International Journal of Recent Engineering Science (IJRES)
ISSN: 2349 – 7157 / doi: 10.1445/23497157/IJRES-V6I6P104

Volume 6 Issue 6, 17-23, Nov-Dec 2019
© 2019 Seventh Sense Research Group®

Design and Development of Communication Salvage upon Encrypted Information in Cloud Computing

S. Ravichandran[1], R. Rajkumar[2]
[1](Computer Science Department, Annai Fathima College of Arts & Science, Madurai, India)
[2](Information Technology Department, Annai Fathima College of Arts & Science, Madurai, India

Abstract

Cloud computing is a technology that consumes data centres to provide information and requests. It provides resources, for instance, centralized storage, processors and network bandwidth to users. Moreover, it allows consumers to access their files consuming any machine among internet link. The largest establishments, almost Cloud computing, are privacy and security. One of the answers to improve cloud security is that all the sensitive information is encrypted before subcontracting. As a result, the data cannot be accessed using the plaintext keyword search. Because multiple data files are stored in the cloud server, it is necessary to provide Keyword-based search service and ranking of results to users. This paper describes the principle of various searching techniques that are used to search over encrypted information. Minimal performances like searchable encryption scheme permit cloud users to search the desired data over encrypted information. But these techniques focus on exact keyword matches in which users can usage a little distinct Keyword. Possibly, several search techniques like the Multi-keyword ranked search scheme can rank search effects constructed on the relevancy. The Multi-keyword ranked search is used to build an efficient index to improve the search efficiency and blind storage system that does not allow other search users to access the manuscript. It also discusses how to develop the exploration facility and ranking solutions essentially.

Keywords - *Sensitive data, Plaintext keyword search, Multi- keyword ranked search; Privacy-preserving; Cloud computing*

I. INTRODUCTION

Cloud Computing has been proposed because the next- generation equipment which supports on-demand top quality applications and services from a distributed pool of configurable calculating properties. Cloud computing is computing in which great sets of remote servers are networked to allow integrated data storage and online access to calculating stuff. Among the help of cloud amenities, more and more problematic information for instance emails, personal health records are stored on the Cloud, due to its long list of unprecedented advantages including like on-demand self-service, ubiquitous network access, location individual resource pooling, rapid fallback plasticity. An objective of cloud computing is to permit users to require benefits from all the prevailing technologies, bereft of the must for deep awareness about all of them. The Cloud aims to chop costs, and helps the consumers specialize in their core business rather than being inhibited by IT obstacles. Data security has become a vital issue when using third party infrastructures like Cloud. Data Storage using Cloud provides various benefits for the purchasers. Cloud service providers physically secure the IT hardware like servers and routers against unauthorized access, interference, floods, theft are sufficiently robust to diminish interruption, respectively. The Cloud provides high reliability and space for storing. Albeit cloud services offered by Cloud Service Providers (CSPs) are trusted based on the license agreement, there may need an opportunity of knowledge leakage. The Cloud vendors guarantees that applications available as a service through the Cloud (SaaS) are covered employing specifying, designing, imposing, checking out and retaining suitable request safety tactics inside the manufacturing atmosphere the safety is obtainable by the trusted third parties could also be distorted by distinctive procedures. To protect data privacy and unsolicited combat accesses, cloud providers encrypt that each one sensitive data before outsourcing and only authorized users have entree to their information. Furthermore, digital identities and authorizations should be protected as must any information that the contributor gathers or creates about customer activity within the Cloud, so on provide end-to-end data confidentiality assurance within the Cloud. Three sorts of cryptography techniques are utilized in Cloud, namely: Symmetric key cryptography, Asymmetric key cryptography and Hash algorithm. Symmetric key cryptography uses one key, whereas asymmetric key cryptography uses two keys. However, the hash algorithm doesn't use any key to encrypt and decrypt data.

To construct a secure cloud computing system, securities at infrastructure, application software stages and service platforms have to be learned. Data encryption is one of sufficient means to achieve data security in cloud computing. However, data encryption made and decryption is a very challenging task, assumed that there might be a significant quantity of outsourced information files. Moreover, data owners may share their outsourced data with a significant number of consumers who might want only to recover evident precise information files they are interested in through a specified period. One of the most popular ways to do so is keyword-based search. The keyword search technique permits consumers to recover files of interest selectively and has been widely used in plaintext search scenarios. It provides various features such as user registration, users authentication, uploading information with encryption, transfer information with decryption, creation of individual file program for each consumer. This offers an easy to use environment with sophisticated techniques. The proposed scheme will provide maximum efficiency by consuming secure along with practical procedures. Regrettably, data encryption, which restricts user's ability to perform a keyword search and more demands the safety of keyword secrecy, kinds the conventional plaintext search methods fail to retrieve encrypted cloud data.

