ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data

Suvarna Dandekar[1], Vrushali Pandit[2], Snehal Kurumkar[3], Sunita Khamkar[4], Tarte V.G[5]

Student, Dept. of CS, Parikrama College of Engineering , Kashti, Savitribai Phule Pune University, Pune, India[1234]

Prof. Dept. of CS, Parikrama College of Engineering, Kashti, Savitribai Phule Pune University, Pune, India[5]

ABSTRACT: With the appearance of distributed computing, information proprietors are spurred to outsource their perplexing information administration frameworks from neighborhood locales to the business open cloud for extraordinary adaptability and financial reserve funds. In any case, for securing information protection, touchy information must be scrambled before outsourcing, which obsoletes customary information usage in light of plaintext watchword seek. Accordingly, empowering an encoded cloud information seek administration is of fundamental significance. Considering the vast number of information clients and archives in the cloud, it is important to permit various watchwords in the inquiry demand and return records in the request of their significance to these catchphrases. Related takes a shot at searchable encryption concentrate on single catchphrase inquiry or Boolean watchword look, and seldom sort the indexed lists. In this paper, surprisingly, we characterize and take care of the testing issue of protection saving multi-catchphrase positioned look over scrambled information in distributed computing (MRSE). We build up an arrangement of strict protection necessities for such a safe cloud information use framework. Among different multi-watchword semantics, we pick the effective similitude measure of "direction coordinating," i.e., however many matches as could be expected under the circumstances, to catch the pertinence of information records to the inquiry question. We assist use "inward item comparability" to quantitatively assess such similitude measure. We first propose an essential thought for the MRSE in light of secure internal item calculation, and afterward give two altogether enhanced MRSE plans to accomplish different stringent necessities in two diverse danger models. To enhance look experience of the information seek administration, we advance extend these two plans to bolster more hunt semantics. Exhaustive examination researching protection and proficiency sureties of proposed plans is given. Probes this present reality information set further show proposed conspires to be sure present low overhead on calculation and correspondence.

KEYWORDS: sensitive data; multi-keyword ranked search; latent semantic

I. INTRODUCTION

Figuring assets are shared by numerous clients. The advantages of cloud can be stretched out from individual clients to associations. The information stockpiling in cloud is one among them. The virtualization of equipment and programming assets in cloud invalidates the money related speculation for owning the information stockroom and its upkeep. Numerous cloud stages like Google Drive, cloud; SkyDrive, Amazon S3, Dropbox and Microsoft Azure give stockpiling administrations. Security and protection concerns have been the real difficulties in distributed computing. The equipment and programming security components like firewalls and so forth have been utilized by cloud supplier. These arrangements are not adequate to shield information in cloud from unapproved clients due to low level of straightforwardness. Since the cloud client and the cloud supplier are in the distinctive trusted area, the outsourced information might be presented to the vulnerabilities. In this way, before putting away the profitable information in cloud, the information should be encoded [2]. Information encryption guarantees the information privacy and trustworthiness. To save the information protection we have to outline a searchable calculation that takes a shot at scrambled information. Numerous scientists have been adding to looking on scrambled information. The inquiry procedures might be single catchphrase hunt or multi watchword seek. In gigantic database the pursuit may bring about numerous archives to be coordinated with catchphrases. This causes trouble for a cloud client to experience all reports and have most pertinent records. Seek in view of positioning is another arrangement, wherein the records are positioned taking into account their pertinence to the catchphrases [3]. Practical searchable encryption strategies help the cloud

DOI: 10.15680/IJIRCCE.2016. 0403245

clients particularly in pay-as-you utilize model. The analysts consolidated the rank of reports with numerous catchphrase hunts to think of effective monetarily reasonable searchable encryption systems. In searchable encryption related writing, calculation time and calculation overhead are the two most as often as possible utilized parameters by the specialists as a part of the space for breaking down the execution of their plans. Calculation time (likewise called "running time") is the timeframe required to perform a computational procedure for instance looking a watchword, producing trapdoor and so on. Calculation overhead is identified with CPU usage regarding asset assignment measured in time.

