International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

A Novel Architecture of multi-keyword ranked search over encrypted cloud data supporting synonym queries

[1]GovardhanPothu & [2]Mr Annapareddy V N Reddy*
[1]SV College of Engineering
[2]M.Tech, (Ph.D); KL University, SV College of Engineering

Abstract:
As an enhancement we enhance the existing system and in this paper we propose an effective approach to solve the problem of multi-keyword ranked search over encrypted cloud data supporting synonym queries. The main contribution of this paper is summarized in two aspects: multi-keyword ranked search to achieve more accurate search results and synonym-based search to support synonym queries. Meanwhile, existing search approaches over encrypted cloud data support only exact or fuzzy keyword search, but not semantics-based multi-keyword ranked search. Therefore, how to enable an effective searchable system with support of ranked search remains a very challenging problem.

Index Terms—Cloud computing; searchable encryption; privacy-preserving; keyword search; ranked search

1. 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 [2], [3]. 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, for example, e-mails, personal health records, photo albums, tax documents, financial transactions, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud [4]; this, however, obsoletes the traditional data utilizationservice 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. On the one hand, to meet the effective data retrieval need, the large amount of documents demand the cloud server to perform result relevance ranking, instead of returning undifferentiated results. Such ranked search system enables data users to find the most relevant information quickly, rather than burdensomely sorting through every match in the content collection [5]. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation, however, should not leak any keyword related information. On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse results. As a common practice indicated by today's web search engines (e.g., Google search), data users may tend to provide a set of keywords instead of only one as the indicator of

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

their search interest to retrieve the most relevant data. And each keyword in the search request is able to help narrow down the search result further. "Coordinate matching" [6], i.e., as many matches as possible, is an efficient similarity measure among such multi-keyword semantics to refine the result relevance, and has been widely used in the plaintext information retrieval (IR) community. However, how to apply it in the encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirements like the data privacy, the index privacy, the keyword privacy, and many others (see Section 3.2). In the literature, searchable encryption [7], [8], [9], [10], [11], [12], [13], [14], [15] is a helpful technique that treats encrypted data as documents and allows a user to securely search through a single keyword and retrieve documents of interest. However, direct application of these approaches to the secure large scale cloud data utilization system would not be necessarily suitable, as they are developed as crypto primitives and cannot accommodate such high service-level requirements like system usability, user searching experience, and easy information discovery. Although some recent designs have been proposed to support Boolean keyword search [16], [17], [18], [19], [20], [21], [22], [23], [24] as an attempt to enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality (see Section 7). Our early works [25], [26] have been aware of this problem, and provide solutions to the secure ranked search over encrypted data problem but only for queries consisting of a single keyword. How to design an efficient encrypted data search mechanism that supports multi-keyword semantics without privacy breaches still remains a challenge open problem

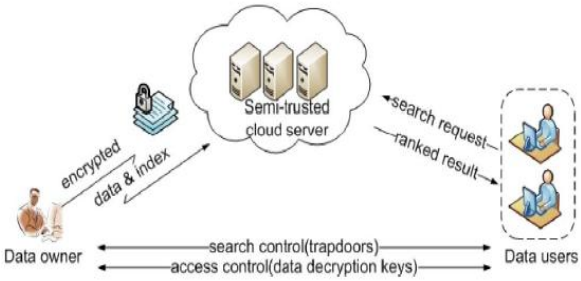

Fig. 1. Architecture of the search over encrypted cloud data.

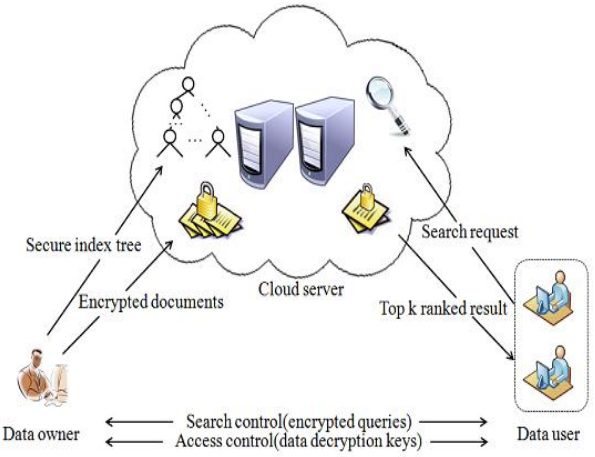

Fig. 2.Architecture of the search over encrypted cloud data.

