Cloud Data Retrieval for Multi Keyword based on Data Mining Technology

Kavya G
P.G Scholar, Dept. of Computer Science & Engineering
RajaRajeswari College of Engineering
Bangalore, Karnataka, India

K.S. Rajesh
Associate Prof. Dept of Computer Science & Engineering
RajaRajeswari College of Engineering
Bangalore, Karnataka, India

***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 (SSE) going allows retrieval of encrypted data above cloud. Here focus on addressing data privacy problems using SSE. 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. To eliminate the leakage, I propose a two-round searchable encryption (TRSE) 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 SSE. 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.**

***Index Terms*— cloud, data privacy, holomorphic encryption, preserving encryption.**

I. INTRODUCTION

The cloud computing a critical pattern for advanced data services, has become a necessary feasibility factor for data users to outsource data. There are several Controversies on privacy however, have been incessantly presented as outsourcing of personal sensitive information including emails, health history, personal files, and personal messages is explosively expanding. Reports of data loss and privacy breaches in cloud computing systems appear from time to time The main problem is treat on data privacy roots 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 data and the communication between users and the cloud at will, lawfully or unlawfully. 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 from the cloud. The clients are assured that The Cloud infrastructure is very robust and that information is very strong will always be available at any time and that information available when they want to access. Computing services need to be highly reliable, scalable, robust, strong and autonomic to support universal access, dynamic discovery. In particular clients indicate the required service level through "Quality of Service" parameters, which are noted in SLAs traditional with providers. Fall these paradigms, the recently emerged Cloud computing paradigm appears to be the most promising paradigms.

II. LITERATURE SURVEY

Here we describe how the earlier technologies were used to prevent the data.

A. Clearing the clouds away from the true potential and obstacles posed by this computing capability

"A View of Cloud Computing" by Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, Dav id Patterson, Ariel Rabkin, Ion Stoica." Cloud computing, the long-held dream of computing as a utility, has the potential to transform a large part of the IT industry, making software even more attractive as a service and shaping the way IT hardware is designed and purchased. Developers with innovative ideas for new Internet services no longer require the large capital outlays in hardware to deploy their service or the human expense to operate it. They need not be concerned about overprovisioning for a service whose popularity does not meet their predictions, thus wasting costly resources, or under provisioning for one that becomes wildly popular, thus missing potential customers and revenue. Moreover, companies with large batch-oriented tasks can get results as quickly as their programs can scale, since using 1,000 servers for one hour costs no more than using one server. Our goal of this article is to reduce that confusion by clarifying terms, providing simple figures to quantify comparisons between of cloud and conventional computing, and identifying the top technical and non-technical obstacles and opportunities of

cloud computing. Armbrust is a more detailed version of this article.

B. *Efficient Secure Ranked keyword search Algorithms over outsource cloud data*

Ms. Mayura R. Girme1, Prof.G.M. Bhandari21,2Department of Computer Science and Engineering Bhivarabai Sawant Institute of Technology & Research (BSIOTR) they define this paper solve the effective yet secure ranked keyword search over encrypted cloud data. used order preserving symmetric encryption to protect the cloud data. Even though there are lots of searching techniques available, they are not giving efficient search results. For example the search results returned 40 records and in those 30 records are relevant and the remaining 10 records result contains irrelevant data. This paper mainly focuses on searching methods which will improve the efficiency of searching. We used both keyword search and concept based search methods in order to retrieve the relevance search criteria. This method will retrieve the documents based on broader conceptual entities, which will improve the efficiency of ranked keyword search. Traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing cloud environment, they may suffer from the following two main drawbacks. On the one hand, for each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved file in order to find ones most matching their interest, which demands possibly large amount of post processing overhead On the other hand, invariably sending back all files solely based on presence/absence of the keyword further incurs large unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm.

