International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

An efficient approach secret key for multi-keyword ranked search over encrypted cloud data

A.Pramod kumar & Veerababu kinnera
1Assistant Professor, Matrusri Engineering College, Hyderabad, T.S, India
2Student, Saifabad PG College ,Masabtank, Hyderabad, T.S, India

ABSTRACT: *Data mining focuses on generating a useful knowledge from the data sources rather than a mere data extracting technology. As increasing popularity of cloud computing, more and more data owners are motivated to their data to cloud servers for great convenience and reduced cost in data management. So, sensitive data should be encrypted before outsourcing for, privacy requirements, which obsoletes data utilization like keyword- based document retrieval. A secret key generation for multi-keyword ranked search over encrypted cloud data, which simultaneously dynamic update operations like modify and view of documents. Specifically, the vector space model and the widely-used TF IDF model are combined in the index construction and query generation. The secure kNN algorithm is utilized to encrypt the index and query vectors, and mean while ensure correct relevance score calculation between encrypted index and query vectors. In order to provide security, the secrete key is generated for each data owners individual electronic mail. That secrete key can be copied and pasted for further login. When the user wants to change their password the secret key also been updated.*

KEYWORDS- Multi-keyword ranked search, dynamic update, dynamic secret key, privacy preserving, cloud computing

I. INTRODUCTION

Cloud computing is witnessing rapid innovations in the recent years. It has two main tasks storing and accessing data and programs by means of Internet rather than usage of a computer's hard drive. The entity cloud presents an extensive range of services. It reduces the complexity of the networks, makes provision for customization, scalability, efficiency etc. Besides, the information stored on cloud is generally not easily lost. Because of its on-demand nature, you could typically buy cloud computing the same way we would buy electricity, telephone services, or Internet access from a utility company. It is so easy with the cloud because one can add extra services (or take them away) at a moment's notice as the business needs change.

As cloud technology is becoming more and more wide-spread, the challenges((like leaking of sensitive data [1], hacking [2], unencrypted data at risk [3–5]involved in maintaining the technology is also increasing. Cloud security, the policies, technologies, controls etc that are used to protect the data, the various applications on the cloud and the associated infrastructure, is becoming an integral field of research in the field of Network Security, and more broadly in Computer Security.The evolution of the Cloud Security policies is equally important to keep up with the cloud issues. As a kind of emerging business computational prototype, Cloud Computing distributes computation task on the resource pool which consists of a large number of computers and accordingly the application systems gain the computation working strength, the storage space and software service according to its demand. The working of cloud computing can be viewed by two distinctive features. One is the cloud infrastructure which is the building block for the upper layer cloud application. Cloud computing has achieved two important goals for the distributed computing by the means of three technical methods.

High Scalability the cloud infrastructure can be expanded to very large scale even to thousands of servers and high availability so that the services are available even when quite a number of servers fail. The present-day achievements in data, mobile, wireless and Internet technologies cannot be magnified. And hence Cloud computing is an emerging commercial model that promises to eliminate the need for maintaining expensive computing facilities by companies and institutes alike. Cloud computing technology makes it possible develop and host an application design for the internet where information technology (IT) related facilities are provided

"as a service"; allowing clients to access technology-enabled services more economically and flexibly on a pay-as-you-use basis. Cloud Computing applications are cloud based services also known as Software as a Service (SaaS). These applications can do everything from keeping track of notes to accounting. Cloud applications give operatives access to their information from anywhere around the globe and must required an Internet connection. This ensures team work, allowing collaborated working as multiple people can view and edit the same information at once. Cloud applications also allow enterprises to push new developments to all users at once, ensuring all round benefit at the same time. Among the many incentives for using cloud, organizations are looking into ways to assess some of the applications they plan on employing into their environment through the use of a cloud. The adoption of a hybrid cloud approach consents for testing application workloads that can provide a comfortable environment without an initial investment.

An organization would seek to have the additional capacity and availability of an environment when needed on a pay-as you-use basis. With cloud computing, there are now readily available environments personalized for your needs often combining automated provisioning of physical and virtualized resources.

