PRESERVING PRIVACY OF MULTIPLE DATA IN A CLOUD OVER A SECURED NETWORK

[1] Ms. Alpha Vijayan, [2] Ms. Josmy George
[1] Senior Assistant Professor, [2] Assistant Professor
[1] Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

Abstract : In cloud computing, to protect privacy when outsourcing database management system, confidential data has to be encryptedwhich however renders traditional query processing methods inapplicable. First the confidentiality requires hiding the value and the relative order of the attribute in records from the cloud server, but computing a range query requires comparing the values of this attribute. It defines and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing Establish a set of strict privacy requirements for such a secure cloud data utilization system.Among various multi-keyword semantics, choose the efficient similarity measure of coordinate matching.This further use inner product similarity to quantitatively evaluate such similarity measure. First propose a basic idea for the multi-keyword ranked search over encrypted cloud data schemes based on secure inner product computation, and then give two significantly improved multi-keyword ranked search over encrypted cloud data schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, further extend these two schemes to support more search semantics.

IndexTerms **- Ball tree index ,coordinate matching, CloudStack, k-nearest neighbour, infrastructure convergence, service level agreements.**

I. INTRODUCTION

Cloud computing is no doubt an effective approach to deal with big data, through providing on-demand high quality services from powerful and configurable computing resources. One application of cloud computing is database as a service, where the data owner outsources complex database management systems into the cloud server. To protect data privacy against attacks from the cloud server, confidential data must be encrypted before being uploaded to the cloud server. This, however, makes it difficult to perform traditional query processing operations. This problem faces two challenges. The code is then followed by two threads that will run in parallel. Dirty COW exploit is a race condition vulnerability; this means that certain events have to occur in a specific order that are fairly unlikely to happen under normal circumstances. This attack that exploits the race condition and escalates local user privilege. Infrastructure as a Service (IaaSwe should utilize not only a cloud management module but also a network management module. However, it is difficult to check the duration time and to observe the digested information about the resources. To investigate these problems in a cloud computing environment, a cloud service infrastructure based on open-source software, namely, CloudStack is designed and deployed. This model regularly stores the usage data for computing resources based on Hadoop and HBase. In addition, the model analyses the raw data for virtual machines and makes an effective recommendation regarding the consumption of computing resources. This linear scan is not acceptable for large databases.

Ball tree index is adopted to retrieve such points while pruning as many data points as possible. Briefly, a ball tree is a binary tree such that each non-leaf node represents a ball and has two child nodes. A data point belonging to a parent goes to the child ball whose center is closer to the data point. All data points are only stored at the leaf nodes. To build such a ball tree, we could keep separating the data point space into two partitions (left and right child) recursively until the number of data points in some partition is below a predefined threshold and we make this partition as a leaf node. This threshold is called as max leaf size .

To meet the effective data retrieval need, the large amount of documents demand the cloud server to perform result. relevance ranking, instead of returning undifferentiated results. Such ranked search system enables data users to find the most relevant information quickly, rather than burdensomely sorting through every match in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation, however, should not leak any keyword related information. On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse results. As a common practice indicated by today's web search engines, data users may tend to provide a set of keywords instead of only one as the indicator of their search interest to retrieve the most relevant data. And each keyword in the search request is able to help narrow down the search result further. "Coordinate matching", i.e., as many matches as possible, is an efficient similarity measure among such multi-keyword semantics to refine the result relevance, and has been widely used in the plaintext information retrieval (IR) community. However, how to apply it in the encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirements like the data privacy, the index privacy, the keyword

privacy, and many others. The data owner could easily employ the traditional symmetric key cryptography to encrypt and then outsource data.

II. MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA

Encryption is a perfect way to protect data privacy which also gives rise to the difficulty in executing regular query on such encrypted data. To keep the advantage of computing ability on the server side, it is not reasonable to download the entire encrypted data and execute the query on the data after decryption, the cloud server has to have the ability to execute query directly over the encrypted data and return the correct results to the user. There have been considerable interests in querying encrypted data and various queries are considered. Briefly, the works include equality test, range query search, aggregation query computation as well as keyword based query search or similarity query search.

