NOVEL MEDICAL DATA SHARING AND MINING TECHNIQUE FOR ELECTRONIC HEALTH RECORD SYSTEM

Dr. R. Gunasundari[1], *L. Thara*[2], *V. Sreelakshmi*[3]

Abstract

By stepping into the era of big data, Internet users usually choose to upload their personal data to remote cloud servers such that they can reduce the cost of local data management and maintenance. But they also throw up additional challenges. Unluckily, the thrive on cloud services can cause a bottleneck if data end up parked on several clouds and thus still need to be moved to be shared. The proposed system leads to a trend that data owners prefer to remotely outsource their data to clouds for the enjoyment of the high-quality retrieval and storage service without worrying the burden of local data management and maintenance. However, secure share and search for the outsourced data is a formidable task when it comes to a medical department, which may easily incur the leakage of sensitive personal information about victims in a hospital. Efficient data sharing and searching with security is of critical importance. This system, suggests a novel medical data sharing and mining system for such medical records. When compared to current systems which support either searchable attribute-based functionality or attribute-based proxy re-encryption, our new primitive supports both abilities and provides flexible keyword update service. Specifically, the system enables a data user to efficiently share the patient data to a specified group of other users matching a sharing policy here the data users are known as doctors and meanwhile, the data will maintain its searchable property but also the corresponding search keyword(s) can be updated after the data sharing. The new mechanism is applicable to many real-world applications, such as electronic health record systems. It is also proved chosen cipher text secure in the random oracle model.

Keywords : cloud, mining, secure, encryption, privacy, search, share.

I. Introduction

More and more enterprises/individuals are beginning to outsource their resident data to the cloud servers with the rapid growth of cloud computing. In addition to individuals, many industries and research institutions also follow the trend to remotely store commercial and scientific data to clouds to enjoy high speed data process and retrieval service. However, it meanwhile unavoidably encounters with many unpredictable security and privacy challenges. They face enormous security and privacy risks under open networks and not fully reliable cloud environments. That is, there may be data leakage or disclosure, data

[1]Associate Professor & Head, Department of Information Technology, Karpagam University, E-mail : gunasoundar04@gmail.com,

[2] Research Scholar in Computer Science, Karpagam University, E-mail : kltharavijay@gmail.com,

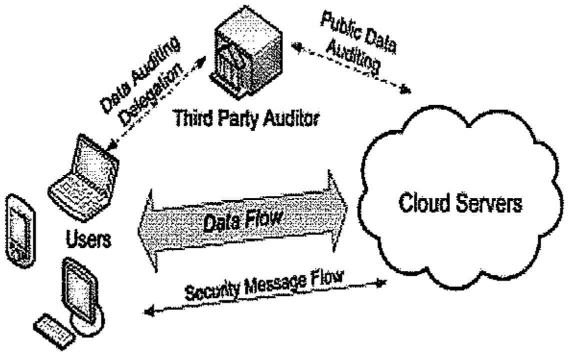
[3] PG Student, Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamil Nadu, INDIA. E-mail sreeaugust01@gmail.com

corruption or loss, and user privacy breach when outsourcing their data to a public cloud or using their outsourced data.

In recent times, several studies were conducted to address these risks, and a series of results were offered to enable data and privacy protection in unreliable cloud environments. The standing systems can be categorized into two types: (i) Searchable symmetric key encryption and (ii) Searchable Public Key Encryption (SPKE). Here, if we regard an attribute as a search keyword, the privacy of the keyword cannot be achieved as the system is built in the attribute publicly known model. It is unknown that if we can employ anonymous ABE technique to yield both data share and search as well as keyword privacy. The drawback is that, Data sharing and keyword privacy cannot be maintained at the same time.

