Asian Journal of Engineering and Applied Technology
ISSN: 2249-068X (P) Vol. 3 No. 1, 2014, pp. 9-15

DOI: https://doi.org/10.51983/ajeat-2014.3.1.707

Survey on Enhancing Privacy Data Sharing Between Multi-Owners in Cloud for Dynamic Groups

A.Balaji[1], M.Lokesh[2], M.Sakthivel[3], and P.Anbumani[4]

[1, 2, 3] Department of CSE, [4] Assistant Professor, Department of CSE,
Tagore Institute of Engineering and Technology, Tamil Nadu, India
E-mail: balaji81192@gmail.com
(Received on 18 March 2014 and accepted on 15 June 2014)

Abstract **- With the character of low maintenance, cloud computing provides an economical and efficient solution for sharing group resource among cloud users. Unfortunately, sharing data in a multi-owner manner while preserving data and identity privacy from an untrusted cloud is still a challenging issue, due to the frequent change of the membership. In this paper, we propose a secure multiowner data sharing scheme, named Mona, for dynamic groups in the cloud. By leveraging group signature and dynamic broadcast encryption techniques, any cloud user can anonymously share data with others. Meanwhile, the storage overhead and encryption computation cost of our scheme are independent with the number of revoked users.**

Keywords: **Data Sharing, Cloud Users**

I. Introduction

CLOUD computing is recognized as an alternative to traditional information technology due to its intrinsic resource-sharing and low-maintenance characteristics. In cloud computing, the cloud service providers (CSPs), such as Amazon, are able to deliver various services to cloud users with the help of powerful data centres. By migrating the local data management systems into cloud servers, users can enjoy high-quality services and save significant investments on their local infrastructures. One of the most fundamental services offered by cloud providers is data storage. Let us consider a practical data application. A company allows its staffs in the same group or department to store and share files in the cloud. By utilizing the cloud, the staffs can be completely released from the troublesome local data storage and maintenance. However, it also poses a significant risk to the confidentiality of those stored files. Specifically, the cloud servers managed by cloud providers are not fully trusted by users while the data files stored in the cloud may be sensitive and confidential, such as business plans. To preserve data privacy, a basic solution is to encrypt data files, and then upload the encrypted data into the cloud . Unfortunately, designing an efficient and secure data sharing scheme for groups in the cloud is not an easy task due to the following challenging issues. This work extends into government based application as follows, Government agencies and other organizations often need to publish micro data, e.g., medical data or census data, for research and other purposes. Typically, such data are stored in a table, and each record (row) corresponds to one individual. Each record has a number of attributes, which can be divided into the following three categories:

1) Attributes that clearly identify individuals. These are known as explicit identifiers and include, e.g., Social Security Number.

2) Attributes whose values when taken together can potentially identify an individual. These are known as quasi-identifiers, and may include, e.g., Zip code, Birth-date, and Gender.

3) Attributes that are considered sensitive, such as Disease and Salary. When releasing micro data, it is necessary to prevent the sensitive information of the individuals from being disclosed.

Two types of information disclosure have been identified in the literature: identity disclosure and attribute disclosure. Identity disclosure occurs when an individual is linked to a

particular record in the released table. Attribute disclosure occurs when new information about some individuals is revealed, i.e., the released data make it possible to infer the characteristics of an individual more accurately than it would be possible before the data release. Identity disclosure often leads to attribute disclosure. Once there is identity disclosure, an individual is re identified and the corresponding sensitive values are revealed. Attribute disclosure can occur with or without identity disclosure. It has been recognized that even disclosure of false attribute information may cause harm. An observer of a released table may incorrectly perceive that an individual's sensitive attribute takes a particular value and behaves accordingly based on the perception. This can harm the individual, even if the perception is incorrect. While the released table gives useful information to researchers, it presents disclosure risk to the individuals whose data are in the table. Therefore, our objective is to limit the disclosure risk to an acceptable level while maximizing the benefit. This is achieved by anonymizing the data before release. The first step of anonymization is to remove explicit identifiers. However, this is not enough, as an adversary may already know the quasi-identifier values of some individuals in the table. This knowledge can be either from personal knowledge (e.g., knowing a particular individual in person), or from other publicly available databases (e.g., a voter registration list) that include both explicit identifiers and quasi-identifiers. A common anonymization approach is generalization, which replaces quasiidentifier values with values that are less-specific but semantically consistent. As a result, more records will have the same set of quasi-identifier values.