Cloud can offer the leeway of storing files and accessing, retrieving and recovering them from any web enabled interface. We have high availability, speed, scalability and security for the environment at all times. There is also the possibility to store the data either on or off premises depending on the regulatory compliance requirements. Yet another benefit resulting from the use of cloud based on the cost effectiveness of a Disaster Recovery (DR) solution that provides for a faster recovery from a network of different physical locations at a much lower cost that the traditional DR site with fixed resources, a much higher cost and rigid procedures. Cloud-based backup can be one of the solutions where we can automatically dispatch data to any location across the wire without any issues of security, availability and capacity

II. RELATED WORKS

In this section, we have studied previous research papers related to the privacy preserving data mining (PPDM) and query processing over encrypted data. The brief review of existing related work is as follows:

S. De Capitani di Vimercati, S. Foresti, and P. Samarati[1] Proposed a Ensuring proper privacy and protection of the information stored, communicated, processed, and disseminated in the cloud as well as of the users accessing such information is one of the grand challenges of our modern society. As a matter of fact, the advancements in the Information Technology and the diffusion of novel paradigms such as data outsourcing and cloud computing, while allowing users and companies to easily access high quality applications and services, introduce novel privacy risks of improper information disclosure and dissemination. The different aspects of the privacy are characterized in to privacy risk in the cloud, privacy risk for users, privacy risks for stored data and privacy risk for data access.

P. Samarati and S. De Capitani di Vimercati [2] Approach Data outsourcing is an emerging paradigm that allows users and companies to give their (potentially sensitive) data to external servers that then become responsible for their storage, management, and dissemination. Although data outsourcing provides many benefits, especially for parties with limited resources for managing an ever more increasing amount of data, it introduces new privacy and security concerns. The main issues that needs to be addressed for guaranteeing proper protection and access to outsourced are, data protection, query execution, private access, data integrity and correctness, access control enforcement and private collaborative computation. A first solution used for preventing a server from accessing data stored on its own machines consists in encrypting the data before outsourcing them.

H. Hu, J. Xu, C.Ren,and B. Choi [3] Proposed Query processing that preserves both the data privacy of the owner and the query privacy of the client is a new research problem. It shows increasing importance as cloud computing drives more businesses to outsource their data and querying services. However, most existing studies, including those on data outsourcing, address the data privacy and query privacy separately and cannot be applied to this problem. A holistic and efficient solution

that comprises a secure traversal framework and an encryption scheme based on privacy homomorphism has been proposed. The framework is scalable to large datasets by leveraging an index-based approach.

Based on that, a secure protocol for processing typical queries such as k-nearest-neighbor queries (kNN) on R-tree index has been proposed. P. Paillier [4] approach composite residuosity class problem is a significant computational methods applied to public-key cryptography. A new trapdoor permutation and two homomorphic probabilistic encryption schemes computationally comparable to RSA has been proposed. The proposed cryptosystems, based on usual modular arithmetic, are provably secure under appropriate assumptions in the standard model. The new trapdoor mechanism is based on composite residuosity in contrast to prime residuosity. The trapdoor provides a new cryptographic building-block for conceiving public-key cryptosystems.

M. S. Islam, M. Kuzu, and M. Kantar cioglu [5] Implement Remote data storage offers reduced data management overhead for data owners in a cost effective manner. Sensitive documents, however, need to be stored in encrypted format due to security concerns. But, encrypted storage makes it difficult to search on the stored documents.

Various protocols have been proposed for keyword search over encrypted data to address this issue. Most of the available protocols leak data access patterns due to efficiency reasons. A simple technique to mitigate the risk against the proposed attack at the expense of a slight increment in computational resources and communication cost was proposed. The proposed mitigation technique is generic enough to be used in conjunction with any searchable encryption scheme that reveals data access pattern. The disclosure of data access patterns during „search over encrypted text". Pose a potential vulnerability. Formalized model that can be used to launch an inference attack utilizing this vulnerability and empirically show their efficiency in successfully predicting query identities. The "hiding access pattern" is extremely important in encrypted keyword search and therefore is a necessary characteristic of a secure encrypted search scheme.

E. Shi, J. Bethencourt, T,-H, Chan, D. Song and A.Perrig [6] Proposed a scheme allows a network gateway to encrypt summaries of network flows before submitting it to an untrusted repository. When network intrusions are suspected, an authority can release a key to an auditor, allowing the auditor to decrypt flows whose attributes fall within specific ranges. However, the privacy of all irrelevant flows is still preserved. The security for Multi-dimensional Range Query over Encrypted Data and prove the security of our construction under the decision bilinear Diffie-Hellman and decision linear assumptions in certain bilinear groups was defined formally. Multi-dimensional Range Query over Encrypted Data implies a solution to its dual problem, which enables investors to trade stocks through a broker in a privacypreserving manner.

R. Agrawal and R.Srikant[7] Design a privacy preserving data mining was addressed for a scenario in which two parties owning confidential databases wish to run a data mining algorithm on the union of their databases, without revealing any unnecessary information. The generic protocols in such a case are of no practical use and therefore more efficient protocols are required.

III. PROPOSED WORK

The search process of the UDMRS scheme is a recursive method upon the tree, named as "Greedy Depth first Search (GDFS)" algorithm.