II. System Model of Cloud Data Services

A cloud is referred to as an untrusted cloud environment when its resources and services are open for public use and communication is performed over a untrusted network. Generally, a public cloud (e.g., Amazon AWS, Microsoft Azure, and Google Cloud Platform) is not fully trusted by users. [2] Therefore, although the benefits of this new cloud service paradigm are tremendous, serious security risks and privacy challenges are raised under untrusted cloud environments. The report from the Cloud Security Alliance (CSA) [9] shows that the data security problems are among the top threats in the cloud. Additionally, when users use cloud data services, privacy breaches often occur due to undesirable interference from internal and external adversaries. Thus, it can be seen that cloud data services are intrinsically not secure from the viewpoint of cloud users.



The user is the party that consumes the cloud data, for example, retrieving the specific data, getting the data computing results, and accessing the shared data. Cloud Servers / Cloud Service Provider (CSP), who has significant resources and expertise in building and managing distributed cloud storage servers, owns and operates live Cloud Computing systems. A third party may be involved to provide security functionalities in the cloud, such as an Attribute Authority (AA) and Third Party Auditor (TPA). An AA is a trusted key authority in an attribute-based access control system [3]. It takes charge of generating attribute keys for users according to their identity and updating or revoking users' attribute keys when their roles change. A TPA is a semi-trusted party from the viewpoint of users in a public data verification Scheme [7],[10]. It may be trusted to check the correctness of data stored in the cloud on behalf of users, but it has no privilege to access the actual content of data.

III. Overview of Security Solutions

In unreliable cloud environments, a challenging problem is to empower fine-grained [4] implementations of data access and accomplish secure data sharing among large-scale users. Apparently, since an owner does not trust the cloud, traditional access control mechanisms, normally depending on a trusted server, are not apt for cloud data sharing. To address this challenge and impose owner controlled access control, data should be encrypted by the owner before outsourcing it to the cloud, and then the owner can implement fine-grained access control over the encrypted data by securely distributing keys. Based on this idea, access control based on encryption is proposed. Two typical mechanisms in this direction are access control based on selective encryption [5] and Attribute- Based Encryption (ABE) [5], respectively.

To propose a secure system, we start with Attribute-Based Encryption (ABE) with a significant reason that it provides fine-grained expressiveness in medical data share and search. Here the system stores all medical reports shared by the patient's will be stored in a free cloud, that is drop box cloud service and it provides free storage only up to 2GB. After storing the data to a cloud server, the data owner that is doctor usually needs two necessary operations: one is data searching, and the other is data sharing with lab members.

Advantages of Drop Box

➢ Here they always securely back up your work so you don't have to worry about ever losing the data's.

➢ Dropbox saves your files throughout all of your devices. Know that all your work is readily available to you no matter what device you are using at the moment or where you are.

➢ Any deleted files can be restored from the cloud, and the deleted files remains in cloud for 30days.

➢ Dropbox will also allow the user to share their media files like photos and videos directly to the social websites like face book, Google circle etc.

From the above discussions, we can see the importance of secure searching and sharing for encrypted data in remote cloud storage scenario. Protecting the privacy of search data and keyword but also supporting efficient encrypted data sharing in the context of Attribute-Based Encryption that is an interesting problem in this literature. This motivates our work.

The system proposes the following techniques :

- ATTRIBUTE-BASED KEYWORD SEARCH (ABKS)
- ATTRIBUTE-BASED PROXY RE-ENCRYPTION (ABPRE)
- KEYWORD SEARCH

3.1 Attribute - Based Keyword Search (ABKS)

To hide search contents as well as searched keywords from cloud server, we introduced the notion of Public Key Encryption (PKE) with keyword search, in which a user delivers a special token associated with keyword(s) to the server such that the server can use the token to allocate all encrypted data with the same keyword(s). The server, however, knows nothing about the keyword(s) and the data. To make it clear and understandable we will consider that a doctor is searching about a symptom, now the token will retrieves the diseases related to that symptom keyword. This is how the mining process takes place in searching a keyword.

