International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Survey on Privacy Preservation for Multi-Keyword Search on Data Network

Nilima S. Kasar[1], Vitthal S. Phad[2]

[1]Department of Computer Engineering, Smt. Kashibai Navale College of Engineering (SKNCOE), Savitribai Phule Pune University, Pune, India

[2]Assistant Professor, Department of Computer Engineering, Smt. Kashibai Navale College of Engineering (SKNCOE), Savitribai Phule Pune University, Pune, India

Abstract: ***In recent years, growth of private and semi-private information has grown rapidly on information network so, mechanisms to search such information have failed in privacy preservation. In today's emerging information networks, it is essential to find a solution for accessing the data from large number of providers while maintaining a privacy . So, there is a strong need to develop a technique which provides user required information in ranked order while preserving the users privacy. This paper analyzes different methods concerning about privacy of data in the public and private networks and also, analyzes a system based on public cloud platform where data sharing is takes place between public and private systems on a network through authentication and encryption . This system works on local private server which controls multiple systems. This paper is motivated by the lack of tools for an access controlled multi-keyword search of data on the public and private network.***

Keywords: Encryption, Decryption, Ranking algorithm

1. Introduction

In this age of cloud computing, data users, while enjoying a lots of benefits from the cloud like cost effectiveness and data availability etc. are simultaneously disincline or even cause resilience to use the clouds because they lose data control. The recent research and industrial efforts towards giving back the data control to cloud users have given birth to a variety of multi-domain cloud platforms, most eminent information networks. In such an information network, a data owner can continue to have the full control of their data by using trusted service providers or using a personal server which is administrated by themselves. The main aspect of such information network is that it does not need mutual trusts between servers, that is, an owner only needs to trust his/her personal server and nothing more. In this paper it analyzes the matter of providing an economical search mechanism that respects privacy concerns of the taking part content suppliers. Here, one solution is to create a centralized index of content that works in conjunction with an access management implementing search protocol across networked suppliers. The centralized index itself provides study and quantitative privacy guarantees that hold albeit the complete index is created public. The degree of privacy provided by the index will to be tuned to suit the wants of the suppliers, and overhead incurred by the search protocol is proportional to the degree of privacy provided. This technique of providing economical search over access-controlled content preserves the vital attractiveness of private data sharing — every supplier has complete management over the knowledge it shares: what quantity is shared, once it's shared, and with whom it's shared. In information networks, private personal records of an individual owners are stored on autonomous service providers system ,they stored the records on behalf of these individual owners by imposing strict access control rules for information sharing. Such information networks have the following salient features: 1) Providers, do not mutually trust each other which are in different administrative domain; 2) The responsibility of providers is to protect owners' privacy.

2. Motivation

Nowadays, as there is a large collection of data and creation of large distributed repositories, search over this data while respecting access control is critical. Other problem here is that of to ensure privacy of the content owners by maintaining an efficient index of distributed content. For avoiding privacy leakages of access-control through the index requires the index hosting site must be fully trusted and secured by all participating content providers, but nowadays this is almost impractical because large competing organizations or individuals wish to share theirs contents. Worse case is that when index host is compromised by hackers this could lead to a complete and devastating privacy loss. So, an efficient privacy-preserving search therefore requires an index structure that prevents data privacy breaches even if the index is made public. A process of a PPI construction must addresses the correctness of the resulting structure and also resist the privacy violations during this process. To gain a relevant information is also required while preserving a privacy.

Hence there is a need to preserve privacy of data in information networks.

3. Literature Survey

LIGHT — Light weight Hash Tree (LIGHT) [2] a query-efficient yet low-maintenance indexing scheme. LIGHT employs a novel naming mechanism and a tree summarization strategy for graceful distribution of its index structure. It shows through analysis that it can support

various complex queries with near-optimal performance. Extensive experimental results also demonstrate that, compared with state of the art over-DHT indexing schemes, LIGHT saves 50-75 percent of index maintenance cost and substantially improves query performance in terms of both response time and bandwidth consumption. In addition, LIGHT is designed over generic DHTs and hence can be easily implemented and deployed in any DHT-based P2P system.

