

Top-K Retrieval of Encrypted Cloud Data by Using Secure Multi-Keyword

K. Manoj Kumar* E. Purushotham²

¹M.Tech Student, SITAMS Chittoor, A.P., India.

²Associate Professor, SITAMS Chittoor, A.P., India.

ABSTRACT

Cloud computing is a promising pattern for data outsourcing and high quality data services. The data owner has a collection of n files to outsource onto the cloud server in encrypted form. To achieve this, the data owner needs to build a searchable index from a collection of keywords and then outsources both the encrypted index and encrypted files onto the cloud server. The authorized data user at first generates a query request and the cloud server sends relevant files to the data user. To eliminate the information leakage, a two-round searchable encryption (TRSE) scheme has been proposed that supports top-k multi-keyword retrieval. Homomorphic encryption and Vector space model are employed that involve in ranking. Since ranking is done on user-side based on order-preserving encryption (OPE) efficiency in retrieval of file is improved. The files are ranked in the order of relevance by users interest and only the files with the highest relevance are sent back to users.

Keywords: TRSE, OPE, Homomorphic

1. INTRODUCTION

Cloud computing, a crucial pattern for advanced knowledge service, has become a necessary practicability for knowledge users to source knowledge. Controversies on privacy, however, are ceaselessly given as outsourcing of sensitive data together with e-mails, health history and private photos is explosively developing. Reports of data loss and privacy breaches in cloud computing systems seem from time to time. The most threat on knowledge privacy roots within the cloud itself. Once users source their personal knowledge onto the cloud, the cloud service suppliers are ready to management and monitor the information and therefore the communication between users and therefore the NSA program, operating with AT&T and Verizon, that recorded over ten million phone calls and therefore the bigger powers it offers to telecommunication firms to watch user activity. to confirm privacy, users typically write the info before outsourcing knowledge utilization. However, notwithstanding the encrypted knowledge utilization

is feasible; users still ought to communicate with the cloud and permit the cloud operates on the encrypted knowledge that doubtless causes outflow of sensitive data.

Furthermore, in cloud computing, information homeowners could share their outsourced information with different of users, who might want to solely retrieve the information files they're curious about. one in all the foremost common ways in which to try and do therefore is through keyword-based retrieval. Keyword-based retrieval could be a typical information service and wide applied in plaintext situations, within which users retrieve relevant files in a very file set supported words. it seems to be a troublesome work in cipher text situation thanks to restricted operations on encrypted information. Besides, to enhance practicability and save on the expense within the cloud paradigm, it's most popular to urge the retrieval result with the foremost relevant files that match users' interest rather than all to be graded within the order of connection by users' interest and solely the files with the best relevance area unit sent back to users. A series of searchable regular encoding (SSE) schemes are projected to alter search on cipher text. Ancient point schemes alter users to firmly retrieve the cipher text, however these schemes support solely mathematician with the queried keyword of those files within the result. to enhance security while not sacrificing potency, schemes given in show that they support top-k single keyword retrieval below numerous eventualities. The authors of created makes an attempt to resolve the matter of top-k multikeyword over encrypted cloud information. These schemes, however, suffer from 2 problems—Boolean illustration and the way to strike a balance between security and potency, within the former, files are hierarchic solely by the amount of retrieved keywords.

Entrusting all the work to the user could be natural thanks to avoid information escape. However, the restricted machine power on the user aspect and also the high machine overhead precludes data security. the difficulty of secure multikey word top-k retrieval over encrypted cloud data, thus, is: the way to build the cloud do additional work throughout the method of retrieval while not data escape. In this paper, we introduce the ideas of similarity connection and

theme lustiness to formulate the privacy issue in searchable coding schemes, and then solve the insecurity downside by proposing a two-round searchable encryption (TRSE) theme. Novel technologies within the cryptography community and knowledge retrieval (IR).

Communities are utilized, as well as homomorphic coding and vector area model. The majority of computing work is finished on the cloud whereas the user takes half in ranking that guarantees top-k multikey word retrieval over encrypted cloud knowledge with high security and sensible potency.

