ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Privateness -Protect Diffuse Keywords Classified Searching Over Cloud Computing

Ali Naji Yasser
Master of Science (Information System),
Nizam College (Autonomous),O.U,
Basheer Bagh, Hyderabad.

T. Ramdas Naik
Assistant Professor Dept, Computer Science (PG)
Nizam College (Autonomous),O.U,
Basheer Bagh, Hyderabad.

ABSTRACT:

The Cloud Computing can provide dynamically scalable resources provisioned as a service over the cyberspace. It transfers stored data as well as service for demand users. Cloud Computing represents one of the most significant shifts in information technology many of us are likely to see in our lifetimes .Cloud defined as a computing machine or group of machine which as a server/s, holds all the applications and document necessary all at one place so that any user can access it from anywhere whoever has a access to that server without actually using his/her physical machine. Any individual user who has permission to access the server can use the server's processing power to run an application, store data, or perform any other computing task. Data can be stored and retrieved anywhere and anytime with the help of the cloud. But retrieving the encrypted data is a challenging one for the user, and the later keyword based approach overcomes that problem. The advancement in cloud computing has motivated the data owners to outsource their data management systems from local sites to commercial public cloud for great flexibility and economic savings. But people can enjoy full benefit of cloud computing if we are able to address very real privacy and security concerns that come with storing sensitive personal information. 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, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE).

We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we 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.

We further use "inner product similarity" to quantitatively evaluate such similarity measure. We 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, we 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.

Keywords:

Cloud computing, encryption, Privateness -Protect, keyword search, MRSE, privacy preserving, Confidential Data.

INTRODUCTION:

Distributed computing is a field of computer science that studies distributed systems. A distributed system is a software system in which components located on networked computers communicate and coordinate their actions by passing messages. The components interact with each other in order to achieve a common goal. There are many alternatives for the message passing mechanism, including RPC-like connectors and message queues. Three significant characteristics of distributed systems are: concurrency of components, lack of a global clock, and independent failure of components. An important goal and challenge of distributed systems is location transparency. Examples of distributed systems vary from SOA-based systems to massively multiplayer online games to peer-to-peer applications.

A computer program that runs in a distributed system is called a distributed program, and distributed programming is the process of writing such programs. Distributed computing also refers to the use of distributed systems to solve computational problems. In distributed computing, a problem is divided into many tasks, each of which is solved by one or more computers, which communicate with each other by message passing.Other typical properties of distributed systems include the following: The system has to tolerate failures in individual computers.The structure of the system (network topology, network latency, number of computers) is not known in advance, the system may consist of different kinds of computers and network links, and the system may change during the execution of a distributed program.Each computer has only a limited, incomplete view of the system. Each computer may know only one part of the input.

Distributed systems are groups of networked computers, which have the same goal for their work. The terms “concurrent computing”, “parallel computing”, and “distributed computing” have a lot of overlap, and no clear distinction exists between them. The same system may be characterized both as “parallel” and “distributed”; the processors in a typical distributed system run concurrently in parallel. Parallel computing may be seen as a particular tightly coupled form of distributed computing, and distributed computing may be seen as a loosely coupled form of parallel computing. Nevertheless, it is possible to roughly classify concurrent systems as “parallel” or “distributed” using the following criteria:

In parallel computing, all processors may have access to a shared memory to exchange information between processors.In distributed computing, each processor has its own private memory (distributed memory). Information is exchanged by passing messages between the processors.

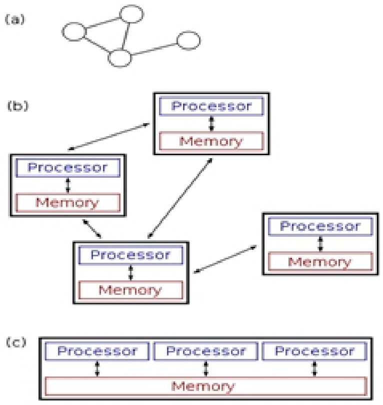

Fig:- distributed and parallel system

The figure on the right illustrates the difference between distributed and parallel systems. Figure (a) is a schematic view of a typical distributed system; as usual, the system is represented as a network topology in which each node is a computer and each line connecting the nodes is a communication link. Figure (b) shows the same distributed system in more detail: each computer has its own local memory, and information can be exchanged only by passing messages from one node to another by using the available communication links. Figure (c) shows a parallel system in which each processor has a direct access to a shared memory.

