**INTRODUCTION**

Cloud computing is a latest model for IT infrastructure. It provides high quality of applications and services on-demand from a shared pool of resources. Most of the time for making profit, unauthorized operations are conducted on outsourced data. To provide security and confidentiality over sensitive information and to block the unauthorized access, new technique is used. In this technique data must be encrypted before storing to cloud servers. This activity is carried out by data owner. But over such encrypted data, traditional approaches for its retrieval are useless. Traditional approach is based on plaintext query search.

To get proper query results in decrypted format, download all data and decrypt it at user side is also impractical. Because, every user wants only query related data instead of all data.

What’s more, getting the huge number of outsourced data and great deal of customers into consideration, it is also difficult to achieve performance and system usability.

Because of all these circumstances, it is very important thing to explore effective search techniques over encrypted outsourced data.

The proposed system concentrates over the encrypted cloud information as well as utilizes the tree-based search. System also supports to dynamic operation and multi keyword ranked search over list of the files. Commonly, the vector space model and term frequency (TF) inverse document frequency (IDF) model are combined for the query generation and index generation to provide multi keyword ranked search. Essential search proficiency is obtained by implementing an index structure that is tree-based recommendation a Greedy Depth-first Search algorithm based over the index tree. This system is flexibly obtains sub linear search time and allows the insertion as well as deletion of files due to specific structure of tree-based index. The query and index vectors are encrypted by implementing the secure KNN algorithm [1].Additionally, ensures the accuracy relevance score computation between the encrypted query and index vectors. We developed two secure search techniques: the fundamental dynamic multi-keyword ranked search (BDMRS) technique is developed for the popular cipher text model and the modified dynamic multi keyword ranked search (EDMRS) technique for the famous background model to avoid various attacks in different threat models.

This system consist three various entities: cloud server, data owner and data user. Data Owner has a list of files to be encrypted and stored over the cloud. User is allowed to search over this encrypted information. In this system, owner of data first generates the secure searchable index tree and after that creates encrypted file. This index tree and encrypted files are stored over the cloud server. Data owner has responsibility of key distribution to the authorized users that is needed for file decryption. On the basis of query request for specific document from user, cloud server executes searching over the index tree and the list of encrypted top k ranked outcomes is given to user. At the end, user can decrypt the obtained files through utilizing secret key provided by owner of data [2], [6].

Our system gives capability such as user revocation to the data owners. Data owner performs user revocation at the time when one of the users leaves the organization. Due to privacy system requires user revocation, user who leaves the organization with secret key that is threat for the privacy of data. So, privacy is maintained by the data owner through changing the key of all users.

**GOAL AND OBJECTIVE**

The goal of this project is to propose novel techniques which perform multi keyword ranked search on data which is encoded in cloud, these method concurrently supports active update actions like insert and delete the data on the cloud.

**The Objective of the proposed application is as follows:**

* System should achieve Precision and privacy.
* System should be efficient.
* System should be dynamic.
* Search efficiency.
* Produce top k-ranking results