We define an equivalence class of an anonymized table to be a set of records that have the same values for the quasi-identifiers. To effectively limit disclosure, we need to measure the disclosure risk of an anonymized table. To this end, Samarati and Sweeney introduced k-anonymity as the property that each record is indistinguishable with at least k-1 other records with respect to the quasi-identifier. In other words, k-anonymity requires that each equivalence class contains at least k records. While k-anonymity protects against identity disclosure, it is insufficient to prevent attribute disclosure. To address this limitation of k-anonymity, One problem with l-diversity is that it is limited in its assumption of adversarial knowledge. As we shall explain below, it is possible for an adversary to gain information about a sensitive attribute as long as she has information about the global distribution of this attribute. This assumption generalizes the specific background and homogeneity attacks used to motivate ' diversity. Another problem with privacy-preserving methods, in general, is that they effectively assume all attributes to be categorical; the adversary either does or does not learn something sensitive. Of course, especially with numerical attributes, being close to the value is often good enough. In this paper, we propose a novel privacy notion called "closeness." We first formalize the idea of global background knowledge and propose the base model t-closeness which requires that the distribution of a sensitive attribute in any equivalence class to be close to the distribution of the attribute in the overall table (i.e., the distance between the two distributions should be no more than a threshold t). This effectively limits the amount of individual-specific information an observer can learn.

II. Cloud Data Record Linkage

The last few decades have witnessed a tremendous increase in the use of computerized databases for supporting a variety of business decisions. The data needed to support these decisions are often scattered in heterogeneous distributed databases. In such cases, it may be necessary to link records in multiple databases so that one can consolidate and use the data pertaining to the same real world entity. If the databases use the same set of design standards, this linking can easily be done using the primary key (or other common candidate keys).

However, since these heterogeneous databases are usually designed and managed by different organizations (or different units within the same organization), there may be no common candidate key for linking the records. Although it may be possible to use common non key attributes (such as name, address, and date of birth) for this purpose, the result obtained using these attributes may not always be accurate. This is because non key attribute values may not match even when the records represent the same entity instance in reality. Record linkage techniques have been widely used in real-world situations - such as health care immigration and census where all the records are available

locally. However, when the matching records reside at a remote site, existing techniques cannot be directly applied because they would involve transferring the entire remote relation, thereby incurring a huge communication overhead.

As a result, record linkage techniques do not have an efficient implementation in an online, distributed environment and have mostly been confined to either local master files or to matching data from various sources in a batch processing mode. The databases exhibiting entity heterogeneity are distributed, and it is not possible to create and maintain a central data repository or warehouse where pre-computed linkage results can be stored. A centralized solution may be impractical for several reasons.

First, if the databases span several organizations, the ownership and cost allocation issues associated with the warehouse could be quite difficult to address. Second, even if the warehouse could be developed, it would be difficult to keep it up-to-date. As updates occur at the operational databases, the linkage results would become stale if they are not updated immediately. This staleness maybe unacceptable in many situations, For instance, in a criminal investigation, one maybe interested in the profile of crimes committed in the last 24 hours within a certain radius of the crime scene. In order to keep the warehouse current, the sites must agree to transmit incremental changes to the data warehouse on a real-time basis.

