IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 3, Ver. IV (May-Jun. 2014), PP 75-79
www.iosrjournals.org

Secure Keyword Search Using TRSE Scheme over Cloud Data

[1]A. Libana Mary, [2]Ms. R.Asmetha Jeyarani, M.Tech
[1]II M.E (CSE) Shivani Engineering College, Trichy, Tamil Nadu, India
[2]Asst.Professor/CSE Shivani Engineering College, Trichy, Tamil Nadu, India

Abstract: *Cloud computing is a pattern which is mainly necessary for outsourcing the data and also for sharing the computing resources. According to privacy it has been presented as an outsourcing of sensitive information. The searchable symmetric encryption is mainly focus on addressing the data privacy issues. The aspect of similarity relevance and scheme robustness has been formulated by the privacy issues. The server side ranking based on order-preserving encryption and it's inevitably leaks the data privacy has to be observed and it can control in the proposed scheme. Two-round searchable encryption scheme is mainly used to eliminate the leakage that supports top-k multi keyword retrieval and a two main model has been used. The vector space model is helps to provide sufficient search accuracy. The homomorphic encryption which involves the user in ranking while the majority of computing works is done on the server side by operations only on cipher text.*

Index Terms: *TRSE, Top-K, order Preserving Encryption, Vector Space Model, Homomorphic Encryption*

I. Introduction

Network security has become more important to personal computer users, organizations, and the military. Network security involves the authorization of access to data in a network, which is controlled by the network administrator. Users choose or are assigned an ID and password or other authenticating information that allows them access to information and programs within their authority. Network security is becoming of great importance because of intellectual property that can be easily acquired through the internet. There are currently two fundamentally different Networks, data networks and synchronous network comprised of switches. The internet is considered as a data network. Since the current data network consists of computer based routers, information can be obtained by special programs, such as Trojan horses planted in the routers. The synchronous network that consists of switches does not buffer data and therefore are not threatened by attackers. That is security is emphasized in data networks, such as the internet and other networks that link to the internet.

Attacks can also interfere with the system's intended function, such as viruses, worms and Trojans. The method can be caused by denial of service DoS attack. When new viruses emerge, the antivirus is updated to be able to guard against those threats. This process is the same for firewalls and intrusion detection systems. The security threats and internet protocol were analysed to determine the necessary security technology.

In cloud computing, data owners may share their outsourced data with a number of users, who might want to only retrieve the data files they are interested in. One of the most popular ways to do so is through keyword-based retrieval. Keyword-based retrieval is a typical data service and widely applied in plaintext scenarios, in which users retrieve relevant files in a file set based on keywords. However, it turns out to be a difficult task in cipher text scenario due to limited operations on encrypted data. To improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevance are sent back to users.

TABLE 1: ATTACK METHOD AND SECURITY TECHNOLOGY

Computer security attributes	Attack methods	Technology for internet security
Confidentiality	Eavesdropping Hacking, Phishing, DOS, IP spoofing	IDS, Firewall, Cryptographic Systems, IP Sec, SSL
Integrity	Viruses, Worm Trojans, Eaves dropping,DOS, Ip spoofing	IDS, Firewall, IP Sec and SSL, Anti-Malware Software
Privacy	Email Bombing, Spamming, Hacking,DOS and Cookies	IDS,Firewall, IP Sec and SSL, Anti-Malware Software
Availability	DOS,Email Bombing,Spamming,System Boot Record Infectors	IDS,Firewall, Anti-Malware Software

Common internet attacks methods are broken down into categories. Some attacks gain system knowledge or personal information, such as eavesdropping and phishing. Attacks can also interfere with the system's intended function, such as viruses, worms and Trojans. The method can be caused by denial of service DoS attack. When new viruses emerge, the antivirus is updated to be able to guard against those threats. This process is the same for firewalls and intrusion detection systems. The security threats and internet protocol were analysed to determine the necessary security technology. The security technology is mostly software based, but many common hardware devices are used. The current development in network security is not very impressive. Most of the current security algorithms are computational intensive and require substantial processing power.