SS-PPI — This SS-PPI [3] is novel privacy preserving index deliberation, which, in conjunction of circulated access control-implemented search protocols which, gives hypothetically ensured protection of content security. Contrasted with existing system this proposal highlights with a progression of distinct components: (an) it joins access control arrangements in the privacy preserving file, which moves forward both search effectiveness and attack resilience; (b) it utilizes a quick file development convention by means of a novel utilization of the secret sharing plan in a completely conveyed way (without trusted outsider), requiring just consistent (normally two) round of correspondence; (c) it gives data theoretic security against colluding adversaries during record development and inquiry replying.

PPI —PPI is a mapping function built on the set of documents D which are shared among the set of providers p1, p2, . . . , pn. It accepts a query Q and returns a subset of providers M that may contain matching documents. PPI must act like a traditional index: over time the record must return results for indistinguishable queries unless index contents itself has changed. Additionally for any question q whose outcomes are a subset of another query q, the outcome set returned for q must be a subset of that returned for q .These behavioral prerequisites prevents attacks that endeavor security breaks by filtering through of false positives. The PPI [4] must be executed with consideration: a naive implementation could without much of a stretch yield more data than is given by the PPI definition. For instance, the indexing host power total all share content locally and preprocess it to appear a list with genuine positives alone; the false positives as required by the definition being embedded into results at inquiry time. Notice that for this situation the appeared index itself does not relate to PPI definitions. A public disclosure of the emerged record would bring about Provable Exposure of component suppliers. Rather, author require that an emerged file ought not yield any more data than that got from executing a comprehensive rundown of inquiries against the PPI.

Fairplay — Fairplay [5] is an undeniable framework that executes an non specific secure function evaluation (SFE). Fairplay includes a high level procedural definition dialect called SFDL customized to the SFE paradigm; a compiler of SFDL into an one-pass Boolean circuit displayed in a dialect called SHDL and Bob/Alice programs that evaluate the SHDL. This system enables us to present the first evaluation of an overall SFE in real settings, as well as examining its components and identifying potential bottlenecks. It provides a test-bed of ideas and enhancements concerning SFE, whether by replacing parts of it, or by integrating with it.

MPC[5] — Multi Party Computation also known as secure computation or Secure multi-party computation. It is a subbranch of cryptography with the objective of creating methods for parties to jointly compute a function over their inputs but keeping those inputs as a private. In an MPC, a given number of participants, p1, p2,... pN, each have private data, respectively d1, d2... dN. Participants want to compute the value of a public function on that private data: F (d1, d2,...,dN) while keeping their own inputs private.

TF-IDF — This stands for term frequency–inverse document frequency. It is a numerical statistic that is intended to reflect how important a word is to a document in a collection. It is often used as a weighting factor in information retrieval and text mining. Variations of the tf–idf weighting scheme are often used by search engines as a central tool in scoring and ranking a document's relevance given a user query. Tf-idf can be successfully used for stop-words filtering in various subject fields including text summarization and classification.

Table 1: Survey Table

Sr.no	Paper	Technique and method used	Advantages	Limitations
1	A Adaptive Key Recovery Attacks on NTRU-based Somewhat Homomorphic Encryption Schemes[11]	This also implemented the homomorphic evaluation of AES, showing that it offers advantages against the BGV scheme	It is obvious that a homomorphic encryption scheme cannot have security of ciphertexts under adaptive attacks.	It shows that building CCA1-secure homomorphic schemes is not trivial.
2	Efficient Privacy Preserving Secure ODARM Algorithm in Horizontally Distributed Database[12]	It makes use of a quality function that avoids patterns uncorrelated with the target.	Direct application of sequential algorithms to distributed databases is not effective; it requires a large amount of communication overhead.	The Horizontal distributed database is more secure than Vertical distributed database.
3	Towards Collaborative Search in Digital Libraries Using Peer-to-Peer Technology[13]	These prototype implementations allows for the easy exchange of strategies for query routing (i.e., selecting the peers to which the query is sent) as well as for merging the results returned by different peers.	It only supports single-term exact-match queries and does not support any form of ranking.	The goal of finding high-quality search results with respect to precision and recall cannot easily be reconciled with the design goal of unlimited scalability

