International Journal of Science Engineering and Advance Technology

A Novel Multi-keyword Ranked Search System In encrypted and Synonym Queries supported Cloud

Ramachandra Rao Netinti [1], S.Madhri [2]
[1]PG Scholar, Pydah College of Engineering, Kakinada, AP, India, E-mail: netintiramchandra@gmail.com.
[2]Assistant Professor, Pydah College of Engineering, Kakinada, AP, India.

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

Key words: Cloud computing, searchable encryption, privacy-preserving, keyword search, ranked search.

I. Introduction

The Distributed computing of a field computer science that studies distributed systems. The distributed system is a software system in which components located on networked computers communicate and the coordinate their actions by passing through 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 lack of a global clock, and independent failure of components characteristics of distributed systems are: concurrency of components. An important goal and challenge of distributed systems is location transparency. Examples of online games to peer-to-peer applications. A computer program that runs in a distributed system is called a distributed program, programs. Distributed computing also refers to the use and distributed programming is the process of writing such of distributed systems to solve computational problems. Each of which is solved by one or more computers, which communicate with each other by message passing. In distributed computing, a problem is divided into many tasks, the word *distributed* in terms such as "distributed system", "distributed programming", and "distributed algorithm" originally referred to computer networks where geographical area. The terms are nowadays used in a much wider sense, individual computers were physically distributed within some even referring to autonomous processes each other by message passing. While there is no single that run on the same physical computer and interact with definition of a distributed system, the following defining properties are commonly used:

- There are several autonomous computational entities, each of which has its own local memory.
- The entities communicate with each other by message passing.

In this article, the large computational entities are called *computers* or *nodes*. Alternatively, each computer may have its own user with individual needs, and the purpose of the distributed system communication services to the users. Other typical properties to coordinate A distributed system may have a common goal, such as solving the use of shared resources or provide of distributed systems include the following:

- The system has to tolerate failures in individual computers.
- The structure of the system (network topology, network latency, the system may consist of different kinds of computers and network links, and the system may number of computers) is not known in advance, change during the execution of a distributed program.
- Each computer may know only one part of the input. Has only a limited, incomplete view of the system. Each computer

Distributed systems are groups of networked computers, which have "parallel computing", and "distributed them. The same system may be characterised both as "parallel" and "distributed"; the processors in a typical distributed computing" have a lot of overlap, and no clear distinction exists between system run concurrently in parallel. Parallel computing may be seen as a particular tightly coupled form of distributed computing, the same goal for their work. The terms "concurrent 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 shared memory to exchange information between processors computing, all processors may have access.
- In distributed computing, each processor has its own private memory (distributed memory). Information is exchanged by passing messages between the processors.

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.

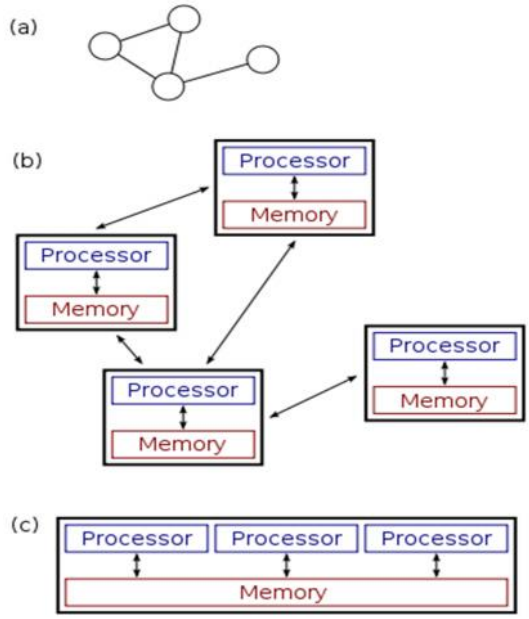


Figure (c) shows a parallel system in which each processor has a direct access to a shared memory. The situation is further complicated by the traditional uses of the terms parallel and distributed *algorithm* that do not quite match the above definitions of parallel and distributed *systems*; see the 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.

II.System Analysis

Existing System:

The effective data retrieval need, the large amount of documents demand ranking, information quickly, rather than burdensomely sorting the cloud server to perform result relevance through instead of returning undifferentiated results. Such ranked search system enables data users to find the most relevant every match in the content collection. Ranked search can also "pay-as-you-use" cloud paradigm. For privacy protection, back only the most relevant data, which is highly desirable in the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system such ranking operation, however, should not leak any keyword related information. On the other hand, to improve elegantly eliminate unnecessary network traffic by sending to support multiple keywords search, as single keyword search often yields far too coarse results.

Disadvantages Of Existing System:

The encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirements. On enrich the search flexibility; they are still not adequate to provide users with acceptable result ranking functionality

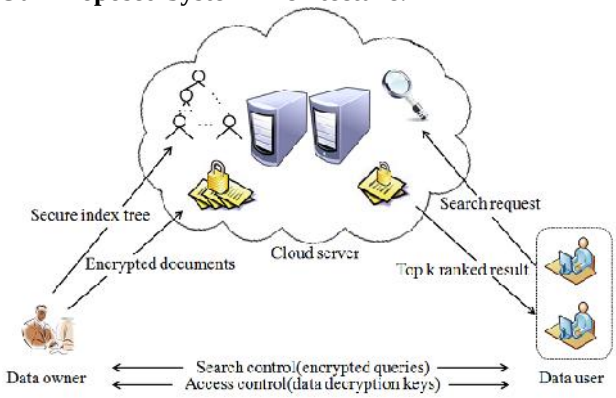
Proposed System:

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 Of Proposed System:

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

Our Proposed System Architecture:



III. Implementation

Modules Description

Data Owner Module

This module helps the module helps the owner to upload his file with encryption using RSA algorithm. Owner to register

those details and also include login details. This ensures the files to be protected from unauthorized user.

Data User Module

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 Module:

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 Module

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

IV. Privacy-Preserving And Efficient Mrse

To efficiently achieve multi-keyword ranked search, we propose to employ "inner product similarity" [6] to quantitatively evaluate the efficient similarity measure "coordinate matching." Specifically, Di is a binary data vector for document Fi where each bit to compare the similarity of different documents to the query. But, to preserve strict system wise privacy, data vector Di, query vector Q and their inner product Di_Q should not be exposed to the cloud server. In this section, we first propose a basic idea for the MRSE using secure inner product computation, which is adapted from a secure KNN technique, and then show how to significantly improve it to be privacy-preserving against different threat models in the MRSE framework in a step-by-step manner. We further discuss supporting more search semantics and dynamic operation.

Results

Screen Shots

V. 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 UN trusted.

VI. 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," 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-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 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 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. 31st Int'l Conf. Distributed Computing Systems (ICDCS '10), pp. 383-392, June 2011.