- Based on the UDMRS scheme, build the basic dynamic multi-keyword ranked search (BDMRS) scheme with using the secure kNN algorithm.
- The BDMRS scheme can defend the Index Confidentiality and Query Confidentiality in the identified cipher text model.

Module Description:

1) Search Process of UDMRS Scheme
2) BDMRS Scheme
3) EDMRS Scheme
4) Individual Key and Password Update

Search Process of UDMRS scheme The search process of the UDMRS scheme is a recursive process upon the tree, named as "Greedy Depth first Search (GDFS)" algorithm.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

Construct a result list denoted as RList, whose element is defined as ⟨RScore; FID ⟩. Here, the RScoreis the relevance score of the document fFIDto the query, which is calculated according to Formula

(1). The RListstores the k accessed documents with the largest relevance scores to the query. The elements of the list are ranked in downward order according to the RScore, and will be updated timely through the search process.

2). RScore(Du;Q) – The function to calculate the application score for query vector Q and index vector Du stored in node u.

kthscore– The smallest relevance score in accessible RList, which is initialized as 0.

hchild– The child node of a tree node with upper relevance score.

lchild– The child node of a tree node with lower bearing score.

Since the possible largest bearing score of documents rooted by the node u can be predict, only a part of the nodes in the tree are access through the search process.

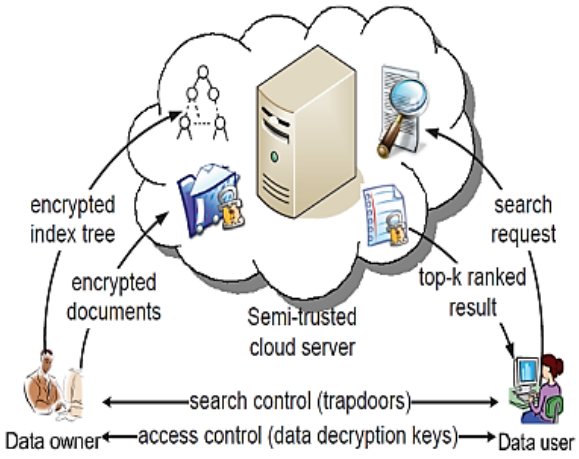

Fig-1 System Architecture

Algorithm 1

Input: the document collection F = {f1,f2,.., fn} with the identifiers FID = {FID|FID = 1,2,..,n}.

Output: the index tree T for each document fFIDin F do
Construct a leaf node u for fFID, with u.ID =GenID (), u.Pl =u.Pr = null, u.FID = FID, and
D[i] = for i = 1,…..m;
Insert u to CurrentNodeSet;
end for
while the number of nodes in CurrentNodeSet is larger than 1
do
if the number of nodes in Current NodeSet is even, i.e. 2h then for each pair of nodes u′ and u″ in CurrentNodeSet do
Generate a parent node u for u′ and u″, with u.ID = GenID(), u.Pl = u′, u.Pr = u″, u.FID =0 and D[i] = max{u′.D[i]; u″.D[i]} for each i = 1,…..,m;
Insert u to TempNodeSet;
end for
else
for each pair of nodes u′ and u″ of the former (2h − 2) nodes in CurrentNodeSet
do
Generate a parent node u for u′ and u″;
Insert u to TempNodeSet;
end for
Create a parent node u1 for the (2h − 1)-th and 2h-th node, and then create a parent node u for u1 and the (2h + 1)-th node;
Insert u to TempNodeSet;
end if
Replace CurrentNodeSet with TempNodeSet and then clear TempNodeSet;
end while
return the only node left in CurrentNodeSet, namely, the root of index tree T ;

Algorithm 2 GDFS (index tree node u)
1: if the node u is not a leaf node then
2: if RScore(Du,Q) >kth score then
3: GDFS(u:hchild);
4: GDFS(u:lchild);
5: else
6: return
7: end if
8: else
9: if RScore (Du,Q) >kth score then

10:Delete the element with the smallest relevancescore from RList;
11: Insert a new element RScore(Du,Q); u.FID⟩andsort all the elements of RList;
12: end if
13: return
14: end if

BDMRS Scheme: The UDMRS scheme based on construct the basic dynamic multi-keyword ranked search (BDMRS) scheme by using the secure kNN algorithm. The BDMRS scheme is designed to achieve the goal of privacy preserving in the known cipher text model, and the four algorithms included are described as follows:

SK ← Setup() Initially, the data owner generates the secret key set SK, including

1) a randomly generate m-bit vector S where m is equal to the cardinality of vocabulary, 2) two (m×m) invertible matrices M1 and M2. Namely, SK = {S;M1;M2}.

