VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

MULTI-KEYWORD RANKED SYNONYM SUPPORTING SEARCH IN ENCRYPTED CLOUD DATA

Minnu C. Tomy and Vidhya S. S.
Vimal Jyothi Engineering College, Chemperi, Kannur, Kerala, India
E-Mail: minnuctomy@gmail.com

ABSTRACT

Migration of users into the cloud environment is increased with the high popularity of services provided by the cloud providers. When large number users are outsourcing their files into the cloud environment privacy becomes the most important issue. As a result the clients outsource their data/information after encryption. The searching and retrieval of data becomes most complex when the files are stored in the encrypted format. In the previous works Multi-keyword ranked search over encrypted data supporting synonym queries is proposed is implemented to assure the privacy enhanced searching method. The ranking technique is used to retrieve the most similar values over the encrypted data files. However clients cannot assure that whether the all retrieved results are correctly ranked or not. The rank test method can be implemented to find out the files are having similar fields or not. The proposed system in this work is used to retrieve the files with the most similarity values. To achieve fried man rank testing mechanism is used which tends to check the integrity of ranked files. The experimental tests conducted were proves that the proposed methodology provides better result than the existing approaches in terms of improved privacy and retrieval rate.

Keywords: cloud computing, encryption, search.

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.

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.

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. Cloud bursting 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.

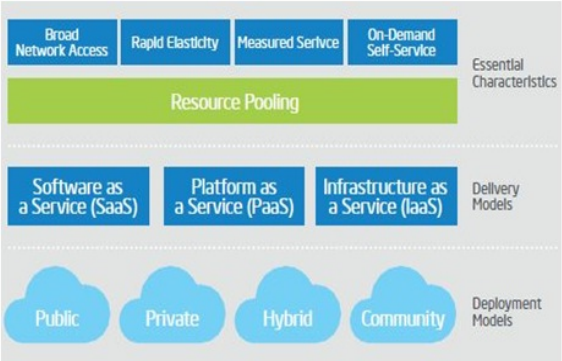

Figure-1. NIST cloud architecture.

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.

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.

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.

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 [4]. 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 [5]. 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.

PROBLEM DEFINITION

Storing and retrieval is the important application in the today's world which needs to be concerned more in future. In the previous researches, privacy and security is given strongly by encrypting the data before storing into the cloud environment. The encrypted data storage in cloud will make the retrieval process as the most complex one which will lead to performance degradation by not retrieving the similar files. The existing work, overcomes this problem by enabling the search retrieval over an encrypted cloud data base. However the existing work doesn't concentrate on retrieving the similar documents based on knowledge what the user tried to expose. It only looks for a submitted keywords and synonyms over a database and will retrieve the information. This methodology will lead to a retrieval of documents based on occurrence of terms and not based on the concept. This difficulty is overcome in this work by introducing the new methodology for identifying the semantic relationship among the keywords submitted and the documents stored in the cloud database.

SYSTEM ANALYSIS

The proposed scheme consists of the following modules.

- Initialization Process
- Index Generation process
- Query Generation Process
- Searching Process
- Protection of sensitive information
- Rank Integrity Checking
- Performance Evaluation

System Model

The 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.

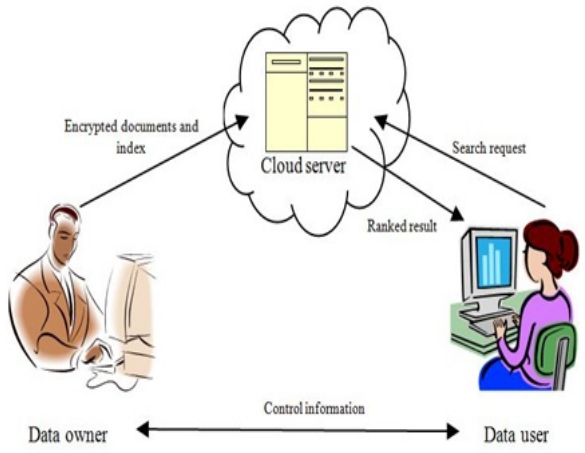

Figure-2. System model.

Proposed System

Fried man ranking test is implemented in this work to improve the search efficiency of the files. This mechanism is used to improve the search retrieval process by testing the rank fixed by the cloud servers. Whenever the user enters the multi key word for file retrieval, the trapdoor will be generated. Through that trap door value, the cloud server will match the user's keyword query with the searchable index of the encrypted files and will retrieve the most alike files. In this procedure, in order to assure retrieved files are similar to the queries and the result is retrieved by checking all the files in the group, rank test mechanism is used. The Fried man rank test mechanism is implemented in this work which can ensure that all files from the group is ranked properly and only the similar files are retrieved. This is done by ranking each block in file and comparing with the other files.
Fried man rank test is computed as follows:

$$F_R = \frac{12}{rc(c+1)} \sum_{j=1}^{c} R_j^2 - 3r(c+1)$$

Where,

R_j^2 =Square of the total of the ranks for group j (j=1, 2… c)

r = number of blocks

c = number of groups

Advantages of Proposed System

1. "Coordinate matching" by inner product similarity.
2. Reduce the communication cost.
3. Improving the search time and results for data based on the query.
4. Enhance the privacy for data

CONCLUSIONS

It motivates and solves the problem of secure multi keyword top-k retrieval over encrypted data in the cloud. In this work, a novel framework is proposed for the problem of multi-keyword ranked search over encrypted cloud data with supporting synonym queries, and to establish a variety of privacy requirements. Amongst various multi-keyword semantics, the efficient similarity measure is coordinate matching, in order to effectively get the relevance of outsourced documents to the query keywords, and use "inner product similarity" to quantitatively evaluate such similarity measure It propose Privacy enhanced Rank test based query retrieval are used. In this mechanism the ranked queries that retrieved by the cloud server will be tested for similarity measure. This will output whether the files from the most similarity level that are retrieved. The privacy over mechanism is also provided by restricting the cloud server to learn the information from the data set. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments shows this

proposed scheme introduces low overhead on both computation and communication.

REFERENCES

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

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

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

[4] P.Niranjan Reddy and Y.Swetha. 2013. "Techniques for Efficient Keyword Search in Cloud Computing," (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4, No. 1.

[5] Muhammad Sajid Khan, Chengliang Wang, Ayesha Kulsoom and Zabeeh Ullah. 2013. "Searching Encrypted Data on Cloud," IJCSI International Journal of Computer Science Issues, Vol. 10, No. 6, No 1.