A Survey Paper on Secure Keyword Search for Outsourced Cloud Data

Prof. Kalpesh Prajapati

Assistant Professor
Department of Computer Engineering
Gujarat Power Engineering & Research Institute
Mehsana-384460, Gujarat, India.

Abstract: **To advance cloud computing, the community must take proactive measures to ensure security. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data. But on other hand consumers are facing serious difficulties that how to search the most suitable services from cloud. It is desirable to store data on data storage servers in encrypted form to reduce security and privacy risks. Although cloud based services offer many advantages, privacy of the outsourced data is a big concern. To mitigate this concern, it is desirable to outsource sensitive data in an encrypted form but cost of encryption process would increase the heavy computational overhead on thin clients such as resource-constrained mobile devices. However, several keyword searchable encryption method have been developed in last decade where each method have their own advantages and disadvantages. This paper gives an overview of the state of cryptographic data search and describes the different approaches of recently developed for secure keyword basis data search reported in literature.**

Keywords*: *Secure data search, cloud computing, cloud storage, data outsourcing. Information retrieval

I. INTRODUCTION

The amount of data generated by individuals and enterprises is rapidly increasing. With the emerging cloud computing paradigm, the data and corresponding complex management tasks can be outsourced to the cloud for the management flexibility and cost savings. Unfortunately, as the data could be sensitive, the direct data outsourcing would have the problem of privacy leakage. The encryption can be used, before the data outsourcing, with the concern that the operations can still be accomplished by the cloud. Cloud outsource storage is one of important services in cloud computing. Cloud users upload data to cloud servers to reduce the cost of managing data and maintaining hardware and software. To ensure data confidentiality, users can encrypt their files before uploading them to a cloud system. However, retrieving the target file from the encrypted files exactly is difficult for cloud server [1] [2] [3].

Rapidly increasing need of computational efficiency and the scalable storage technique which can handle a tons of data leads to work with the cloud computing. That offers the individuals and organizations to write the applications for the cloud platforms and storage the huge amount of data. But the storage on the local disk increases the overhead of maintenance and complexity of cloud servers. In order to manage the data the storage of data can be performed on the third party servers. These servers are specifically designed for storage services and frequent data access [4].

But data storage on the third party server leads to harm in privacy and security aspects of data therefore the cryptographic techniques are utilized for enhancing the security and privacy of data and data owner. The cryptographic data storage alters the data format using the mathematical techniques and recovers it on demand basis. But due to this nature of cryptography the traditional data retrieval processes are not able to deliver the user query relevance data.

This paper gives an overview of secure data search over cryptographic cloud and their approaches and describes some of the commonly used techniques to solve complex secure search method in cloud computing.

II. BACKGROUND

The background of a study is an important part of our survey paper. It provides the context and purpose of the study. Hence there is need for background study that contribute to prepare proposed system in future.

A. Cloud computing

Cloud computing is a computing paradigm, where a big pool of systems are associated in confidential or public networks, to provide dynamically scalable infrastructure for purpose, data and file storage. With the arrival of this technology, the cost of computation, application hosting, content storage and release is reduced considerably. Cloud computing is a practical approach to experience direct cost remuneration and it has the impending to convert a data center from a capital-intensive set up to a variable priced environment. The idea of cloud computing is based on a very primary major of, reusability of IT capabilities'. The difference that cloud computing carry compared to conventional concepts of “grid computing”, “distributed computing”, “utility computing”, or “autonomic computing” is to widen horizon across governmental boundaries. Forrester defines cloud computing as [5].

B. Cloud Architectural Components

Cloud service models are commonly divided into SaaS, PaaS, and IaaS that exhibited by a given cloud infrastructure. It's helpful to add more structure to the service model stacks: Figure 1 shows a cloud reference architecture [6] that makes the most important security-relevant cloud components explicit and provides an abstract overview of cloud computing for security issue analysis.