Cloud Network Security:

Cloud computing, a critical pattern for advanced data service, has become a necessary feasibility for data users to outsource data. It has been incessantly presented as outsourcing of sensitive information including e-mails, health history and personal photos is explosively expanding. Reports of data loss and privacy breaches in cloud computing systems appear from time to time. When users outsource their private data onto the cloud, the cloud service providers are able to control and monitor the data and the communication between users and the cloud.

In cloud computing, data owners may share their outsourced data with a number of users, who might want to only retrieve the data files they are interested in. One of the most popular ways to do so is through keyword-based retrieval. Keyword-based retrieval is a typical data service and widely applied in plaintext scenarios, in which users retrieve relevant files in a file set based on keywords. However, it turns out to be a difficult task in cipher text scenario due to limited operations on encrypted data. To improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevance are sent back to users.

Essential Characteristics:

On demand service: Get computing capabilities as needed automatically.
Broad Network Access: Service available over the net using desktop, laptop, PDA, mobile phone
Resource pooling: Provider resources pooled to server multiple clients
Rapid Elasticity: Ability to quickly scale in/out service
Measured service: control, optimize services based on metering

Top 5 Security Risks of Cloud Computing:

Cloud computing can offer small businesses significant cost-saving benefits namely, pay-as-you-go access to sophisticated software and powerful hardware the service does come with certain security risks. When evaluating potential providers of cloud-based services, top five security concerns should be in mind.

Secure data transfer:

All of the traffic travelling between the network and whatever service accessing in the cloud must traverse the Internet. The data is always travelling on a secure channel; only connect your browser to the provider via a URL that begins with https. Also, data should always be encrypted and authenticated using industry standard protocols, such as IPSec, that have been developed specifically for protecting Internet traffic.

Secure software interfaces:

The CSA recommends that be aware of the software interfaces, or APIs, that are used to interact with cloud services. Reliance on a weak set of interfaces and APIs exposes organizations to variety of security issues related to confidentiality, integrity, availability, and accountability, says the group in its Top Threats to Cloud Computing

Secure stored data:

The data should be securely encrypted when it's on the provider's servers and while it's in use by the cloud service. The few cloud providers assure protection for data being used within the application or for disposing of data. The providers should securely dispose the data, for example, by deleting the encryption key.

User access control:

Data stored on a cloud provider's server can potentially be accessed by an employee of that company. First, consider carefully the sensitivity of the data that are allowing out into the cloud. Second, ask providers for specifics about the people who manage the data and the level of access they have to it.

Data separation:

Every cloud-based service shares resources, namely space on the provider's servers and other parts of the provider's infrastructure. Hypervisor software is used to create virtual containers on the provider's hardware for each of its customers. But CSA notes that attacks have target the shared technology inside Cloud Computing environments.

II. Related Work

The various clustering algorithm has been used for predicting the high dimensional data. The high dimensional data are regularly affected by the curse of dimensionality.

A. Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions

A single-user SSE scheme includes a client that wants to store a private document collection on an honest-but-curious server. The security techniques are Access Pattern, Search Pattern, Trace, and Non-adaptive security for SSE, Non-adaptive semantic security, Multi-User Searchable Encryption. Allows a party to outsource the storage of his data to another party in a private manner, while maintaining the ability to selectively search over it. Private-key storage outsourcing allows clients with either limited resources or limited expertise to store and distribute large amounts of symmetrically encrypted data at low cost. Since regular private-key encryption prevents one from searching over encrypted data, clients also lose the ability to selectively retrieve segments of their data. To address this, several techniques have been introduced for provisioning symmetric encryption with search capabilities the resulting construct is typically called searchable encryption.

The new adversarial models for SSE are:

First, on- adaptive only considers adversaries that make their search queries without taking into account the trapdoors and search outcomes of previous searches.Second, adaptive considers adversaries that choose their queries as a function of previously obtained trapdoors and search outcomes. All previous work on SSE (with the exception of oblivious RAMs) falls within the non-adaptive setting.