II. LITERATURE SURVEY

With the benefit of capacity as an administration numerous endeavors are moving their significant information to the cloud, since it costs less, effortlessly versatile and can be gotten to from anyplace at whatever time. The trust between cloud client and supplier is vital. We utilize security as a parameter to set up trust. Cryptography is restricted of setting up trust. Searchable encryption is a cryptographic strategy to give security. In writing numerous specialists have been taking a shot at creating productive searchable encryption plans. In this paper we investigate a percentage of the compelling cryptographic strategies taking into account information structures like CRSA and B-Tree to upgrade the level of security, thus trust. We attempted to actualize the hunt on scrambled information utilizing Azure cloud platform.[2]
Distributed computing is producing part of enthusiasm to give answer for information outsourcing and amazing information administrations. More foundation, associations and enterprises are investigating the likelihood of having their applications, information and their IT resources in cloud. As the information and there by the cloud's size builds seeking of the applicable information is relied upon to be a test. To conquer this test, look record is made to help in speedier inquiry. Be that as it may, seek Index creation and calculation has been mind boggling and tedious, prompting cloud-down time there by ruining the quickness in responding to information demand for mission basic necessities. Center of this paper is to clarify how in the proposed framework, reusability of hunt list is lessening the intricacy of pursuit record calculation. Look list is proposed to be made utilizing parameters like likeness importance, client positioning and plan heartiness. Client positioning certifications why an expression or a sentence or a watchword is utilized much of the time as a part of the transferred information. The proposed framework guarantees that the reusability of hunt record idea, exceptionally diminishes cloud down time while keeping up the security utilizing searchable symmetric encryption (SSE).The client asked for document is recovered from the cloud, utilizing Two-round searchable encryption (TRSE) plan that backings top-k multi-watchword recovery. [1]
These days, more individuals are persuaded to outsource their neighborhood information to open cloud servers for incredible accommodation and lessened expenses in information administration. Be that as it may, regarding security issues, touchy information ought to be encoded before outsourcing, which obsoletes conventional information use like catchphrase based archive recovery. In this paper, we exhibit a safe and effective multi-catchphrase positioned look plan over encoded information, which also underpins dynamic overhaul operations like erasure and insertion of reports. In particular, we develop a file tree in view of vector space model to give multi-catchphrase look, which in the interim backings adaptable overhaul operations. Also, cosine comparability measure is used to bolster precise positioning for query output. To enhance seek productivity, we promote propose an inquiry calculation in light of "Insatiable Depth-first Traverse Strategy". In addition, to ensure the pursuit protection, we propose a safe plan to meet different security necessities in the known ciphertext danger model. Investigates the genuine word dataset demonstrate the adequacy and productivity of proposed plan. [3]

III. PROBLEM STATEMENT

We characterize and take care of the testing issue of security saving multi-catchphrase positioned seek over scrambled information in distributed computing (MRSE). We set up an arrangement of strict protection necessities for such a safe cloud information usage framework. Among different multi-catchphrase semantics, we pick the productive comparability measure of "direction coordinates," i.e., whatever number matches as could be expected under the circumstances, to catch the pertinence of information reports to the pursuit inquiry. We advance use "inward item comparability" to quantitatively assess such closeness measure. As an exceptional instance of adjustment, the operation of erasing existing reports present less calculation and correspondence cost since it just requires to upgrade the record recurrence of all the catchphrases contained by these archives.

IV. OBJECTIVE

I. 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.
II. 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.
III. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.
IV. The proposed schemes indeed introduce low overhead on computation and communication.
V. The proposed schemes introduce nearly constant overhead while increasing the number of query keywords.

Therefore, our schemes cannot be compromised by timing-based side channel attacks that try to differentiate certain queries based on their query time.

VI. EXISTING SYSTEM AAPROACH

The existing techniques on keyword-based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical. All these multi keyword search schemes retrieve search results based on the existence of keywords, which cannot provide acceptable result ranking functionality. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval.

VII.PROPOSED SYSTEM

A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data We construct a special tree-based index structure and propose a “Greedy Depth-first Search” algorithm to provide efficient multi-keyword ranked search. The proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme.

I. Abundant works have been proposed under different threat models to achieve various search functionality,
II. Recently, some dynamic schemes have been proposed to support inserting and deleting operations on document collection.
III. This paper proposes a secure tree-based search scheme over the encrypted cloud data, which supports multi keyword ranked search and dynamic operation on the document collection.

