Cloud Computing: Ranked Keyword Search

Sonia Malik[*1], Ms. Sapna Aggarwal[*2]
[1]M.Tech, Jind Institute of Engineering & Technology, KUK,Haryana (India)soniyamalik90@gmail.com
[2] Assistant Professor, Jind Institute of Engineering & Technology, KUK,Haryana (India)
er.sapna.aggarwal@gmail.com

Abstract:
Cloud computing becomes more widely prevailed storage for outsourced data which may contain more sensitive information such as credit card numbers, passwords, e-mails, personal health records etc. As the data owners cannot risk their unencrypted outsourced data so as the cloud servers. The cloud server may fail to keep up the integrity of the cloud data due to hacking or entry of unauthorized entities. While searching the data in the cloud the attackers prefer the keyword which is not secured properly. The existing technique resolves the optimization complexities in ranked keyword search and its effective utilization of remotely stored encrypted cloud data. But it limits the further optimizations of the search results by preventing cloud server to interact with cloud users to maintain the integrity of actual owner's keyword and the data associated with it. The aim is to define a framework which enhances the accuracy of the ranked keyword search by secured machine learning, which does not affect the data integrity. Introducing new and interactive access permissions allows only specific group of people to guide the search engine. This technique lists the exact or necessary search results for any encrypted keyword. Due to this learning the privacy of the keyword does not get to be violated because, the owner of the encrypted keyword has some lists of users to whom only the machine should learn for secured and improved search results.

General Terms*: Efficient Ranked Keyword Search, Search engine in Cloud, Security in Search engine.*
Keywords:
Search in cloud, secured search engines, Inter cloud communication in cloud search engines.

1. **Introduction:-**

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service.

However, cloud computing technology challenges many traditional approaches to datacenter and enterprise application design and management. The effectiveness and efficiency of traditional protection mechanisms are being reconsidered as the characteristics of this innovative deployment model can differ widely from those of

traditional architectures. An alternative perspective on the topic of cloud security is that this is but another, although quite broad, case of "applied security" and that similar security principles that apply in shared multi-user mainframe security models apply with cloud security. It is The very nature of cloud computing-based services, private or public, that promote external management of provided services. This delivers great incentive to cloud computing service providers to prioritize building and maintaining strong management of secure services. Security issues have been categorized into sensitive data access, data segregation, privacy, bug exploitation, recovery, accountability, malicious insiders, management console security, account control, and multi-tenancy issues. Solutions to various cloud security issues vary, from cryptography, particularly public key infrastructure (PKI), to use of multiple cloud providers, standardization of APIs, and improving virtual machine support.

ISSN 2454-308X
9 770024 543081

2. Cloud Models:-

- Delivery Models
  – SaaS(Software as a Service)
  – PaaS (Platform as a Service)
  – IaaS (Infrastructure as a Service)
- Deployment Models
  – Private cloud
  – Community cloud
  – Public cloud
  – Hybrid cloud

We propose one more Model: Management Models (trust and tenancy issues)

– Self-managed

-3rd party managed (e.g. public clouds and VPC)

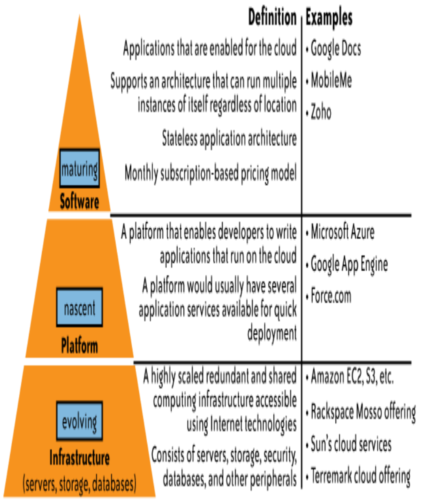
Definition
Examples
Applications that are enabled for the cloud
Supports an architecture that can run multiple instances of itself regardless of location
Stateless application architecture
Monthly subscription-based pricing model
• Google Docs
• MobileMe
• Zoho
maturing
Software
A platform that enables developers to write applications that run on the cloud
A platform would usually have several application services available for quick deployment
• Microsoft Azure
• Google App Engine
• Force.com
nascent
Platform
A highly scaled redundant and shared computing infrastructure accessible using Internet technologies
Consists of servers, storage, security, databases, and other peripherals
• Amazon EC2, S3, etc.
• Rackspace Mosso offering
• Sun's cloud services
• Terremark cloud offering
evolving
Infrastructure
(servers, storage, databases)