A few years ago, the health insurance companies and the medical providers in the area agreed to automate the entire process of claims filing, handling, payment, and notification. In the automated process, medical service provider files health insurance claims electronically using information (about patients and services provided) stored in the provider database. A specialized computer program at the insurance company then processes each claim, issues payments to appropriate parties, and notifies the subscriber. The insurance companies request that their subscribers inform them of the existence of (and changes in) secondary coverage. However, many subscribers forget to send the appropriate notification to update the subscriber database.

To complicate things further, different employers use different calendars for open enrollment some use the calendar year, others use the fiscal year, and many academic institutions use the academic year. Furthermore, subscribers often change jobs, and their insurance coverage's change accordingly. With rapid economic growth and the proliferation of double income families around the city, each insurance company receives several thousand updates per day to their subscriber database. However, since not all updates are propagated across the companies, stories of mishandled claims are quite common.

III. Cloud Data Sharing and Privacy

Storing data in the cloud has become a trend. An increasing number of clients store their important data in remote servers in the cloud, without leaving a copy in their local computers. Sometimes the data stored in the cloud is so important that the clients must ensure it is not lost or corrupted. While it is easy to check data integrity after completely downloading the data to be checked, downloading large amounts of data just for checking data integrity is a waste of communication bandwidth. Hence, a lot of works have been done on designing remote data integrity checking protocols, which allow data integrity to be checked without completely downloading the data. Remote data integrity checking is independently propose RSA-based methods for solving this problem. After that Shah et al. propose a remote storage auditing method based on pre-computed challenge-response pairs. Recently many works focus on providing three advanced features for remote data integrity checking protocols: data dynamics, public verifiability and privacy against verifiers. The protocols in support data dynamics at the block level, including block insertion.

The new data storage paradigm in "Cloud" brings about many challenging design issues which have profound influence on the security and performance of the overall system. One of the biggest concerns with cloud data storage is that of data integrity verification at untrusted servers. For example, the storage service provider, which experiences Byzantine failures occasionally, may decide to hide the data errors from the clients for the benefit of their own. What is more serious is that for saving money and storage space the service provider might neglect to keep or deliberately delete rarely accessed data files which belong to an ordinary client. Consider the large size of the outsourced electronic data and the client's constrained resource capability, the core of the

problem can be generalized as how can the client find an efficient way to perform periodical integrity verifications without the local copy of data files. Although schemes with private auditability can achieve higher scheme efficiency, public auditability allows anyone, not just the client (data owner), to challenge the cloud server for correctness of data storage while keeping no private information. Then, clients are able to delegate the evaluation of the service performance to an independent third party auditor (TPA), without devotion of their computation resources. In the cloud, the clients themselves are unreliable or may not be able to afford the overhead of performing frequent integrity checks. In Cloud Computing, the remotely stored electronic data might not only be accessed but also updated by the clients, e.g., through block modification, deletion, insertion, etc. Unfortunately, the state of the art in the context of remote data storage mainly focus on static data files and the importance of this dynamic data updates has received limited attention so far.

IV. Privacy & Security Law

In the world of cloud computing, data is collected for a wide array of purposes, from people in different jurisdictions, and according to the policies of organizations that may differ widely in their business models, culture and technology applications. When data resides and is processed in the cloud, what data protection and privacy laws apply? Is data stored in the cloud transferred internationally? How is that determination made? The authors explain what cloud computing is, how it is used, and delve into some of the special privacy challenges raised by computing in the cloud. They raise a governance model based on the principle of accountability as a possible way forward.

Companies are moving to the cloud, and for good reason. Cloud computing makes it possible for companies to relocate their IT and its management and maintenance outside of their organizations. Because payment for many cloud computing services is based on a utility (like electricity, or mobile phone) or subscription model, customers only pay for what they use. Freed from the need to acquire service and maintain their IT infrastructure, businesses become more nimble, better able to adapt to changing market demands and to take advantage of service more effectively and economically provided by others. Cloud computing empowers companies to redirect resources toward core functions and competencies.