To enrich search functionalities, conjunctive keyword search over encrypted data have been proposed. These schemes incur large overhead caused by their fundamental primitives, such as computation cost by bilinear map, for example, or communication cost by secret sharing, for example, as a more general search approach, predicate encryption schemes are recently proposed to support both conjunctive and disjunctive search. Conjunctive keyword search returns all-or-nothing," which means it only returns those documents in which all the keywords specified by the search query appear; disjunctive keyword search returns undifferentiated results, which means it returns every document that contains a subset of the specific keywords, even only one keyword of interest. In short, none of existing Boolean keyword searchable encryption schemes support multiple keywords ranked search over encrypted cloud data while preserving privacy as we propose to explore in this paper.

Two new schemes have been proposed to support more search semantics which improve the search experience of the multi-keyword ranked search over encrypted cloud data (MRSE) scheme, and also study the dynamic operation on the data set and index which addresses some important yet practical considerations for the MRSE design. On a different front, the research on top-k retrieval in database community is also loosely connected to our problem. Besides, Cao et. al. proposed a privacy-preserving graph containment query scheme which solves the search problem with graph semantics.

With focus on the index and query, the MRSE (Multi-keyword Ranked Search over Encrypted cloud data) system consists of four algorithms as follows: Taking a security parameter ‘ as input, the data owner outputs a symmetric key as SK. . BuildIndex ðF; SKÞ. Based on the data set F, the data owner builds a searchable index I which is encrypted by the symmetric key SK and then outsourced to the cloud server. After the index construction, the document collection can be independently encrypted and outsourced. . Trapdoor ðfWÞ. With t keywords of interest infW as input, this algorithm generates a corresponding trapdoor Tew. Query ðTeW ; k; IÞ. When the cloud server receives a query request as (TeW , k), it performs the ranked search on the index I with the help of trapdoor TeW , and finally returns FeW, the ranked id list of top-k documents sorted by their similarity with fW. Neither the search control nor the access control is within the scope of this paper. While the former is to regulate how authorized users acquire trapdoors, the later is to manage users' access to outsourced documents.

To solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict systemwise privacy in the cloud computing paradigm, we should utilize not only a cloud management module but also a network management module. However, it is difficult to check the duration time and to observe the digested information about the resources. To investigate these problems in a cloud computing environment, we designed and deployed the cloud service infrastructure based on open-source software, namely, CloudStack. The proposed model regularly stores the usage data for computing resources based on Hadoop and HBase. In addition, the model analyses the raw data for virtual machines and makes an effective recommendation regarding the consumption of computing resources. 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 multikeyword 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 in two threat models with increased attack capabilities.

2.1 Advantages

- Its great flexibility and industrial savings are prompting both individuals and enterprises to source their local complicated data administrated system into the cloud.
- MRSE schemes to achieve different stringent privacy requisites in two various threat designs.
- Service provider did not know about the original content in the data owner that is the encrypted data most secure.
- Standard deviation is normal to be smaller so as to earn high rigor indicating the good purity of reclaimed documents.

III SYSTEM ARCHITECHTURE

The system architecture and data flow are given above in figures1 and 2 respectively.