Figure 2.1: Impact of cloud computing on the governance structure of IT organizations

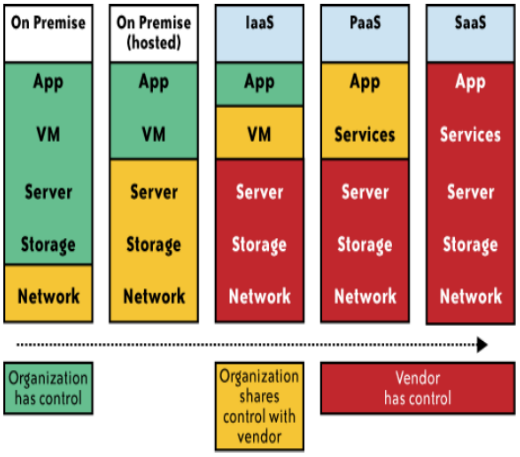

Figure 2.2: Delivery Models Behavior

If cloud computing is so great, why isn't everyone doing it?

- The cloud acts as a big black box, nothing inside the cloud is visible to the clients
- Clients have no idea or control over what happens inside a cloud
- Even if the cloud provider is honest, it can have malicious system admins who can tamper with the VMs and violate confidentiality and integrity
- Clouds are still subject to traditional data confidentiality, integrity, availability, and privacy issues, plus some additional attacks

Companies are still afraid to use clouds

- Most security problems stem from:
  - Loss of control
  - Lack of trust (mechanisms)
  - Multi-tenancy
- These problems exist mainly in 3rd party management models
  - Self-managed clouds still have security issues, but not related to above

3. Problem Statement:-

The existing technique resolves the optimization complexities in ranked keyword search and its effective utilization of remotely stored encrypted cloud data. But it limits the further optimizations of the search results by preventing search engine to interact with cloud users to maintain the integrity of actual owner's keyword and the data associated with it. Consider an encrypted cloud data hosting service involving three different entities, as illustrated in Fig 1data owner, data user, and cloud server. Data owner has a collection of n data files that he wants to outsource on the cloud server in encrypted form while still keeping the capability to search through them for effective data utilization reasons.

Now consider the secure ranked keyword search problem as follows: the search result should be returned according to certain ranked relevance criteria (e.g., keyword frequency-based scores, as will be introduced shortly), to improve file retrieval accuracy for users without prior knowledge on the file collection. We primarily consider an "honest-but-curious" server in our model, which is consistent with most of the previous searchable encryption schemes.

The problem with the techniques available for implementing search engine in an environment consists of sensitive outsourced cloud data can be summarized as:

a. Lacking of effective mechanisms to ensure the file retrieval accuracy is very difficult.

b. Security is not addressed fully and limits search engine's accuracy.

The ranked keyword search over encrypted data is to achieve economies of scale for Cloud Computing. This process start from the review of existing searchable symmetric encryption schemes and provides the definitions and framework for this proposed ranked searchable symmetric encryption. Searchable encryption allows data owner to outsource his data in an encrypted manner while maintaining the selectively search capability over the encrypted data. In order to achieve more efficient solutions, almost all the existing works on searchable encryption literature resort to the weakened security guarantee, i.e., revealing the access pattern and search pattern but nothing else. Result, i.e., which files have been retrieved.

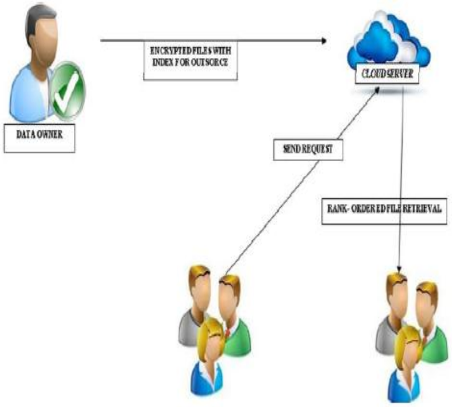
ENCRYPTED FILES WITH INDEX FOR OUTSORCE
CLOUD SERVER
DATA OWNER
SEND REQUEST
RANK- ORDERED FILE RETRIEVAL

Figure 3: Ranked keyword searching in cloud computing

The search pattern includes the equality pattern among the two search requests (whether two searches were performed for the same keyword), and any information derived thereafter from this statement.

