A Literature Review on Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data

Sakshi Shrivastava[1] & Prof. Pankaj Singh[2]
[1]M-Tech Research Scholars, [2]Research Guide, Deptt. of Computer Science Engineering
SAM College of Engineering and Technology, Bhopal

Abstract: Information Technology has always been considered a major pain point of enterprise organizations, from the perspectives of both cost and management. The information technology industry has experienced a dramatic shift in the past decade factors such as hardware commoditization, open-source software, virtualization, workforce globalization, and agile IT processes have supported the development of new technology and business models. Cloud computing now offers organizations more choices regarding how to run infrastructures, save costs, and delegate liabilities to third-party providers. It has become an integral part of technology and business models, and has forced businesses to adapt to new technology strategies. Accordingly, the demand for cloud computing has forced the development of new market offerings, representing various cloud service and delivery models. These models significantly expand the range of available options, and task organizations with dilemmas over which cloud computing model to employ.

Keywords - Privacy, Search, Keywords, Encryption, Cloud Data.

I. INTRODUCTION

Cloud computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources [1]. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. To protect data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, e.g., emails, personal health records, photo albums, tax documents, financial transactions, etc., may have to be encrypted by data owners before outsourcing to the commercial public cloud [2]; this, however, obsoletes the traditional data utilization service based on plaintext keyword search. The trivial solution of downloading all the data and decrypting locally is clearly impractical, due to the huge amount of bandwidth cost in cloud scale systems. Moreover, aside from eliminating the local storage management, storing data into the cloud serves no purpose unless they can be easily searched and utilized. Thus, exploring privacy-preserving and effective search service over encrypted cloud data is of paramount importance. Considering the potentially large number of on-demand data users and huge amount of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability and scalability.

The idea of providing a centralized computing service dates back to the 1960s, when computing services were provided over a network using mainframe time-sharing technology. In 1966, Canadian engineer Douglass Parkhill published his book The Challenge of the Computer Utility, in which he describes the idea of computing as a public utility with a centralized computing facility to which many remote users connect over networks.

Definition

There are countless definitions and interpretations of cloud computing to be found from multiple sources. The term “cloud computing” itself likely comes from network diagrams in which cloud shape are used to describe certain types of networks, either the Internet or internal networks. Some sources refer to cloud computing as a set of applications delivered as services combined with the data enter hardware and software that enables the applications. Others say that cloud computing is a business model rather than a specific technology or service.

Cloud computing consists of both technological and business components. Certain cloud-enabling technologies significantly helped to form the cloud, and it is unlikely that cloud computing could have existed without them. These more closely in the next, but it is worth mentioning that cloud-enablers such as open-source software, virtualization, distributed storage, distributed databases, and monitoring systems are the cornerstones of cloud infrastructure.

Cloud computing assumes that every software application or system component becomes a service or part of a service. Therefore, the architecture of new or existing systems might have to be changed to become cloud compatible. As such, in order to realize the value of the cloud and enable it for an organization, businesses must typically make major structural adjustments to internal IT organizations and evangelize cloud philosophy to employees. Depending on the type of cloud used by an

organization, this may also create competition within the company. It is typical that people resist change, so cloud evangelists often face resistance within their organizations.

Definition of Terms

The following sub encompasses a definition of key concepts as a basis of this research. All concepts are explained later in detail with their specific sources, so that in this study the references do not occur.

Cloud Computing: Form of cost-efficient and flexible usage of IT services. The services are offered just-in-time over the internet and are paid per usage.

Clusters: Locally distributed units with the same kind of hardware and operating systems being capable of processing a large amount of data collaboratively.

Grids: Globally distributed units with different operating systems and hardware being capable of processing a large amount of data collaboratively.

Hybrid Cloud: A mixture of a private and public cloud.

Infrastructure as a Service: Users being able to use servers, storage, network settings on-demand from other providers on a pay-per-use basis.

Platform as a Service: Developers being able to build their own applications offered on development platforms that are maintained and secured by other providers.

Private Cloud: Clouds that are used in a private network providing more security.

Public Cloud: Clouds that can publicly run anywhere in the world.

Scalability: Refers to the performance of handling growing amounts of work.

Software as a Service: Users can utilize software being offered over the internet without worrying about its maintenance, back-ups or security.

Supercomputers: Machines assembled with a lot of processors that are merged into 1 machine with high performance capabilities.

Traditional IT Outsourcing: Ordinary way of a company to choose an external tender to take care of their IT resources with physical assured locations.

Utility Computing: The very idea of computing resources being offered as a service.

Virtualization: With virtualization servers are utilized more efficiently enabling one server to be used by several customers.

Clouds

Together with virtualization, clouds can be defined as computers that are networked anywhere in the world with the availability of paying the used clouds in a pay-per-use way, meaning that just the resources that are being used will be paid. In the following the types of clouds will be introduced.

