IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, May 2022
Impact Factor: 6.252

An Analysis of Several Keyword-Based Search Methods for Encrypted Data

Dr. M. Mohamed Ismail1 and Dr. P. Rizwan Ahmed[2]
Associate Professor, Department of Computer Science[1]
Asst. Professor & Head, Department of Computer Applications[2]
Mazharul Uloom College, Ambur, Tamil Nadu, India

Abstract: *Cloud computing provides several attractive benefits for users like on-demand computing, pay as per use. It brings great convenience to consumers; where shared resources, data and information are provided to computers on-demand and consumer has to pay as per use. Ideally for these services, consumers should be in a position to verify the charges billed to them. 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 such as mail servers and file servers in encrypted form to reduce security and privacy risks. Encryption is the process of encoding messages or information in such a way that only authorized parties can read it. Multiple encryption techniques are invented to encrypt the data. That way encrypted search has become an important problem in security. This is due to a combination of three things: (1) search is now the primary way we access our data; (2) we are outsourcing more and more of our data to third parties; and (3) we trust these third parties less and less. Because of this, the problem of encrypted search is now of interest to many sub-fields in computer science (e.g., databases, security, cryptography, privacy). Some existing methods are more practical than others, some are more secure than others and some are more functional or flexible. These schemes do not support verifiability of search result. To save computation cost or download bandwidth it is viewed as, selfish cloud server only conducts a fraction of search operation or semi-honest-but-curious server return a part of result. To tackle such challenges, multiple search scheme are invented.*

Keywords: Consumer-centric cloud computing; privacy preserving; verifiable search; Encryption; Decryption

I. INTRODUCTION

Encryption is the process of encoding messages or information in such a way that, it enables only authorized parties to read it. Multiple encryption techniques have been invented, to encrypt the data. This has lead to a critical problem for search on the encrypted data. There are two kinds of key-based encryption algorithms, symmetric encryption algorithms (also known as secret key algorithms) and asymmetric encryption algorithms (or called as public key algorithms). The difference is that symmetric encryption algorithms use the same key for encryption and decryption (or the decryption key is easily derived from the encryption key), whereas asymmetric encryption algorithms use a different key for encryption and decryption, and the decryption key cannot be derived from the encryption key. Here are four main risks posed by cloud storage services:

1. **Accidental data loss:** Users can share sensitive information stored in the cloud with anyone, opening the door to common errors such as sending the link to a document to the wrong person.
2. **Data theft:** Hackers know that cloud storage services are modern day pots of gold, making them a common target for attack. Of course, all reputable cloud storage services take security very seriously. However, users themselves create easy theft opportunities, such as by using simple passwords for their cloud storage accounts rather than the complex hard-to-crack ones that are enforced within your organization.
3. **Storage provider vulnerabilities:** Cloud storage providers have full access to your data and control where it is stored, making your data subject to security and technical issues with the providers themselves.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, May 2022
Impact Factor: 6.252

4. **Poor user practices:** Your users are mobile, and the use of cloud storage services is increasing rapidly. Users often find it quicker and easier to access files from the cloud rather than using a VPN to connect to the corporate network. If they aren't given a safe, approved way to use such services they invariably search for work-around, without consideration for the security of your corporate data.

In today's world most of the communication is done using electronic media. Data security plays a vital role in such communication. Hence there is a need to protect data from malicious attacks. This can be achieved by Cryptography. The earlier encryption algorithm is Data Encryption Standard (DES) which has several loopholes such as small key size and sensible to brute force attack etc. and it can not be provide high level, efficient and exportable security. These loopholes overcome by a new algorithm called as Advanced Encryption Standard (AES).

However, it is a very difficult to search the most suitable services or products for ordinary consumers, as there are so many services and products in cloud. Meanwhile, data outsourcing enables the data owner and the cloud service provider not in a same trusted domain, making the data owner not manage data in real time. It is a common practice to encrypt sensitive information before outsourcing.

The plain text of 128 bits is given as input to encryption block in which encryption of data is made and the cipher text of 128 bits is throughout as output. The key length of l28bits, l92bits or 256bits is used in process of encryption. The AES algorithm is a block ciphers that user the same binary key for both encryption and decryption of data blocks.

1. **Cryptography -** Cryptography is the science of secretcodes, enabling the confidentiality of communication through an insecure channel. It protects against unauthorized parties by preventing unauthorized alteration of use. Generally speaking, it uses a cryptographic system to transform a plaintext into a cipher text. using most of the time a key. It has different Encryption and Decryption algorithms to do so.
2. **Cipher Text -** This is the scrambled message producedas output from Encryption algorithm. It depends on the plaintext and the secret key. For a given message, two different keys will produce two different cipher texts.
3. **Encryption -** Encryption is the process of convertingdata, in plain text format into a meaningless cipher text by means of a suitable algorithm. The algorithm takes secret key and plain text as input and produces cipher text.
4. **Decryption -** Decryption is converting the meaninglesscipher text into the original information using decryption algorithms. The decryption algorithm is inverse of encryption algorithm. This takes key and cipher text as input and produces original plain text.
5. **Symmetric Key Cryptography -** It uses the same secret(private) key to encrypt and decrypt its data. It requires that the secret key be known by the party encrypting the data and the party decrypting the data.
6. **Asymmetric Key Cryptography -** Asymmetric usesboth a public and private key. This allows for distribution of your public key to anyone with which they can encrypt the data they want to send securely and then it can only be decoded by the person having the private key.

