ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 2, February 2016

# A Survey on Searchable Encryption Schemes in Storage Servers

Samya Ali[1], Geethu P C[2]

PG Scholar, Department of Computer Science and Engineering,
Vidya Academy of Science and Technology
Thrissur, Kerala, India[1]

Asst. Professor, Department of Computer Science and Engineering,
Vidya Academy of Science and Technology
Thrissur, Kerala, India[2]

**Abstract:** Searchable encryption is a new concept used to keep privacy while storing data in an untrusted third party. This approach can be used in a lot of applications such as mail servers, files systems, data bases management etc. Searchable encryption consists on storing encrypted data, retrieving it without any leak of information and keeping anonymity while retaining information confidentiality. This paper discusses about various searchable encryption schemes to retrieve encrypted documents stored in an untrusted server in a secure manner.

**Keywords**: Searchable Encryption, Privacy, Keyword Search, Data retrieval

## I. Introduction

Nowadays modern life requires much more electronic and computer resources to comply tasks that terminal equipment cannot solely fulfill. These tasks can be storage services, software computing tasks, financial services, multi-party data access. Actually, users need access to their data everywhere, with unlimited storage capacity and availability every time. All these features are not provided by local data storage. Consequently to avoid this hindrance, users are obliged to trust a third party to perform these applications. In fact, users are already used to store their messages in mail servers; furthermore they want to store more data in the outsourced servers to take advantage of huge space. To ensure security to the stored documents, these documents are encrypted before storing it in an untrusted server.

However, data encryption makes data utilization effectiveness and efficiency a very challenging task given that there could be a large amount of outsourced data files. Encryption makes it impossible for both insiders and outsiders to access the data without the keys but at the same time removes all search capabilities from the data owner. One trivial solution to re-enable searching functionality is to download the whole database, decrypt it locally, and then search for the desired results in the plaintext data. For most applications, this approach would be impractical. Another method lets the server decrypt the data, runs the query on the server side, and sends only the results back to the user. This allows the server to learn the plaintext data being queried and hence makes encryption less useful. Instead, it is desirable to support the fullest possible search functionality on the server side, without decrypting the data, and thus, with the smallest possible loss of data confidentiality. This is called searchable encryption.

This paper is divided into three sections. Section 2 describes about searchable encryption schemes and the different searchable techniques and section 3 concludes the paper.

## II. Searchable Encryption (SE)

Searchable encryption is a recent concept that performs searches on encrypted data without any leak of information. The main idea is to be able to perform an encrypted query without having to download the whole encrypted data. Indeed, searchable encryption is composed of two steps:

- Storing a special encryption of data on the untrusted third party (Store phase).
- Make an encrypted search query to retrieve the desired information (Search phase).

Searchable encryption scheme is of two types: Symmetric Searchable Encryption (SSE) and Asymmetric Searchable Encryption (ASE).

In SSE, a user encrypts the data using symmetric/private key encryption schemes (e.g. AES) before outsourcing it to the untrusted server. This setting is appropriate when the user that searches over the data is also the one who generates it. The main advantage of this setting is the efficiency, but it lacks of functionality as it can only be used for a single user scenario.

In ASE, a user encrypts the data using asymmetric/public key encryption schemes before outsourcing it to the server. This setting is appropriate for a scenario where the user searching over the data is different from the user who generates it. For example, multiple users can use the public key of a certain user to encrypt and upload the data, however; only the owner with the corresponding private key can generate the search token and therefore can perform a search over the encrypted data. The main advantage of ASE is its functionality whereas the drawback is inefficiency.

There are different techniques for searching and retrieving data from a storage server and they are described in the following section.

### A. *Boolean Symmetric Searchable Encryption*

Tarik Moataz describes the searching of cloud data over encrypted format in [1]. Here, Boolean symmetric searchable encryption‖ (BSSE) technique is used for searching. BSSE uses the Boolean expression query to perform queries which are composed of conjunction, disjunction and negation of keywords. In this method, the main process is the Gram- Schmidt orthogonalization process. This process first encodes the keywords that used in queries and labels the inner products to perform the searching of data. This method is fully randomized. The search is linear and the increase of labels size implies a longer computation phase for each document. But it only focuses on simple keyword matching and used only for searching Boolean queries.