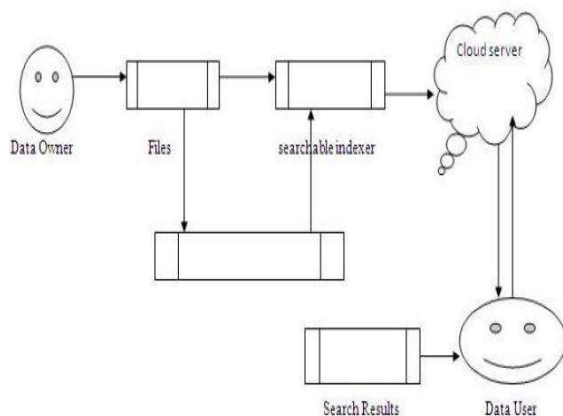


Fig 1: System Architecture

The rest of this paper is organized as follows: We provide scenario and related background in Section 2, and then we give the security definitions and problems with existing schemes in Section 3. In Section 4, we present the detailed description of the proposed searchable encryption scheme. In Section 5, we discuss two main issues of our scheme. Sections 6 and 7 give the security analysis and performance analysis, respectively. Related works are reviewed in Section 8. Section 9 concludes this paper.

2. LIFE CYCLE MODEL

- Waterfall approach was 1st SDLC Model to be used wide in software system engineering to make sure success of the project. In falls model, typically, the end result of 1 part acts because the input for future part consecutive.

Following is a diagrammatic representation of different phases of waterfall model

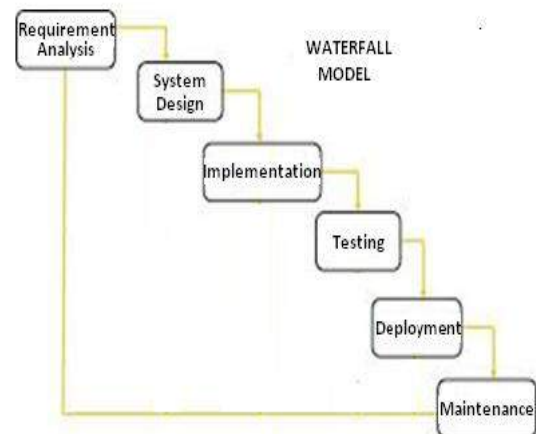


Fig 2: Different Phases of Waterfall Model

The sequential phases in Waterfall model are:

- Requirement Gathering and analysis: All attainable needs of the system to be developed square measure captured during this part and documented. Such needs embody files to upload onto cloud.
- System style: System Design helps in shaping overall system design. In this part, UML diagrams like use case, category and sequence diagrams are designed.
- Implementation: With inputs from system style, the system is 1st developed in little programs known as units, that square measure integrated within the next part.
- Integration and Testing: All the units developed within the implementation part square measure integrated into a system when testing of every unit. Post integration the complete system is tested for any faults and failures.
- Deployment of system: during this part, the merchandise is deployed to the user so files square measure uploaded onto the cloud, which can be accessed by different users.
- Maintenance: There square measure some problems that come back up within the consumer surroundings. to repair those problems patches square measure free. Conjointly to boost the potency of the project some higher techniques are used.

Hence, our project endures of these phases of falls model.

3. PRELIMINARIES

3.1 Scenario

We contemplate a cloud knowledge processing system [ADP system | ADPS | system] hosting data service, as illustrated in Fig. 1, during which 3 totally different entities are involved: cloud server, data owner, and knowledge user. The cloud server hosts third-party knowledge storage and retrieve services. Since knowledge might contain sensitive data, the cloud servers cannot be totally entrusted in protective knowledge. For this reason, outsourced files should be encrypted. Any reasonably data outpouring that might have an effect on knowledge privacy are considered unacceptable.

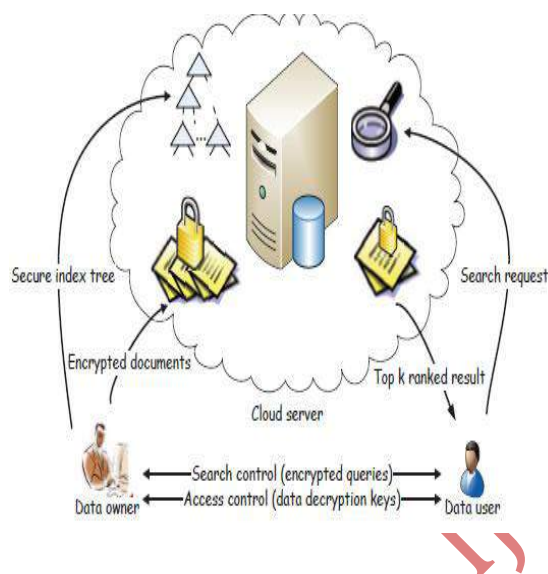


Fig 3: Scenario of retrieval of encrypted cloud data.

