Survey on Keyword Search in Encrypted Cloud Data

Minnu C Tomy
PG Scholar
Vimal Jyothi Engineering College
Chemperi, Kannur, Kerala

Vidhya S S
Asst. Professor
Vimal Jyothi Engineering College
Chemperi, Kannur

Abstract—**Today cloud computing becomes widespread due to its enormous services. Nowadays more and more sensitive data is being centralized into the cloud by users. The key barrier to pervasive uptake of cloud computing is the lack of trust in clouds by potential customers. In order to maintain the security of sensitive user data in the non trusted servers, the data or information must be encrypted before they are uploaded to the cloud. The sensitive data can be stored in an encrypted format thereby it become available only to authenticated users. However, this raises a new challenge for accessing encrypted data efficiently. Keyword based search is a best solution for this issue. Traditional keyword search techniques over encrypted cloud data support only exact or fuzzy keyword search. A multi-keyword ranked search over encrypted data supporting synonym query is very effective and efficient. This paper surveys different searching mechanisms in encrypted cloud data.**

Keywords— Cloud Computing; Encryption; Search

I. INTRODUCTION

Cloud computing has changed the way industries approach IT, enabling them to become more agile, introduce new business models, offer more services, and trim down IT costs. Cloud computing technologies can be implemented in a wide variety of architectures, under diverse service and deployment models, and can coexist with many technologies and software design methods. The cloud computing background continues to realize explosive growth. Yet for security professionals, the cloud presents a huge dilemma: How do you embrace the benefits of the cloud while maintaining security controls over your organizations' assets? It becomes a question of balance to determine whether the increased risks are truly worth the agility and economic benefits. Maintaining control over the data is paramount to cloud success. A decade ago, enterprise data typically resided in the organization's physical infrastructure, on its own servers in the organization's data center, where one could segregate sensitive information in individual physical servers. Today, by virtualization and the cloud, data may be under the organization's logical control, but physically stored in infrastructure owned and managed by a different entity. This shift in control is the number one reason new approaches and techniques are required to ensure organizations can maintain data security. When an outside party owns, controls, and supervises infrastructure and computational resources, how can you be assured that business or regulatory data remains private and secure, and that your organization is protected from damaging data breaches? This makes cloud data security essential.

II. CLOUD COMPUTING

The importance of Cloud Computing is increasing and it is receiving a growing consideration in the scientific and industrial communities. The NIST (National Institute of Standards and Technology) proposed the following definition of cloud computing: "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 provider interaction. This cloud model promotes availability" [1]. The cloud improves collaboration, agility, scalability, availability, ability to adapt to variations according to demand, speed up development work, and provides potential for cost reduction through optimized and efficient computing. Cloud Computing combines a number of computing ideas and technologies such as Service Oriented Architecture (SOA), Web 2.0, virtualization and other technologies with reliance on the Internet, supporting common business applications online through web browsers to satisfy the computing needs of users, while their software and data are maintained on the servers.

A. Cloud Delivery Models are:

• Private Cloud– Cloud infrastructure is provisioned for use by a single organization that comprises multiple tenants. Private clouds may be operated on- or off-premises and are behind the company firewall.

• Public Cloud– A cloud service provider offers services to multiple businesses, academic institutions, government agencies, and other organizations with access via the Internet.

• Hybrid Cloud– Hybrid clouds combine two cloud delivery models that remain unique as entities, but they are bound together by technology that enables data and application portability. Cloudbursting is an example of one way enterprises use hybrid clouds to balance loads during peak demand periods.

• Community Cloud – Cloud infrastructure is provisioned for the exclusive use of a specific community of user organizations with shared computing requirements such as security, policy, and compliance.

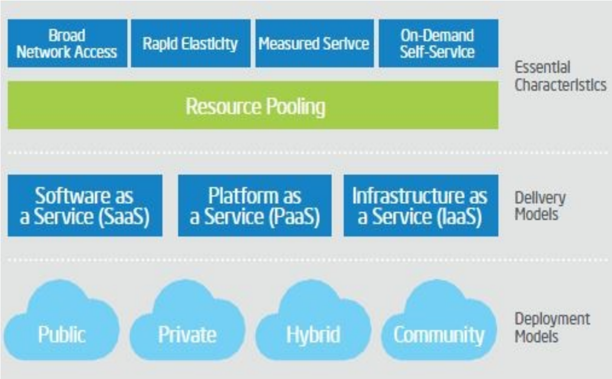

Fig 1.NIST Cloud Architecture

B. *The Service layers for these delivery models are:*

• Infrastructure as a service (IaaS) – Cloud infrastructure is the collection of hardware and software that enables the essential characteristics of the cloud. IaaS allows users to self-provision these resources in order to run platforms and applications.

