Privacy Preserving Search over Encrypted Data on Cloud

Rajendra Hiray[1] Ajinkya Ingle[2] Kedar Jangam[3] Ravina Dhage[4] Prof. R. T. Umbare[5]
[1,2,3,4,5]JSPM Rajarshi Shahu College of Engineering Tathawade, Pune, India

Abstract— Cloud computing provides people and enterprises Brobdingnagian computing power and ascendable storage capacities to support a range of huge data applications in domains like health care and scientific analysis, thus additional and additional knowledge homeowners square measure concerned to to source their knowledge on cloud servers for excellent convenience in knowledge management and mining. However, knowledge sets like health records in electronic documents typically contain sensitive data that brings regarding privacy problems if the documents square measure free shared to partially untrusted third-parties in cloud. A sensible and wide used technique for knowledge privacy preservation is to encrypt data before outsourcing to the cloud servers, that but reduces knowledge utility and makes many ancient knowledge analytic operators like keyword-based top-k document retrieval obsolete. During this paper, we tend to investigate the multi-keyword top-k search problem for massive encryption against privacy breaches, associated arrange to determine an efficient and secure answer to the current drawback. Specifically, for the privacy concern of question knowledge, we tend to construct a special tree-based index structure and style a random traversal formula, which makes even a similar question to provide totally different visiting ways on the index, and would possibly to boot maintain the accuracy of queries unchanged under stronger privacy. For up the question efficiency, we tend to tend to propose a gaggle multi-keyword top-k search theme supported plan of partition wherever a gaggle of tree-based indexes square measure made for all documents. Finally, we tend to mix these ways along into an efficient and secure approach to deal with our projected top k similarity search. In-depth experimental results on real-life knowledge sets demonstrate that our projected approach will significantly improve the potential of defensive the privacy breaches, the measurability and the time efficiency of question process over the progressive ways.

Key words: Cloud Data Sharing, CP-ABE, Key Management, Security, Efficiency

I. INTRODUCTION

CLOUD computing has emerged as a disruptive trend in both IT industries and research communities recently, its salient characteristics like high scalability and pay-as you-go fashion have enabled cloud consumers to purchase the powerful computing resources as services according to their actual requirements, such that cloud users have no longer need to worry about the wasting on computing resources and the complexity on hardware platform management. Nowadays, more and more companies and individuals from a large number of big data application shave outsource their data and deploy their services into cloud servers for easy data management, efficient data mining and query processing tasks. Data encryption has been widely used for data privacy preservation in data sharing scenarios, it refers to mathematical calculation and algorithmic scheme that transform plaintext into cypher text, which is a non-readable form to unauthorized parties. A variety of data encryption models have been proposed and they are used to encrypt the data before outsourcing to the cloud servers. However, applying these approaches for data encryption usually cause tremendous cost in terms of data utility, which makes traditional data processing methods that are designed for plaintext data no longer work well over encrypted data .

Data encryption has been wide used for knowledge privacy preservation in knowledge sharing situations, it refers to mathematical calculation and algorithmic theme that remodel plaintext into cypher text that may be a non-readable type to unauthorized parties. A spread of knowledge secret writing models have been planned [3], [4], [5] and that they square measure accustomed write in code the data before outsourcing to the cloud servers. However, applying these approaches for encryption typically cause tremendous price in terms of knowledge utility, that makes traditional processing strategies that square measure designed for plaintext knowledge now not work spill encrypted knowledge. Data encryption has been wide used for knowledge privacy preservation in knowledge sharing situations, it refers to mathematical calculation and algorithmic theme that remodel plaintext into ciphertext that may be a non-readable type to unauthorized parties. A spread of knowledge secret writing models have been planned [3], [4], [5] and that they square measure accustomed write in code the data before outsourcing to the cloud servers. However, applying these approaches for encryption typically cause tremendous price in terms of knowledge utility, that makes traditional processing strategies that square measure designed for plaintext knowledge now not work spill encrypted knowledge.