Public Clouds

A public cloud encompasses the traditional concept of cloud computing, having the opportunity to use computing resources from anywhere in the world. The clouds can be used in a so-called pay-per-use manner, meaning that just the resources that are being used will be paid by transaction fees.

Private Clouds

Private clouds are normally data enters that are used in a private network and can therefore restrict the unwanted public to access the data that is used by the company. It is obvious that this way has a more secure background than the traditional public clouds. However, managers still have to worry about the purchase, building and maintenance of the system.

Hybrid Clouds

As the name already reveals, a hybrid cloud is a mixture of both a private and public cloud. This can involve work load being processed by an enterprise data centre while other activities are provided by the public cloud.

Below an overview of all three clouds computing types is illustrated.

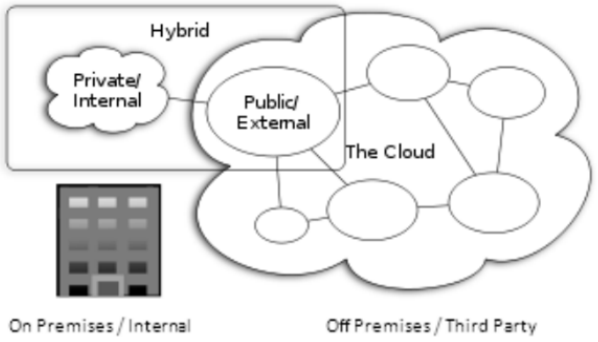

Fig. 1: Cloud Computing Types

The answer is to present cryptographic schemes that enable searching on encrypted data without leaking any information to the untrusted server.

- These techniques are provably secure. The techniques provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext given only the cipher text. The techniques provide controlled searching, so that the untrusted server cannot search for a word without the user's authorization. The techniques support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The techniques also support query isolation, meaning that the untrusted server learns nothing more than the search result about the plaintext.
- These schemes are efficient and practical. The algorithms we present are simple and fast. More specifically, for a document of length n, the encryption and search algorithms only need O (n) number of stream cipher and block cipher operations. the schemes introduce essentially no space and communication overhead. They are also flexible and can be easily extended to support more advanced searches.

II. LETERATURE SURVEY

N. Cao, C. Wang, M. Li, K. Ren and W. Lou, [1] With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, for the first time, define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). Authors first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, authors further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.

N. Cao, C. Wang, M. Li, K. Ren and W. Lou, [2] With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, for the first time, author define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE). Authors establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, authors choose the efficient similarity measure of "coordinate matching", i.e., as many matches as possible, to capture the relevance of data documents to the search query. Authors further use "inner product similarity" to quantitatively evaluate such similarity measure. Authors first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.

Ning Cao, S. Yu, Zhenyu Yang, W. Lou and Y. T. Hou, [3] With the increasing adoption of cloud computing for data storage, assuring data service reliability, in terms of data correctness and availability, has been outstanding. While redundancy can be added into the data for reliability, the problem becomes challenging in the "pay-as-you-use" cloud paradigm where authors always want to efficiently resolve it for both corruption detection and data repair. Prior distributed storage systems based on erasure codes or network coding techniques have either high decoding computational cost for data users, or too much burden of data repair and being online for data owners. In this paper, the design a secure cloud storage service which addresses the reliability issue with near-optimal overall performance. By allowing a third party to perform the public integrity verification, data owners are significantly released from the onerous work of periodically checking data integrity. To completely free the data owner from the burden of being online after data outsourcing, this paper proposes an exact repair solution so that no metadata needs to be generated on

the fly for repaired data. The performance analysis and experimental results show that their designed service has comparable storage and communication cost, but much less computational cost during data retrieval than erasure codes-based storage solutions. It introduces less storage cost, much faster data retrieval, and comparable communication cost comparing to network coding-based distributed storage systems.

Dawn Xiaoding Song, D. Wagner and A. Perrig, [4] It is desirable to store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. But this usually implies that one has to sacrifice functionality for security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query, without loss of data confidentiality. Authors describe the cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. Author's techniques have a number of crucial advantages. They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext when only given the ciphertext; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the user's authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The algorithms presented are simple, fast (for a document of length n, the encryption and search algorithms only need O(n) stream cipher and block cipher operations), and introduce almost no space and communication overhead, and hence are practical to use today.

