ISSN: 2321-7782 (Online)
Special Issue, December 2013
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

A Survey on Cloud Data Security

C. Rakini[1]
PG Student
Department of CSE
Velammal Engineering College
Anna University
India

M.S. Murali Dhar[2]
Faculty
Department of CSE
Velammal Engineering College
Anna University
India

Dr. R. Manimegalai[3]
Professor
Department of CSE
Velammal Engineering College
Anna University
India

Abstract: ***The Cloud Computing concept provides dynamically scalable resources provisioned as a service over the cyberspace. It transfers storage as well as service for demand users. Large amount of data can store in the cloud. Cloud provider encrypts the confidential data and stores it in the cloud so that only the authenticated users can access the data. Thus the keyword privacy is maintained. Present traditional searchable encryption techniques are using keyword search to find out the results with encrypted format. It can support Boolean search and show the results as huge amount of files. These search results are not utilized efficiently in user's side. Searching is very difficult in encrypted data. To overcome this problem secure ranked keyword search over encrypted cloud data is proposed. Ranked search greatly increases system usability by altering search result relevance ranking instead of sending irrelevant results, and further ensures the file retrieval accuracy. We used order preserving symmetric encryption to protect the cloud data Even though there are lots of searching techniques available, they are not providing efficient search results. In this paper we focus on different searching techniques and at the end a better solution is identified.***

Keywords: Cloud computing, Ranked Keyword Search, Searchable Encryption, Boolean Search, Order Preserving Symmetri Encryption, Confidential Data.

I. INTRODUCTION

Cloud computing is the collection of virtualized and scalable resources, capable of hosting application and providing required services to the users with the "pay only for use" strategy where the users pay only for the number of service units they consume. It allows consumers and businesses to use applications without installation and access their personal files at any computer with internet Access. Cloud Computing has attracted the giant companies like Google, Microsoft, and Amazon and considered as a great influence in today's Information Technology industry [1, 2].

II. DEVELOPMENT MODELS OF CLOUD

- *Public cloud*

This infrastructure is available for public use alternatively for a large industry entity and is closely-held by an organization selling cloud services.

- *Private cloud*

The private cloud infrastructure is operated for the exclusive use of an organization. The cloud may be managed by that organization or a third party. Private clouds may be either on- or off premises.

- *Hybrid cloud*

A hybrid cloud combines multiple clouds (private, community of public) where those clouds retain their incomparable identities, but are limit together as a unit. A hybrid cloud may pass standardized or ownership access to data and applications, as well as application portability.

- *Community cloud*

A community cloud is one where the cloud has been organized to serve a common function or purpose.

III. TYPES OF SERVICES

Cloud computing can be defined as the supply of computing services via the Internet such as Applications(SaaS), Platforms, Infrastructure (IaaS), Process orchestration and integration.

- Software as a Service (SaaS)

Software‘s are provided as a service to the consumers according to their requirement, change consumers to use the services that are hosted on the cloud server.

- Platform as a Service (PaaS)

Clients are provided platforms access, which capable them to put their own customized software‘s and other applications on the clouds.

- Infrastructure as a Service (IaaS)

Pricing, storage, broad network access, and other basic computing resources are granted, enables consumers to manage the operating systems, applications, storage, and network connectivity.

IV. LITERATURE REVIEW

In Cloud Computing, to protect data privacy and combat unasked accesses, sensitive data has to be encrypted before outsourcing [4] so as to give end-to-end data confidentiality assurance in the cloud. However, data encryption get effective data utilization a very challenging task given that there could be a large amount of outsourced data files. Besides, in Cloud owners may share their outsourced data with a large number of users, who might want to only retrieve definite specific data files they are interested in during a given period. One of the most familiar ways to do so is through keyword-based search.

Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search method. This existing searchable scheme will support only Boolean keyword search, which will combine words and phrases using the words AND, OR, NOT operators.[3].So we survey on different searching techniques and identify the better solution for this problem.

SEARCHING TECHNIQUES

a) Searchable Encryption

Ning Caoy et al. [4] has proposed method which allow users to securely search complete encrypted data through keywords, these method support only Boolean search, without capturing any relevant data. This approach suffers from two main drawbacks when directly applied in the context of Cloud Computing. First one, users, who do not necessarily have pre-knowledge of the encrypted cloud data, have to post process every got file in order to find ones most matching their interest;

another drawback, invariably getting all files containing the queried keyword further incurs unnecessary network traffic, when retrieve more than one files.

b) Single Keyword Searchable Encryption

NingCaoy et al. [4] presented a Traditional single keyword searchable encryption schemes usually build an encrypted searchable index such that its content is hidden to the server unless it is given appropriate trapdoors generated via secret key(s). Our early work solves secure ranked keyword search which utilizes keyword frequency to rank results instead of returning undifferentiated results. However, it only supports single keyword search. Where anyone with public key can write to the data stored on server but only authorized users with private key can search. Public key solutions are usually very computationally expensive however.