II. LITERATURE SURVEY

A. Paper 1: A Secure and Dynamic Multi Keyword Ranked Search Scheme over Encrypted

The major aim of this paper is to resolve the matter of multi-keyword hierarchical search over encrypted cloud knowledge (MRSE) at the time of protective actual technique wise privacy within the cloud computing construct. Knowledge holder's area unit inspired to source their tough knowledge management systems from native sites to the business public cloud for big flexibility and monetary savings. But for protecting knowledge privacy, sensitive knowledge got to be encrypted before outsourcing, which performs ancient knowledge utilization supported plaintext keyword search. As a result, permitting Associate in Nursing encrypted cloud knowledge search service is of supreme significance. Visible of the massive range of information users and documents within the cloud, it's essential to allow many keywords within the search demand and come back documents within the order of their acceptable to those keywords. Similar mechanism on searchable cryptography makes centre on single keyword searcher Boolean keyword search, and infrequently type the search results. within the middle of various multi-keyword linguistics, deciding the well-organized similarity live of

coordinate matching, it means as several matches as doable, to capture the suitable knowledge documents to the search question. notably, we consider dot product similarity i.e., the number of question keywords shows in a document, to quantitatively estimate such match live that document to the search question. Through the index construction, each document is connected with a binary vector as a sub-index wherever every bit characterize whether or not matching keyword is contained within the document.

B. Paper 2: Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data

The innovation in cloud computing has encouraged the data owners to outsource their data managing system from local sites to profitable public cloudfor excessive flexibility and profitable savings. But people can like full benefit of cloud computing, if we are able to report very real secrecy and security concerns that come with loading sensitive personal information. Allowing an encrypted cloud data search facility is of great significance. In view of the huge number of data users, documents in the cloud, it is important for the search facility to agree multi keywords query and arrange for result compare.

C. Paper 3: Secure Indexes

A secure index is a data structure that allows a query with a trapdoor for a word x to test in O(1) time only if the index contains x; The index reveals no information about its contents without valid trapdoors, and trapdoors can only be generated with a secret key. Secure indexes are a natural extension of the problem of constructing data structures with privacy guarantees such as those provided by oblivious and history independent data structures. In this paper, we formally define a secure index and formulate a security model for indexes known as semantic security against adaptive chosen keyword attack (ind-cka). We also develop an efficient ind-ckasecure index construction called z-idx using pseudo-random functions and Bloom filters, and show how to use z-idx to implement searches on encrypted data. This search scheme is the most efficient encrypted data search scheme currently known; It provides O(1) search time per document, and handles compressed data, variable length words, and boolean and certain regular expression queries. The techniques developed in this paper can also be used to build encrypted searchable audit logs, private database query schemes, accumulated hashing schemes, and secure set membership tests.

D. Paper 4: Fuzzy Identity-Based Encryption.

A Fuzzy IBE scheme can be applied to enable encryption using biometric inputs as identities; the error-tolerance property of a Fuzzy IBE scheme is precisely what allows for the use of biometric identities, which inherently will have some noise each time they are sampled. Additionally, we show that Fuzzy-IBE can be used for a type of application that we term "attribute-based encryption". In this paper we present two constructions of Fuzzy IBE schemes. Our constructions can be viewed as an Identity-Based Encryption of a message under several attributes that compose a (fuzzy) identity. Our IBE schemes are both error-tolerant and secure against collusion attacks. Additionally, our basic construction does not use random oracles. We prove the security of our schemes under the Selective-ID security model.

III. EXISTING SYSTEM

In existing System, the keyword-based search is such one widely used applications, and its traditional processing methods cannot be directly applied to encrypted data. Therefore, howto process such queries over encrypted data and at the same time guarantee data privacy becomes a hot research topic. Fortunately, many methodologies based on searchable encryption have been studied. The single keyword search is not smart enough to support advanced queries and the boolean search is unrealistic since it causes high communication cost. Therefore, more recent works like focus on the multi keyword ranked search, which is more practical in pay-asyou-go cloud paradigm.

A. Existing System Disadvantages:

- Single keyword search is not smart enough to support advanced queries.
- Boolean search is unrealistic since it causes high communication cost.

IV. OBJECTIVE

1) Big data encryption against privacy break.
2) Improve the capability of defending the privacy break.
3) Improve scalability and the time efficiency of query processing.

