ISSN: 2321-7782 (Online)
Volume 3, Issue 6, June 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Secure Multi Keyword Ranked Search for Cloud Data

P. Suresh[1]
Research Scholar,
Department of Computer Science,
H.H. The Rajahs College (Autonomous),
India

R. Malathi[2]
Assistant Professor,
Department of Computer Science,
H.H. The Rajahs College (Autonomous),
India

Abstract: ***Cloud computing is store and retrieve on data. To access the cloud data we must have internet connection .In this paper, for the first time data owner, commercial public stored data in cloud. Cloud data is full privacy and efficiency. This is cloud management provider to maintenance, multi-keyword ranked search encrypted cloud data. MRSE keyword is to implement to access measurement of different person "coordinate matching" with help on similarity measurement techniques.***

Keywords: Cloud, Encryption, Similarity-Measurement, Third Party Auditor.

I. INTRODUCTION

Cloud computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store their data in to the cloud so as to enjoy the on-demand high quality applications and service from a shard pool of configurable computing resources. cloud computing is type of store data .this is type of computing that relies on sharing computing resources rather than having local server or personal devices to handle. They are 3 type of store-age data in cloud. cloud computing provides the way to share distributed resource and services. They are provide PaaS, SaaS, IaaS service . In this method some important security services including authentication, encryption and decryption compression are provided in cloud computing system. The existing system user to audit the cloud storage for very high communication and cost.

In proposed system, in this paper very efficient method to follow service provider .so managed access any user in data. MRSE method to implement in this paper. This module is used to help the user to get accurate result based on the multiple keyword concepts. The users can enter the multiple words query, the server is going to split that query into a single word after search that word file in our database.

Finally, display the matched word list from the database and the user gets the file from that list. Cloud service provider that offers customers storage or service available via private (or) public network cloud. Among different multi key word semantics the choose the efficient principle of "coordinate matching" by inner product similarity.

A similarity measure is a function which computes the degree of similarity between a pair of vector or documents. There is a large number of similarity measures proposed in the literature, because the best similarity measure doesn't exist. so in this paper very efficient MRSE query model effective data retrieval.

Section 4. Section 5 explains the proposed work results and accuracy rate of the Tamil character recognition. The final section provides the scope of the research and conclusion.

II. RELATED WORK

Ning Cao, Member et al., define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). 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. We further use "inner product similarity" to quantitatively evaluate such similarity measure. 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. [1]

Jyoti et al Multi-process or multi-threaded models can rarely cope with thousands of connections because of memory limits, system scheduler limits, and lock contention everywhere. Event-driven models do not have these problems because implementing all the tasks in user-space allows a finer resource and time management. The down side is that those programs generally don't scale well on multi-processor systems. That's the reason why they must be optimized to get the most work done from every CPU cycle. On security aspects Encryption alone for Cloud privacy. Their classification hierarchy of Cloud Computing is not standard model and has few shortcomings, as we would discuss duly. [2]

The previous work [1] mainly focused on providing privacy to the data on cloud in which using multi-keyword ranked search was provided over encrypted cloud data using efficient similarity measure of co-ordinate matching. The previous work [4] also proposed a basic idea of MRSE using secure inner product computation. There was a need to provide more real privacy which this paper presents. In this system, stringent privacy is provided by assigning the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user in order to protect the user's data on cloud from the CSP and the third party user. Thus, by hiding the user's identity, the confidentiality of user's data is maintained.[3]

Ruturaj Desai et al propose new scheme to solve the problem of multi keyword search over encrypted data using trusted third party in cloud computing. User will encrypt their data locally. Before encrypting data, the index will be created. Trusted third party will use all these indexes to search data similar to the search query of user. Using these search results, cloud server will send encrypted document to the user. he previous work mainly focused on providing privacy to the data on cloud in which using multi -keyword ranked each was provided over encrypted cloud data using efficient similarity measure of co-ordinate matching. [4]

Ankatha Samuyelu Raja et al this paper focuses on multi keyword search based on ranking over an encrypted cloud data (MRSE). The search uses the feature of similarity and inner product similarity matching. The experimental results show that the overhead in computation and communication are considerably low. [5].