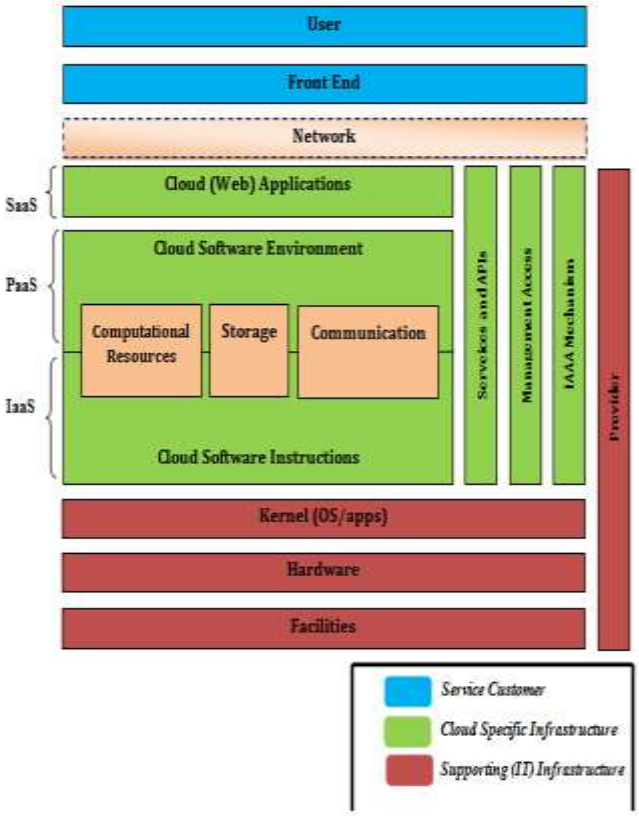

Figure 1: the Cloud Reference Architecture

Software as a Service (SaaS) Cloud consumers release their applications in a hosting environment, which can be accessed through networks from various clients (e.g. Web browser, PDA, etc.) by application users. Cloud consumers do not have control over the cloud infrastructure that often employs multi-tenancy system architecture, namely, different cloud consumers' applications are organized in a single logical environment in the SaaS cloud to achieve economies of scale and optimization in terms of speed, security, availability, disaster recovery and maintenance. Examples of SaaS include SalesForce.com, Google Mail, Google Docs, and so forth.

C. Information Retrieval

The effectiveness of information retrieval systems is measured by comparing performance on a common set of queries and documents. The meaning of the term information retrieval can be very broad. Just getting a credit card out of your wallet so that you can type in the card number is a form of information retrieval. However, as an academic field of study, information retrieval might be defined thus [7]:

"Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)".

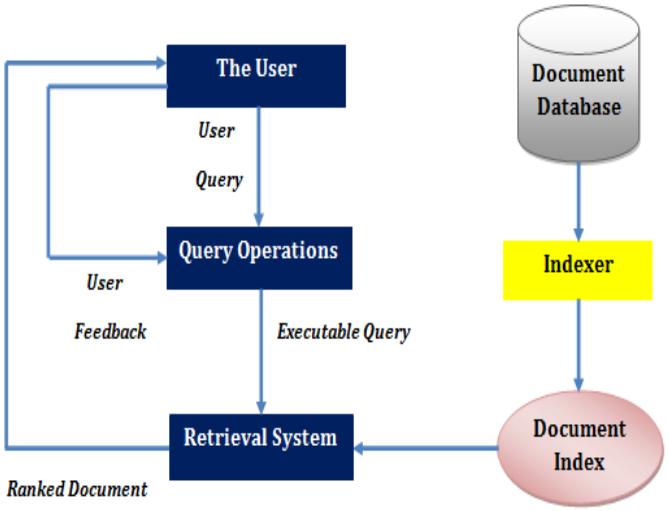

Figure 2 Information Retrieval Systems

Figure 2 show the basic block of the IR system. The field of information retrieval also covers supporting users in browsing or filtering document collections or further processing a set of retrieved documents. Given a set of documents, clustering is the task of coming up with a good grouping of the documents based on their contents. It is similar to arranging books on a bookshelf according to their topic. Given a set of topics, standing information needs, or other categories (such as suitability of texts for different age groups), classification is the task of deciding which classes, if any, each of a set of documents belongs to. It is often approached by first manually classifying some documents and then hoping to be able to classify new documents automatically [8].

D. Secure Keyword Search

Exploring privacy-preserving and effective search service over encrypted cloud data is of paramount importance. Considering the potentially large number of on demand data users and huge amount of outsourced data documents in the cloud, the problem is particularly challenging project is extremely difficult to meet also the requirements of performance, system usability, and scalability.

