SURVEY ON: CLOUD DATA RETRIEVAL FOR MULTIKEYWORD BASED ON DATA MINING TECHNOLOGY

[1]Kavya G, [2]K.S. Rajesh

[1]P.G Scholar, Dept. of Computer Science & Engineering RajaRajeswari College of Engineering, Bangalore, Karnataka
[2]Associate Prof. Dept. of Computer Science & Engineering RajaRajeswari College of Engineering, Bangalore, Karnataka

Abstract: **Cloud computing has emerging as a pattern for data outsourcing and it provide high-quality of data services. It concerns of very sensitive information on cloud and it causes potentially the privacy problems. Encryption protects data security to some level, but at the cost of compromised efficiency. Searchable symmetric encryption going allows retrieval of encrypted data above cloud. Here focus on addressing data privacy problems using. For the first time, formulate the privacy issue from the characteristic of similarity relevance and scheme robustness. Here observe that server-side position based on order preserving encryption inevitably disclosures data privacy. To eliminate the leakage, I propose a two-round searchable encryption scheme that supports multikeyword top-key retrieval in the cloud. In TRSE, Employ a vector model and holomorphic encryption. The vector space model helps to provide sufficient search correctness, and the holomorphic encryption it enables users to involve in the ranking while the major of computing work is done on the server side by operations only on cipher text .As a result, the information leakage here focus on addressing data privacy problems using . For the first time, formulate the privacy issue from the characteristic of similarity relevance and scheme robustness. Here observe that server-side position based on order preserving encryption (OPE) inevitably disclosures data privacy Addressing data privacy problems using Searchable Symmetric Encryption in Cloud Data and its Relative services with concern to leakage avoidance for efficient query solution retrieval.**

Keywords: **cloud, data privacy, holomorphic encryption, preserving encryption.**

I. INTRODUCTION

The cloud computing a critical pattern for advanced data services, has become a necessary feasibility for user to outsource data. There are several Arguments on privacy however, the privacy have been incessantly presented as outsourcing of personal sensitive information including the e-mails, health history, personal files, and personal messages is expanding explosively . The records of data losing and privacy breaks in cloud computing systems appear from time to time The main problem is treat on data privacy starts in the cloud itself When client upload or outsource their private data onto the cloud the cloud service providers are able to control and monitor the between cloud users communication about the data will happen properly or improperly. The latest model to emerge is that of Cloud computing which potentials reliable services delivered through other generation may be next-generation data centers that are built on virtualized compute and storage technologies. Clients will be able to access applications and data from a “Cloud” anywhere in the world on demand when they want the information they will access since the cloud. The customer are very confident about the cloud infrastructure, that is very strong and very robust the information always available when they want to access the information at any time the information is ready to access. Computing services need to be highly reliable, scalable, robust, strong and autonomic to support worldwide access, dynamic discovery. In particular clients indicate the required service level through “Quality of Service” that will write usually with service provider these examples the recently emerged Cloud computing paradigm appears to be the most promising paradigms.

II. RELATED WORK

The cloud computing system consider the data service, which mainly involved the three entities which they involved data user, data owner and cloud server. The cloud retrieves the data services and server host the third party data storage services that contain the important information are the sensitive information the cloud server protecting the data that data cannot trusted fully the cloud data are entrusted, the files are must be encrypted the data privacy are unacceptable and that kind of leakage that data would affect the privacy as un acceptable. The companies and the individuals are in wide range for them the serious cause that occurring frequently on the top most is the loss and theft of the laptops and the subsequent data discloser that has been ranked in top most frequently occurring incidents. To control this threat data confidentiality is important the present technologies not satisfy the requirements. The emerging action on the internet access and cloud computing by using the “cloud shedder”. The attackers can get the junk or unwanted files have obtained in the hard drive by using the cloud shredder the genuine user can access the data in the file the security services are transparent. There are some of the expectations as below:

1. The data loss will happen only the laptop are out of owners control for temporarily that not in use ,at that time the user should logged off the system or shut shutdown the server system.
2. Whenever the mobile network and the Wi-Fi using the cloud storage services that will access the data, cloud that will have connection absence such as the cloud or transmission failure in this scheme.
3. The Cloud storage is not trusted one we can’t able trust that network fully the files content are the curious with the other adversaries of the network on the cloud. The cloud are the not fully trusted on the network.
4. The most frequently the very large scheme not considered exceptionally one the scheme confidential files and documents in the slides and the trip are the carried are used confidently used file