**Literature Review**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Sr.No*** | ***Reference No.***  ***Concerned Author(s)& Year*** | ***Concept used*** | ***Performance Evaluation Parameter*** | ***Claims by Concerned Author(s)*** | ***Our Findings*** |
| *1.* | Yuzhe Tang and Ling Liu | Privacy preserving Indexes | Semantic Meanings | e-MPPI for providing the distributed document search along with quantitatively differentiated privacy preservation | An MPC-reduction technique based on the efficient use of secret sharing schemes. We also discovered common-term vulnerability and proposed a term-mixing solution. |
| *2.* | Ning Cao, Cong Wang, Ming Li, KuiRen, and Wenjing Lou | Privacy preserving multi-keyword ranked search over encrypted cloud data (MRSE) | Coordinate Matching, Inner Product  Similarity | 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. | MRSE using secure inner product computation. |
| *3.* | Y. Tang, L. Liu, A. Iyengar, K. Lee, and Q. Zhang | Anidentity-mixing protocol against the attack in e-PPI. | Effectiveness the e-PPI in terms of delivering quantitative privacy protection. Performance of our index construction protocol. | Proposed e-PPI construction protocol is the first without any trusted third party and/or trust relationships between providers. | The construction protocol for ǫ-PPI without any trusted party involved. |
| *4.* | Yuzhe Tang, Shuigeng Zhou | a Low maintenance Hash Tree, for efficient data indexing over DHTs. | maintenance cost, performance for exact-match queries | A Low maintenance Hash Tree, for efficient data indexing over DHTs. LHT employs a novel naming function and a tree summarization strategy to gracefully distribute its index structure. | LHT can save up to 75%(at least 50%) maintenance cost, and achieves better performancein exact-match and range query processing. |
| *5.* | Randy Baden, Adam Bender | Attribute-based encryption (ABE), Online social networks (OSNs) | Privacy in OSNs | Persona provides an effective means of creating applications in whichusers, not the OSN, define policy over access to private data. | Persona hides user data with attribute-based encryption (ABE), allowing users to apply fine-grained policies over who may view their data. |
| *6.* | K.S.Sureh, Mrs.SaritaChowdary, T. Balachary | private-key cryptography and symmetric Encryption | Privacy & Security | The personal health record system needs security against attackers and hackers. Scalable and Secure sharing includes basic securities to protect the information from unauthorized access and loss. | Paper proposed the new approach for existing PHR system for providing more security using symmetric encryption which plays an important role because these are unique. |
| *7.* | Y. Tang, T. Wang, and L. Liu | SS-PPI, a novel privacy-preserving index abstraction | Privacy Protection and Execution Efficiency | Focus is on addressing the privacy concerns of contentproviders; that is, the search should not reveal the specific associationbetween contents and providers | It incorporates access control policies in the privacypreserving index, which improves both search efficiency and attack resilience; |
| *8.* | M. Bawa, R. J. Bayardo, Jr, R. Agrawal, and J. Vaidya | Privacy-preserving Index a distributed access-control enforcing search protocol | Privacy-preserving index (PPI) | The new index provides strong and quantifiable privacy guarantees that hold even if the entire index is madepublic. | Content providers maintain complete control in defining access groups and ensuring its compliance |
| *9.* | A. Ben-David, N. Nisan, and B. Pinkas | Secure Multi-Party Computation | Number of computation players, Size of the circuit and General run time | The BMR protocol is modified in a novel way and considerably improved its performance by using the Ben-Or-Goldwasser-Wigderson (BGW) protocol for the purpose of constructing gate tables. | The performance of fast machines is dramatically reduced if even a single player is using a weak machine. The reason for this is that in every communication round the fast players have to wait until the weakest player finishes its computation and sends its results. |
| *10.* | S. Zerr, E. Demidova, D. Olmedilla, W. Nejdl, M. Winslett, and S. Mitra | r-confidential Zerber indexing | Response Size for the DFM Index, Efficiency in Query Answering, | A tunable r-confidentiality measure, as the degree of information from inaccessible documents an index can leak, given an adversary compromises the index and possesses some background knowledge on the corpus and/or language statistics. | Zerber, an r-confidential global inverted index for sensitive documents. Zerber relies on a centralized set of largely untrusted index servers and offers resistance against inappropriate information disclosure even if k-1 index servers are compromised. |

**PROBLEM DEFINITION**

* Maintainability, accessibility are the functions of cloud, also the main motivation is the privacy attract towards the distinct organization for storing their data on the cloud Server.
* Sensitive data is encrypted before putting on cloud to maintain the confidentiality. To retrieve multi keyword query and ranking result, we use tree based search scheme.
* To sustain the thousands of records encoded keys of the multiple users and recovering the sensitive document when the client is revoked is one of the challenging tasks for the organization

**RESEARCH METHODOLOGY**

Cloud data owners prefer to outsource documents in an encrypted form for the purpose of privacy preserving. Therefore it is essential to develop efficient and reliable ciphertext search techniques. One challenge is that the relationship between documents will be normally concealed in the process of encryption, which will lead to significant search accuracy performance degradation. Also the volume of data in data centers has experienced a dramatic growth. This will make it even more challenging to design ciphertext search schemes that can provide efficient and reliable online information retrieval on large volume of encrypted data.

The next generation web called the Semantic Web will help the user to retrieve the useful data that is stored on the cloud in the form of ontology and make the data visible to the user which is hidden behind the cloud. The aim of the proposed ranking algorithm is to provide users the result set of relevant data.