### B. *Ranked Keyword Search*

C Wang describes the searching of cloud data using Secure Ranked Keyword Search method in [2]. The traditional keyword search fully concentrates on the Boolean Search only which is used for searching Boolean queries. In this method, ranking technique is used for searching an encrypted data. The Order Preserving Mapping technique is also used. This Order Preserving Mapping technique protects the sensitive score information. This method is highly efficient. But this searching leads to collision in the network. Here, encrypted files are highly processed after main searching and these files are post processed.

In [3], C Wang proposed a method to overcome the above issue. It describes the searching and retrieving of the encrypted cloud data through ranked keyword search by using Ranked Searchable Symmetric Encryption, Order Preserving Symmetric Encryption and One Many Order Preserving Mapping. These methods are used for the purpose of ranked keyword search and guarantees security and performance. In this method, there is only minimum communication and computation overhead and it avoids network traffic and unwanted file retrieval. But this method does not support multiple keywords and it leads to increase in the search time and cost.

In [4], Swaminathan proposed the method of Confidentiality Preserving Rank-Order technique for the searching of encrypted cloud data. Confidentiality-preserving baseline model is created and this model prevents the untrusted data centre from learning information about the query and the document collection. Confidentiality Preserving Rank-Order technique creates a framework by using secure index, encrypted domain search and ranked retrieval for retrieving the cloud data over large collections of document. Depending upon encrypted search queries, the documents are securely ordered and the most relevant documents are retrieved from the encrypted data collection. This method is highly efficient, with high search accuracy for wide range of applications and it protects document/query confidentiality against an outside intruder. But it is designed only for unencrypted data and also the complexity of this searching is high and protecting communication link is little more complex.

### C. *Multi Keyword Ranked Search*

N. Cao proposed the searching of encrypted cloud data using Privacy-Preserving Multi-keyword Ranked Search(MRSE) in [5]. Here, the basic concept used is co-ordinate matching. Co-ordinate matching obtains the similarity between search query and data documents. Inner product similarity is also used to describe the Multi-keyword Ranked Search over Encrypted Cloud Data (MRSE). Here four modules of searching are performed over encrypted cloud data. The four modules in this method are; Encrypt module, Client module, Multi-Keyword module, and Admin module. The features of this method are, Multi-keyword Ranked Search, Privacy-Preserving, and high efficiency. It eliminates unnecessary traffic and improves search accuracy. The disadvantages of this method are single keyword search with ranking and Boolean keyword search with ranking are not possible. It is not suitable for large scale cloud data and it provides much less semantics and this scheme is developed as crypto primitives.

*D. Fuzzy Keyword Search*

Jin Li proposed the searching of encrypted cloud data through Fuzzy keyword searching in [6]. The techniques which are used to obtain the fuzzy keyword search are Wild card-Based Technique, Gram-Based technique and Symbol-Based trie Traverse Search Scheme. Wild card-Based Technique is used to prevent the problem of editing operations at the same position key words. In this, the edit operation problem is solved by Edit Distance which includes Substitution, Deletion & Insertion. Gram-Based technique constructs the fuzzy set depends on grams. The gram of a string is a substring. This substring acts as a signature for high quality efficient approximate search. In Symbol-Based trie-Traverse Search Scheme, the main idea behind is that all trapdoors sharing a common prefix may have common nodes. All fuzzy words in the trie are finding by a depth-first search. Fuzzy keyword search service aims at accommodating various types and representation inconsistencies in different user searching inputs. It increases the searching effectiveness, improve the space efficiency, optimize time efficiency and size of the fuzzy keyword set is controllable too. But large storage is complex. It supports only Boolean keyword search and does not support the ranked search problem.