While conventional searchable encryption schemes permit a consumer to tightly search over encrypted data complete keywords without initial decrypting it, and these procedures maintain only predictable Boolean keyword search exclusive of imprisoning any significance of the files in the search result. Generally, cloud server assigns ranks to manuscripts to make the hunt as faster. This ranked search system allows data consumers to find the most relevant information very quickly, rather than a complex sorting through every compliment in the substance group. Ranked Search techniques can also elegantly eliminate unnecessary network traffic by transfer spinal only the most applicable information, which is too required in the "pay- as-you-use" cloud paradigm. For privacy protection, such ranking operation should not leak any keyword associated data. Alternatively, to recover the hunt solution accurateness along with to improve the consumer searching knowledge, it is also essential for such grading system to maintain numerous keywords hunt, equally faced to single keyword search. For ranking operation, the method proposed by employs a secure "k-nearest neighbour (k-NN)" to achieve certain search functionality. This could return not only the exact matching files but also the documents with the conditions hidden semantically related to the question Keyword that such a solution is not secure. To improve security, it increases the data confidentiality and access patterns. The privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE) consuming reliable inside product calculation which is modified from a reliable k-nearest neighbour (kNN) Technique and then showed how to significantly improve it to be secrecy- conserving alongside distinctive hazard patterns in the MRSE essential in a stage-by-stage technique.

As a standard practice indicated by today's web search engines (such as Google search), information consumers may incline to deliver a group of keywords in its place of only one as the indicator of their search curiosity to recover the most related information and every Keyword in the hunt application is equal to assist narrow down the hunt solution supplementary. The "Coordinate matching" is an efficient similarity measure among such multi-keyword semantics to enhance the answer importance and has been broadly consumed within the plaintext information salvage (IR) community and further consumes "inner product similarity" to formalize such principle for similarity measurement quantitatively.

II. INFORMATION SALVAGE INCLOUD

A. System Model

The system model has three entities, as depicted in Figure1: the data owner, the data user and the cloud server. A Data owner has a group of data records. A set of distinct keywords is extracted from the data collection. The data owner will construct an encrypted searchable index from the data. Then, the data owner uploads both the encrypted index and the encrypted data collection to the cloud server. Data user provides keywords for the cloud server, respectively. The cloud server simply transmits back top-files that are greatest appropriate to the hunt autopsy. So, the data consumers and Data owners share keys using a key distribution unit.

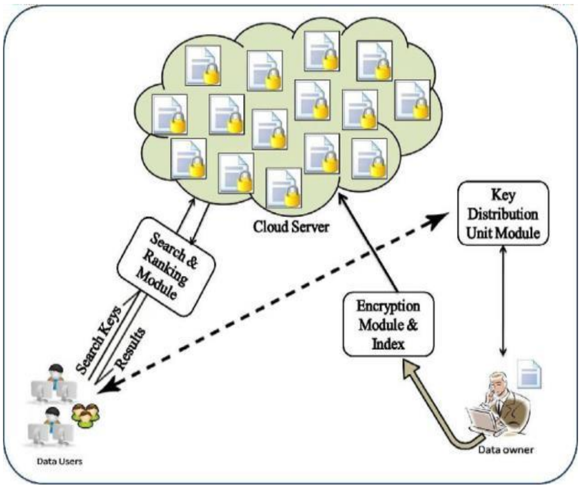

Figure1. The architecture of Keyword-based ranked search over encrypted Cloud

B. Information Salvage Techniques

Information salvage is a process of obtaining required information resources from a collection of information resources. Hunts can be constructed on full-text indexing or metadata. Information salvage systems then rank documents. Most of the information salvage systems assign a numeric score to each manuscript and rank manuscripts by this result. Information salvage models typically express the salvage performance of the system in terms of two quantities: precision and recollection. The precision remains the proportion of the number of related manuscripts recovered in the system to the whole number of manuscripts recovered. Recollection is the part of the number of appropriate manuscripts retrieved for a query to the numeral of manuscripts pertinent to that query in the entire document set. Numerous models had been aimed for information salvage performances. The most widely consumed models in information salvage performances are the conjecture network model, the vector space model and the probabilistic model.

a) Vector Space Model

The vector space model represents text by a vector of terms. Latent Semantic Indexing (LSI) is the way of representing terms and documents in a term- document space, is considered a vector-space information salvage model. The classes of techniques in vector-space models are similar to "formal, feature-based, individual, partial match" salvage techniques. They model the documents as sets of terms that can be separately slanted and controlled, respectively. It performs queries by comparing the representation of the query to each document in the space and can retrieve documents that do not contain one of the search terms.