III. PROPOSED WORK

Algorithm Steps

The clustering algorithm using the above similarity concept is carried out in the following manner :-for each point of the data set (x_1,x_2,....x_n) list the k-nearest neighbors (eg. using education distance measure) order number (1,2...k).regard each point as it own zeroth neighbor in this way the first entry in each neighborhood table row is label indicating which point the list belongs to once the neighborhood list have been tabulated, the raw data can be discarded. the computation can be entirely integer. Step (2) provides setup an image label of length n, with each entry initially set to the first entry of the corresponding neighborhood row. step (3) provide all possible pairs of neighborhood rows are tested in following manner. replace both label entries by the smaller of the two existing entries if both zeroth neighbors (the points being tested) are found in found in both neighborhood rows and at least k_t neighbor matches exist between the two rows (k_t is referred to as the similarity threshold).also replace all appearances of the higher label with the lower label if the above test is successful. Step (4) provide the cluster under the k,k_t selections are now indicated by identical labeling of the points to the cluster.

Step(5) provide re -computation with new k and k_t can be carried out simply by returning to step2.the first selection of k should be the largest the investigator will ever require so that the original vector data need not be recalled. Because step 2 and 3 are integer operation on a data set of site n*(k+1).where k is usually <<n and may even be smaller than l. re-calculation for pair of (k,k_t) until the grouping are meaningful to the user is extremely fast and does not require the recall of the raw data. Propagating label changes as required in step 3 can be made much faster than table search by using a list linked algorithm that forms and utilizes chaining information about each group no matter how the members of the group are coffered throughout the complete set of label in the table. Although the Euclidean metric is mentioned in step 1, the method is by no means restricted to this measure and any suitable measure can be used.

Even step 1 is not as computationally expensive as it may appear. two temporary data arrays, i,xi each of size k+1,one integer and one real, are used for the generation of each neighborhood list now, initialize all i entries to zero and all xi entries to a very large number plus a number corresponding to the order in the array so that the last number is the array so that the largest and the number are monotonically increasing from the first entry to the (k+1) entry. only when a point in [x1,x2...xn] is closer (eg..Using Euclidean distance metric then the last entry pushed down out of the array and the new distance squeezed into the list to monotonic order. The label of the new point replace the corresponding I entry. when all pones have been tested for the list, transfer I entries to the corresponding main (non temporary) neighbor list array, do this for all points in {x1,x2...xn}.all k nearest neighbors can be found relatively computational expense over that of finding the nearest neighbor list row generation, only n distance measures need be made; when l is large this represents the largest computational expense of the algorithm. The computational complexity of calculating the near neighbor table is of the order (n)/2 L+ (k) operations where is a relatively small factor to allow for the extra overhead of testing for all k near neighbor for each point. Strictly speaking, only n(n-1)/2 distance measures are necessary but in the algorithm implemented redundant distance calculation were tolerated to enhance programming simplicity and conserve storage. as these operations are on real (floating point) data they are relatively expensive, especially if the computer used has no floating point hardware. the computing complexity for one pass of the neighborhood table to explore cluster used for neighborhood size k and "belongingness" threshold k_t os of the order of at most (n(n-1)/2)(k+1)2 integer comparisons plus a data order and threshold dependent cost of the link listing procedure that is evoked only when matches of sufficient vote are detected. Considerable saving result in testing for the mutual appearance of the neighborhood and abandoning further consideration on failure of this test. Some actual times for typical data sets will be given later to illustrate the relative times for step 1 and step 3 of the above algorithm.

Explanation

Since our goal is the provide a definition of the intuitive concept of similarity, we first clarify our intuitions about similarity. Intuition:1 the similarity between a and b is related to their commonality. The more commonality they share, the more similarity they are intuition: 2 the similarity between a and b related to the differences between them. the more differences they have, the less similarity they are. Intuition:3 the maximum similarity between a and b is reached when a and b identical ,no matter how much commonality they share. Our goal is to arrive at a definition of similarity that captures the above intuition. However they are many alternative ways to define similarity that would be consistent with the intuitions. in this section ,we first make a set of additional assumptions about similarity that we believe to be reasonable. a similarity measure can be derived from those assumptions. in order to capture the intuition that the similarity of two objects are related to their commonality, we need a measure of commonality. our first assumption is.assumptio:1 the commonality between a and b is measured by I(common (A,B)) where common (A,B) is a proposition that states the commonalities between A and B;I(s) is the amount of information contained in a propositions s. for example A is an orange and B is an apple, the proposition that states the commonality between A and B is fruit (A) and fruit (B).the information contained in a statement is measured by the negative logarithm of the probability of the statement. therefore I(common(A,B)=-log p(fruit (A) and fruit(B)).we also need a measure of the differences between two objects. since knowing both the commonalities and the differences between A and B means