*E. Practical Techniques*

The different techniques for searching the encrypted cloud data without any loss of data confidentiality is described in [7]. It uses pseudo random function**,** pseudo random generator and sequential scan. This scheme uses sequential scan to provide the security for the resulting system. It has following schemes. The Basic scheme controlled searching, support for hidden searches and the final scheme. It is provably secure, efficient and practical. It supports controlled and hidden search and query isolation. This method is simple and fast. It has no space and communication overhead. It is very flexible. But the seque

ntial scan is not efficient if data size is large. The scheme is too slow in searching for a large number of documents. If we search too often, server may be able to learn some information.

*F. Conjunctive Search*

P. Golle proposed the Conjunctive Keyword searching of Encrypted cloud data in [8]. This method defines a security model which describes a protocol that allows conjunctive keyword queries with linear communication cost, amortized linear communication cost and constant cost. In this protocol, amortized linear communication cost uses standard hardness assumption and constant cost uses new hardness assumption. In the total number of documents stored on the server, the capabilities size for conjunctive queries is linear. But the most of the communication cost between the user and the server can be done online. Each capability consists of two parts, Proto-capability and Query part. The user must give server capabilities to determine the correct set of document by using set intersection and meta keywords. The conjunctive search is based on two schemes: Conjunctive search scheme with constant online and communication cost. The advantage of conjunctive search scheme with constant communication cost is communication cost is linear. But the drawbacks are it allows secure conjunctive search with small capabilities only and it can solve the problem of secure boolean search on encrypted data only partially. The complete solution requires the ability to do disjunctive keyword search securely both across and within keyword fields.

In [9], the objective is to achieve an efficient conjunctive keyword searching of encrypted data. Here two schemes are used to achieve an efficient conjunctive keyword searching of cloud data namely, Shamir secret sharing and bilinear pairing. The first scheme is based on Shamir Secret Sharing and provides the most efficient search technique in this context. Although the size of its trapdoors is linear in the number of documents being searched, the overhead remains reasonable in practice. Nonetheless, to address this limitation alternative method based on bilinear pairings is proposed. This bilinear pairings yield constant size trapdoors. This latter construction is not only asymptotically more efficient than previous secure conjunctive keyword search schemes in the symmetric setting, but also it incurs significantly less storage overhead. Additionally, unlike most previous work, these constructions are proven secure in the standard model.

In the previous conjunctive keyword search scheme, the communication and storage costs linearly depend on the number of stored data in the database, and hence it is not really suitable for a large scale database. Jin Wook Byun proposed an efficient conjunctive keyword search scheme over encrypted data in aspects of communication and storage costs in [10]. Here the storage cost of a user and the communication cost between a user and a data supplier are reduced to the constant amounts. It defines security model for a conjunctive keyword search scheme and proves that the proposed scheme is secure under the Decisional Bilinear Diffie-Hellman (DBDH) assumption in the random oracle model. The main advantages of this method are it is more efficient than the previous one and suitable for large scale database. But it is used only in large scale databases.

### G. Symmetric Encryption

R. Curtomola described the searching and retrieving of the outsourced data by using multi user Searchable Symmetric Encryption (SSE) in [11]. This method focuses active research, several security definitions and constructions, which is achieved by Non-adaptive setting and Adaptive adversary. This method is used because it is more efficient and guarantees more security than all previous constructions. It supports multi user setting and does not require authentication. But it is not suitable for large scale cloud data and cannot accommodate high level requirements.

M. Naveed proposed the blind storage scheme to achieve a searchable encryption in [12]. It allows the data owner to store the data more securely and makes it visible only to the data owner. A blind storage allows the data owner to save the set of files on a remote server in such a way that the server does not know about the contents of the files what the owner is stored. It supports adding new files, updating or deleting existing files. The server will know only the file name of what the data owner is uploading. Hence the blind storage leaks only little information to the server. By storing the data in the blind storage the data owner can prevent other data users to unaware about the content of the file. However data will be saved in the form of fixed blocks and each block will be indexed in order to know about the file. Moreover the server will know about the existence of the file (and its size not the name used by the data user to refer to the file or its contents) only when the data user retrieves it later. It helps in achieving fully adaptive security with no server side computation.