1. Design Goals:

To enable ranked search for effective utilization of outsourced cloud data under the aforementioned model, our system design should simultaneously achieve security and performance guarantees as follows. . Multi-keyword ranked search. To design search schemes which allow multi-keyword query and provide result similarity ranking for effective data retrieval, instead of returning undifferentiated results. . Privacy-preserving. To prevent the cloud server from learning additional information from the data set and the index, and to meet privacy requirements specified in Section Efficiency. Above goals on functionality and privacy should be achieved with low communication and computation overhead

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 10, October 2015
Available at http://internationaljournalofresearch.org

2. Background and Building Blocks:

2.1 System & Threat Model

Consider a cloud data hosting service that involves three entities: data owner, cloud server and data user. The data owner may be an individual or an enterprise, who wishes to outsource a collection of documents D = (D1, D2, . . . , Dn) in encrypted form C = (C1, C2, . . . , Cn) to the cloud server and still preserve the search functionality on outsourced data. Ci = ES[Di] is the encrypted version of the document Di computed using a semantically secure encryption scheme E with a secret key S. To enable multi-keyword ranked search capability, the data owner constructs searchable index I that is built on m distinct keywords K = (k1, k2, . . . , km) extracted from the original dataset D. Both I and C are outsourced to the cloud server. To securely search the document collection for one or more keywords $K^{-} \in K$, the authorized data user uses search trapdoor (distributed by the data owner) that generates the search request to the cloud server. Once the cloud server receives such request, it performs a search based on the stored index I and returns a ranked list of encrypted documents $L \subseteq C$ to the data user. The data user then uses the secret key S, securely obtained from the data owner, to decrypt received documents L to original view. We assume a honest-but-curious model for the cloud server. The cloud server is honest, that is, it is always available to the data user and it correctly follows the designated protocol specification, and it provides all services that are expected. The curious cloud server may try to perform some additional analysis to breach the confidentiality of the stored data. In the rest of the paper, the cloud server and the adversary are the same entity. That way, the adversary has access to the same set of information as the cloud server. For this work, we are not concerned about the cloud server being able to link a query to a specific user; nor are we concerned about any denial-of-service attacks.

2.2 Notations and Preliminaries:

Let D = (D1, D2, . . . , Dn) be a set of documents and K = (k1, k2, . . . , km) be the dictionary consisting of unique keywords in all documents in D, where $\forall$ i $\in$ [1,m] ki$\in$ {0, 1} $*$. C = {C1, C2, . . . , Cn} is an encrypted document collection stored in the cloud server. Ii is a searchable index associated with the corresponding encrypted document Ci. If A is an algorithm then a $\leftarrow$ A(. . .) represents the result of applying the algorithm A to given arguments. Let R be an operational ring, we write vectors in bold, e.g. v $\in$ R. The notation v[i] refers to the i-th coefficient of v. We denote the dot product of u, v $\in$ R as u $\otimes$ v = P i=1 u[i] $\cdot$ v[i] $\in$ R. We use bxe to indicate rounding x to the nearest integer, and bxc, dxe (for x > 0)to indicate rounding down or up.

Cryptographic Notations:

A private-key encryption scheme is a set of three polynomial-time algorithms SKE = (Gen, Enc, Dec) such that Gen is a probabilistic algorithm that takes a security parameter k and returns a secret key Ksecret; Enc is a probabilistic algorithm that takes a key Ksecret and a message m, and outputs a ciphtertext ξ; Dec is a deterministic algorithm that takes a secret key Ksecret and a ciphertext ξ, and outputs m if Ksecret is the valid secret key. We say that SKE is CPA-secure if the ciphtertexts it outputs do not reveal any partial information about the original plaintext to an adversary that can adaptively query an encryption oracle. We also make use of use pseudo-random function (PRF), which is a polynomial-time computable function that cannot be distinguished from random functions by any probabilistic polynomial-time adversary. We refer the reader for formal definitions of semantic security, CPA-security and PRFs. We now review definitions related to homomorphic encryption. Our definitions are based on Gentry's works and but we slightly relax the definition of decryption correctness, to allow a negligible probability of error.

3. Secured Multi-keyword Ranked Search over Encrypted Cloud Data:

In cloud computing data possessor are goaded to farm out their complex data management systems from local sites to the commercial public cloud for greater flexibility and economic savings. To ensure safety of

stored data, it is must to encrypt the data before storing. It is Shiba Sampat Kale et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7093-7096 www.ijcsit.com 7093 necessary to invoke search with the encrypted data also. The specialty of cloud data storage should allow copious keywords in a solitary query and result the data documents in the relevance order. In, main aim is to find the solution of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. A variety of multi- keyword semantics are available, an efficient similarity measure of "coordinate matching" (as many matches as possible), to capture the data documents' relevancy to the search query is used. Specifically "inner product similarity", i.e., the number of query keywords appearing in a document, to quantitatively evaluate such similarity measure of that document to the search query is used in MRSE algorithm. The main limitation of this paper was, the user's identity (ID) is not kept hidden. Due to this, whoever puts the data on Cloud Service Provider was known. This may be risky in some situations where confidentiality of data needs to be maintained. Hence, this drawback is overcome in the proposed system.

3.1.Privacy Preserving Keyword Searches on Remote Encrypted Data:

Consider the problem: a user U wants to store his files in an encrypted form on a remote file server S. Later the user U wants to efficiently retrieve some of the encrypted files containing specific keywords, keeping the keywords themselves secret and not to endanger the security of the remotely stored files. For example, a user may want to store old e-mail messages encrypted on a server managed by Yahoo or another large vendor, and later retrieve certain messages while travelling with a mobile device. In , solutions for this problem under well-defined security requirements are offered. The schemes are efficient as no public-key cryptosystem is involved. Indeed, the approach is independent of the encryption method chosen for the remote files. They are incremental too. In that, user U can submit new files which are secure against previous queries but still searchable against future queries. From this, the main theme taken is of storing data remotely on other server and retrieving that data from anywhere via mobile, laptop etc.