The data is very sensitive the servers are storing in the database by protecting the data encrypted the plaintext information by hiding by storing the assured the data. the loss of the data and ability to how it will use useless in the sense, the attacker will use the encryption will use the sensitive data to protect by the user the attacker will compromised data from the database the data is protected from the attacker it will remains safe .the decryption keys are protected properly When the encryption was done correctly or exactly the encryption will protected the data that is very sensitive data it will reduces or decreases the amount of the identity and the data owners decreases the liability. The cipher text property describes successfully the cryptographic techniques, the large amount of data are distributed and stored in the the outsourcing allows the client to store the data. The symmetric encryption has been proposed by providing the capabilities the research is the searchable encryption , [2]The technologies that allows the service to the customer they allows the different type of services high customizable and the expertise flexibly integrate customer services and delivered the highly customizable services data mash up is the application that allows special type of services integrated the data from the provider depending on the technologies and the user request it is going to integrated with the data from the multiple sources bring that challenges Data providers contain the sensitive data join the multiple private data that is sets together with the sensitive information to the other data provider. The integrated data and the identification of the data from the individuals and could potentially sharpen the identification of individuals and, there reveal their specific person information on the sensitive information that was not available before the mash up the data from the different sources are mash up in this technology where the technologies that allows the service to the customer they allows the different type of services high customizable and the expertise flexibly integrate customer services and delivered the highly customizable services data mash up is the application.

When the traditional model privacy was such as the high-dimensional data of the data, would suffer from the known as the of high dimensionality problem curse , the data analysis is resulting in unnecessary data for further data analysis, in the social network the problem is privacy that is mash up with the online services the privacy was resolve a privacy problem in a real-life mash up application for the online advertising industry in social networks where the algorithm to address the aforementioned challenges so they propose a service-oriented architecture along with a privacy-preserving data mash up. The challenges on real-life data that suggest that our proposed architecture and algorithm is very effective for preserving simultaneously on both information and the privacy utility on the mash up data. [1]The real applications in the project require parties that transfer the sensitive are very important data or the information securely in the real word, in this article the system provides the exact match capabilities to provide that is secure anonymous data base search this provides the secure and the private search where the Boolean files are filtered in the encryption the efficient parties execute the exact search match this presents the normal search match for the preserving the guarantees the engineering secure search system where it shows, this article further considerations for the more general settings of the search engines for the framework for the process where it contain the existing work similarity in the error tolerance and the hamming distance and primitive, for engineering usable private secure search systems.