**System Architecture**

A hierarchical clustering method is proposed to support more search semantics and also to meet the demand for fast cipher text search within a big data environment. The proposed hierarchical approach clusters the documents based on the minimum relevance threshold, and then partitions the resulting clusters into sub-clusters until the constraint on the maximum size of cluster is reached. In the search phase, this approach can reach a linear computational complexity against an exponential size increase of document collection. In order to verify the authenticity of search results, a structure called minimum hash sub-tree is designed. We extend this notion of semantic similarity to consider inherent relationships between concepts using ontologies. We propose ranking algorithm with multi keyword and ontology.

**Expected Outcome:**

This scheme introduced the method which is more protected, proficient and dynamic approaches are used. System provides the accurate multi keyword ranked search, the dynamic insertion, deletion of documents and the user revocation. In user revocation, data owner modify the key of all users when any one of them leave organization. It increases the security of organization.

**TECHNOLOGY AND ASSOCIATED PLATFORM**

**Hardware Specification**

Hard Disk: 80 GB

RAM: 512 MB

Processor: Intel Pentium 4 and above

**Software Specification**

Technology Used: Core Java.

Tools: JDK 1.5 or above, Netbeans

Operating System: Windows XP or above

**CONCLUSION**

This paper studies various techniques of searching in the encrypted cloud data storage. We have systematically presents the security and data utilization issues in the cloud storage related to all available searching techniques. Thus identified the main issues that are to be satisfied for secured data utilization are keyword privacy, Data privacy, Index privacy, Query Privacy, Fine-grained Search, Scalability, Efficiency, Result ranking, Index confidentiality, Query confidentiality, Query Unlinkability, semantic security and Trapdoor Unlinkability. Most of the searching techniques mainly focus on security and some on data utilization.

**ACTIVITY PLAN**

* Phase 1:
  + Literature Survey
  + Analysis of literature data
  + Collection and Study of relevant data set
* Phase 2:
  + UI Design and File Selection, Encryption & Upload Process
  + Providing TF-IDF Value
* Phase 3:
  + Multi-Keyword Search Result in Ranked Order
  + Key Distribution Center (KDC) for allowing decryption of data received from private server, at client side.
  + Analysis and Interpret the output of algorithm
* Conclusion
* Preparation of Project Report

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Activity** | **Sept-Oct 2016** | **Nov-Dec 2016** | **Jan-Feb 2016** | **Mar-Apr 2016** | **May 2016** |
| Literature Survey |  |  |  |  |  |
| Critical analysis and comparison of technologies studied and results achieved in research. |  |  |  |  |  |
| Modelling and design and dataset searching or creation |  |  |  |  |  |
| Implementation of module 1  UI Design and File Selection, Encryption & Upload Process |  |  |  |  |  |
| Implementation of module 2  Providing TF-IDF Value |  |  |  |  |  |
| Implementation of module 3  Multi-Keyword Search Result in Ranked Order |  |  |  |  |  |
| Implementation of module 3  Key Distribution Center (KDC) for allowing decryption of data received from private server, at client side. |  |  |  |  |  |
| Conclusions |  |  |  |  |  |
| Preparation of Project Report |  |  |  |  |  |

**BIBLIOGRAPHY**

[1]Yuzhe Tang and Ling Liu, Fellow, IEEE, “Privacy-Preserving Multi-Keyword Search

in Information Networks, ”*IEEE Transactions On Knowledge And Data Engineering, Volume 27, Issue 9, 2015*

[2] Ning Cao, Cong Wang, Ming Li, KuiRen, and Wenjing Lou, “Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data”,*Proc. IEEE Infocom, Volume 3, Issue 8, 2014*

[3]YuzheTang ,Ling Liu , ArunIyengar, Kisung Lee, Qi Zhang, “E-PPI: Locator Service in Information Networks with Personalized Privacy Preservation”, *IEEE Transactions On Knowledge And Data Engineering, Volume 7,Issue 6, 2015*

[4] Yuzhe Tang and Shuigeng Zhou, “LHT: A Low-Maintenance Indexing Scheme over DHTs”, *The 28th International Conference on Distributed Computing Systems, Volume 10, Issue 3, 2008*

[5] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, “Persona: An Online Social Network with User-Defined Privacy”,*ACM SIGCOMM, Volume 9, Issue 6, 2009*

[6]K.S.Sureh, Mrs. Sarita Chowdary, T. Balachary. “ A Cloud Based System for Patient Health Records Using Symmetric Encryption”, *International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11, 2013*