VIII. PROPOSED SYSTEM ALGORITHMS

1. Algorithm to provide efficient multi-keyword ranked search .
2. The secure kNN algorithm is utilized to encrypt the index and query vectors.
3. Propose a “Greedy Depth-first Search” algorithm based on this index tree.
4. Algorithm achieves better-than-linear search efficiency but results in precision loss.
5. The LSH algorithm is suitable for similar search but cannot provide exact ranking.
6. {I′s ; ci} ← GenUpdateInfo (SK; Ts; i; up type)) This algorithm generates the update information {I′s ; ci} which will be sent to the cloud server.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

IX. SYSTEM ARCHITECTURE

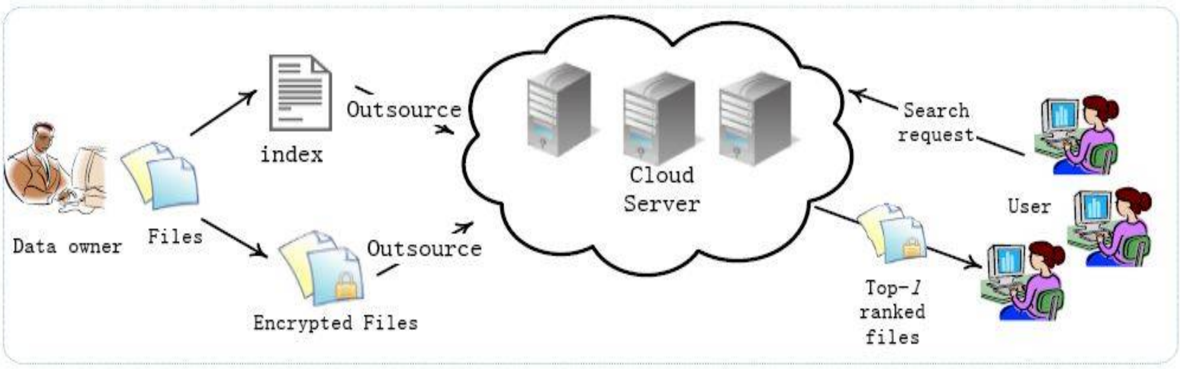

Fig No 01. System Architecture

The cloud server both follows the designated protocol specification but at the same time analyzes data in its storage and message flows received during the protocol so as to learn additional information.

The designed goals of our system are following:

Latent Semantic Search: We use statistical techniques to estimate the latent semantic structure, and get rid of the obscuring “noise” [5].

Multi-keyword Ranked Search: It supports both multi-keyword query and support result ranking.

Privacy-Preserving: Our scheme is designed to meet the privacy requirement and prevent the cloud server from learning additional information from index and trapdoor.

1) Index Confidentiality. The TF values of keywords are stored in the index. Thus, the index stored in the cloud server needs to be encrypted;

2) Trapdoor Unlinkability. The cloud server should not be able to deduce relationship between trapdoors.

3) Keyword Privacy. The cloud server could not discern the keyword in query, index by analyzing the statistical information like term frequency.

X. EXPERIMENTAL SETUP

In this section, we show a thorough experimental evaluation of the proposed technique on a real dataset: the MED dataset.

F-measure that combines precision and recall is the harmonic mean of precision and recall[8]. Here, we adopt F-measure to weigh the result of our experiments.

$$F = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

For a clear comparison, our proposed scheme attains score higher than the original MRSE in F-measure. Since the original scheme employs exact match, it must miss some similar words which is similar with the keywords. However, our scheme can make up for this disadvantage, and retrieve the most relevant files. Fig .2 shows that our method achieves remarkable result.