4.	Content-Based Retrieval in Hybrid Peer-to-Peer Networks[14]	Each leaf node is modeled as a digital library running an effective content-based text retrieval algorithm	Name based retrieval is not sufficient when digital libraries are large or when file-naming conventions are uncertain.	These results demonstrate that content-based retrieval is more accurate
5.	Privacy-Preserving in Outsourced Transaction Databases from Association Rules Mining[15]	It has considered that the attacker knows the domain of items and their exact frequency and can use this knowledge to identify cipher items and cipher item sets.	The particular drawback attacked in this paper is outsourcing of pattern mining inside company privacy.	An encryption scheme, called Rob Frugal, is proposed that is based on 1–1 substitution ciphers for items and adding fake transactions.

4. Proposed Work

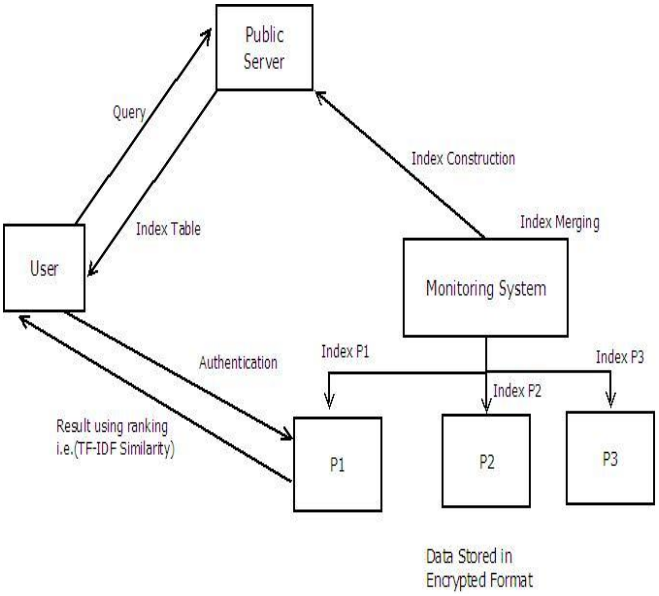

Figure 1: Architecture of proposed system

In our system we are independent on trust-based systems. Here our output for multi-keyword query search is ranked data while preserving the privacy of data owners.

First, the privacy preservation of private data which is present on private servers is done by encrypting this data. Secondly, the indexing is done on private data on private servers then the resulting output index table is provided to public server after this, the user send a query to public server and get a required ranked result.

Our contribution lies in encrypting the private data (refer fig 4.1) and giving required data in ranking order to the user.

5. Conclusion

This paper presented an all-inclusive survey of methods which are concerning about data privacy. The main features, the advantages and disadvantages of each privacy preserving algorithms and access control methods are described. As per survey, a strong need to develop privacy-preserving data access methods for public and private networks because user privacy is challenging issues nowadays. Paper proposed a system through which user can get an access to required data in ranked order using PPI and encryption techniques.

References

[1] Yuzhe Tang and Ling Liu , Fellow , ―Privacy Preserving Multi-Keyword Search in Information Network", IEEE transactions on knowledge and data engineering ,vol. 27, no. 9, Sept 2015.

[2] Yuzhe Tang, Shuigeng Zhou, ―LIGHT: A Query-Efficient Yet Low-Maintenance Indexing Scheme over DHTs", School of Computer Science, Fudan University and Shanghai Key Lab of Intelligent Information Processing, Shanghai, China, 2010.