To search over encryption data multiple search techniques are invented which are described below.

II. RELATED WORK

In the paper of [1] they proposed an efficient verifiable keyword -based semantic search scheme. Comparing to most of the existing searchable encryption schemes, the proposed scheme is more practical and flexible, better suiting users different search intensions. Moreover, the proposed scheme protects data privacy and supports verifiable searchability, in the presence of the semi honest server in the cloud computing environment.

In the paper of [2] proposed an effective approach to solve the problem of synonym based multi-keyword ranked search over encrypted cloud data. The main contributions are summarized in two aspects; synonym based search and similarity ranked search. The search results can be synonyms of the predefined keywords, not the exact or fuzzy matching keywords, due to the possible synonym substitution and/or her lack of exact knowledge about the data. The vector space model is adopted combined with cosine measure, which is popular in information retrieval field, to evaluate the similarity between search request and document.

In the paper [3] solved the problem of multi-keyword ranked search over encrypted cloud data and establish a variety of privacy requirements. Among various multi keyword semantics, we choose the efficient similarity measure of coordinate matching i.e. as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords and use inner product similarity to quantitatively evaluate such similarity measure. For meeting the challenge

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, May 2022
Impact Factor: 6.252

of supporting multi-keyword semantic without privacy breaches, they proposed a basic idea of MRSE using secure inner product computation. Then they have given two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models.

In the paper [4] proposed the first verifiable SSE scheme, which offers data privacy. Verifiable searchability and efficiency, in the presence of an unusually strong adversarial server in cloud computing environment. The rigorous security analysis together with through experimental evaluations on a resource-constrained device using real data sets confirms that the proposed VSSE realizes our design goals and is a promising solution to mediate the conflicts between data usability and data privacy in such scenario.

I reviewed related work and illustrate the difference of different keyword-based search techniques. Li at al [6] firstly proposes a fuzzy keyword search scheme over encrypted cloud data. Wang et al [5] proposed a secure ranked search scheme. This scheme supports only single keyword search while fuzzy keyword search scheme tacks the problem of minor typos and format inconsistencies. Chai et al [4] propose a verifiable search scheme which can prove correctness and completeness of result efficiently. Based on VSSE and fuzzy keyword search, Wang et al [6] proposes a scheme supporting both verification and fuzzy search, but the scheme ignores result ranking.

One is that most of these schemes supports only exact keyword search. That means returned result is completely dependent on whether query terms users enter match pre-set keywords. The other one is that most existing searchable schemes assume that the cloud server is honest-but-curious. However, Chai et al [4] notice that the cloud may be selfish to save its computation or download bandwidth. That is, the cloud server might conduct only a fraction of search of search operation or return a part of result honestly.

Besides, Fu et al [7] recently proposed a multi-keyword search scheme in encrypted cloud environment which can achieve synonym query. The main contribution of the scheme is that it solves the problem of synonym search.