The different models for private search are:

Common to all three models is a server (sometimes called the database) that stores data and a user that wishes to access, search, or modify the data while revealing as little as possible to the server. The important differences between these three settings are:

Private-key searchable encryption:

Searching on private-key-encrypted data, the user himself encrypts the data, so he can organize it in an arbitrary way and include additional data structures to allow for efficient access of relevant data. The data and the additional data structures can then be encrypted and stored on the server so that only someone with the private key can access it.

Public-key searchable encryption:

The searching on public-key-encrypted data, users who encrypt the data (and send it to the server) can be different from the owner of the decryption key. In a typical application, a user publishes a public key while multiple senders send e-mails to the mail server. Anyone with access to the public key can add words to the index, but only the owner of the private key can generate "trapdoors" to test for the occurrence of a keyword.

Single-database PIR:

In single-database private information retrieval a user can retrieve data from a server containing unencrypted data without revealing the access pattern and with total communication less than the data size. This was extended to keyword searching, including searching on streaming data. The data in PIR is always unencrypted; any scheme that tries to hide the access pattern must touch all data items. Otherwise, the server learns information: namely, that the untouched item was not of interest to the user. Thus, PIR schemes require work which is linear in the database size.

Thus the problem has been visited incase of searchable symmetric encryption, which allows a client to store its data on a remote server in such a way that it can search over it in a private manner.

B.Secure Ranked Keyword Search over Encrypted Cloud Data

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, thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. Straightforward yet ideal construction of ranked keyword search under the state-of-the-art searchable symmetric encryption (SSE)

security definition, and demonstrate its inefficiency.Cloud Computing enables cloud customers to remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources.

The benefits brought by this new computing model include but are not limited to: relief of the burden for storage management, universal data access with independent geographical locations, and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc.

Besides, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios.

The problem has to be solved are:

1) The problem of secure ranked keyword search over encrypted cloud data, and provide such an effective protocol, which fulfills the secure ranked search functionality with little relevance score information leakage against keyword privacy.

2) Thorough security analysis shows that our ranked searchable symmetric encryption scheme indeed enjoys "as-strong as-possible" security guarantee compared to previous SSE schemes.

3) Extensive experimental results demonstrate the effectiveness and efficiency of the proposed solution. To enable ranked searchable symmetric encryption for effective utilization of outsourced cloud data under the aforementioned model, our system design should achieve the following security and performance guarantee.

Specifically, we have the following goals:

i) Ranked keyword search: to explore different mechanisms for designing effective ranked search schemes based on the existing searchable encryption framework;

ii) Security guarantee: to prevent cloud server from learning the plaintext of either the data files or the searched keywords, and achieve the as-strong-as-possible security strength compared to the existing searchable encryption schemes;

iii) Efficiency: above goals should be achieved with minimum communication and computation overhead.

c. Confidentiality-Preserving Rank-Ordered Search

A new framework for confidentiality preserving rank-ordered search and retrieval over large document collections. The practical techniques for relevance scoring methods and cryptographic techniques, such as order preserving encryption, to protect data collections and indices and provide efficient and accurate search capabilities to securely rank-order documents in response to a query.

Cryptographic encryption protects data from compromise due to theft or intrusion. In addition to outsider attacks, security measures should also be taken against potential insider attacks. For example, when information storage is out sourced to a third-party data center, system administrators and other personnel involved may not be trusted to have decryption keys and access the content of the data collections. When an authorized user remotely accesses the data collection to search and retrieve desired documents, the large size of the collections often makes it infeasible to ship all encrypted data to the user's side, and then perform decryption and search on the user's trusted computers. New techniques are needed to encrypt and organize the data collections in such a way as to allow the data center to perform efficient search in encrypted domain.