[3] Y. Tang, T. Wang, and L. Liu, ―Privacy preserving indexing for ehealth information networks," in Proc. 20th ACM Int. Conf. Inf. Knowl. Manage., 2011, pp. 905–914.

[4] Mayank Bawa, Roberto J. Bayardo Jr • Rakesh Agrawal, Jaideep Vaidya, ―Privacy-preserving indexing of documents on the network", The VLDB Journal DOI 10.1007/s00778-008-0129-7.

[5] A. Ben-David, N. Nisan, and B. Pinkas, ―Fairplaymp: A system for secure multi-party computation," in Proc. ACM Conf Comput.Commun. Security, 2008, pp. 257–266.

[6] Zhun-Ga Liu, Quan Pan, ―A New Incomplete Pattern Classification Method Based on Evidential Reasoning", School of Automation, Northwestern Polytechnical University, Xi'an 710072, China, 2015.

[7] Preethi Mathew, Dr. S. Sasidhar Babu, ―Secure Fuzzy Multi-Keyword Ranked Search over Encrypted Cloud Data", M. Tech Student, Dept. of CSE, SNGCE, Kolenchery, Kerala, India, 2015.

[8] Y. Tang, L. Liu, A. Iyengar, K. Lee, and Q. Zhang, ―e-PPI: Locator service in information networks with personalized privacy preservation," in Proc. IEEE 34th Int. Conf. Distrib. Comput. Syst., Madrid, Spain, Jun. 30–Jul. 3, 2014, pp. 186–197.

[9] Randy Baden, Adam Bender, ―Persona: an online social network with user-defined privacy", University of Maryland, College Park, MD, USA, 2009.

[10] K.S.Sureh, Mrs. SaritaChowdary, T. Balachary,―A Cloud Based System for Patient Health Records Using Symmetric Encryption", Dept of CSE, MLRIT, Dundigal, Hyderabad, India, 2013.

[11] Ricardo Dahab, Steven Galbraith, and Eduardo Morais, ―Adaptive Key Recovery Attacks on NTRU-based Somewhat Homomorphic Encryption Schemes", 1 Institute of Computing, University of Campinas, Brazil, 2013.

[12] Priyanka.G , Premkumar. M, ―Efficient Privacy Preserving Secure ODARM Algorithm in Horizontally Distributed Database", Department of Computer

Science and Engineering, Sri Shanmugha College of Engineering and Technology, Sankari, India, 2015.
[13] Matthias Bender, Sebastian Michel, –Towards Collaborative Search in Digital Libraries Using Peer-to-Peer Technology", Max-Planck-Institute fur Informatik, 66123 Saarbr¨ucken, Germany, 2004
[14] Jie Lu, Jamie Callan, –Content-Based Retrieval in Hybrid Peer-to-Peer Networks", School of Computer Science Carnegie Mellon University Pittsburgh, PA, 2003.
[15] Ms. Deokate Pallavi B., Prof. M.M. Waghmare, –Privacy-Preserving in Outsourced Transaction Databases from Association Rules Mining", Student of ME (Information Technology), DGOFFE, University of Pune, Pune, 2014.

Author Profile

Nilima Subhash Kasar Research Scholar at Smt. Kashibai Navale College of Engineering (SKNCOE), Savitribai Phule Pune University. She has received B.E. in Information Technology Engineering from Pune University, Pune. Currently she is pursuing M.E. in Computer Engineering from Smt. Kashibai Navale College of Engineering (SKNCOE), Savitribai Phule Pune University, Pune, India. Her area of interest are Information Security and Cloud Computing.

Prof. Vitthal S. Phad received the B.E. degree in Information Technology Engineering from SGGSIET Nanded in 2006 and M.E. Degree in Computer Engineering from SRTMU Nanded in 2011. His area of research are Image processing ,Information Security and Network Security. He is working as Assistant Professor in Department of Computer Engineering, SKNCOE Pune, India.