However, they are more flexible than inverted indices, since each term can be separately slanted, permitting that term to become less or more significant inside a manuscript or the whole manuscript group. This model is associated with exact and lexical matching techniques. Using placing terms, manuscripts, and queries in a term-document galaxy and calculating comparisons among the queries and the manuscripts or terms, permit the solutions of a query to designate graded conferring to the similarity compute consumed. Since words often have multiple meanings, it is difficult for a similar philological method to distinguish among dual manuscripts that distribute a given the word, but use it differently, without understanding the context in the word used. Moreover, a lexical matching technique provides no ranking or straightforward grading system.

b) Probabilistic Model

The probabilistic salvage model remains constructed on the Possibility Grading Notion, which statuses that data salvage system is assumed to grade the manuscripts constructed on their possibility of significance to the query it also describes the uncertainty in the depiction of the information required and the manuscripts. This probabilistic approach in information salvage technique that contains relevant and non-relevant documents is used to estimate the probability of a term appearing in a relevant document, and that decides whether documents are relevant or not. Users start with information needs, in which they translate into query representations. Those are manuscripts, which are transferred into manuscript depictions. Constructed on these two representations, a system tries to satisfy documents information needs. In the vector space models of information salvage, matching is done in a formally defined manner but semantically lacking inexactness of index terms. One of the most probabilistic salvage methods is the statistical distribution of the conditions in together the applicable and non- applicable manuscripts.

c) Inference Network

In this model, a document instantiates a term with confident potency, and accept from many terms is gathered by a given query to calculate the corresponding of a numeric notch for the manuscript. This potency of instantiation of a term for a manuscript can be considered as the weight of the term in the manuscript. Popular approaches include fuzzy sets, symbolic reasoning and diversity of probability designs. Dual implication patterns constructed on probabilistic approaches are of specific curiosity: Dempster-Shafer theory of evidence and Bayesian inference links. The Bayesian inference link remains a DAG(directed acyclic dependency graph) in which nodes represents propositional variables, or constants and edges represents dependence relations between propositions as shown in Fig.2

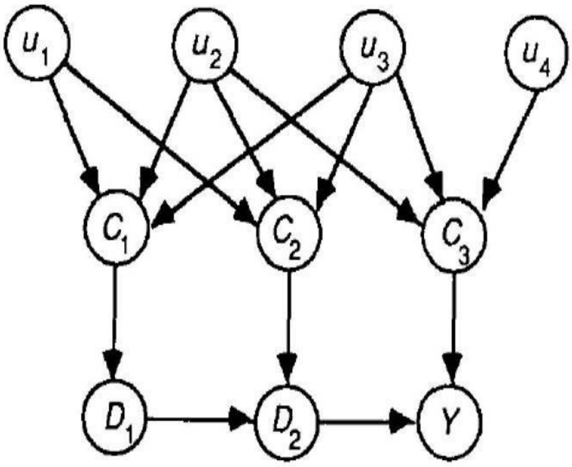

Figure3. Representation of DAG in Bayesian inference network

The Dempster-Shafer theory of evidence is used as an alternative method for evaluating these kinds of probabilistic inference links. Before calculating the confidence related with a query given a set of indicators. Similarly, Dempster-Shafer can be viewed as computing the probability that the indication would permit us to prove the query. Hereafter the Bayesian and Dempster-Shafer models are different and can lead to distinctive solutions, respectively. Entirely the three information salvage techniques support Keyword-based searches.