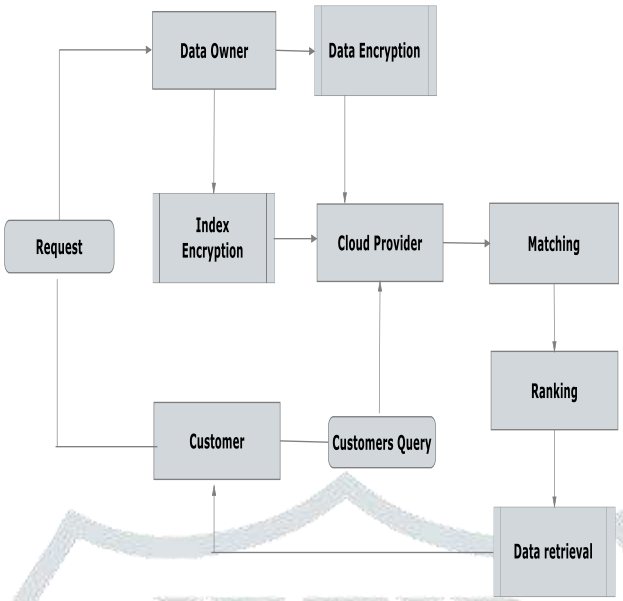


Fig .1: System architecture

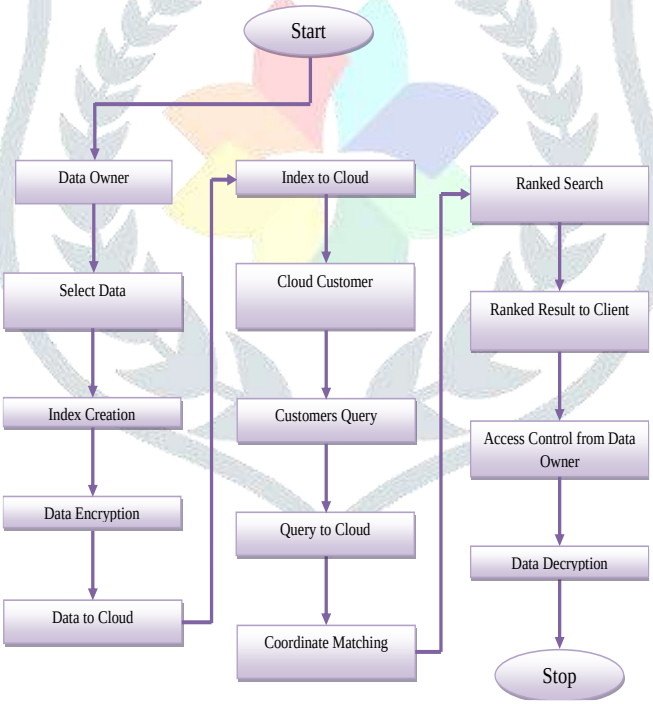


Fig .2: Data flow

IV CLOUD COMPUTING TECHNOLOGY

Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility (like the electricity grid) over a network. Cloud computing provides computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services. Parallels to this concept can be drawn with the electricity grid, wherein end-users consume power without needing to understand the component devices or infrastructure required to provide the service.

Cloud computing is different from hosting services and assets at ISP data center. It is all about computing systems are logically at one place or virtual resources forming a Cloud and user community accessing with intranet or Internet. So, it means Cloud could reside in-premises or off-premises at service provider location. There are types of Cloud computing like 1. Public Clouds2. Private Clouds 3. Inter-clouds or Hybrid Clouds, say Mr.B.L.V. Rao- CIO and IT Leaders and expert in cloud computing.

Cloud computing describes a new supplement, consumption, and delivery model for IT services based on Internet protocols, and it typically involves provisioning of dynamically scalable and often virtualized resources. It is a byproduct and consequence of the ease-of-access to remote computing sites provided by the Internet. This may take the form of web-based tools or applications that users can access and use through a web browser as if the programs were installed locally on their own computers.

Cloud computing providers deliver applications via the internet, which are accessed from browsers, desktop and mobile apps, while the business software and data are stored on servers at a remote location. In some cases, legacy applications (line of business applications that until now have been prevalent in thin client Windows computing) are delivered via a screen-sharing technology, while the computing resources are consolidated at a remote data center location; in other cases, entire business applications have been coded using web-based technologies such as AJAX.

At the foundation of cloud computing is the broader concept of infrastructure convergence (or Converged Infrastructure) and shared services. This type of data center environment allows enterprises to get their applications up and running faster, with easier manageability and less maintenance, and enables IT to more rapidly adjust IT resources (such as servers, storage and networking) to meet fluctuating and unpredictable business demand.

Most cloud computing infrastructures consist of services delivered through shared data-centers and appearing as a single point of access for consumers' computing needs. Commercial offerings may be required to meet service-level agreements (SLAs), but specific terms are less often negotiated by smaller companies.

4.1 Cloud Working Progress

Most websites and server-based applications run on particular computers or servers. What differentiates the cloud from the way those are set up is that the cloud utilizes the resources from the computers as a ***collective virtual computer***, where the applications can run independently from particular computer or server configurations. They are basically floating around in a "cloud of resources", making the hardware less important to how the applications work.

With broadband internet, the need to have the software run on your computer or on a company's site is become less and less essential. A lot of the software that people use nowadays are completely web-based. The cloud takes advantage of that to bring it to the next level. To understand how does cloud computing work, imagine that the cloud consists of layers — mostly the **back-end** layers and the **front-end** or user-end layers. The front-end layers are the ones you see and interact with. When you access your email on Gmail for example, you are using software running on the front-end of a cloud. The same is true when you access your Facebook account. The back-end consists of the hardware and the software architecture that fuels the interface you see on the front end.

Because the computers are set up to work together, the applications can take advantage of all that computing power as if they were running on one particular machine. Cloud computing also allows for a lot of flexibility. Depending on the demand, you can increase how much of the cloud resources you use without the need for assigning specific hardware for the job, or just reduce the amount of resources assigned to you when they are not necessary. The transition from being very 'personal hardware dependent' to a world where resources are shared among the masses is creeping up on us slowly and unobtrusively. Very many people have already transitioned to using a cloud environment for most of their time in front of the computer without even realizing it.