4. Framework Definition:-

This process defines and solves the problem of effective yet secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. It first gives a straightforward yet ideal construction of ranked keyword search searchable symmetric encryption (RSSE)[6] security definition, and demonstrates its inefficiency. To achieve more practical performance, the process then propose a definition for ranked searchable symmetric encryption, and give an efficient design by properly utilizing the existing cryptographic primitive, order-preserving symmetric encryption (OPSE).Relevance Score Calculation[6],

Score (t, Fd) = 1 / |Fd| □ (1+ln fd,t),
Where, t- is the term searched by the user,
FD, t- denotes the Term Frequency (TF) of the term t in file FD,
In- denotes the natural logarithm of TF of the file FD,

| FD|- denotes the length of the file. Therefore, the same relevance score appearing in different lists of the index I will be mapped to different "bucket" in R. Combining this with the one-to-many mapping will randomize the encrypted values from an overall point of view. The sampling during an OPSE operation is a function belonging to O (log M), and is at most 5 log M +12 on average Thus, mitigation of the useful information revealed to the cloud server This process also aim to develop the more efficient in ranked keyword search and provide more security; using the process of TDT4 mechanism and privilege technique. TDT4 mechanism is used for provide the efficient ranked keyword search. In this process information retrieval, a ranking function is used to calculate relevance scores of matching files to a given search request.

Design Goals:
To enable ranked keyword search for effective utilization of outsourced cloud data under the aforementioned model, our system design should achieve the following security and performance guarantee.
a) Ranked keyword search: For efficient searching process the process use the mechanism of Topic detection and tracking 2004. The search time includes fetching the posting list in the index, decrypting, and rank ordering each entry.
b) Security guarantee: For providing the security in the cloud server, this process uses the privilege method.

Mechanisms for Implementation: Topic detection and tracking:
TDT refers to automatic techniques for finding topically related material in streams of data techniques that could be quite valuable in a wide variety of applications where efficient and timely information access is important. For example, a lot of useful information could be gleaned from a multitude of news sources, but no one has the time to watch, listen to, or read carefully each of the many news sources available. Tasks can vary in focus and size from hypothetical applications to enabling technologies. In brief, the goal of each of the tasks is:

a. Topic Tracking– detect stories that discuss a target topic,
b. Link Detection– detect whether a pair of stories discuss the same topic,
c. Topic Detection– detect clusters of stories that discuss the same topic,
d. First Story Detection– detect the first story that discusses a topic, and
e. Story Segmentation– detects story boundaries.

Privilege:
A privilege is a special entitlement to immunity granted by the state or another authority to a restricted group, either by birth or on a conditional basis. It can be revoked in certain circumstances.

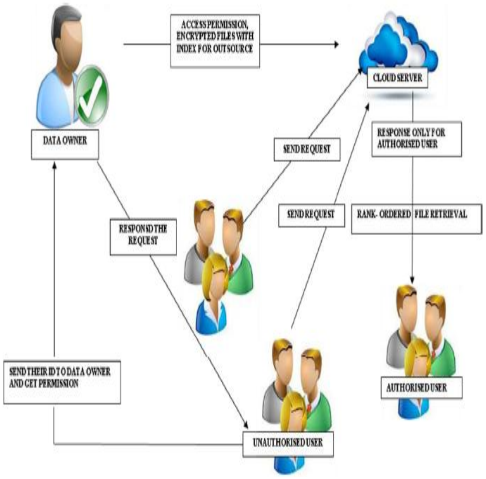
ACCESS PERMISSION, ENCRYPTED FILES WITH INDEX FOR OUTSOURCE
DATA OWNER
CLOUD SERVER
SEND REQUEST
RESPONSE ONLY FOR AUTHORISED USER
SEND REQUEST
RANK- ORDERED FILE RETRIEVAL
RESPONSD THE REQUEST
SEND THEIR ID TO DATA OWNER AND GET PERMISSION
AUTHORISED USER
UNAUTHORISED USER

Figure 4: Mechanisms for Implementation: Topic detection and tracking

5. Results and Discussion:-

Execution process:

a. Owner uploads the file in to the cloud server, and set the privilege to the particular user for easily access data.

b. And give the particular permission like write, read or both for providing the security. Here the user's are separated by authorized user and unauthorized user. Authorized user is the owner permitted person and unauthorized user is unpermitted person.

c. So authorized user easily access the data from the cloud server by using the ranked efficient keyword search by the mechanism of TDT4 mechanism. Unauthorized user asks the permission to access the data.

d. After the data owner permission, then only the authorized user access the data in the cloud server.