The requirements of balancing privacy and confidentiality with efficiency and accuracy pose significant challenges to the design of search schemes for a number of search scenarios. This problem has attracted interests from the cryptography community in recent years to investigate theories and techniques for "searchable encryption." Advances in information retrieval have gone well beyond Boolean searches; scoring schemes have been widely employed to quantify and rank-order the relevance of a document to a set of query terms .To explore a framework to securely rank-order documents in response to a query, and develop techniques to extract the most relevant document(s) from a large encrypted data collection.

During the search process, the query terms are encrypted to prevent the exposure of information to the data center and their intruders, and to confine the searching entity to only make queries within an authorized scope. Utilizing term frequencies and other document information, we apply cryptographic techniques such as order-preserving encryption to develop schemes that can securely compute relevance scores for each document, identify the most relevant documents, and reserve the right to screen and release the full content of relevant documents.

III. Conclusions

From the literature survey, we discussed about the existing ranking method for cluster ranking the data's. Thus The TRSE scheme generate the fully homomorphic encryption, which fulfils the security requirements of multi keyword top-k retrieval over the encrypted cloud data. By security analysis, the proposed scheme guarantees

data privacy. According to the efficiency evaluation of the proposed scheme over a real data set, extensive experimental results demonstrate that our scheme ensures practical efficiency.

References

[1] Jiadi Yu, Peng Lu, Yanmin Zhu, Guangtao Xue and Minglu Li "Towards Secure Multi- Keyword Top-k Retrieval over Encrypted Cloud Data" - IEEE TRANSACTIONS ON DEPEDABLE AND SECURE COMPUTING, 2013

[2] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A.Konwinski, G. Lee, D. Patterson, A. Rabkin, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58,2010.

[3] M. Arrington, "Gmail Disaster: Reports of Mass Email Deletions,"http://www.techcrunch.com/2006/12/28/gmaildisasterreportsof-mass-email-deletions/, Dec. 2006.

[4] Amazon.com, "Amazon s3 Availability Event: July 20, 2008,"http://status.aws.amazon.com/s3-20080720.html, 2008.

[5] RAWA News, "Massive Information Leak Shakes Washington over Afghan War,"http://www.rawa.org/ temp/runews/2010/08/20/ massive-information-leak-shakes-washington-overafghan-war.html, 2010.

[6] AHN, "Romney Hits Obama for Security Information Leakage,"http://gantdaily.com/2012/07/25/romney-hits-obama-forsecurity-information-leakage/, 2012.

[7] Cloud Security Alliance, "Top Threats to Cloud Computing," http://www.cloudsecurity alliance.org, 2010.

[8] C. Leslie, "NSA Has Massive Database of Americans'PhoneCalls,"http://usatoday30.usatoday.com/news/washington/2006-05-10/, 2013.

[9] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,"Proc. ACM 13th Conf. Computer and Comm. Security (CCS), 2006.

[10] 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), 2010.

[11] 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: Advances in Database Technology (EDBT), 2009.

[12] M. van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully Homomorphic Encryption over the Integers," Proc. 29th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques, H. Gilbert,pp. 24-43, 2010.

[13] O. Regev, "New Lattice-Based Cryptographic Constructions,"J. ACM, vol. 51, no. 6, pp. 899-942, 2004.

[14] "NSF Research Awards Abstracts 1990-2003," http://kdd.ics.uci.edu/databases/nsfabs/nsfawards.html, 2013.

[15] "20 Newsgroups," http://kdd.ics.uci.edu/databases/ 20newsgroups/20newsgroups.html, 2013.

[16] S. Gries, "Useful Statistics for Corpus Linguistics," A Mosaic of Corpus Linguistics: Selected Approaches, Aquilino Sanchez Moises Almela, eds., pp. 269-291, Peter Lang, 2010.

[17] C. Gentry, "Fully Homomorphic Encryption Using Ideal Lattices,"Proc. 41st Ann. ACM Symp. Theory of computing (STOC), pp. 169-178, 2009.

[18] D. Dubin, "The Most Influential Paper Gerard Salton Never Wrote," Library Trends, vol. 52, no. 4, pp. 748-764, 2004.

[19] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

[20] D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, "Public- Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (Eurocrypt), 2004.