• Platform as a service (PaaS) – PaaS enables users to adapt legacy applications to a cloud environment or develop cloud-aware applications using programming languages, services, libraries, and other developer tools.

• Software as a service (SaaS) – Users can run applications via multiple devices on cloud infrastructure.

III. Security in Cloud

Although there are a lot of benefits to adopting Cloud Computing, there are also some considerable barriers to acceptance [2]. One of the most major barriers to adoption is the security, followed by issues regarding compliance, privacy and authorized matters. Since Cloud Computing represents a relatively new computing model, there is a huge deal of uncertainty about how security at all levels (network, host, application, data levels, etc.) can be achieved and how application security is moved to Cloud Computing. That uncertainty has consistently led information executives to state that security is their number one concern with Cloud Computing. Security concerns relate to risk areas such as external data storage, dependency on the public internet, lack of control, multi-tenancy and integration with internal security.

Compared to conventional technologies, cloud has many specific features, such as its great scale and the fact that resources belonging to cloud providers are entirely distributed, heterogeneous and completely virtualized. Conventional security mechanisms such as identity, authentication, and authorization are no longer enough for clouds in their current form. Because of the cloud service models employed, the operational models, and the methodologies used to enable cloud services, cloud computing may present different risks to an association than traditional IT solutions. Regrettably, integrating security into these solutions is often perceived as making them more rigid.

The cloud computing research community, particularly the Cloud Security Alliance, has recognized security issues in cloud. In its Top Threats to Cloud Computing Report (Ver.1.0) [3], it listed seven top threats to cloud computing:

1. Abuse and nefarious use of cloud computing
2. Insecure application programming interfaces
3. Malicious insiders
4. Shared technology vulnerabilities
5. Data loss or leakages
6. Account, service and traffic hijacking
7. Unknown risk profile.

A. *Traditional Techniques for Protecting Data in the Cloud*

Traditional models of data protection have often focused on network centric and perimeter security, often with devices such as firewalls and intrusion detection systems [4]. But this approach does not provide sufficient protection against advanced persistent threats, privileged users, or other insidious types of security attacks. Many enterprises use database audit and protection (DAP) and Security Information and Event Management (SIEM) solutions to gather together information about what is happening. But monitoring and event correlation alone do not translate into data security. At a time when regulation and compliance issues are at an all-time high, it’s dangerous to assume that monitoring, collecting, and storing logs can protect the organization from security threats, as they are reactive controls. In today’s environment, both data firewalls and data security intelligence are essential to adequately protect the enterprise from new and different types of attacks. While many organizations have implemented encryption for data security, they frequently overlook inherent weaknesses in key management, access control, and monitoring of data access. If encryption keys are not sufficiently protected, they are susceptible to theft by wicked hackers. Vulnerability also lies in the access control model; thus, if keys are appropriately protected but access is not sufficiently controlled or robust, wicked or compromised personnel can attempt to access sensitive data by assuming the identity of an authorized user.

Security and privacy of data stored in the cloud are major setbacks in the field of Cloud Computing. This has always been a significant aspect of quality of service. Data theft can happen either by an insider or an outsider in the cloud.

IV. SEARCH IN ENCRYPTED CLOUD DATA

As Cloud Computing becomes widespread, more sensitive information are being transferred into the cloud, such as emails, individual health records, confidential videos and photos, business finance data, government documents, etc. By doing so i.e., storing their data into the cloud, the data owners/providers can be relieved from the burden of data storage space and maintenance so as to enjoy the on- demand high excellence data storage service [5]. However, the truth that data providers and cloud server are not in the similar trusted domain may put the outsourced data at risk, as the cloud server might no longer be fully trusted in such a cloud environment because of a number of reasons, they are: the cloud server may leak information content to unauthorized entities or it may be hacked. It follows that sensitive data typically should be encrypted prior to outsourcing for data privacy and combating unwanted accesses.