DOI: 10.15680/IJIRCCE.2016. 0403245

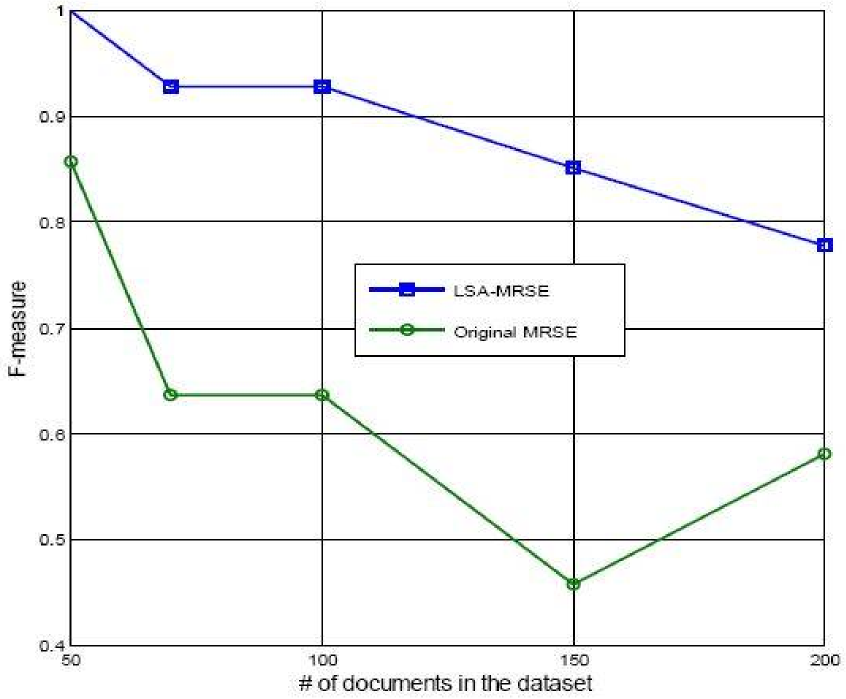

Fig. 2. Comparison of two schemes

EXSISTING AND PROPOSED COMPARISON-

In this section we present comparison result of Single Key word Search Ranked search and Multi Keyword Ranked Search over a Encrypted Data on Cloud as shown in following figures .In this Result Each Ranked search and Multi Keyword Ranked Search Over An Encrypted Data On Cloud. In this Result Existing System is Single Keyword Search System and proposed System is nothing but MRES System.

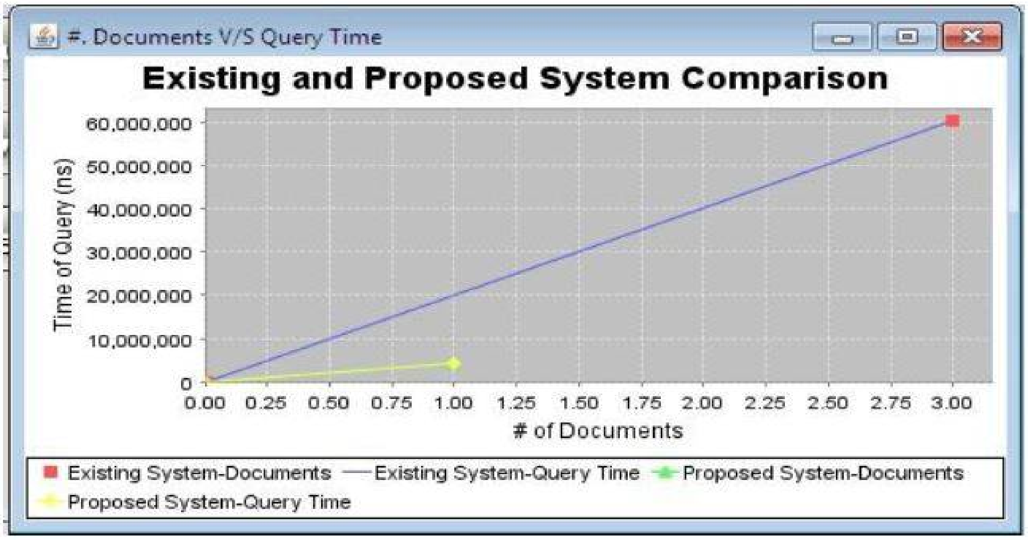
#. Documents V/S Query Time
Existing and Proposed System Comparison
60,000,000
50,000,000
40,000,000
30,000,000
20,000,000
10,000,000
0
Time of Query (ns)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
# of Documents
🟥 Existing System-Documents
Existing System-Query Time
Proposed System-Documents
Proposed System-Query Time

Fig 3: Comparison Graph- No. Of Documents V/S Query Time

It is a Comparison graph of Existing System and Our System. The graph is plotted Number of Documents that the respective system's search result returned and Time required to return the documents; in respective System. As shown in the graph Our system requires less time which is less than around 5 ns with most specific result of number of document equal to one which is less than the documents returned by existing system which are three and time required is around 6ns