III FACE IMAGE ANNOTATION

3.1 Architecture of Search over Cloud Data

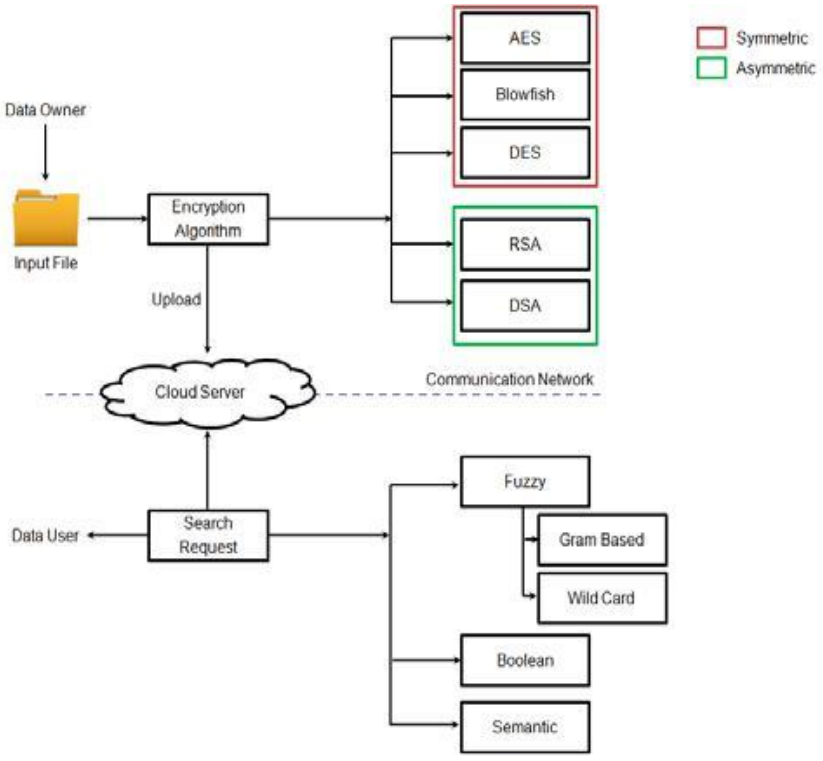

Figure: Architecture of search over cloud data

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, May 2022
Impact Factor: 6.252

The architecture in search services involves three different entities: the data owner, the data user and the cloud server. The data owner has a collection of data documents to be outsourced to the cloud server in the encrypted form. To enable the searching capability over encrypted documents for effective data utilization, the data owner, before outsourcing, will first build an encrypted searchable index and then outsource both the index and the encrypted document collection to the cloud server. To search the document collection for given keywords, an authorized user acquires a corresponding trapdoor through search control mechanisms. Upon receiving from a data user, the cloud server is responsible to search the index and return corresponding set of encrypted documents.

3.2 Processing Steps of Encryption

1. SubBytes - A non-linear substitution step where each byte is replaced with another according to a lookup table (S-box).
2. ShiftRows - A transposition step where each row of the state is shifted cyclically a certain number of times.
3. MixColumns - A mixing operation which operates on the columns of the state, combining the four bytes in each column.
4. AddRoundKey - Each byte of the state is combined with the round key; each round key is derived from the cipher key using a key schedule.

3.3 Processing Steps of Search Technique

Setup - In this algorithm the data owner initiates the scheme to generate the random key and a secret key.
GenIndex - To improve the search efficiency, a symbol-based tree to store elements in a finite symbol set is built.
Preprocess:

- The data owner scans the plaintext document collection D and extracts the distinct keywords of D, denoted as W;
- The data owner computes the score of all distinct keywords on basis of presence in number of documents from collection.

GenQuery - When the user inputs the query terms Q, first builds term similarity tree TST(Q,v,m) andexecutes keyword semantic extension, getting the extended query.
Search - Upon receiving the search request, the cloud server performs the search operation over the index G. The search is principally to find a path in G according to the search request, from the root node to the leaf node. The existence of a path indicates that the queried words happens at least one of the targeted data files.
Verify & Rank - When the user receives the ranked outcome from the cloud server, he can verify the correctness and completeness of search result.

IV. CONCLUSION

After the study of techniques for encrypting files, it is observed that: Among Symmetric key encryption algorithms, AES provides better security in less time and among asymmetric encryption algorithms, RSA gives better security in only single round. And Semantic search technique over encrypted data returns more relevant files on seacrh.

REFERENCES

[1]. Zhangjie Fu, JiangangShu, Xingming Sun and Nigel Linge, "Smart Cloud Search Services : Verifiable Keyword-based Semantic search over Encrypted Cloud Data," *IEEE Trans.*, 2014.
[2]. 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 Trans. ConsumerElectron.,*vol. 60, no. 1, pp. 164-172, 2014.
[3]. S.N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, "Privacy-preserving multikeyword ranked search over encrypted cloud data," *Proceedings of IEEE INFOCOM 2011*, pp. 829-837, 2011.
[4]. Q. Chai and G. Gong, "Verifiable Symmetric Searchable Encryption for Semi-Honest-but-Curious Cloud Servers," *Proceedings of IEEE International Conference on Communications (ICC'12)*, pp. 917-922, 2012.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, May 2022
Impact Factor: 6.252

[5]. C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "Secure ranked keyword search over encrypted cloud data," *Proceedings of IEEE30th International Conference on Distributed Computing Systems (ICDCS)*, pp. 253-262, 2010.

[6]. J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. J. Lou, "Fuzzy keyword search over encrypted data in cloud computing," *Proceedings of IEEE INFOCOM 2010*, San Diego, CA, USA, pp. 1-5, 2010.

[7]. *Z. Fu, X. Sun, N. Linge, and L. Zhou, "Achieving Effective Cloud Search* Services: Multi-keyword Ranked Search over EncryptedCloud Data Supporting Synonym Query," *IEEE Trans. ConsumerElectron.,*vol. 60, no. 1, pp. 164-172, 2014.

[8]. *G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, and K. Miller, "WordNet:*An online lexical database," *Int. J. Lexicograph*. vol.3,no. 4, pp. 235–244,1990.