4.2 Public cloud

Public cloud describes cloud computing in the traditional mainstream sense, whereby resources are dynamically provisioned to the general public on a fine-grained, self-service basis over the Internet, via web applications/web services, from an off-site third-party provider who bills on a fine-grained utility computing basis.

4.3 Community cloud

Community cloud shares infrastructure between several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally. The costs are spread over fewer users than a public cloud (but more than a private cloud), so only some of the benefits of cloud computing are realized.

4.4 Hybrid cloud

Hybrid cloud is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models. Briefly it can also be defined as a multiple cloud systems which are connected in a way that allows programs and data to be moved easily from one deployment system to another.

4.5 Private cloud

Private cloud is infrastructure operated solely for a single organization, whether managed internally or by a third-party and hosted internally or externally. They have attracted criticism because users "still have to buy, build, and manage them" and thus do not benefit from lower up-front capital costs and less hands-on management, essentially "[lacking] the economic model that makes cloud computing such an intriguing concept".

As cloud computing is achieving increased popularity, concerns are being voiced about the security issues introduced through adoption of this new model. The effectiveness and efficiency of traditional protection mechanisms are being reconsidered as the characteristics of this innovative deployment model differ widely from those of traditional architectures.

We should utilize not only a cloud management module but also a network management module. However, it is difficult to check the duration time and to observe the digested information about the resources. To investigate these problems in a cloud computing environment, we designed and deployed the cloud service infrastructure based on open-source software, namely, CloudStack. The proposed model regularly stores the usage data for computing resources based on Hadoop and HBase. In addition, our model analyzes the raw data for virtual machines and makes an effective recommendation regarding the consumption of computing resources. Some of the key characteristics of cloud computing are:

- **On-demand Self Service**

A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.

- **Broad Network Access**

Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms.

- **Resource Pooling**

The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or data center). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

- **Measured Service**

Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service. Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service.

- **Selection of Provider**

A good service provider is the key to good service. So, it is imperative to select the right service provider. One must make sure that the provider is reliable, well-reputed for their customer service and should have a proven track record in IT- related ventures. As cloud computing, has taken hold, there are six major benefits that have become clear, 1) Anywhere/anytime access - It promises "universal" access to high-powered computing and storage resources for anyone with a network access device. 2) Collaboration among users -cloud represents an environment in which users can develop software based services and from which they can deliver them. 3) Storage as a universal service - the cloud represents a remote but scalable storage resource for users anywhere and everywhere. 4) Cost benefits - the cloud promises to deliver computing power and services at a lower cost.

V MODULES

5.1 Binary Data Generation

Data owner select the data and create the bit vector for that data. Using that bit vector of the data the binary data is generated. The binary data is the index for the data in the data owner. The bit vector is the bytes form of the data in the data owner. The bit

vector is converted into the binary data. These bit vector and the binary data are ready for the data ciphering. Since it was a Linux kernel based bug, anyone running the Linux kernel below 4.8.3 were susceptible to the attacks. The only way to prevent the attack is to patch the current version of kernel or to run a newer Linux with the updated kernel which is not vulnerable anymore.

Users can utilize SystemTap as a mitigation technique. It is a tool for dynamically instrumenting Linux kernel based Operating Systems. System Administrators can use SystemTap to extract, filter and summarize data in order to diagnose performance or functional problems. Antivirus software has the potential to detect elevating permission attacks but cannot completely prevent the attack. Hence the best prevention method is to upgrade the Linux kernel.

5.2 Data Ciphering

In this model, the cloud server is supposed to only know encrypted data set C and searchable index I, both of which are outsourced from the data owner. Known background model. In this stronger model, the cloud server is supposed to possess more knowledge than what can be accessed in the known ciphertext model. Such information may include the correlation relationship of given search requests (trapdoors), as well as the data set related statistical information. As an instance of possible attacks in this case, the cloud server could use the known trapdoor information combined with document/keyword frequency to deduce/identify certain keywords in the query.

Then the data owner have to encrypt the original data and send it to server. And then encrypt the binary data or the index and send it to server. Service provider did not know about the original content in the data owner. These index are used to refer the data in the service provider. It gives more security in the server side, so that the attackers can't use the data. Our system must prevent Server from learning any additional correspondence between plaintext values and ciphertext values except those obtained by prior knowledge. That is, we must protect the plaintext values for any encrypted records or queries from being disclosed to Server.