C. "Retrieval of Encrypted cloud data using multikeyword"

Rajesh Kumar, Dr. K. Rubasoundar authors proposed the retrieval of encryption Cloud Computing represents one of the most significant shifts in information technology many of us are likely to see in our lifetimes. For the protection of data privacy, sensitive data has to be encrypted before outsourcing, which makes effective data utilization a very challenging task. Here searchable symmetric encryption (SSE) used to secure and retrieve the data from the cloud. In this work, we focus on addressing data privacy issues using SSE. The concept was formulating the privacy issues of the stored data and to retrieve the data from the cloud in a sequence manner. We observe that server-side ranking based on order-preserving encryption (OPE) inevitably leaks data privacy. By using OPE we find a Boolean search. To avoid the leakage of data, here we propose a two-round searchable encryption (TRSE) scheme that supports top-k multikeyword retrieval. In TRSE, we employ a vector space model and homomorphism encryption. As a result, information leakage can be eliminated and the stored data is secured. This is proposed scheme guarantees high security and practical efficiency.

D. Single-Database Private Information Retrieval From Fully Homomorphic Encryption

In this the authors Xun Yi, Mohammed Golam Kaosar, Russell Paulet, and Elisa Bertino, Fellow, IEEE (2013) are discus about the private information retrieval that allows a user to retrieve the data without revealing the data to the database server in this they concentrated about the PIR facilities protocols with the configuration protocols work as if the user sends the index of a bit or a block to the database server and then receives the bit or the block from the database server. Security analysis has shown that the generic single-database PIR is semantically secure if the underlying FHE scheme is semantically secure. We have implemented our practical PBR protocol based on a variant of the DGHV somewhat homomorphic encryption scheme for a database composed of 10,000 elements of size of 200k bits. On an Intel Core2 Duo CPU E4600 with clock speed of 2.40 GHz, our experiment has shown that our PBR protocol is practical. Compared with existing PIR and PBR protocols, our PIR and PBR protocols are conceptually simpler. Our practical PBR protocol has lower computation complexity but higher communication complexity than existing PBR protocols. Overall, our practical PBR protocol is more e efficient than existing PBR protocols in terms of total protocol execution time when a high-speed network is available. Our future work will further improve efficiency of PIR and PBR protocols from variants of existing FHE schemes, such as

E. Syslog

C.Lonvick, The BSD syslog protocol, the authors The syslog process was one such system that has been widely accepted in many operating systems. Flexibility was designed into this process so the operations staffs have the ability to configure the destination of messages sent from the processes running on the device. No stringent coordination is required between the transmitters and the receivers. Indeed, the transmission of syslog messages may be started on a device without a receiver being configured, or even actually physically present. In another dimension, the syslog process could be configured to forward the messages across a network to the syslog process on another machine. the syslog protocol provides a transport to allow a machine to send event notification messages across IP networks to event message collectors - also known as syslog servers.

But it have some drawback. It doesn't protect log data while transferring and at end-point. It will not accept more than 1k size of message. Message is not in proper format.

F. K. Kent and M. Souppaya.Guide to Computer Security Log Management, nist publication 800-92. This publication seeks to assist organizations in understanding the need for sound computer security log management. It provides practical, real-world guidance on developing, implementing, and maintaining effective log management practices throughout an enterprise. The guidance in this publication covers several topics, including establishing log management infrastructures, and developing and performing robust log management processes throughout an organization. The publication presents log management technologies from a high-level viewpoint, and it is not a step-by-step guide to implementing or using log management technologies.

III. PROPOSED SYSTEM

In this paper we concentrated issues and problem regarding insecurity is solved using 2-round search encryption scheme by utilizing newer cryptography and information retrieval community solutions, we devise and use vector space model for mapping relevancy. For encryption a homographic 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. In the proposed scheme, the majority of computing work is done on the cloud while the user takes part in ranking, which guarantee multikeyword security multikeyword retrieval over encrypted cloud data with high security and practical efficiency. Propose the concepts of similarity relevance and scheme robustness. To Perform the first attempt to formulate the privacy issue in searchable encryption, and we show server-side ranking based on order preserving encryption (OPE) inevitably. 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. Experiment and testing sequels will be performed to compare and improve our method of solution proposal.

