "Cloud computing for mobile users: Can offloading computation save energy?" proposes 4 approaches to saving energy and extending battery life time in mobile devices,

1. Adopting new generation of semiconductor technology
2. avoiding energy wastage
3. executes programs slowly
4. eliminating computation all together

The paper proposes offloading of computation to save energy. The disadvantage of offloading is the high risk of privacy and security. The computation offloading depends on the wireless network communication, which will cause reliability concern. The wireless communication prevented by limited connectivity and low energy efficiency. Data storage is another reliability problem.

In 2010 A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone," proposed to measure the power consumption and energy efficiency of mobile devices. This paper presents the power consumption rate of different usage scenarios and analyses the contribution of different components in the mobile devices to overall power consumption.

In 2012 C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," proposed a one-to-many mapping OPE. They implemented a complicated algorithm for security protection. However performance and energy consumption would a problem since their algorithm was complicated and need much computing resource. This scheme proposed one round trip search scheme.

V. Conclusion And Future Work

This paper is the study of searching over encrypted data with traffic and energy efficiency over mobile cloud storage. The proposed system deal with the single keyword search using OPE algorithm and TF-IDF table. It is more efficient compared to the traditional system. The basic security challenges of mobile cloud storage such as statistic information leak, keyword files association leak and secure information acquisition are handled by this system.

The proposed system only handles the single keyword search. To improve the accuracy of the result multi-keyword search should implement. The multi-keyword search improves accuracy of result and can reduce the traffic by reducing top-k relevant files.

The current OPE algorithm is very simple and it does not support the multi-keyword search. Hence multi-keyword search using a powerful algorithm with efficiency can be implementing as future work.

Reference

[1] D. Huang, "Mobile cloud computing," IEEE COMSOC Multimedia Communications Technical Committee (MMTC) E-Letter, 2011.

[2] 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.

[3] Y. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Applied Cryptography and Network Security. Springer, 2005, pp. 391–421.

[4] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," Parallel and Distributed Systems, IEEE Transactions on, vol. 23, no. 8, pp. 1467–1479, 2012.

[5] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in Proceedings of the 2004 ACM SIGMOD international conference on Management of data. ACM, 2004, pp. 563–574.

[6] K. Kumar and Y. Lu, "Cloud computing for mobile users: Can offloading computation save energy?" Computer, vol. 43, no. 4,pp. 51–56, 2010.

[7] A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone," in Proceedings of the 2010 USENIX conference on USENIX annual technical conference. USENIX Association, 2010, pp. 271–284.

[8] Wikipedia, http://en.wikipedia.org/wiki/tf-idf. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 1, pp. 222–233, 2014.

[9] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in INFOCOM, 2011 Proceedings IEEE. IEEE, 2011, pp. 829–837.

[10] A. Aizawa, "An information-theoretic perspective of tf-idf measures," Information Processing and Management, vol. 39, pp. 45–65, 2003.

[11] Jian Li, Ruhui Ma, Haibing Guan, "TEES: An Efficient Search Scheme over Encrypted Data on Mobile Cloud", IEEE Transactions on cloud computing. 2015