[7] Yuzhe Tang,Ting Wang, Ling Liu, ShicongMeng, and BalajiPalanisamy. “Privacy-Preserving Indexing for eHealth Information Networks”,*ACMCIKM’11, Volume 2, Issue4, 2011*

[8] MayankBawa, Roberto J. Bayardo Jr., Rakesh Agrawal, “Privacy-Preserving Indexing of Documents on the Network”,*Proceedings of the 29th VLDB Conference, Volume 4, Issue 6, 2003*

[9]Assaf Ben-David, Noam Nisan, Benny Pinkas, “FairplayMP – A System for Secure Multi-Party Computation”,*ACM CCS, Volume 7, Issue 5, 2008*

[10] SergejZerr, Elena Demidova, Daniel Olmedilla, Wolfgang Nejdl, Marianne Winslett2 and SoumyadebMitra, “Zerber: r-Confidential Indexing for Distributed Documents”,*ACM EDBT, Volume 8, Issue 6,2008*

[11] Ming Li Member, IEEE, Shucheng Yu, Member, IEEE, Yao Zheng, Student Member, IEEE, KuiRen, Senior Member, IEEE, and Wenjing Lou, Senior Member, IEEE, “Scalable and Secure Sharing of Personal Health Records in Cloud Computing using”,*IEEE Transactions On Parallel And Distributed Systems Volume 24, Issue 1,2013*

[12] Preethi Mathew, Dr. S. SasidharBabu , “Secure Fuzzy Multi-Keyword Ranked Search

over Encrypted Cloud Data”,*International Journal of Innovative Research in Computer and Communication Engineering, Volume 3, Issue 8, 2015*

[13] C. Gentry, “Fully homomorphic encryption using ideal lattices,” *41st Annal. ACM Symp. Theory Computer., Volume 10 ,Issue 6,2009*

[14] A. Ben-David, N. Nisan, and B. Pinkas, “Fairplaymp: A system forsecure multi-party computation,” *ACM Conf. Computer Commuication Security, Volume 6,Issue 9,2008*

[15] W. Henecka, S. Kogl, A.-R.Sadeghi, T. Schneider, and I. Wehren- €berg, “TASTY: Tool for automating secure two-partycomputations*,” ACM Confernce Computer Communication Security, Volume 7, Isssue 3 ,2010*

[16] I. Damgard, M. Geisler, M. Krøigaard, and J. B. Nielsen,“Asynchronous multiparty computation: Theory andimplementation,” *International Conference Practice Theory Public Key Cryptography, Volume 5, Issue 4,2009*

[17] M. Kantarcioglu and C. Clifton, “Privacy-preserving distributed mining of association rules on horizontally partitioned data,” *IEEE Transaction Knowledge Data Engineering., Volume 16, Issue 9, 2015*

[18] Y. Tang, J. Xu, S. Zhou, and W. Lee, “m-LIGHT: Indexing multi-dimensional data over DHTs,” *IEEE International Confernce Distributed Computing System., Volume 7,Issue 2,2009*

[19 ] Y. Tang, S. Zhou, and J. Xu, “LIGHT: A query-efficient yet low-maintenance indexing scheme over DHTs,” *IEEE Transcation Knowledge Data Engineering., volume 22, Issue 1, 2010.*

[20] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubrama-niam, “L-diversity: Privacy beyond k-anonymity,” *International Confernce Data Engineering, Volume 2,Issue 5,2006*

[21] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibratingnoise to sensitivity in private data analysis,” *Confernce Theory Cryptography, Volume 2,Issue 6,2006*

[22] M. K. Reiter and A. D. Rubin, “Crowds: Anonymity for web trans-actions,” *ACM Transcation Information System Security, Volume 1, Issue 1,1998.*

[23]B. Bloom, “Space/time trade in hash coding with allowable errors”, *Communications of ACM, Volume 2,Issue 7,1970.*

[24] R. J. Bayardo Jr., A. Somani, D. Gruhl, and R. Agrawal.Youserv , “A web hosting and content sharing tool for the masses”,*Conference onWorld Wide Web (WWW), Volume 5,Issue 7, 2002*

[25] M. K. Reiter and A. D. Rubin. Crowds:, “Anonymity for Web transactions”, *ACM Transactions on Information and System Security, Volume 5,Issue 3, 1998.*