knowing what A and B are, we assume. assumption 2:the differences between A and B is measured by I(description (A,B) is a proposition that describe what A and B are intuition 1 and 2 state that the similarity between two objects are related to their commonalities and differences. we assume that commonalities and differences are the only factors. assumption 3: the similarity between A and B, sim(A,B),is function of their commonalities and differences. that is sim(A,B)=f(I common (A,B)) the domain of f is {(x,y)1x≥0,y>0,y≥x}.intuition 3: state that the similarity measure reaches a constant maximum when the two objects are identical we assume the constant is 1.assumption 4: the similarity between a pair of identical objects is 1.when A and B are identical, knowing their commonalities means knowing what they are ie..I (common (A,B))=I(description (A,B).therefore, the function f must have the property :$\forall x > 0, f(x,x) = 1$ when there is commonality between A and B,we assume their similarity is 0,no matter how different they are for example, the similarity between "depth -first search" and " leather sofa" is neither nor lower than the similarity between "rectangle" and interest rate".assumption:5 $\forall y > 0, f(0,y) = 0$ suppose two objects A and B can viewed from two independent perspectives .their similarity can be computed separately from each perspective. we assume that the overall similarity of the two documents is a weighted average of their similarity computed from different perspective. $\forall x1 \leq y1, y2 \leq y2: f(x1 + x2, y1 + y2) = y1$/y1+y2f(x,y1)+y2/y1+y2+(x1+x2) from the above assumption, we can proved the following theorem. similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are sim(A,B)=log p(common(A,B))/log p(description(A,B)).

IV. METHODOLOGY USED

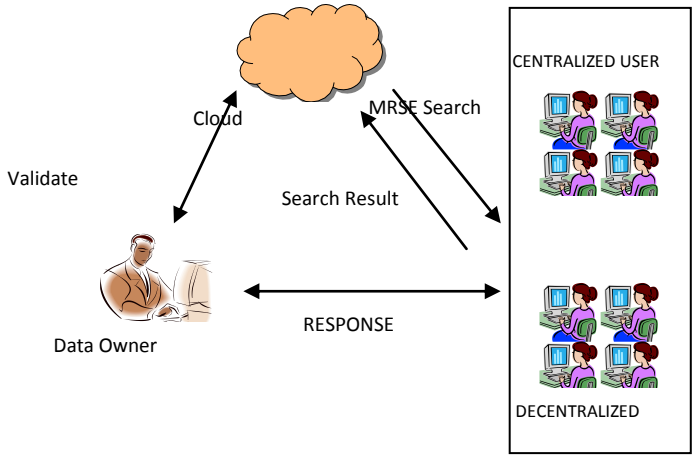

Fig 1. Multi Keyword Ranked Search over Cloud Data Architecture

Cloud computing is defined as type of computing that relies on sharing computing resources rather than having local servers (or) personal device to handle applications. This is uses network of large group of servers typically running low-cost consumer technology, with specialized connections to spread data processing across them. This shard it infrastructure contains large pools of system that are linked together. Cloud computing adopts concepts from (SOA) service oriented architecture that can help the user break these problem into services that can be integrated to provide solution. Cloud computing share characteristics with are client server model, grid computing, mainframe computing, utility computing, peer-to-peer.

Cloud computing some different characteristics are cost, maintenance, reliability and security. Cost is very low in maintenance cloud account. Maintenance of cloud computing application is easier, because they do not need to be installed on each user's computer and can be accessed from different place. Reliability purpose improves with use of multiple redundant sites, which makes well-designed cloud computing suitable for business continuity and disaster recovery. Security main purpose of can improve due to centralization of date, increased security focused resources. But concerns can persist about loss of control over certain sensitive for stored kernels. Cloud computing provides offer their service according to several fundamental models are IAAS, PAAS, SAAS. Infrastructure as a service in the most basic cloud service model &according to the IETF (internet engineering task force).provides of IAAS offer computer. Physical (or) virtual machine and other resources. PAAS in the model, cloud providers deliver a computing platform. Typically including operating systems programming

language execution environment, database and web server. Application developer can develop and run their software solution. on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software. SaaS this model using business. Users are providers access to application software and database. SaaS is sometimes referred to as "on-demand software" and is use basis or using a subscription fee. Cloud clients cloud computing using networked client devices, such as desktop computer, laptops, tablets and smart phones and any ether net enables device such as home automation. Deployment model provides in cloud computing private, public, and hybrid. Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally (or) by a third -party and hosted either internally (or) externally. Undertaking a private cloud project requires a significant level and degree of engagement and requires the organization to reevaluate decisions about existing resources. Public cloud when the services are rendered over a network that is open for public use.