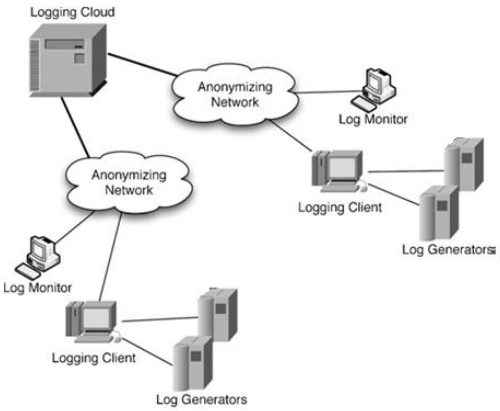

Fig1. System Architecture

The fig1 system architecture shows the how the system is designed it contains Log Generators: These are the computing devices that generate log data. In this system we are having two log generator. The function of log generator is to generate the log in batches and at particular time interval it has to send it to respective Log Client ,Logging Client or Logging Relay: The logging client is a collector that receives groups of log records generated by one or more log generators, and prepares the log data so that it can be pushed to the cloud for long term storage. We use the terms logging client and logging relay interchangeably. The logging client or relay can be implemented as a group of collaborating hosts. For simplicity this system has single logging client. Logging Cloud: The logging cloud provides long term storage and maintenance service to log data received from different logging clients belonging to different organizations, The logging cloud is maintained by a cloud service provider Log Monitor: These are hosts that are used to monitor and review log data. They can generate queries to retrieve log data from the cloud. Based on the log data retrieved, these monitors will perform further analysis as needed. They can also ask the log cloud to delete log data permanently, or rotate logs. Since the logging client and log monitor operate independent of each other, they can communicate only in an asynchronous manner. This means that if a logging client wants to send some data to the log monitor (or vice versa), the sender cannot expect the receiver to be online to receive the data. As a result the sender has to publish the data in some location and the receiver needs to retrieve the data from there when needed a use case diagram at its simplest is a representation of a user's interaction with the system and depicting the specifications of a use case. A use case diagram can portray the different types of users of a system and the various ways that they interact with the system. This type of diagram is typically used in conjunction with the textual use and will often be accompanied by other types of diagrams as well.

IV. CONCLUSION

In today's scenario internet is one of the fastest growing sources of information and huge amount of data resides on it. In this project we identified validating the data being added to the internet as a major problem. Data can be validated either syntactically or semantically. Many syntactic validators are available today and they are very efficient. Only few syntactic validators are available today and they are highly inefficient. Most of these syntactic validators use tableaux reasoning algorithm for validation. This algorithm will compute the deductive closure of all the given facts or axioms. Major draw backs of this method are:
It is highly non-deterministic since there are a great number of different possibilities of construction. Model developed by tableaux reasons can be extremely large, even for relatively small ontologies. We considered these drawbacks of tableaux reasons and the inefficiency of syntactic data validators in this project and we developed an ontological tool that is domain independent and will perform the specific task of data validation syntactically. This tool will not compute the deductive closure of the facts; instead the tool will only compute a set of inferences on the given facts in the ontology. These computed inferences are stored in the database for further use. Whenever needed some of the inferences are used to validate the data instantly as soon as the user enters the data. Finally we can conclude that with this approach we were able to address both of the drawbacks of the tableaux algorithm by eliminating the calculation of deductive closure of the axioms. The model thus developed is highly efficient even for large ontologies. Performance of this tool is significantly high when compared to the performance of the tableaux reasons.

V FEAUTER WORK

Currently this is implemented based on one organization in future we can scale to multiple organizations and we can provide log monitor and log Client as SaaS. Current

implementation of the logging client is loosely coupled with the operating system based logging. In the future, we plan to refine the log client implementation so that it is tightly integrated with the OS to replace current log process. In addition, to address privacy concerns current implementation allows access to log records that are indirectly identified by upload-tag values. We plan to investigate practical homomorphic encryption schemes that will allow encryption of log records in such a way that the logging cloud can execute some queries on the encrypted logs without breaching confidentiality or privacy. This will greatly reduce the communication overhead between a log monitor and the logging cloud needed to answer queries on logs.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. 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] 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.

[11] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

[12] D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, "Public-Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (Eurocrypt), 2004.

[13] A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A.L. Varna, S. He, M. Wu, and D.W. Oard, "Confidentiality-Preserving Rank-Ordered Search," Proc. Workshop Storage Security and Survivability, 2007.

[14] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multikeyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, 2011.

[15] H. Hu, J. Xu, C. Ren, and B. Choi, "Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism," Proc. IEEE 27th Int'l Conf. Data Eng. (ICDE), 2011.