P. Xu, H. Jin, Q. Wu and W. Wang, [5] Public-key encryption with keyword search (PEKS) is a versatile tool. It allows a third party knowing the search trapdoor of a keyword to search encrypted documents containing that keyword without decrypting the documents or knowing the keyword. However, it is shown that the keyword will be compromised by a malicious third party under a keyword guess attack (KGA) if the keyword space is in a polynomial size. Authors address this problem with a keyword privacy enhanced variant of PEKS referred to as public-key encryption with fuzzy keyword search (PEFKS). In PEFKS, each keyword corresponds to an exact keyword search trapdoor and a fuzzy keyword search trapdoor. Two or more keywords share the same fuzzy keyword trapdoor. To search encrypted documents containing a specific keyword, only the fuzzy keyword search trapdoor is provided to the third party, i.e., the searcher. Thus, in PEFKS, a malicious searcher can no longer learn the exact keyword to be searched even if the keyword space is small. Authors propose a universal transformation which converts any anonymous identity-based encryption (IBE) scheme into a secure PEFKS scheme. Following the generic construction, author's instantiate the first PEFKS scheme proven to be secure under KGA in the case that the keyword space is in a polynomial size.

III. PROBLEM IDENTIFICATION

The problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, authors choose the efficient similarity measure of "coordinate matching,", as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use "inner product similarity" to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, they proposed a basic idea of MRSE using secure inner product computation. Then MRSE schemes to achieve various stringent privacy requirements in two different threat models.

IV. CONCLUSION

Remote and un-trusted storage systems allow clients with limited resources to store and distribute large amounts of data at low cost. In order to preserve confidentiality, the remotely-stored data must be encrypted prior to transmission. Unfortunately, encryption restricts a client's ability to selectively access segments of her data, especially when she wishes to only retrieve specific content. To address this dilemma, a number of techniques have been recently proposed for achieving a less stringent storage model, one based on the notion of secure, delegated, searchable encryption. Intuitively, in order to provide secure searchable encryption schemes, most of these approaches associate an index with each document that, when combined with a trapdoor for a keyword, returns information signifying the association of the keyword with the document. Informally.

REFRENCES

[1] N. Cao, C. Wang, M. Li, K. Ren and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," in IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222-233, Jan. 2014.

[2] N. Cao, C. Wang, M. Li, K. Ren and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," INFOCOM, 2011 Proceedings IEEE, Shanghai, 2011, pp. 829-837.

[3] Ning Cao, S. Yu, Zhenyu Yang, W. Lou and Y. T. Hou, "LT codes-based secure and reliable cloud storage service," INFOCOM, 2012 Proceedings IEEE, Orlando, FL, 2012, pp. 693-701.

[4] Dawn Xiaoding Song, D. Wagner and A. Perrig, "Practical techniques for searches on encrypted data," Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on, Berkeley, CA, 2000, pp. 44-55.

[5] P. Xu, H. Jin, Q. Wu and W. Wang, "Public-Key Encryption with Fuzzy Keyword Search: A Provably Secure Scheme under Keyword Guessing Attack," in IEEE Transactions on Computers, vol. 62, no. 11, pp. 2266-2277, Nov. 2013.

[6] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and Efficiently Searchable Encryption," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

[7] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Paillier, and H. Shi, "Searchable Encryption Revisited: Consistency Properties, Relation to Anonymous Ibe, and Extensions," J. Cryptology, vol. 21, no. 3, pp. 350- 391, 2008.

[8] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing," Proc. IEEE INFOCOM, Mar. 2010.

[9] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, "Public Key Encryption That Allows PIR Queries," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

[10] P. Golle, J. Staddon, and B. Waters, "Secure Conjunctive Keyword Search over Encrypted Data," Proc. Applied Cryptography and Network Security, pp. 31-45, 2004.

[11] L. Ballard, S. Kamara, and F. Monrose, "Achieving Efficient Conjunctive Keyword Searches over Encrypted Data," Proc. Seventh Int'l Conf. Information and Comm. Security (ICICS '05), 2005.

[12] D. Boneh and B. Waters, "Conjunctive, Subset, and Range Queries on Encrypted Data," Proc. Fourth Conf. Theory Cryptography (TCC), pp. 535-554, 2007.

[13] R. Brinkman, "Searching in Encrypted Data," PhD thesis, Univ. Of Twente, 2007.

[14] Y. Hwang and P. Lee, "Public Key Encryption with Conjunctive Keyword Search and Its Extension to a Multi-User System," Pairing, vol. 4575, pp. 2-22, 2007.

[15] J. Katz, A. Sahai, and B. Waters, "Predicate Encryption Supporting Disjunctions, Polynomial Equations, and Inner Products," Proc. 27th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2008.

[16] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption," Proc. 29th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT '10), 2010.

[17] E. Shen, E. Shi, and B. Waters, "Predicate Privacy in Encryption Systems," Proc. Sixth Theory of Cryptography Conf. Theory of Cryptography (TCC), 2009.

[18] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing," Proc. 31st Int'l Conf. Distributed Computing Systems (ICDCS '10), pp. 383- 392, June 2011.

[19] 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 '10), 2010.

[20] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 8, pp. 1467-1479, Aug. 2012.