Generally public cloud service providers like Amazon, Microsoft and Google own and operate the infrastructure at their data center and access is generally via the internet.

Hybrid cloud is a composition of two (or) more clouds. (Private, public (or) community) that remain distinct entities but are bound together, offering the benefits of multiple deployment models. Hybrid cloud can also mean the ability to connect collation, managed and l or dedicated services with cloud resources. Security and privacy cloud computing poses privacy concerns because the service provider can access the service provider can access the data that is on the cloud at a time. It could accidentally (or) deliberately altering r (or) even delete in information.

He future of cloud computing is bright for the companies that implement the technology now. While these are some trends that are expected in the future, the future is not limited to these trends, remain abreast of the latest development to help your company maintain a competitive advantage. multi keyword search ranked in cloud the objective of keyword research is to generate, with precision and recall, large number of terms that are highly yet no obvious to the given input keyword. Process of keyword research involves brainstorming and the use of keyword research tools. Keyword density is the percentage of times a keyword (or) private appears on a web page compared to the total number of words on the page. Main concept of algorithm similarity measurement in computer science, a similarity between two objects. although no single definition of a similarity measure exists, usually measure are in some sense the inverse of distance matrices are they take on large values for similar objects and either zero(or)a negative value for very dissimilar objects. The concept of similarity is fundamentally important in almost every scientific field. For example in mathematics, geometric method for assessing similarity is used in studies of congruence and homothetic as well as in allied field such as trigonometry. Similarity measure different types are distance based, feature based, probabilistic based similarity measurements.

V. Experimental Results

User authentication is a process that allows a device to verify the identity of someone who connects to a network resource. There are many technologies currently available to a network administrator to authenticate users.

Challenges of Authentication in the Cloud

Public clouds are generally defined as being external to the companies that use it. It is where software-as-a-service (SaaS)-type applications (typically Web-based) run.The problem with this lack of cohesion is it makes it difficult to affect things that we take for granted in the private cloud.

We are just beginning to solve some of these problems. For example, the Security Assertion Markup Language and Active Directory Federation Services protocols allow identity federation. That makes it possible for apps that run in public clouds to authenticate users using corporate credentials.

When we look at authentication and authorization aspects of cloud computing, most discussions today point towards various forms of identity federation and claims-based authentication to facilitate transactions between service end points as well as intermediaries in the cloud. Even though they represent another form of paradigm shift from the self-managed and explicit implementations of user authentication and authorization, they have a much better chance at effectively managing access from the potentially large numbers of online users to an organization's resources.

Encrypt and Cloud Storage

This module is used to help the server to encrypt the document using AES Algorithm and to convert the encrypted document to the password protected with activation code and then activation code send to the user for download.

Encrypt using AES

Encryption is the process of converting plaintext to cipher-text (had to understand) by applying mathematical transformations. These transformations are known as encryption algorithms and require an encryption key.

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

Cloud storage

Uploading, from a customer computing platform to a key store of the cloud computing platform, a cloud-based encryption key based on a customer-based encryption key, the cloud-based encryption key and customer-based encryption key being able to encrypt or decrypt customer data used by an application server running on the cloud computing platform;

Accessing, according to an encryption or decryption mechanism, the unlocked cloud-based encryption key to encrypt or decrypt customer data stored on a database of the main memory and used by the application server.

Multi-keyword Ranked Search

In cloud computing data possessor are goaded to farm out their complex data management systems from local sites to the commercial public cloud for greater flexibility and economic savings. To ensure safety of stored data, it is must to encrypt the data before storing. It is necessary to invoke search with the encrypted data also. The specialty of cloud data storage should allow copious keywords in a solitary query and result the data documents in the relevance order. In main aim is to find the solution of multi-keyword ranked search while preserving strict system-wise privacy in the cloud computing paradigm. A variety of multi- keyword semantics are available, an efficient similarity measure of "coordinate matching" (as many matches as possible), to capture the data documents' relevancy to the search query is used. Specifically "inner product similarity", i.e., the number of query keywords appearing in a document, to quantitatively evaluate such similarity measure of that document to the search query is used in MRSE algorithm.

Decryption using AES

Decryption is the reverse process of getting back the original data from the cipher-text using a decryption key. In Symmetric cryptology- The encryption key and the decryption key could be the same as in symmetric or secret key cryptography, The key can different as in asymmetric or public key cryptography.