c) Ranked Keyword Search

Cong Wang et al [5] discuss The major disadvantage of above mentioned techniques gets the better of in ranked keyword search. This system enables data users to find the most related information rapidly, rather than burdensome sorting through every match in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data. For privacy protection, such ranking function, however, should not leak any keyword relevant information. Another One, to improve search result accuracy as well as enhance user searching experience, it is also essential for such ranking system to support multiple keywords search.

d) Boolean Keyword Searchable Encryption

Ning Caoy et al. [4] focused a Boolean Keyword Searchable Encryption given without capturing any relevance of the files in the search result. When directly applied in large co-operative data outsourcing cloud environment, they may affect from the following two main drawbacks. First drawback, for each search request, users does not have pre-knowledge of the encrypted cloud data and go through every retrieved file in order to find one of the most matching their interest; second drawback, invariably sending back all files solely based on presence or absence of the keyword further incurs large unnecessary network traffic. In short, lacking of effective mechanisms to ensure the file retrieval accuracy is a significant drawback of existing searchable encryption schemes in the context of Cloud Computing. so conjunctive keyword search over encrypted data have been proposed for this problem. Conjunctive keyword search returns “all-or-nothing”, which means it only returns those documents in which all the keywords specified by the search query appear.

e) Conjunction of keyword search

Deepa P et al [6] has discussed this technique; conjunction of keywords is implemented for searching. The conjunctive keyword search mechanism will retrieve most efficient and relevance of data files. The conjunctive keyword search automatically creates ranked results so that the searching is efficient and flexible. This technique uses the wildcard based method and gram based method for constructing fuzzy keyword sets and symbol based trie- traverse scheme for generating a multi way tree to store the fuzzy keyword sets generated. This reduces the storage overhead. The Edit distance concept used for quantifies the keyword similarity.

f) Keyword based search

Kiruthigapriya Sengoden et al. [3], proposed a keyword search technique which allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. This existing searchable scheme will support only Boolean keyword search, which will combine words and phrases using the words AND, OR, NOT operators. This leads to following drawbacks.

- Non relevant data search result
- Large unnecessary network traffic is occurred.

- Decrease the efficiency and File retrieval accuracy.

So this kind of plaintext search method fails for cloud data. In order to improve the efficiency of ranked keyword search we used concept based searching techniques for file retrieval in which search words are conceptually related to the topic.

g) Concept Based Searching

Kiruthigapriya Sengoden et al. [3], have proposed concept based searching techniques return a list of files that not only contain the exact search terms, but also search words are conceptually related to the topic, which provides a wider search scope capability. So the combination of both keyword searches along with concept search produce the relevant search result which greatly improve the efficiency of search.

h) Fuzzy Keyword Search

Jin Li et al[7] have proposed this method, It enhances system usability when searching input exactly matches. Keywords are measured using edit distance and fuzzy keyword sets are making. Straight forward and wild card based are the two approaches are dealt with edit distance .In straight forward approach edit distance are calculated where all the forms of keywords are to be listed .Based on this indexing is built .Trapdoor are shared between user and the owner While retrieving file user computes the trapdoor based on the request, server matches with index table and return all potential identifiers.

i) Searchable Symmetric Encryption

Reza Curtmola et al [8] presented a Searchable symmetric encryption (SSE) allows a party to outsource the storage of its data to another party (a server). SSE schemes enable users to securely retrieve the cipher text, but these method support only Boolean keyword search, i.e., whether a keyword subsists in a file or not, without regarding the difference of relevance with the queried keyword of these files in the result. To improve security without sacrificing efficiency, schemes presented in [9], [10], [24] show that they support top-k single keyword retrieval under various scenarios in a secret manner, while maintaining the ability to selectively search complete it. so We introduce new adversarial models are non-adaptive and adaptive.

Non adaptive

It considers adversaries that make their search queries without taking into account the trapdoors and search outcomes of previous searches.

Adaptive

It chooses their queries as a function of previously obtained trapdoors and search outcomes. All previous work on SSE (with the exception of oblivious RAMs) falls within the non-adaptive setting.

j) Ranked Searchable Symmetric Encryption

P. Naresh K. et al [9] have focused RSSE.Ranked Searchable Symmetric Encryption framework is used to support rank search which built over the SSE cryptographic primitive. This technique uses KeyGen algorithm for generating public/private key pair and BuildIndex algorithm to generate index file containing keywords educed from file. The file collection and the index file are outsourced after encryption with frequency based relevance score. While the retrieval phase uses TrapdoorGen algorithm generates a trapdoor using the user's request. Upon the user request server runs SearchIndex algorithm which searches the files based on ids and relevance scores and sent files to the user. But RSSE has huge communication overhead when ranking is on user side and two round trip times is taken. Therefore efficient RSSE frame work uses Order Preserving Symmetric Encryption Scheme (OPSE) .It supports deterministic property in which a random coin generator and sampling function implemented. OPSE is used instead of encrypting scores in RSSE and in retrieval phase OPSE values are much more relevant. They provide better efficiency while retrieving files with top-k retrieval.

DATA PROTECTION

a) ORDER-PRESERVING SYMMETRIC ENCRYPTION

Alexandra Boldyreva et al [10] Order-preserving symmetric encryption (OPE) is a deterministic encryption scheme (aka. cipher) whose encryption function preserves numerical ordering of the plaintexts. OPE has a long account in the form of one-part codes, which are lists of plaintexts and the matching cipher texts, both ordered in alphabetical or numerical order so only a single copy is required for efficient encryption and decryption[3].