The situation is further complicated by the traditional uses of the terms parallel and distributed algorithm that do not quite match the above definitions of parallel and distributed systems; see the section Theoretical foundations below for more detailed discussion. Nevertheless, as a rule of thumb, high-performance parallel computation in a shared-memory multiprocessor uses parallel algorithms while the coordination of a large-scale distributed system uses distributed algorithms.

LITERATURE SURVEY:

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength.

Once these things are satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

“AUTHORS:S. Kamara and K. Lauter”

We consider the problem of building a secure cloud storage service on top of a public cloud infrastructure where the service provider is not completely trusted by the customer. We describe, at a high level, several architectures that combine recent and non-standard cryptographic primitives in order to achieve our goal. We survey the benefits such architecture would provide to both customers and service providers and give an overview of recent advances in cryptography motivated specifically by cloud storage.

AUTHORS:Y.-C. Chang and M. Mitzenmacher”

We consider the following problem: a user $\mathcal{U}$ wants to store his files in an encrypted form on a remote file server $\mathcal{S}$. Later the user $\mathcal{U}$ wants to efficiently retrieve some of the encrypted files containing (or indexed by) specific keywords, keeping the keywords themselves secret and not jeopardizing 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 this paper, we offer solutions for this problem under well-defined security requirements. Our schemes are efficient in the sense that no public-key cryptosystem is involved. Indeed, our approach is independent of the encryption method chosen for the remote files. They are also incremental, in that $\mathcal{U}$ can submit new files which are secure against previous queries but still searchable against future queries.

AUTHORS: J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou

As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud. For the protection of data privacy, sensitive data usually have to be encrypted before outsourcing, which makes effective data utilization a very challenging task. Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords and selectively retrieve files of interest, these techniques support only exact keyword search. That is, there is no tolerance of minor typos and format inconsistencies which, on the other hand, are typical user searching behavior and happen very frequently. This significant drawback makes existing techniques unsuitable in Cloud Computing as it greatly affects system usability, rendering user searching experiences very frustrating and system efficacy very low. In this paper, for the first time we formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy.

Fuzzy keyword search greatly enhances system usability by returning the matching files when users’ searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. In our solution, we exploit edit distance to quantify keywords similarity and develop an advanced technique on constructing fuzzy keyword sets, which greatly reduces the storage and representation overheads. Through rigorous security analysis, we show that our proposed solution is secure and privacy-preserving, while correctly realizing the goal of fuzzy keyword search.

PROBLEM STATEMENT:

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. 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. Draw backs:

- The encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirement.

- On enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality

Problem Solution:

In this paper, for the first time, we define and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system wise privacy in the cloud computing paradigm. Among various multi-keyword semantics, we 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.

Specifically, we use "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. During the index construction, each document is associated with a binary vector as a sub-index where each bit represents whether corresponding keyword is contained in the document.

The search query is also described as a binary vector where each bit means whether corresponding keyword appears in this search request, so the similarity could be exactly measured by the inner product of the query vector with the data vector. However, directly outsourcing the data vector or the query vector will violate the index privacy or the search privacy.

To meet the challenge of supporting such multi keyword semantic without privacy breaches, we propose a basic idea for the MRSE using secure inner product computation, which is adapted from a secure k-nearest neighbor (kNN) technique , and then give two significantly improved MRSE schemes in a step-by-step manner to achieve various stringent privacy requirements.

ADVANTAGES:

- Search result should be ranked by the cloud server according to some ranking criteria.

- To reduce the communication cost.

IMPLEMENTATION:

Data Owner:

This module helps the owner to register those details and also include login details. This module helps the owner to upload his file with encryption using RSA algorithm. This ensures the files to be protected from unauthorized user.

Data User:

This module includes the user registration login details. This module is used to help the client to search the file using the multiple key words concept and get the accurate result list based on the user query. The user is going to select the required file and register the user details and get activation code in mail email before enter the activation code. After user can download the Zip file and extract that file.

Encryption:

This module is used to help the server to encrypt the document using RSA Algorithm and to convert the encrypted document to the Zip file with activation code and then activation code send to the user for download.

Rank Search:

These modules ensure the user to search the files that are searched frequently using rank search.This module allows the user to download the file using his secret key to decrypt the downloaded data. This module allows the Owner to view the uploaded files and downloaded files

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 dataset 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.