I ← GenIndex (F; SK) First, the unencrypted index tree T is built on F by using T ← Build Index Tree(F).

Secondly, the data owner generate two random vectors {Du′;Du″} for index vector Du in each node u, according to the secret vector S. specially, if S[i] = 0, Du′[i] and Du″[i] will be set equal to Du[i]; if S[i] = 1, Du′[i] and Du″[i] will be set as two random values whose sum equals to Du[i].

Finally, the encrypted index tree I is build where the node u supplies two encrypted index vectors Iu= {MT1 Du′;MT2Du″}.

TD ← GenTrapdoor(Wq; SK) With keyword set Wq, the unencrypted query vector Q with length of m is generated. If wi∈ Wq, Q[i] stores the normalize IDF value of wi; else Q[i] is set to 0. Similarly, the query vector Q is split into two random vectors Q′ and Q″. The disparity is that if S[i] = 0, Q′[i] and Q″[i] are set to two accidental values whose sum equals to Q[i]; else Q′[i] and Q″[i] are set as the same as Q[i]. Finally, the algorithm returns the trapdoor

TD = {M−11 Q′;M−12 Q″}.

Relevance Score ← SRScore(Iu;TD) With the trapdoor TD, the cloud server computes the significance score of node u in the index tree I to the query. Note that the relevance score calculated from encrypted vectors is equal to that from unencrypted vectors.

EDMRS Scheme: The BDMRS scheme can defend the Index Confidentiality and Query Confidentiality in the known cipher text model. However, the cloud server is able to link the similar search requests by track path of visited nodes.

In addition, in the known backdrop model, it is possible for the cloud server to recognize a keyword as the normalize TF division of the keyword can be exactly obtained from the final calculated significance scores. The primary cause is that the relevance score calculated from I_u and TD is exactly equal to that from D_u and Q. A heuristic method to further improve the security is to break such exact equality. Thus introduce some tunable arbitrariness to upset the relevance score calculation. In addition, to ensemble different users' preference for higher accurate ranked results or better sheltered keyword privacy, the randomness is set adjustable.

IV. CONCLUSION

Search retrieval is the most essential task in the cloud based search engines where the multi cloud users attempts to retrieve the contents based on their need. In this case security and privacy becomes the most important issue where the users need to submit their information to retrieve the result as per their requirements. In this work, privacy preserved search retrieval based on multi keyword search is introduced which will retrieve the contents in terms of user submitted query in the Vector space model. Along with these, the individual key and Password update based security scheme is supported to enable the multi keyword search retrieval in the more secured manner. The experimental tests conducted were proves that the proposed approach provides better result than the existing work in terms of improves search retrieval accuracy and the privacy.

REFERENCES

[1] S. De Capitani di Vimercati, S. Foresti, and P. Samarati, "Managing and accessing data in the cloud: Privacy risks and approaches," in proc.7th Int.Conf. Risk Security Internet Syst., 2012, pp. 1-9.

[2] P. Samarati and S. De Capitani di Vimercati, "Data protection in Outsourcingscenarios: Issues and direction," in proc. Of ASIACCS, China, Apr. 2010.

[3] H. Hu, J. Xu, C.Ren,and B. Choi, "Processing Private queries over untrusted data cloud through privacy homomorphism," in Proc, IEEE 27th Int ,Conf,Data Eng., 2011, pp. 601-612.

[4] P. Paillier, "Public key cryptosystems based on composite degree residuosity classes," in Proc, 17th Int, Conf, Theory Appl, Cryptographic Techn., 1999, pp.223-238.

[5] M. S. Islam, M. Kuzu, and M. Kantar cioglu, "Access pattern disclosure on searchable encryption: Ramification, attack and mitigation," in NDSS, 2012.

[6] E. Shi, J. Bethencourt, T,-H, Chan, D. Song and A.Perrig "Multidimensional range query over encrypted data," in IEEE Symposium on Security and Privacy (SP"07). IEEE, 2007, PP. 350-364.

[7] R. Agrawal and R.Srikant, "Privacy-Preserving data mining," ACM Sigmod Rec., vol. 29,pp. 439-450,2000.

[8]. Wenhai Sun, Wenjing Lou, Y. Thomas Hou, and Hui Li" Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking". 2014, Volume: 25.

[9]. Yakoob. Sk., Dr. V Krishna Reddy and C. Dastagiraiah "Analysis of Keyword Searchable Methodologies in Encrypted Cloud Data".

[10]. Zhihua Xia, Xinhui Wang, Xingming Sun and Qian Wang "A Secure and Dynamic Multi- keyword Ranked Search Scheme over Encrypted Cloud Data" IEEE Transactions 2015.