Efficiency of the ranked keyword search:-

This type of ranked keyword search enhances the efficient usage of outsourced files by providing Inter cloud communication constantly between data owners and users. So that the cloud server learn nothing from the data uploaded by data owners.

The search time is not affected while fetching the posting list in the index, decrypting, and rank ordering each entry

6. Benefits:-

Ranked keyword search: to explore different mechanisms for designing effective ranked search.

a. Provide more security to the data owner, by means of Inter-cloud communication through e-mail.

b. Authentication of both search results and the outsource process enables search engine to be more robust than before in cloud environment

c. Privilege method is used for the security. So process has the more security compared to the existing system.

7. Related Work:-

Data Security and Storage

- Data remanence
  - Inadvertent disclosure of sensitive information is possible
- Data security mitigation?
  - Do not place any sensitive data in a public cloud
  - Encrypted data is placed into the cloud?
- Provider data and its security: storage
  - To the extent that quantities of data from many companies are centralized, this collection can become an attractive target for criminals
  - Moreover, the physical security of the data centre and the trustworthiness of system administrators take on new importance.

Minimize Loss of Control: IDM Issues in Cloud Computing:-

- Cloud introduces several issues to IDM
  - Users have multiple accounts associated with multiple service providers.
  - Lack of trust
    - Use of Trusted Third Party is not an option
    - Cloud hosts are untrusted
  - Loss of control
    - Collusion between Cloud Service

-Sharing sensitive identity information between services can lead to undesirable mapping of the identities to the user.

Migration of Search Engine Process into the Cloud:-

Working of Search Engine is divided into two part, query dependent module and query independent module.

Query Independent Module consists of Crawler, indexer and Repository.

Query Independent module does not depend upon the user but is an ongoing continuous process.

Query Dependent module consists of Query Module and a Ranking Module. Each Component on query dependent module forms a Part of Cluster.

Whenever user fires a query its goes to a particular cluster depending on the type and data set required by the query and that is determined by cloud controller.

Conclusion:-

In this paper, as an initial attempt, we motivate and solve the problem of supporting efficient ranked keyword search for achieving effective utilization of remotely stored outsourced data in a cloud.

We first give the framework definition to provide secure search facility for the sensitive data stored in cloud environment. We also investigate some further enhancements of our ranked search mechanism, including the efficient support of relevance score dynamics. We are looking forward to the extensive experimental results which will demonstrate the efficiency of our solution.

By enabling a search result authentication mechanism that can detect unexpected behaviors of cloud server like saving cost when handling large number of search requests, software bugs and internal/external attacks.

In future we will support to score dynamics. Score Dynamics adding newly encrypted scores for newly created files, or modifying old encrypted scores for modification of existing files in the file collection.

References:-

1. Akassh A. Mishra, Chinmay Kamat (April 2011),"Migration
of search engine process into the cloud"

2. Armbrust.M.Fox. A, Griffith. R, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, (Feb 2009) "Above the clouds: A berkeley view of cloud computing," University of California, Berkeley.

3. Boneh .D. G. D. Crescenzo, R. Ostrovsky, and G. Persiano, (2004), "Public key encryption with keyword search," in Proc. of EUROCRYP'04, volume 3027 of LNCS.

4. Y.-C. Chang and M. Mitzenmacher, (ACNS 2005), "Privacy preserving keyword searches on remote encrypted data".

5. Chien Chin Chen and Meng Chang Chen, (SIGIR 2012) "TSCAN: A Content Anatomy Approach to Temporal Topic Summarization".

6. Cong Wang, Ning Cao, Kui Ren, Wenjing Lou "En abling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data.

.

7. Song.D,Agner and Perrig. A, (IEEE 2000), "Practic al techniques for searches on encrypted data," in Proc. of IEEE Symposium on Security and Privacy'00.