3.2 Cryptographic Cloud Storage:

When the benefits of using a public cloud infrastructure are clear, it introduces significant security and privacy risks. In fact, it seems that the biggest obstacle to the adoption of cloud storage (and cloud computing in general) is concern over the confidentiality andintegrity of data. In, an overview of the benefits of a cryptographic storage service, for example, reducing the legal exposure of both customers and cloud providers, and achieving regulatory compliance is provided. Besides this, cloud services that could be built on top of a cryptographic storage service such as secure backups, archival, health record systems, secure data exchange and e-discovery is stated briefly.

3.3. Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data:

On one hand, users who do not necessarily have prior knowledge of the encrypted cloud data, have to post process every retrieved file in order to find ones most matching their interest; On the other hand, invariably retrieving all files containing the queried keyword further incurs unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm. This paper has defined and solved the problem of effective yet secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency) thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. For the first time, the paper has defined and solved the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud

data utilization system to become a reality. The proposed ranking method proves to be efficient to return highly relevant documents corresponding to submitted search terms. The idea of proposed ranking method is used in our proposed system in order to enhance the security of data on Cloud Service Provider.

3.4 Providing Privacy Preserving in Cloud Computing:

Privacy is an important issue for cloud computing, both in terms of legal compliance and user trust and needs to be considered at every phase of design. The paper tells the importance of protecting individual's privacy in cloud computing and provides some privacy preserving technologies used in cloud computing services. Paper tells that it is very important to take privacy into account while designing cloud services, if these involve the collection, processing or sharing of personal data. From this paper, main theme taken is of preserving privacy of data. This paper only describes privacy of data but doesn't allow indexed search as well as doesn't hide user's identity. Thus, these two drawbacks are overcome in our proposed system.

3.5 Privacy Preserving Data Sharing With Anonymous ID Assignment:

In this paper, an algorithm for anonymous sharing of private data among N parties is developed. This technique is used iteratively to assign these nodes ID numbers ranging from 1 to N. This assignment is anonymous in that the identities received are unknown to the other members of the group. In , existing and new algorithms for assigning anonymous IDs are examined with respect to trade-offs between communication and computational requirements. These new algorithms are built on top of a secure sum data mining operation using Newton's identities and Sturm's theorem. The main idea taken from this paper is of assigning anonymous ID to the user on the cloud.

3.6 Enabling Efficient Fuzzy Keyword Search over Encrypted Data in Cloud Computing:

In this paper, main idea is to formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. This basic idea is taken but it is for multi-keyword raked search (MRSE scheme) in our proposed system. In, design of secure cloud storage.

H.Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing: Achieving finegrainedness, scalability, and data confidentiality of access control simultaneously is a problem which actually still remains unresolved. The paper addresses this challenging open issue by, on one hand, defining and enforcing access policies based on data attributes, and, on the other hand, allowing the data owner to delegate most of the computation tasks involved in fine-grained data access control to untrusted cloud servers without disclosing the underlying data contents. In, authors have proposed a privacy-preserving public auditing system for data storage security in Cloud Computing scheme is proposed. It utilizes the homomorphic linear authenticator and random masking to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process, which eliminates the burden of cloud user from the tedious and possibly expensive auditing task, it also alleviates the user's fear of his/her outsourced data leakage.

4. CONCLUSION

In this paper, for the first time we define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching," i.e., 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, we propose a basic idea of MRSE using secure inner product computation. Then, we give two improved MRSE schemes to achieve various stringent

privacy requirements in two different threat models. We also investigate some further enhancements of our ranked search mechanism, including supporting more search semantics, i.e., TF

IDF, and dynamic data operations. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world data set show our proposed schemes introduce low overhead on both computation and communication. In our future work, we will explore checking the integrity of the rank order in the search result assuming the cloud server is untrusted.

REFERENCES

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. 829-837, Apr, 2011.

[2] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A Break in the Clouds: Towards a Cloud Definition," ACM SIGCOMM Comput. Commun.Rev., vol. 39, no. 1, pp. 50-55, 2009.

[3] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM, pp. 693-701, 2012.

[4] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.

[5] A. Singhal, "Modern Information Retrieval: A Brief Overview," IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001.

[6] I.H. Witten, A. Moffat, and T.C. Bell, Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kaufmann Publishing, May 1999.

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

[8] E.-J. Goh, "Secure Indexes," Cryptology ePrint Archive, http:// eprint.iacr.org/2003/216. 2003.

[9] Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.

[10] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06), 2006.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

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

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

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

[25] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked

GovardhanPothu, SV College of Engineering

MrAnnapareddy V N REDDY M.Tech, (Ph.D); KL University, SV College of Engineering