III. KNOWLEDGE-BASED SEARCH TECHNIQUES

A. Plain text keyword search

This paper proposed a plaintext keyword-based search by protecting data privacy, which is the most prevalent techniques are to selectively recover files finished Keyword- constructed hunt instead of retrieving all the encrypted files related to Keyword. This searching technique is impractical in cloud computing technology because the keyword-based search technique allows users to retrieve files rather than resulting in the desired file. The encrypted file restricts data user's to perform a keyword-based search, and it demands the keyword privacy protection. Thus a plaintext keyword search method fails over encrypted cloud data**.**

B. Fuzzy Keyword Search over Encrypted data

This paper proposed a fuzzy keyword searching technique over encrypted cloud information. This search technique enhances the system usability, which returns the input exactly or the closest possible matching files by keyword search semantics when meticulous harmonizing crashes. The searching techniques are used in three methods—namely, Wild card- Based Technique, symbol-based Gram-Based technique and trie-Traverse Search Scheme. A Wild card-Based Technique solves the problem of edit operations at similar location keywords, respectively. A Gram-Based technique is used to develop the fuzzy set depends on grams. A Symbol-Based trie-Traverse Search Scheme is that all trapdoors sharing a mutual precede may have mutual connections. That Technique is highly efficient, and it also increases the searching to improve the space efficiency, optimize time efficiency, size of the fuzzy keyword set and effectiveness is manageable. Then plentiful stowage is difficult, maintain only Boolean keyword hunt and it does not maintain the graded hunt operation.

C. Boolean Symmetric searchable encryption

This paper describes the searching of cloud data over encrypted format by Boolean Symmetric Searchable Encryption (BSSE). The BSSE technique uses the Boolean expression queries for performing negation, conjunction and disjunction of keywords. A significant process consumed in this method is the Gram- Schmidt orthogonalization process which encodes the keywords, and they are used in queries, labels and inner products to perform the searching of data. This method is fully randomized; the search is linear and increases the size of the label, which implies a more extended computation phase for each document.

Table1. Comparisons of Keyword Based search techniques

S.No	Methods	Goal	Techniques	Data sets	Pros / Cons
1	Plaintext Keyword Search	Selectively retrieve files through Keyword-based search	Plaintext keyword search technique	Data privacy, encrypted files, Keyword-based search	**Pros: P**rotecting data privacy **Cons**: Demands the keyword privacy protection
2	Fuzzy Keyword Search	Enhances the system usability by returning the input exactly or the closest possible matching files	Symbol-Based trie- Traverse Search Scheme, Gram- Based Technique, and Wild card-Based Technique	Edit distance, fuzzy search, and trapdoor.	**Pros:** Edit Distance can be implemented. Highly efficient. Growth probing efficiency. **Cons:** Large storage complexities. Maintenance Only Boolean keyword hunt.
3	Boolean Symmetric Searchable Encryption	Support both conjunctive and disjunctive search	Gram-Schmidt orthogonalization process, Labeling and inner products perform searching of data	Secret sharing, Predicate encryption,	**Pros:** Linear search, Randomized, Focus on simple keyword matching **Cons**: Used only for searching Boolean Queries. Lengthier calculation stage.
4	Secure Ranked Keyword Search	Enables the data users to find the most relevant information quickly	Ranking technique and Order and Maintaining Plotting Method respectively.	Ranked Search, Information retrieval, Secure, searchable index	**Pros:** Highly Efficient. **Cons**: Network Traffic occurs, Large amount of Post–Processing of encrypted files
5	Privacy-Preserving Multi-keyword Ranked Search	Analyze the latent semantic association between conditions and manuscripts by LSA. Employ a safe tearing k- NN method to encrypt the directory, can acquire the correct graded solutions.	Latent Semantic Analysis, k-nearest neighbour (k-NN) Technique	Multi-keyword ranked search scheme and Encrypted cloud information respectively	**Pros:** Eliminate traffic. Recover hunt precision. Privacy-preserving multi Keyword used. **Cons**: Not suitable for large scale data.

D. Ranked keyword search over encrypted cloud data

This paper proposed a Secure Ranked Keyword Hunt done encrypted cloud information. The Ranked Search enables the data users to find the most relevant information quickly and results in the most relevance ranking instead of sending undifferentiated results, and ensures the file retrieval accuracy. This Technique defines a statistical measure approach from information retrieval (IR) and text mining that is used to build a secure, searchable index. There are two types of techniques that are used for Ranked keyword search. First, the Ranking technique which is the most securable searching method, which is used for searching an encrypted data in Cloud. Second, the Order Preserving Mapping Technique that protects sensitive data. This Technique is Very Competent. This searching technique leads to a collision in the network. Encrypted files are processed mainly, and the most extensive encrypted files are post-processed.