## III. CONCLUSION

Privacy gets more and more important in the outsourced servers. Users store data on servers without any encryption, consequently this information is public and become vulnerable to malicious attacks. Users are afraid from unauthorized access threats and the loss of data integrity and confidentiality. The main goal of outsourced storage is to provide privacy and to keep confidentiality of the data stored in the servers. In this paper, various searchable encryption schemes have been discussed. The scheme to search over encrypted data without the need to decrypt it provides a secure mechanism to outsource the data in untrusted servers like public clouds more efficiently. This paper discusses about various searchable encryption schemes to retrieve encrypted documents stored in an untrusted server in a secure manner.

## REFERENCES

[1] Tarik Moataz, Abdullatif Shikfa, “Boolean Symmetric Searchable Encryption”, ASIA CCS '13 Proceedings of the 8th ACM SIGSAC symposium on Information computer and communications.

[2] Cong Wang, Ning Cao, Jin Li, Kui Ren, Wenjing Lou, “Secure Ranked Keyword Search over Encrypted Cloud Data”, Distributed Computing Systems (ICDCS), IEEE 30th International Conference, 2010.

[3] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data”, Journal IEEE,Vol.23 Issue.8, 2012.

[4] Swaminathan A, Mao Y, Su G-M, Gou H, Varna AL, He S, M. Wu, Oard D, “Confidentiality Preserving Rank-Order search”, Conference Papers (Proceedings of the ACM workshop on Storage security and survivability) pp.7-12, 2007.

[5] Ning Cao, Cong Wang , Li, Ming , Kui Ren, Wenjing Lou, “Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data”, INFOCOM, Proceedings IEEE April 2011.

[6] Jin Li , Qian Wang ; Cong Wang , Ning Cao , Kui Ren , Wenjing Lou, “Fuzzy Keyword Search Over Encrypted Data in Cloud Computing”, INFOCOM, Proceedings IEEE, March 2010.

[7] Daw Xiaoding Song , Wagner, D. , Perrig A, “Practical Techniques for Searches on Encrypted Data”, Security and Privacy, IEEE Symposium May, 2000.

[8] Philippe Golle, Jessica Staddon, Brent Waters, “Secure Conjunctive Keyword Search over Encrypted Data”, Springer Berlin Heidelberg; ACNS, 2004 pp.31-45.

[9] Lucas Ballard, Seny Kamara, Fabian Monrose, “Achieving Efficient Conjunctive Keyword Searches over Encrypted Data”, ICICS Springer Berlin Heidelberg, pp. 414-426, 2005

[10] Jin Wook Byun, Dong Hoon Lee, Jongin Lim, “Efficient Conjunctive Keyword Search on Encrypted Data Storage System”, Springer Berlin Heidelberg; Third European PKI Workshop: Theory and Practice, EuroPKI 2006, Turin, Italy, Proceedings, pp. 184-196, June 2006.

[11] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: Improved definitions an efficient constructions”, ACM CCS 06 conference, 2006.

[12] M.Naveed,M.Prabhakaran and C.A.Gunter, "Dynamic searchable encryption via blind storage",in proc.IEEE symp.security and.privacy, May2014,pp.639- 654.

## BIOGRAPHY

Samya Ali is a postgraduate research student at Vidya Academy of Science and Technology, Thrissur, India, affiliated to Calicut University. Her area of interest is Security and Privacy in Networks. She received her B.Tech Degree in Computer Science and Engineering from Calicut University, Kerala, India, in 2014.

Geethu P.C is an Assistant Professor at Vidya Academy of Science and Technology, Thrissur, India, affiliated to Calicut University. Her area of interest is Network Security. She received her M.Tech Degree in Computer Science and Engineering from Kerala University, Kerala, India, in 2013.