5.3 Data User Access Control

The program was able to successfully overwrite the contents of the file by escalating the privilege of normal user. Users is a group of users authorized to post a query and receive the answers. Owner encrypts the data and then uploads it to the Server. Proxy serves a bridge between Users and Server.

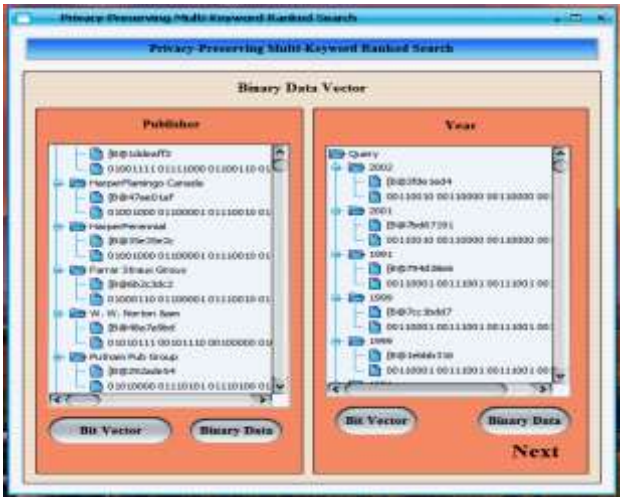
5.4 Data User Query

The data user query is processed by the service provider. The service provider generates the bit vector for the query from the client. Then the service provider converts the bit vector into binary data. Service provider finds the similar data from the index. And send the encrypted data to the data owner. Then the client decrypts the received data by the key from the data owner**.** To design search schemes which allow multi-keyword query and provide result similarity ranking for effective data retrieval, instead of returning undifferentiated results.

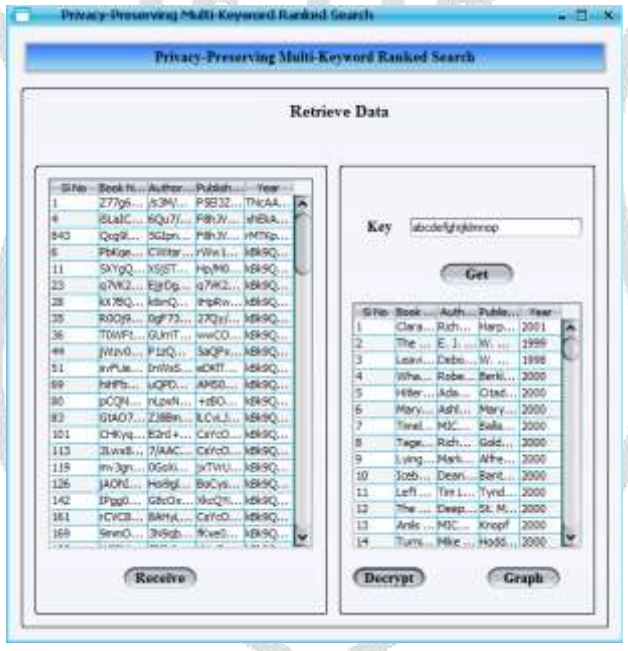
A query from Users will go though the trusted Proxy, which encrypts the query and submits the query to Server. Server computes and returns the answer to Proxy. Proxy then decrypts the answer, and returns the answer to Users. For example, Owner is a hospital, who outsources patient records to the cloud, and Users are various medical research labs, who post queries to retrieve patient records of interests.