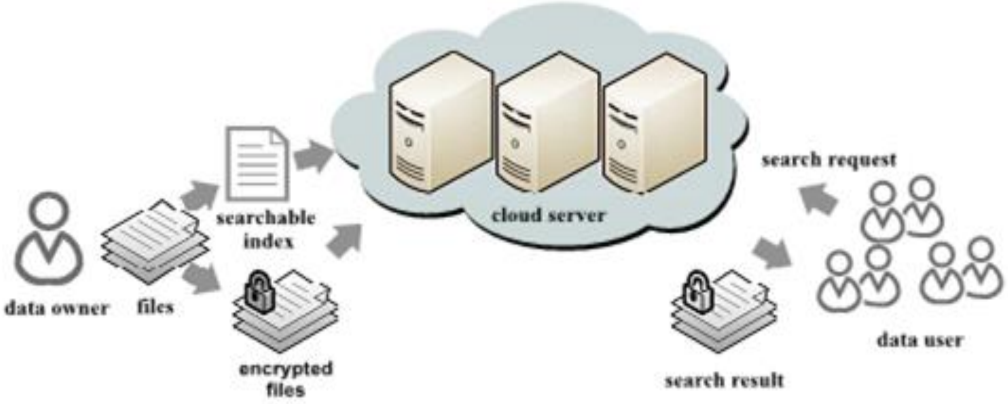
III. PROPOSED WORK

This schemes presented that support the single keyword that retrieval under the various scenarios to improve the security on the top key single keyword the authors are try to improve or solve the top k by using the multikeyword over cloud data.it suffer from two problems how to strike a balance between security and efficiency and Boolean representation. It supports only keyword retrieval without ranking only the Boolean, and supports only the single key words, interaction between the users and sever side and high communication overhead on user side. Here to avoid the leakage of information to the user they preventing the cloud from involving in ranking and entrusting all the work to the user is a natural way to avoid information leakage from the user. However, the high computational overhead prevents the information security from limited computational power on the user side and the. cloud data retrieval for multikeyword data for the top-k retrieval over encrypted cloud data using the data mining technologies issues are, thus, is: it will show the without information leakage How to make the cloud do more work during the process of retrieval. In this project we introduce the concepts of relevance and similarity and scheme robustness to express the privacy issue in searchable encryption schemes, and they proposed a two-round searchable encryption (TRSE) scheme it's going to solve the insecurity problem. The information cloud data retrieval community are employed by using Novel technologies in the cryptography community and, including holomorphic encryption and vector space model.

In the proposed work the part of user work of ranking the highest of computing work on the cloud, which guarantees the data retrieval from top-k multikeyword retrieval over encrypted cloud data by using data mining technologies with high security and practical efficiency.Here we can summarized the contributes:

1. Here inevitably violates data privacy will show server-side ranking based on order-preserving encryption propose the concepts of similarity relevance and scheme robust thus, perform the first attempt to formulate the privacy issue in searchable encryption, and we show) inevitably violates data privacy.
2. Propose a TRSE scheme, employ relevance to support the multikeyword top-k retrieval which fulfils the secure cloud data retrieval from multikeyword top-k over encrypted cloud data from the data mining, specifically, for the first time, they support the relevancy score.
3. The high data privacy is proposed that guarantees thorough analysis on security demonstrates the Furthermore, performance analysis and experimental results show that our scheme is efficient for practical utilization.

Here the two main important issues of the project work the security analysis and receptivity performance analysis the related background scenario then the security definitions and issues with the existing scheme that presents the detailed description of the proposed searchable encryption scheme. The design process the creative that requires the proper and insight and talented skills that required is the part of the design process where the design must be practiced and well developed and learned by the developer or experienced person and the study of the existing system the person should understand the existing system and its fault the design process is involved in the several model developing at the different level the several model has designed the process The design process involves developing several models of the system at different levels of abstraction. The System design is concerned are concentered with the how the functional is to be provided by the different components of the system where the system is developed with the functionalities is to be provided by the different components of the system where it is the very beginning step or first step to move from the domain problems towards the solution of the that particular domain



the system design is basically the bridge between the requirement and the specification and the final solution of satisfying the requirements.it has to be processed through the which requirements are transformed are translated into the representation of the software. that leads to the design very close to the source code of the system. The above fig shows the three different entities are involved: cloud server, data owner, and data user. Where it contain the data owner to upload the files in the encrypted format and it has to maintained the searchable index and it contain the cloud sever it acts has the third party to store the files but it is untrusted, the data user having the authority to access the file but he should be the authorized person to search the request and to search the result in the document in the cloud. In the proposed work in the searchable index data owner creates the index with all the hash code of each keyword in the searchable index when for the multikeyword receives a query from the cloud server, from the encrypted index computes the score from the multikeyword the encrypted index stored the on the cloud the data user score store from the scores return from the files it will store in the data user. The data user going picks or collects the highest top-k scoring files to identify to the request to the server, when the cloud server going to download the files and decrypt that files. Where the retrieval takes the communication between the two round searchable communication between the cloud server and the data user. The two round searchable encryption scheme where the ranking is going done on the user side and the where the scoring calculation is done on the on the sever side.

The frame work of the two round searchable encryption includes the four algorithms that are setup, index build, trapdoorgen score calculate and the rank. Framework of TRSE includes four algorithms:

Setup, IndexBuild, TrapdoorGen, ScoreCalculate ,Rank.

The two round searchable encryption shows that containing the above five algorithms that will mainly help to the user to protect the data in the cloud where mainly using the vector space model and homomarpic encryption that helps the user to protect data and it maintain the loss of the data in the cloud. The algorithm that containing the setup, index,build, trapdoorgen score calculate and the rank.