The data owner features a assortment of n files the info owner features a assortment of n files $C = \{f_1; f_2; \dots; f_n\}$ to source onto the cloud server in encrypted form and expects the cloud server to produce keyword retrieval service to information owner himself or different approved users. To realize this, the info owner must build a searchable index I from a set of l keywords $W = \{w_1; w_2; \dots; w_l\}$ extracted out of C , and so outsources both the encrypted index I_0 and encrypted files onto the cloud server. The data user is permitted to method multikey word retrieval over the outsourced information. The computing power on the user facet is restricted, which suggests that operations on the user facet ought to be simplified. The approved information user at first generates a question $REQ = \{(w'_1, w'_2, \dots, w'_s) | w'_i \in W, 1 < i < s < l\}$. For privacy thought, that keywords the data user has searched should be hid. Thus, the information user squired the question and sends it to the cloud server that returns the relevant files to the

info user. Afterward, the data user will converted into naormal format and builds use of the files.

4. IMPLEMENTATION

4.1 TRSE Design

Existing point schemes use server-side ranking supported OPE to enhance the potency of retrieval over encrypted cloud knowledge. However, server-side ranking supported OPE violates the privacy of sensitive info, that is taken into account uncompromisable within the security-oriented third party cloud computing state of affairs, i.e., security cannot be trade-off for potency. To realize knowledge privacy, ranking has got to be left to the user facet. ancient user -side schemes, however, load serious process burden and high communication overhead on the user facet, because of the interaction between the server and also the user likewise as searchable index return and ranking score calculation.

4.2 TRSE style

Existing SSE schemes use server-side ranking supported OPE to enhance the potency of retrieval over encrypted cloud information. However, server-side ranking supported OPE violates the privacy of sensitive info, that is taken into account uncompromisable within the security-oriented third party cloud computing state of affairs, i.e., security can't be trade-off for potency. To attain information privacy, ranking must be left to the user aspect. ancient user-side schemes, however, load significant process burden and high communication overhead on the user aspect, attributable to the interaction between the server and therefore the user together with searchable index come back and ranking score calculation.

4.3 Framework of TRSE

The framework of TRSE includes four algorithms: Setup, index build, lures doorgen;

4.4 Setup

The info owner generates the key and public keys for the homo morphic encoding theme. The protection parameter λ is taken because the input, the output could be a secret key, and a public key set.

4.5 Index Build(C, PK)

The data owner builds the secure searchable index from the file assortment C . Technologies from IR community like stemming area unit used to create searchable index I from C , so I is encrypted into I_0 with PK , output the secure searchable index I' .

4.6 Trapdoor Gen ($REQ; PK$)

The data user generates secure trapdoor from his request REQ. Vector T' is constructed from user's multi keyword request REQ so encrypted into secure trapdoor T' with public key from , output the secure trapdoor '.

4.7 Score Calculate (T' ; I')

When receives secure trapdoor T' , the cloud server computes the millions of every files in I' with T' and returns the encrypted result vector N back to the info user

4.8 Relevance rating

Some of the multi key word SSE schemes support solely matches a question. Considering the big range of information users and documents within the cloud, it's necessary to permit Multikey word within the search question and come back documents within the order of their connection with the queried keywords.