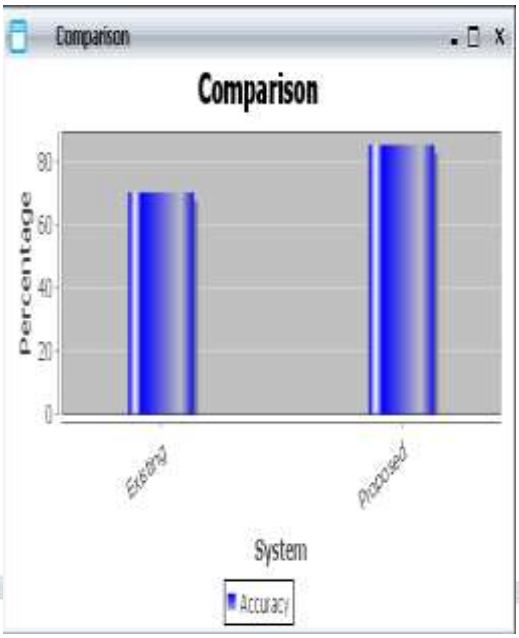
VI RESULTS AND DISCUSSIONS

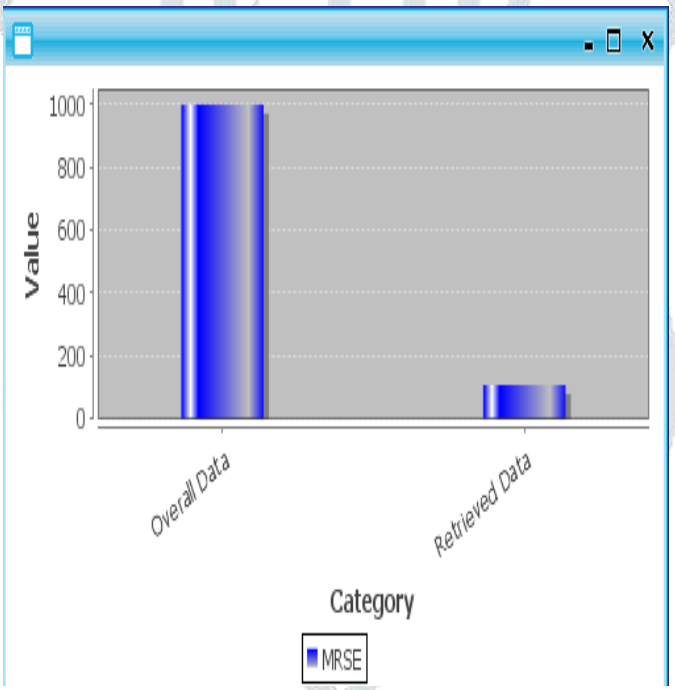
The implementation of the project gives the following results and output.




Privacy-Preserving Multi-Keyword Ranked Search
Privacy-Preserving Multi-Keyword Ranked Search
Inner Product Similarity
Index
Data
Index Searching
Data Searching
Send


Privacy-Preserving Multi-Keyword Ranked Search
Privacy-Preserving Multi-Keyword Ranked Search
Retrieve Data
Key
Get
Receive
Decrypt
Graph





VII CONCLUSION AND FUTURE SCOPE

A major challenge or concern in the cloud computing paradigm is data privacy. On one hand, there is a demand to leverage the powerful resources of the cloud server to provide services to clients. On the other hand, the cloud server must not learn any sensitive information about the data being managed and the queries being answered. We consider the service of answering the class of range query search over numerical data and propose a data encryption scheme to address these requirements. 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 data set show our proposed schemes introduce low overhead on both computation and communication.

REFERENCES

[1] Lei Yang; Qingji Zheng; Xinxin Fan, "RSPP: A reliable, Searchable and Privacy-preserving e-healthcare system for Cloud assisted body area networks", IEEE Infocom 2017-IEEE Conference on Computer Communications, pp.1-9, Oct 2017.
[2] Shivam Gupta, Subhas C. Misra, "Moderating Effect of Compliance, Network and Security on the Critical Success Factors in the Implementation of Cloud ERP", IEEE Transactions on Cloud Computing , pp. 440 – 451, Volume: 4, Issue: 4, Oct.-Dec. 1 2016.
[3] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, "Toward Secure and Dependable Storage Services in Cloud Computing," IEEE Trans. Services Computing, vol. 5, no. 2, pp. 220-232, Apr.-June 2012.
[4] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 8, pp. 1467- 1479, Aug. 2012.
[5] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM, pp. 693-701, 2012.
[6] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing," Proc. 31st Int'l Conf. Distributed Computing Systems (ICDCS '10), pp. 383- 392, June 2011.
[7] N. Cao, Z. Yang, C. Wang, K. Ren, and W. Lou, "Privacypreserving Query over Encrypted Graph-Structured Data in Cloud Computing," Proc. Distributed Computing Systems (ICDCS), pp. 393-402, June, 2011.
[8] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. 829-837, Apr, 2011.
[9] 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 '10), 2010.
[10] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing," Proc. IEEE INFOCOM, 2010.
[11] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," Proc. IEEE INFOCOM, 2010.
[12] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption," Proc. 29th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT '10), 2010.
[13] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing," Proc. IEEE INFOCOM, Mar. 2010.
[14] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.
[15] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k Retrieval from a Confidential Index," Proc. 12th Int'l Conf. Extending Database Technology (EDBT '09), pp. 439-449, 2009.