3.1 Attribute - Based Proxy Re- Encryption (ABPRE)

To efficiently share an encrypted data with others, we introduced PRE whereby a semi trusted proxy can transform an encryption of a message to another encryption of the same message without knowing the message. To employ the notion into ABE setting, we proposed the notion of ABPRE with stronger security. So the patient can send his/her medical report to the doctor by encrypting it using the advanced encryption standard (AES). Now the confidential encrypted medical report can be shared to laboratory authorities for future diagnosing and it can achieved successfully by sending a correct encryption key.

Advantages of AES

- AES algorithm works on the principle of substitution permutation network.
- AES operates on a 4x4 matrix of bytes termed as a state.
- Each round consists of several processing steps, including one that depends on the encryption key.
- A set of reverse rounds are applied to transform cipher text back into the original pain text using the same encryption key.
- Advanced encryption standard not only assures security but also improves the performance in a variety of settings such as smart cards, hardware implementations etc.
- AES is federal information processing standard and there are currently no know non-brute-force direct attacks against AES.
- AES is strong enough to certified for used by the US government for top secret information.

3.3 Keyword Search

Usually, an ABKS supporting keyword search does not simultaneously provide decryption service. This is due to a technical limitation in the construction method of trapdoor token (used for searching). On the other hand, an ABPRE system is not compatible with secure data search. Specifically, if we regard an attribute as a search keyword, the privacy of the keyword cannot be achieved as the system is built in the attribute publicly known model. One might question that if we can influence existing anonymous ABE systems.

The following steps are performed to execute a data mining application through the Data Mining Cloud App [1]

1) The user accesses the Website and submits his/her data mining application, by specifying: location of the input dataset, name of the data mining algorithm, and values of its parameters.

2) The Website inserts a set of tasks into the Task Queue on the basis of the data mining application submitted by the user

3) Each idle Worker picks a task from the Task Queue, and starts its execution on a virtual server.

4) Each Worker gets the input dataset from the location specified by the user. To this end, a file transfer is performed from the Drop box where the dataset is located to the local storage of the virtual server the Worker is running on.

5) After task completion, each Worker puts the result on a Drop box.

Here data mining concept is achieved by Classification and Regression Trees algorithms. It can also be known as CART algorithm. The classification techniques of data mining help to classify the data on the basis of certain rules. This helps to frame policies for the future [1].

Advantages of CART

- Decision trees implicitly perform variables screening are feature selection.
- Decision trees require relatively little effort from users for data preparation.
- Non-linear relationships' between parameters do not affect tree performance.
- The Best feature of using trees for analytics – easy to interpret and explain to executives.

IV. Results and discussion

Data storage integrity is one of the challenging tasks in the cloud. Thus, in [6] author proposes a novel approach for overcome this data integrity issue by using remote data integrity checking protocol, which is based on RSA and HLA signature with the support of public verification. This public verification creates the protocol very flexible. Since the user can direct the data possession to check the TPA. In addition, Tate et al. [8] proposed a scheme for verifying the integrity of multi-user data with the help of trusted hardware. However, these three schemes are not able to preserve the identities of data owners.

It is worth mentioning that all existing public key systems with keyword search fail to guarantee this property. But anyhow the data's get stored in the database under the specific table.

For the first time, a novel and practical notion is introduced, using searchable ABPRE. Our notion guarantees that the keyword search ability of a cipher text can be remained after the sharing of the cipher text. We design a concrete searchable Key-Policy (KP) ABPRE system satisfying the above notion. We also prove the scheme chosen cipher text secure in the Random Oracle Model (ROM).

The scheme is the first of its type supporting the privacy of keyword search but also encrypted data sharing. As of independent interest, our protocol supports keyword update so that a cipher text's keyword can be further updated before the cipher text is shared with others. This property brings a convenience to data owner (who can gain access to the data) in the sense that the cipher text keyword can be freely modified based on data share record.