Searchable encryption is a helpful technique that treats encrypted data as documents and allows a user to securely search through a single keyword and retrieve documents of interest. However, direct application of these approaches to the secure large scale cloud data utilization system would not be necessarily suitable, as they are developed as crypto primitives and cannot accommodate such high service-level requirements like system usability, user searching experience, and easy information discovery. Although some recent designs have been proposed to support Boolean keyword search as an attempt to enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality. Early works have been aware of problem and provide solutions to the secure ranked search over encrypted data problem but only for queries consisting of a single keyword. How to design an efficient encrypted data search mechanism that supports multi-keyword semantics without privacy breaches still remains a challenging open problem [9].

III. LITERATURE SURVEY

The given section provides the understanding about the secure data search by Indexes that are recently contributing in cloud environment therefore a number of research articles and research papers are included in this section.

Sabrina De Capitani di Vimercati et al. [10] investigate issues of data outsourcing and propose an indexing technique for supporting efficient access to encrypted data while preventing possible disclosure of data to users not authorized to access them. Intuitively, this indexing technique accounts for authorizations when producing indexes to ensuring that the different occurrences of the same plaintext value, but accessible by different sets of users, not to be recognizable from their indexes. Authors show that this solution exhibits a limited performance overhead in query evaluation, while preventing leakage of information.

Ming Li et al. [11] identified the importance and challenges of designing privacy assured, flexible and practically efficient search mechanisms for outsourced cloud data services. In particular, authors focus on two representative types of flexible search functionalities: ranked keyword search, and search over structured data. Although these functionalities are already prevalent in information retrieval in the plaintext domain, realizing them in the encrypted domain requires non-trivial effort and is relatively new. In light of this, they first describe several existing technical approaches proposed by us and other researchers, and identify their advantages and limitations. Authors also discuss the open research directions and provide some possible ideas for further investigation. Authors believe the presented results will inspire more research towards making privacy-assured search in the cloud practical and useful.

Cong Wang et al. [12] define and solve the problem of secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy. Specifically, authors explore the statistical measure approach, i.e. relevance score, from information retrieval to build a secure searchable index, and develop a one-to-many order-preserving mapping technique to properly

protect those sensitive score information. The resulting design is able to facilitate efficient server-side ranking without losing keyword privacy. Thorough analysis shows that our proposed solution enjoys "as-strong-as-possible" security guarantee compared to previous searchable encryption schemes, while correctly realizing the goal of ranked keyword search. Extensive experimental results demonstrate the efficiency of the proposed solution.

Syam Kumar Pasupuleti et al. [13] proposed an efficient and secure privacy-preserving approach for outsourced data of resource-constrained mobile devices in the cloud computing. This approach employs probabilistic public key encryption algorithm for encrypting the data and invoke ranked keyword search over the encrypted data to retrieve the files from the cloud. Authors aim to achieve an efficient system for data encryption without sacrificing the privacy of data. Further, this ranked keyword search greatly improves the system usability by enabling ranking based on relevance score for search result, sends top most relevant files instead of sending all files back, and ensures the file retrieval accuracy. As a result, data privacy ensures and computation, communication overheads in reduction. Thorough security and performance analysis, authors prove that this approach is semantically secure and efficient.

Sabrina De Capitani di Vimercati et al. [14] discussed how the use of indexes, typically associated with the encrypted portion of the data, while desirable for providing effectiveness and efficiency in query execution, can - combined with fragmentation - cause potential leakage of confidential (encrypted or fragmented) information. Authors illustrated how the exposure to leakage varies depending on the kind of indexes. Such observations can result useful for the design of approaches assessing information exposure and for the definition of safe (free from inferences) indexes in fragmented data.

Aaron Steele et al. [15] proposed a method that can maintain a similar level of privacy while improving upon the query performance of previous solutions. The motivating principle behind this solution is that if the data owner possesses a small amount of secure local storage, it can be used as a pseudo-index table to improve query performance for selection queries involving conjunctions. Authors offer a heuristic approach for calculating the required storage resources and provide experimental analysis of the scheme.