E. Multi Keyword ranked search over Encrypted Cloud Data

This paper proposed a Privacy-Preserving Multi- Keyword Ranked Search (MRSE) over encrypted cloud data to provide a result similarity ranking for effective data retrieval, rather than frequent indistinguishable solutions. There are two basic concepts used in this method one is Coordinate matching which obtains the similarity between the search query and data documents. And the other is Inner product similarity the number of query keywords performing in a manuscript, to quantitatively assess the similarity of that document to the hunt query. The document that is associated with a binary vector in index construction which indicates the keyword search of interest. Each bit represented as a sub-index that is used to find the corresponding Keyword is contained in the document, where each bit refers to the corresponding Keyword appears in the search request so that that similarity can be measured precisely by the inner product computation. There are four modules used for performing hunt operation in encrypted cloud information. There are four modules, Client Module, Admin Module, Multi-Keyword Module and Encrypt Module. These method solutions in High Efficiency, since it eliminates unnecessary traffic and amends hunt exactitude. Likeness dimensions also simply hunted. The significant disadvantages in this method are Single Keyword search among grading, and Boolean keyword hunt among grading is not conceivable. Also, it is not suitable for large scale cloud data.

IV. CONCLUSION

In this survey, various methods for retrieving information based on keyword search with privacy are discussed. However, Retrieving data in the encrypted cloud data remains a very vital task as of essential safety and privacy difficulties, with numerous necessities like data privacy, index privacy and keyword privacy. Since the data cannot be accessed using plaintext keyword search over large data files. To overcome this problem, it is necessary to provide Keyword-based search service. Among the various searching technique, multi- Keyword ranked search (MRSE) proposes system-wise privacy, but while considering a large scale data, it reduces the ranking and searches service efficiency. Future work would be proposing an efficient MRSE scheme with accurate document ranking.

ACKNOWLEDGEMENTS

The authors are thankful for M. Li, S. Yu, N. Cao and W. Lou for providing the necessary facilities for the preparation of the paper. Also, thanks to IJRES Journal staffs to publish this paper.

REFERENCES

[1] M. Li, S. Yu, N. Cao, and W. Lou, *"Authorized Private Keyword Search over Encrypted Data in Cloud Computing,"* Proc. 31st Int' l Conf. Distributed Computing Systems (ICDCS'10), pp. 383-392, June 2011.

[2] Google, *"Britney spears spelling correction,"* Referenced online at http://www.google.com/jobs/britney.html, June 2009.

[3] D. Song, D. Wagner, and A. Perrig, *"Practical techniques for searches on encrypted data,"* in Proc. of IEEE Symposium on Security and Privacy '00, 2000.

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

[5] Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, *"Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption,"* Proc. 29th Ann. Int' l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT '10), 2010.

[6] C. Wang, Q. Wang, K. Ren, and W. Lou, *"Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,"* Proc. IEEE INFOCOM, 2010.

[7] A. Singhal, *"Modern Information Retrieval: A Brief Overview,"* IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001.

[8] Tarik Moataz, Abdullatif Shikfa, *"Boolean Symmetric Searchable Encryption"*, ASIA CCS '13 Proceedings of the 8th ACM SIGSAC symposium on Information computer and communications security, .pp. 265-276, NY, USA, 2013.

[9] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, *"Secure Ranked Keyword Search over Encrypted Cloud Data,"* Proc. IEEE 30th Int' l Conf. Distributed Computing Systems (ICDCS '10), 2010.

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

[11] S. Yu, C. Wang, K. Ren, and W. Lou, "*Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing,"* Proc. IEEE INFOCOM, 2010.

[12] N. Cao, S. Yu, Z. Yang, W. Lou and Y. Hou, *"LT Codes-Based Secure and Reliable Cloud Storage Service,"* Proc. IEEE INFO- COM, pp. 693-701, 2012.

[13] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, *"Toward Secure and Dependable Storage Services in Cloud Computing,"* IEEE Trans. Services Computing, vol. 5, no. 2, pp. 220-232, Apr.-June 2012.

[14] Gerard Salton, A.Wong, and C. S. Yang. *"A vector space model for information retrieval. Communications of the ACM",* 18(11):613–620, November 1975.

[15] Amira Sallam, Ahmed Moustafa, Ibrahim El-Henawy *"Keyword Search Techniques over Encrypted Outsourcing Data"* International Journal of Engineering Trends and Technology 65.1 (2018): 20-24.

[16] Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, *"Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption,"* Proc. 29th Ann. Int' l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT '10), 2010.

[17] C. Yu, C. Wanger, P. Renul, and P. Lou, *"Realizing Secure, Scalable, and Fine-Grained Data Entry Switch in Cloud Computing,"* Proc. IEEE INFOCOM, 2011.