FUTURE WORK:

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.we will concentration the encrypted data of semantic keyword search in order that we can confrontwiththemore sophisticated search.

REFERENCES:

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-PreservingMulti-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, "ABreak in the Clouds: Towards a Cloud Definition," ACMSIGCOMM 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-BasedSecure 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 Cryptography and Data Security, Jan. 2010.

[5] A. Signal, "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 KaufmannPublishing, May 1999.

[7] D. Song, D. Wagner, and A. Perrig, "Practical Techniques forSearches on Encrypted Data," Proc. IEEE Symp. Security andPrivacy, 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 KeywordSearches 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, "SearchableSymmetric 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, "PublicKey Encryption with Keyword Search," Proc. Int'l Conf. Theory andApplications of Cryptographic Techniques (EUROCRYPT), 2004.

[12] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic andEfficiently Searchable Encryption," Proc. 27th Ann. Int'l CryptologyConf. 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, "SearchableEncryption Revisited: Consistency Properties, Relation to AnonymousIbe, 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, "FuzzyKeyword 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 KeyEncryption That Allows PIR Queries," Proc. 27th Ann. Int'lCryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

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

[17] L. Ballard, S. Kamara, and F. Monrose, "Achieving EfficientConjunctive 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 Querieson Encrypted Data," Proc. Fourth Conf. Theory Cryptography (TCC),pp. 535-554, 2007.

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

[20] Y. Hwang and P. Lee, "Public Key Encryption with ConjunctiveKeyword 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 SupportingDisjunctions, Polynomial Equations, and Inner Products," Proc.27th Ann. Int'l Conf. Theory and Applications of CryptographicTechniques (EUROCRYPT), 2008.

[22] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters,"Fully Secure Functional Encryption: Attribute-Based Encryptionand (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 EncryptionSystems," Proc. Sixth Theory of Cryptography Conf. Theory ofCryptography (TCC), 2009.

[24] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized Private KeywordSearch over Encrypted Data in Cloud Computing," Proc. 31stInt'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 RankedKeyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'lConf. Distributed Computing Systems (ICDCS '10), 2010.

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

[27] W.K. Wong, D.W. Cheung, B. Kao, and N. Mamoulis, "SecurekNN Computation on Encrypted Databases," Proc. 35th ACMSIGMOD Int'l Conf. Management of Data (SIGMOD), pp. 139-152,2009.

[28] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable,and Fine-Grained Data Access Control in Cloud Computing,"Proc. IEEE INFOCOM, 2010.

[29] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-PreservingPublic Auditing for Data Storage Security in Cloud Computing,"Proc. IEEE INFOCOM, 2010.

[30] S. Zerr, E. Demidova, D. Olmedilla, W. Nejdl, M. Winslett, andS. Mitra, "Zerber: r-Confidential Indexing for DistributedDocuments," Proc. 11th Int'l Conf. Extending Database Technology(EDBT '08), pp. 287-298, 2008.

[31] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-kRetrieval from a Confidential Index," Proc. 12th Int'l Conf.Extending Database Technology (EDBT '09), pp. 439-449, 2009.

[32] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Cryptographyfrom Anonymity," Proc. IEEE 47th Ann. Symp.Foundationsof CS, pp. 239-248, 2006.

[33] J. Zobel and A. Moffat, "Exploring the Similarity Space," ACMSIGIR Forum, vol. 32, pp. 18-34, 1998.

[34] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, "Toward Secureand Dependable Storage Services in Cloud Computing," IEEETrans. Services Computing, vol. 5, no. 2, pp. 220-232, Apr.-June 2012.

[35] W.W. Cohen, "Enron Email Data Set," http://www.cs.cmu.edu/~enron/, 2013.

[36] N. Cao, Z. Yang, C. Wang, K. Ren, and W. Lou, "PrivacypreservingQuery over Encrypted Graph-Structured Data in CloudComputing," Proc. Distributed Computing Systems (ICDCS),pp. 393-402, June, 2011.

AUTHORS BIOGRAPHY:

Ali Naji Yasser

pursuing his Master of Science in Information System, from Nizam College (Autonomous),O.U,Basheer Bagh, Hyderabad,India.The Ministry of Education,Republic of Iraq.

T. Ramdas Naik

Assistant Professor Dept, Computer Science (PG), Qualifications : B.E, MCA,M.Tech,(Ph.D) Nizam College (Autonomous),O.U, Basheer Bagh, Hyderabad, India.