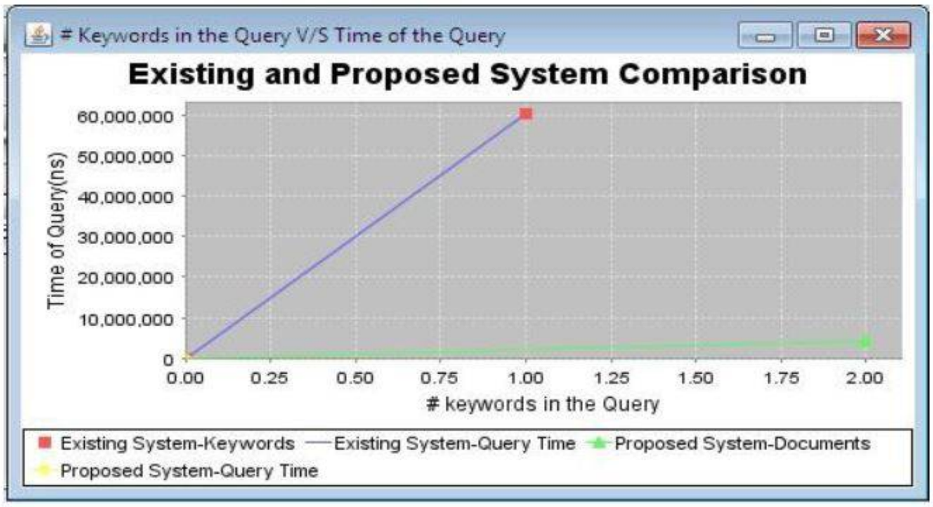

Fig 4: Comparison Graph- No. Of Documents V/S Query Time

It is a Comparison graph of Existing System and Our System. The graph is plotted Number of Documents that the respective system's search result returned and Time required to return the documents; in respective System. As shown in the graph Our system requires less time which is less than around 5 ns with most specific result of number of document equal to one which is less than the documents returned by existing system which are three and time required is around 6ns . Comparison Graph- No. Of Keywords V/S Query Time .It is a Comparison graph of Existing System and Our implemented System. The graph is plotted against Number of Keywords fired in the respective system's search and Time required in respective System. As shown in the graph Our system requires less time which is less than around 5 ns with multiple Keyword Query and existing system requires around 6ns even though a single Keyword query is fired. So Our System Works Better in each and every aspect then existing System.

XI. CONCLUSION

We characterize and take care of the issue of multi-catchphrase positioned seek over scrambled cloud information, and build up an assortment of protection necessities. Among different multi-watchword semantics, we pick the proficient likeness measure of "direction coordinating," i.e., however many matches as would be prudent, to adequately catch the significance of outsourced records to the question catchphrases, and use "internal item similitude" to quantitatively assess such comparability measure. For meeting the test of supporting multi watchword semantic without protection breaks, we propose a fundamental thought of MRSE utilizing secure internal item calculation. At that point, we give two enhanced MRSE plans to accomplish different stringent security necessities in two distinctive danger models. We likewise research some further improvements of our positioned look system, including supporting more hunt semantics, i.e., TF _IDF, and dynamic information operations. Intensive examination exploring security and effectiveness sureties of proposed plans is given, and tests on this present reality information set demonstrate our proposed plans present low overhead on both calculation and correspondence.

REFERENCES

[1] M.Armbrust, "A view of cloud computing",Communications of the ACM,vol.53, no. 4, (2010),pp. 50-58.
[2] D. Boneh, "Public keyencryption with keyword search",Advances in Cryptology-Eurocrypt 2004,Springer, (2004).

[3] R. Curtmola, "Searchable symmetric encryption: improved definitions and efficient constructions",Proceedings of the 13th ACM conference on Computer and communications security,ACM, (2006).

[4] D.X.Song,D. Wagner and A.Perrig,"Practical techniques for searches on encrypted data. in Security and Privacy", 2000. S&P 2000,Proceedings 2000 IEEE Symposium,IEEE, (2000).

[5] C. Wang, "Secure ranked keyword search over encrypted cloud data",Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference,IEEE, (2010).

[6] N. Cao,"Privacy-preserving multi-keyword ranked search over encrypted cloud data",INFOCOM, 2011 Proceedings IEEE,IEEE, (2011).

[7] M.Armbrust, "A view of cloud computing",Communications of the ACM,vol.53, no. 4, (2010),pp. 50-58.

[8] D. Boneh, "Public keyencryption with keyword search",Advances in Cryptology-Eurocrypt 2004,Springer, (2004).

[9] R. Curtmola, "Searchable symmetric encryption: improved definitions and efficient constructions",Proceedings of the 13th ACM conference on Computer and communications security,ACM, (2006).

[10] D.X.Song,D. Wagner and A.Perrig,"Practical techniques for searches on encrypted data. in Security and Privacy", 2000. S&P 2000,Proceedings 2000 IEEE Symposium,IEEE, (2000).

[11] C. Wang, "Secure ranked keyword search over encrypted cloud data",Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference,IEEE, (2010).

[12] N. Cao,"Privacy-preserving multi-keyword ranked search over encrypted cloud data",INFOCOM, 2011 Proceedings IEEE,IEEE, (2011).