4.9 Vector area Model

While tf-idf depicts the burden of one keyword on a file, we tend to use the vector area model to attain a file on multikey word. The vector area model is a pure mathematics model for representing a file as a vector

4.10 Homomorphic encoding

Homomorphic encoding could be a type of encoding that permits specific varieties of computations to be carried out on cipher text and procure an encrypted result that decrypted matches the results of operations performed on the plaintext

4.11 Unpadded RSA

If the RSA public key's modulus and exponent then the encoding of a message is given by . The homomorphic property is then

5. TESTING

5.1 Testing ways

The purpose of testing is to get errors. Testing is that the method of attempting to get each conceivable fault or weakness in a very work product. It provides how to examine the practicality of parts, sub-assemblies, assemblies and or a finished product it's the method of sweat package with the intent of guaranteeing that the software meets its needs and user expectations and doesn't fail in an unacceptable manner. There are numerous varieties of take a look

at. Every take a look at type addresses a selected testing demand.

5.2 Varieties of TESTS

5.2.1 Unit testing:

Unit testing is that the testing of individual package units of the appliance .It is done once the completion of a private unit before integration. .

5.2.2 Integration testing:

Integration tests are designed to check integrated package parts to see if they really run jointly program.

5.2.3 System Testing:

System testing ensures that the complete integrated software meets needs.

5.2.4 White Box Testing:

White Box Testing may be a testing during which the package tester has information of the inner workings, structure and language of the module being tested

5.2.5 Black Box Testing:

Black Box Testing is testing the package with none information of the tested. The planned system has undergone testing; it's free from errors and provides output properly.

6. RESULTS

6.1 Efficiency Improvement

The main charm of the changed FHEI that we tend to use within the TRSE theme is its abstract simplicity compared to Gentry's. This simplicity is achieved at the price of an oversized key size. Though optimizations like standard reduction and compression are often used to scale back the scale of cipher text, the key size remains large for the sensible system. As mentioned in Section five, the user encrypts his trapdoor and sends the cipher text to the cloud server. Therefore, the communication overhead are going to be terribly high if the encrypted trapdoor size is simply too giant. to resolve this drawback and, thus, improve potency, a exchange of the protection of search pattern could also be required unless a replacement coding theme that has additional cheap cipher text size becomes accessible. Researchers from cryptography community have created many makes an attempt to maneuver toward sensible absolutely homomorphic coding over integers.

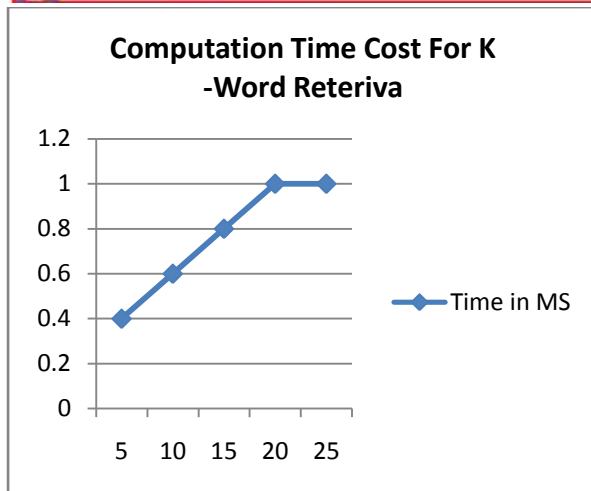


Fig 4: Computation time cost for k-word retrieva

7. CONCLUSIONS

By security analysis, the projected theme ensures that it provides knowledge privacy to all or any the files gift on the cloud server and additionally ensures sensible potency within the retrieval of files. Within the projected system, it'll inspire and solve the matter of security to the files gift on cloud server by creating use of searchable index that comprises keyword and a secret key. Wherever keyword is any word that is provided by the owner and secret key's a random range generated mechanically by a laptop. The projected system additionally guarantees potency by providing ranking to the files and by creating use of top-k retrieval algorithmic program.

REFERENCE:

- [1]. M. Perc, Evolution of the Most Common English Words and Phrases over the Centuries, J. Royal Soc. Interface, 2012.
- [2]. Fuzzy keyword search over encrypted data in cloud computing World Journal of Science and Technology 2012, 2(10):177-185
- [3]. 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),2010.
- [4]. 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.
- [5]. S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS, January 2010, LNCS. Springer, Heidelberg.
- [6]. P. Mell and T. Grance, "Draft nist working definition of cloud computing," Referenced on Jan. 23rd, 2010 Online at <http://csrc.nist.gov/groups/SNS/cloud-computing/index.html>, 2010.
- [7]. Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS'05, 2005.