V. PROPOSED SYSTEM

We investigate the multi-keyword top-k search drawback for giant encoding against privacy break, associated conceive to determine an efficient and secure answer to the present drawback. Specifically, for the privacy concern of question knowledge, we tend to construct a special tree-based index structure and style a random traversal formula, that makes even an equivalent question to supply totally different visiting ways on the index, and may additionally maintain the accuracy of queries unchanged below stronger privacy. For raising the question efficiency, we tend to propose a gaggle multi-keyword top-k search theme supported the thought of partition, wherever a gaggle of tree-based indexes are created for all documents. Finally, we tend to mix these strategies along into associate efficient and secure approach to handle our planned top-k similarity search. In depth experimental results on real-life knowledge sets demonstrate that our planned approach will significantly improve the potential of defensive the privacy breaches, the measurability and therefore the time efficiency of question process over the progressive strategies.

A. Proposed System Advantages

1) Multi-keyword top-k Search.
2) Search efficiency.
3) Privacy Preserving.
4) Index security and Query security.

VI. ALGORITHMS

A. Algorithm 1: AES Algorithm

Algorithm Steps

1) Step 1: Start
2) Step 2: Derive the set of round keys from the cipher key.
3) Step 3: Initialize the state array with the block data (plaintext). .
4) Step 4: Add the initial round key to the starting state array.
5) Step 5: Add the initial round key to the starting state array.
6) Step 6: Perform the tenth and final round of state manipulation.
7) Step 7: Copy the final state array out as the encrypted data (ciphertext).

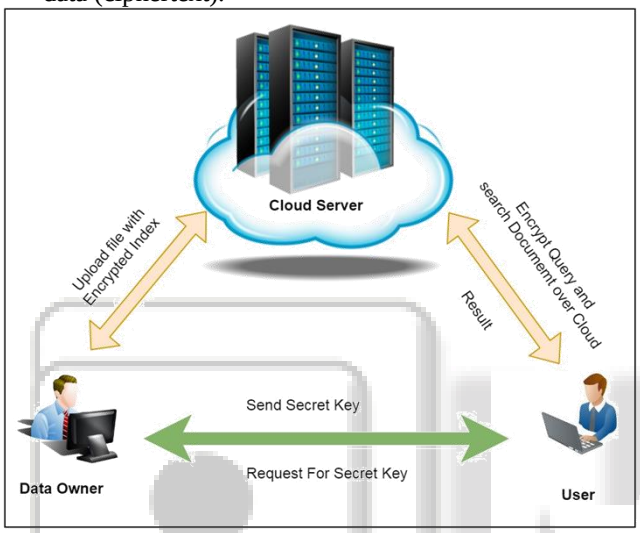

Fig. 1: System Requirement and Specification

B. Hardware Resources Required

1) Processor: Pentium –IV
2) Speed: 1.1 GHz
3) RAM: 256 MB (min)
4) Hard Disk: 20 GB
5) Key Board: Standard Windows Keyboard
6) Mouse: Two or Three Button Mouse
7) Monitor: SVGA

C. Software Resources Required

1) Operating System: Windows 07/08/Above
2) Programming Language: JAVA/J2EE/XML
3) Database: MY SQL

VII. CONCLUSION AND FUTURE SCOPE

We target rising the efficiency and the security of multi-keyword top-k similarity search over encrypted information. At first, we have a tendency to propose the random traversal formula which might bring home the bacon that for 2 identical queries with completely different keys, the cloud server traverses different paths on the index, and also the information user receives different results however with constant high level of question accuracies within the unit of time.

REFERENCES

[1] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions,"in Proceedings of the 13th ACM Conference on Computer and Communications Security. ACM, 2006, pp. 79–88.
[2] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. SP 2000. Proceedings. 2000 IEEE Symposium on, 2000, pp. 44–55.
[3] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003.
[4] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.
[5] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 2, pp. 340–352, 2016
[6] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques forsearches on encrypted data," in Security and Privacy, 2000. SP 2000.Proceedings. 2000 IEEE Symposium on, 2000, pp. 44–55.
[7] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrintArchive,vol. 2003, p. 216, 2003
[8] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keywordsearches on remote encrypted data," in Applied Cryptography andNetwork Security. Springer, 2005, pp. 442–455
[9] Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctivekeyword search and its extension to a multi-user system," inPairing-Based Cryptography–Pairing. Springer, 2007, pp. 2–22.
[10] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keywordsearch over encrypted data," in Applied Cryptography and NetworkSecurity. Springer, 2004, pp. 31–4