V. Conclusion

In this paper we introduce a novel medical data sharing and mining technique that has been designed mainly to avoid the frequency level of patients visiting a specific doctor for their regular medical check-up. The main aim of this system is to give more favor for the patients at the time of their sickness.

The system enables a data owner to efficiently share his/her data to a specified group of users matching a sharing policy and meanwhile, the data will maintain its searchable property but also the corresponding search keyword(s) can be updated after the data sharing. Our system has better efficiency regarding to keyword search and decryption phases. Our solution not only minimizes the computation but also guarantees the trust placed on it in terms of data privacy and identity privacy.

The system has been proposed to help the patient's to get prescription for their ailment directly from the respective doctor in a department. This system is very useful for doctors too for searching a keyword related to any diseases for their reference and research purpose. This system provides a better interaction between both doctors and patients, due to the user friendly environment, and also simple menus helps every basic computer users to perform efficient activities in this system. Even senior citizens can easily use this system.

VI. Scope for Further enhancement

This system can be further improved by developing it as mobile application in different platforms like Android, Windows, IOS, and Blackberry. Features based on the medical essential and technology can be inserted in future versions, which also enhances user friendly methods especially for patient's interaction with the doctor. The limitations faced by this system can be easily rectified in case a standard procedure is adapted. This can be implemented with minor changes to use in various areas.

References

[1] Astha Pareek, Manish Gupta *"Review of Data Mining Techniques in Cloud Computing Database"*, International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-2 Number-2 Issue-4, June, 2012.

[2] *Amazon. 2015. AWS Security Center (2015).* http://aws.amazon.com/security/. Amazon, 2015 & Microsoft Azure Trust Center (2015) http://azure.microsoft.com/en-us/support/trust-center/.

[3] Jingwei Li, Chunfu Jia, Jin Li, and Xiaofeng Chen, *"Outsourcing encryption of*

attribute-based encryption with map reduce", Information and Communications Security, Springer, 191–201, 2012.

[4] Jin Li, Xiaofeng Chen, Jingwei Li, Chunfu Jia, Jianfeng Ma, and Wenjing Lou, *"Fine-grained access control system based on outsourced attribute-based encryption"*, Computer Security (ESORICS'13), Springer, 592–609, 2013.

[5] John Bethencourt, Amit Sahai, and Brent Waters., *"Ciphertext-policy Attribute-Based Encryption"*, Proceedings of IEEE Symposium on Security and Privacy (S&P). IEEE, 321–334, 2007.

[6] Kandukuri .B.R, Paturi .V.R, Rakshit .A, *"Cloud Security Issues,"* IEEE International Conference on Services Computing, 21-25 pp. 517-520, 2009.

[7] Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou, *"Privacy-preserving multi-keyword ranked search over encrypted cloud data"*, IEEE Transactions on Parallel and Distributed Systems (TPDS) 25, 1 (2014), 222–233, 2014.

[8] Tate .S.R, Vishwanathan .R, and Everhart .L, *"Multi-user Dynamic Proofs of Data Possession Using Trusted Harware,"* in ACM CODASPY, pp. 353–364, 2013.

[9] *"The Notorious Nine: Cloud Computing Top Threats in 2013"*, https://downloads.cloudsecurityalliance.org/initiatives/top_threats/TheNotorious_Nine_Cloud_Computing_Top_Threats_in_2013.pdf, Feb. 2013.

[10] Wenhai Sun, Shucheng Yu, Wenjing Lou, Y. Thomas Hou, and Hui Li, "Protecting your right: Attributebased keyword search with fine-grained owner-enforced search authorization in the cloud", Proceedings of IEEE International Conference on Computer Communications (INFOCOM'14), IEEE, 2014.

Author's Biography

Dr. R. Gunasundari received the Ph.D. Degree in Computer Science from Karpagam University, Coimbatore in 2014. She is working as an Associate Prof & Head in the Department of Information Technology, Karpagam University, Coimbatore. She has published 20 National and 21 International papers in various journals. Her broad field of research is in Data mining.