IV. PROBLEM DOMAIN

Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval.

The encryption on data is an effective way to protect the confidentiality of data in cloud. But when it comes to searching, efficiency gets low. A general approach to protect the data confidentiality is to encrypt the data before outsourcing. Searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher-text domain. So far, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc. Among them, multi-keyword ranked search achieves more and more attention for its practical applicability.

Therefore the following issues are occurred during the data retrieval process:

- ✓ Data format is not recognizable thus the data is not analysed on the basis of their contents.
- ✓ Data keywords need to be secure also therefore the large amount of keywords are not stored separately for individual files.
- ✓ Normal cryptographic scenarios for keyword based need to additional encryption and decryption time to secure the data.

V. CONCLUSION

Public clouds are popular nowadays, where they are generally used in the storage and retrieval of the user's information Cloud outsourced storage service reduces the hardware and software maintenance costs of the cloud user. The outsourced storage server is responsible for data management and access control. To ensure the privacy of uploaded data, cloud users encrypt their data before uploading them to the cloud server. This paper suggests an efficient and secure keyword based searching scheme where the user can store and retrieved his files in a secure manner. Therefore, in this paper, survey of various keyword based data retrieval search method for maintain ease of use of end user applications were studied and list out. Hence, the secure search is very important where user stored their data on cloud storage and access through keyword basis.

REFERENCE

[1] Ren-Junn Hwang, Chung-Chien Lu and Jain-Shing Wu, "Searchable Encryption in Cloud Storage", World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering Vol: 8, No: 7, 2014.
[2] Syam Kumar Pasupuleti, Subramanian Ramalingam and Rajkumar Buyya, "An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing", Journal of Network and Computer Applications 64 (2016) 12–22.
[3] Yu, Chia-Mu, Chi-Yuan Chen, and Han-Chieh Chao. "Privacy-preserving multi-keyword similarity search over outsourced cloud data." IEEE Systems Journal 11, no. 2 (2017): 385-394.
[4] Koo, Dongyoung, Junbeom Hur, and Hyunsoo Yoon. "Secure and efficient data retrieval over encrypted data using attribute-based encryption in cloud storage." Computers & Electrical Engineering 39, no. 1 (2013): 34-46.
[5] Sookhak, Mehdi, et al. "Remote data auditing in cloud computing environments: a survey, taxonomy, and open issues." ACM Computing Surveys (CSUR) 47.4 (2015): 65.

[6] B. Grobauer, T. Walloschek, and E. Stöcker, "Understanding Cloud Computing Vulnerabilities". 2011 IEEE Security and Privacy, pp. 50-57.
[7] Buckley, C., Voorhees, E.M. (2004) Retrieval evaluation with incomplete information, in Proc. ACM SIGIR, 25-32.
[8] A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Engineering Bulletin, vol. 24, no. 4, pp. 35–43, 2001
[9] Mithilesh kumar Sharma and S. Karpagam, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", International Journal on Applications in Information and Communication Engineering, Volume 1: Issue 10: October 2015, pp 19-22.
[10] Sabrina De Capitani di Vimercati, Sara Foresti and Pierangela Samarati, "Private data indexes for selective access to outsourced data", In Proceedings of the 10th annual ACM workshop on Privacy in the electronic society, pp. 69-80, ACM, 2011.
[11] Ming Li, Shucheng Yu, Wenjing Lou and Y. Thomas Hou, "Toward privacy-assured cloud data services with flexible search functionalities", 2012 IEEE 32nd International Conference on Distributed Computing Systems Workshops (ICDCSW), pp. 466-470.
[12] Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data", IEEE Transactions on parallel and distributed systems, Volume 23, Number 8 (2012): pp. 1467-1479.
[13] Syam Kumar Pasupuleti, Subramanian Ramalingam and Rajkumar Buyya, "An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing", Journal of Network and Computer Applications 64 (2016): pp. 12-22.
[14] S. De Capitani di Vimercati Sara Foresti and Pierangela Samarati, "On information leakage by indexes over data fragments", 2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW), pp. 94-98.
[15] Aaron Steele and Keith Frikken, "An index structure for private data outsourcing", Data and Applications Security and Privacy XXV (2011): pp. 247-254.