1. **Setup:** the data owner is going to generate the secrete key for the scheme of encryption where owner is generating the public and the private keys for the encryption scheme. the security of the file is more to maintain that they created the setup the security parameter is taken as the input they considered as a input and the output is secret key SK, where input output both are security parameter and a public key set PK.

2. **index Build:** the data owner is going to generate or build the searchable index from the file where they are build the secure searchable index I and that files are the collection contain the index I from C, the index is encrypted into index with the private key where the secure searchable index is outsourcing the output in the searchable index.

3. **Secure trapdoor:** when the receiver or the client receives the secure trapdoor from the data user request REQ. the vector space model that built the user multikeyword request where the data users request REQ and then that is encrypted into the trapdoor where the secure trapdoor encrypted with the public key from the private key the out is going to secure in the trapdoor.

4. **Score calculate:** when the cloud server going to receive the secure trapdoor where the cloud server computes the scores where in each files with the n number with the index with the trapdoor and return the encrypted the cloud data the result of the vector with the back to the data user.

5. **Rank:** when the cloud server going to decrypt the vector by using the secret key SK the cloud server decrypted the files and use the vector get the decrypted files with secure top-K scores.

The whole work of the cloud data retrieval by using the multikeyword can be divided into two types that are by using those two phases the entries process is going work. That two phases are as follows; Initialization phase, Retrieval phase

IV. CONCLUSION

In cloud computing, data is shared among validated users as defined by owner. It relies on keyword based retrieval. This becomes ineffective on encrypted data. In order to improve relevancy ranking method is used on interest and relevant data/file is sent to users. A Boolean keyword search is employed to perform search on encrypted data for keyword mapping and search. Ranking is then performed for presenting and retrieval. There is high probability of compromising Security and Efficiency here with a leakage because of keyword. The proposed solution solved problem by introducing the concepts of relevance mapping and robustness formulation for privacy issue in existing methodology. The issue and problem regarding insecurity is solved using 2-round search encryption scheme by utilizing newer cryptography and information retrieval community solutions. We devised and used vector space model for mapping relevancy for encryption a RSA approach is used. Since all computing is done cloud, the relevancy is dynamically retrieved along user's use. Based on this ranking is mapped and using these to assure the multi keyword retrieval as vector space helps us To perform all these a robust cloud availability is utilized. A server or backend ranking is maintained in order to enable security and privacy of data. Since we retrieve multi-keyword search results and relevance score is mapped accordingly. This helps us to build robust solution that is secure and reliable. This avoids data Leakage.

ACKNOWLEDGEMENT

It's my immense pleasure to express my indebtedness to my guide Associate Prof Rajesh k s department of computer science and Engineering at Rajarajeswari college of Engineering College Bangalore for guiding me at various stages.

REFERENCES

[1] Konwinski, G. Lee, D. Patterson, A. Rabkin, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58, 2010.

[2] M. Arrington, "Gmail Disaster: Reports of Mass Email Deletions,"http://www.techcrunch.com/2006/12/28/gmail-disasterreportsof- massemaildeletions/, Dec. 2006.

[3] Amazon.com, "Amazon s3 Availability Event: July 20, 2008," http://status.aws.amazon.com/s3-20080720.html, 2008

[4] 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), 2010.

[5] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. ACM 13th Conf. Computer and Comm. Security(CCS), 2006.

[6] International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014 Proceedings of International Conference On Global Innovations In computing(ICGICT'14)Organized by Department of CSE,

[7] C. Gentry, "Fully Holomorphic Encryption Using Ideal Lattices,"Proc. 41st Ann. ACM Symp. Theory of computing (STOC), pp. 169-178, 2009.

[8] D. Dubin, "The Most Influential Paper Gerard Salton Never Wrote," Library Trends, vol. 52, no. 4, pp. 748-764, 2004.

[9] Ddd A. Cuyt, V. Brevik Petersen, B. Verdonk, H. Waadeland, and W.B. Jones, Handbook of Continued Fractions for Special Functions. Springer Verlag, 2008.

[10] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, Introduction to Algorithms, pp. 856-887. MIT Press and McGraw-Hill, 2001.