ANALYSIS OF DIFFERENT SEARCHING TECHNIQUES IN CLOUD SECURITY

EXISTING METHODS	DRAW BACK	PROPOSED METHODS	ADVANTAGE
Searchable encryption	Un necessary network traffic can be occurred	Ranked keyword search	Eliminate unnecessary network traffic. Find most relevant information quickly.
Single keyword search	does not support multiple keyword search.	Ranked keyword search	Support multiple keyword search
Boolean keyword search	Provide the irrelevant information.	Conjunction keyword search	Retrieve the highly relevant search results. automatically generates ranked results.
Keyword based search	Non relevant data search result. Decrease the efficiency and File retrieval accuracy.	Concept based search	produce the relevant search result which greatly improve the efficiency of search.
Fuzzy keyword search	Large storage requirements. Lack of efficiency	Wild card based techniques	edit distance can be calculated using substitution, deletion and insertion .
Searchable symmetric encryption	Support Boolean keyword search. Does not provide the accurate documents.	Ranked search symmetric encryption	provide better efficiency while retrieving files with top-k retrieval.

V. CONCLUSION

In the beginning, cloud computing basis and its storage infrastructure is explained. After that different searching techniques in the cloud data are discussed. Each has its own advantages and disadvantages. In future MRSE (**M**ulti keyword **R**anked **S**earch over **E**ncrypted Cloud data) Scheme, this fulfils the secure multi keyword top-k retrieval over encrypted Cloud data. Specifically, for the first time, employ relevance score to support multi keyword top-k retrieval. Novel technologies in the cryptography community and information retrieval (IR) community are employed, including HomoMorphic encryption and vector space model. In the proposed scheme, the majority of computing work is done on the Cloud while the user takes part in ranking, which guarantees top-k multi keyword retrieval over encrypted Cloud data with high security and practical efficiency.

References

1. L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A breakin the clouds: towards a cloud definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50–55, 2009.
2. S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS, Januar2010, LNCS. Springer, Heidelberg.
3. Kiruthigapriya Sengoden, Swaraj Paul "Improving the Efficiency of Ranked keyword Search over Cloud Data" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 3, March 2013
4. Ning Caoy, Cong Wangz, Ming Liy, Kui Renz, and Wenjing Louy" Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data" INFOCOM, 2011 Proceedings IEEE.
5. Cong Wang†, Ning Cao‡, Jin Li†, Kui Ren†, and Wenjing Lou"Secure Ranked Keyword Search over Encrypted Cloud Data" Distributed Computing System,2010 IEE 30th international conference.
6. Deepa P L, S Vinoth Kumar, Dr S Karthik" searching techniques in encrypted Cloud data" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 8, October 2012
7. Jin Li†, Qian Wang† , Cong Wang†, Ning Cao‡ , Kui Ren† , and Wenjing Lou‡ "Fuzzy Keyword Search over Encrypted Data in Cloud Computing" INFOCOM,2010 Proceedings IEEE.
8. Reza Curtmola† Juan Garay‡ Seny Kamara§ Rafail Ostrovsky" Searchable Symmetric Encryption:Improved Definitions and Efficient Constructions" CCS '06 Proceedings of the 13th ACM conference on Computer and communications security.
9. P. Naresh K. Pavan kumar D. K. Shareef " Implementation Of Secure Ranked Keyword Search By Using RSSE" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 2 Issue 3, March – 2013.
10. Alexandra Boldyreva_ Nathan Chenette_ Adam O'Neilly" Order-Preserving Encryption Revisited: Improved Security Analysis and Alternative Solutions" CRYPTO'11 Proceedings of the 31st annual onference on Advances in cryptology.

AUTHOR(S) PROFILE

Rakini.C, is currently pursuing her ME in Computer Science Engineering at Velammal Engineering College which is affiliated to the Anna University of Chennai, Tamil Nadu, India.

Murali Dhar M S, received his Master of Engineering degree in Computer Science and Engineering from Velammal Engineering College, Chennai in 2009. He is pursuing his Ph.D. degree in the research area of Cloud Computing and infrastructure from Anna University, Chennai. Presently he is working as Assistant Professor in the Department of Computer Science and Engineering at Velammal Engineering College, Chennai. His research interests are in various applied/systems topics including Cloud Computing, Distributed Systems, Operating Systems, and Network Security/Resilience.

Dr. Manimegalai, has seventeen years of experience in teaching, research and industry put-together. She has worked as software engineer in DCM Technologies, New Delhi and Xilinx India Technology Services, Hyderabad. Currently she is with Park College of Engineering and Technology as Professor in the Department of Computer Science and Engineering. She is life member in Computer Society of India, Institution of Engineers (India) and Indian Society for Technical Education. She is also a member of IEEE and VLSI society of India. Her areas of interests include Algorithms for VLSI/FPGA Design, Reconfigurable Computing, Natural Language Processing, Cloud Computing and Computer Networks.