However, data encryption makes data utilization effectiveness and efficiency a very challenging task given that there could be a large amount of outsourced data files. Furthermore, in Cloud Computing, data owners/provider may share their outsourced data with a large number of users. The individual users might desire to only retrieve certain precise data files they are interested in throughout a given session. One of the most accepted ways is to selectively retrieve files through keyword-based search as an alternative of retrieving all the encrypted files back which is completely unreasonable in cloud computing scenarios. Such keyword-based search method allows users to selectively retrieve files of interest and has been broadly useful in plaintext search scenarios, such as Google search. Unhappily, data encryption restricts user's ability to perform keyword search and consequently makes the traditional plain text search techniques not suitable for Cloud Computing. In addition to this, data encryption also demands the protection of keyword privacy because keywords usually include significant information related to the data files. Even though encryption of keywords can protect keyword privacy, it further renders the conventional plaintext search techniques useless in this situation.

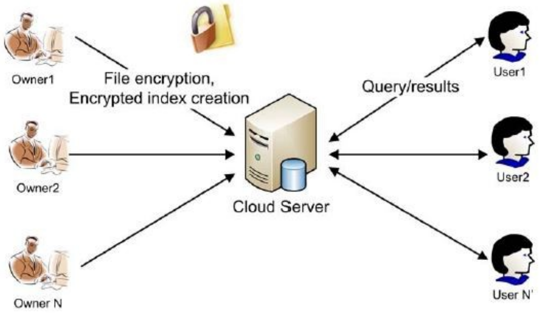

Fig 2.Search in Encrypted Cloud Data

The above figure shows the system model of searching over an encrypted cloud. Which consist of three actors they are cloud server, data owner and user. The data owner, individual or enterprise, has a document collection which will be outsourced into the cloud. Cloud server stores the data. User is the one who interested in the document and search the data in the cloud server.

Existing searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result [6]. Fuzzy keyword search techniques only uses single keywords and typos will produce wrong results. Multi-keyword ranked search over encrypted cloud data provide efficient and effective search result to clients. Searching using multi-keywords is used to help the user to get the accurate result based on the multiple keyword concepts. Ranked search technique greatly improves system usability by normal matching files in a ranked order regarding to certain relevance criteria.

V. CONCLUSION

Cloud Computing offers good service. When these services are used properly, they can reduce cost and management responsibilities in addition to increasing efficiency, agility and performance of an enterprise. Security of data in the cloud is a major problem which makes users anxious about the safety, reliability and efficiency of moving towards cloud computing. Security measures should be incorporated in to cloud to gain customers' trust. There must be a security deployed in each layer of the cloud computing since data leakage is the main concern in cloud. The cryptography technologies offer encryption and decryption of the data and user authentication information to protect it from the unauthorized user or attacker. This method ensures high cloud storage integrity, enhanced error localization and easy identification of misbehaving server. Cryptography is most acceptable solution of security by associated companies in cloud computing environment. In cloud data can be stored both as public as well as private. Various searching strategies are available for both types of data. The private data are stored in the cloud using encryption methods. So only the authenticated members who know the key can access the data. Accessing data from encrypted storage is very difficult. A different type of searching technique is used to search for encrypted data. Fuzzy keyword search, Searchable encryption, Multi-keyword search, Ranked search etc. are some examples of searching methods in encrypted cloud data.

REFERENCES

[1] Peter Mell, Timothy Grance, "The NIST Definition of Cloud Computing," Special Publication 800-145.

[2] Jeon SeungHwan, Yvette E. Gelogo and Byungjoo Park, "Next Generation Cloud Computing Issues and Solutions," International Journal of Control and Automation Vol. 5, March, 2012.

[3] "Top Threats to Cloud Computing Report (Ver.1.0)," Cloud Security Alliance, 2010.

[4] Keiko Hashizume, David G Rosado, Eduardo Fernandez-Medina and Eduardo B Fernandez, "An analysis of security issues for cloud computing," Journal of Internet Services and Applications 2013.

[5] P.Niranjan Reddy,Y.Swetha, "Techniques for Efficient Keyword Search in Cloud Computing," (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (1) , 2013.

[6] C. R. Barde, Pooja Katkade, Deepali Shewale, Rohit Khatale , "Secured Multiple-keyword Search over Encrypted Cloud Data," International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, , Volume 4, Issue 2, February 2014)