Performance Analysis

In this section, we show a thorough experimental evaluation of the proposed technique on a real dataset. The whole experiment is implemented by C# language on a computer with Core 2.83GHz Processor, on Windows 7 system. For the proposed scheme, we will reduce to separate dimensions. The performance of our method is compared with the original MRSE scheme.

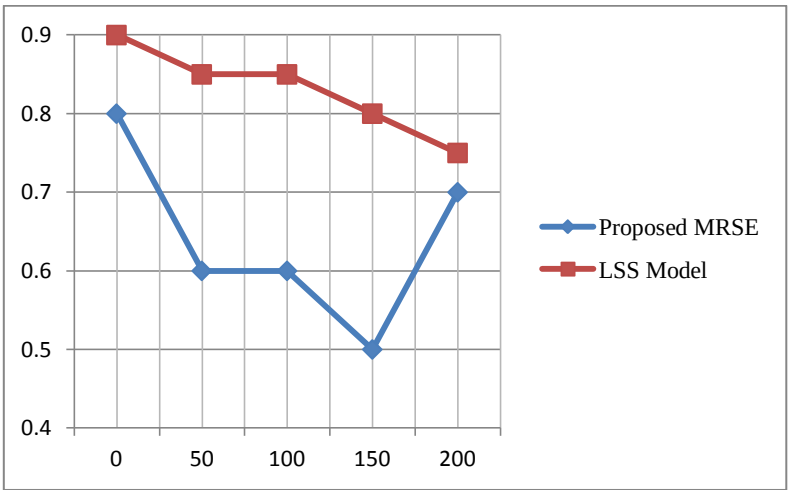

Fig 2. Comparison of Proposed MRSE with LSS Model

In this proposed work we gather 100 sample documents from different web pages and applications to identify the accuracy of the proposed work. The result shown the proposed MRSE will produce high accuracy compare with our existing work of using Latent Semantic Space Model. The measure of similarity between two ranked documents with encryption technique will highly recommended by Proposed MRSE technique with low latency.

VI. Conclusion and Future Work

The first time we define and solve the problem of multi-key word ranked search over encrypted cloud data, and establish a variety of privacy requirement. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., the many matches as possible, to effectively capture similarity between query keyword and outsourced documents, and use "inner product similarity" to quantitatively formalize such a principle for similarity measurement. This paper ,we motivate and solve the problem of efficient and secure ranked multi-keyword search on remotely stored encrypted database model where the database user are protected against privacy violations. We appropriately increase the efficiency of the scheme by using symmetric key encryption method rather than public-key encryption. We are proving that our proposed method satisfies the security requirements. The proposed ranking method highly relevant to the document corresponding to submit searching terms.

As our future work, we will concentrate on the encrypted data of semantic keyword search in order that we can confront with the more sophisticated search.

References

1. "Privacy –Preserving Multi-Keyword Ranked Search over encrypted cloud data" by Ning Cao,Member IEEE,Cong Wang, Member IEEE, Ming Li, Member IEEE, Kui Ren, Senior Member IEEE and Wenjing Lou, Senior Member IEEE.
2. "High Performance Cloud computing with Encryption and Privacy" by Jyothi.
3. "Privacy –Preserving Multi-Keyword Ranked Search with Anonymous ID Assignment over encrypted cloud data" by Shiba Sampat Kale, Prof. Shiva R Lahane.
4. "Privacy Preserving Data Search in cloud computing" by Ruturaj Desai, Nitin R. Thalhar.
5. "Secured Multi Keyword Search over encrypted cloud data" by Ankatha Samuyela Raja, Vasanthi A
6. A. Singhal, Modern information retrieval: A brief overview, IEEE Data Engineering Bulletin, vol. 24, no. 4, pp. 3543, 2001.

7. D. Song, D. Wagner, and A. Perrig, ―Practical techniques for searches on encrypted data,‖ in Proc. of IEEE Symposium on Security and Privacy'00, 2000.
8. C.Yang, "A Fast Privacy-Preserving Multi-keyword Search Scheme on Cloud Data", Cloud and Service Computing (CSC), 2012 International Conference, IEEE, 2012.
9. M.Chuah and W. Hu, "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data", Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference, IEEE, 2011.
10. S.Deshpande, "Fuzzy keyword search over encrypted data in cloud computing", World Journal of Science and Technology, vol. 2, no. 10, 2013.