Cloud computing promises not only cost savings and efficiencies, but the ability to expand and enhance services. Health care delivery provides an important example. Cloud computing would allow a number of hospitals to share infrastructure and link systems to reduce costs and increase efficiencies. By pooling various IT resources into the cloud, hospitals could increase utilization as resources would be delivered only when they are required. The cloud also would provide real-time availability of patient information for doctors, nursing staff and support services, not only nationally but regionally, without regard to country borders. Medical professionals would be empowered to access patient information for consultation and research from any Internet enabled device without special software. Major health care institutions, such as the Cleveland Clinic and Kaiser Permanente, have already entered into partnerships with cloud computing providers to begin to move into the cloud.

Security concerns may be magnified by the dynamic nature of the cloud environment. Indeed, business' ability to benefit from the speed with which the cloud vendors can adjust, develop and change their offerings is one of the cloud's key advantages. That very speed and flexibility may raise concerns that they come at the cost of a certain level of security. The issue of data security features prominently in a European data protection context, where the data controller remains responsible for the collection and processing of personal data, even where the data are processed by a third party. The EU Directive requires the controller to ensure that any third party processing personal data on its behalf takes adequate technical and organizational security measures to safeguard the data. European data protection law requires a contractual provision in between the controller and processor to this effect, and controllers typically seek to monitor whether this obligation is fulfilled by undertaking an audit or conducting due diligence inquiries.

V. Provable Data Possession at Un Trusted Stores

We introduce a model for provable data possession (PDP) that allows a client that has stored data at an untrusted server

to verify that the server possesses the original data without retrieving it. The model generates probabilistic proofs of possession by sampling random sets of blocks from the server, which drastically reduces I/O costs. The client maintains a constant amount of metadata to verify the proof. The challenge/response protocol transmits a small, constant amount of data, which minimizes network communication. Thus, the PDP model for remote data checking supports large data sets in widely-distributed storage systems. We present two provably-secure PDP schemes that are more efficient than previous solutions, even when compared with schemes that achieve weaker guarantees. In particular, the overhead at the server is low (or even constant), as opposed to linear in the size of the data. Experiments using our implementation verify the practicality of PDP and reveal that the performance of PDP is bounded by disk I/O and not by cryptographic computation.

Archival network storage presents unique performance demands. Given that file data are large and are stored at remote sites, accessing an entire file is expensive in I/O costs to the storage server and in transmitting the file across a network. Reading an entire archive, even periodically, greatly limits the scalability of network stores. (The growth in storage capacity has far outstripped the growth in storage access times and bandwidth). Furthermore, I/O incurred to establish data possession interferes with on-demand bandwidth to store and retrieve data. We conclude that clients need to be able to verify that a server has retained file data without retrieving the data from the server and without having the server access the entire file.

VI. Data Privacy Through Optimal K-Anonymization

Industries, organizations, and governments must satisfy demands for electronic release of information in addition to demands of privacy from individuals whose personal data may be disclosed by the process. As argued by Samarati and Sweeney, naive approaches to de-identifying microdata are prone to attacks that combine the data with other publicly available information to re-identify represented individuals. While no record contains any single identifying value, many records are likely to contain unique value combinations. Imagine for instance a represented individual who is the only male born in 1920 living in some sparsely populated area. This individual's age, gender, and zip code could be joined with a voter registry from the area to obtain his name, revealing his medical history. To avoid such so-called linking attacks while preserving the integrity of the released data, Samarati and Sweeney have proposed the concept of k-anonymity. A k-anonymized dataset has the property that each record is indistinguishable from at least k-1 other records within the dataset. The larger the value of k, the greater the implied privacy since no individual can be identified with probability exceeding 1/k through linking attacks alone.

The process of k-anonymizing a dataset involves applying operations to the input dataset including data suppression and cell value generalization. Suppression is the process of deleting cell values or entire tuples. Generalization involves replacing specific values such as a phone number with a more general one, such as the area code alone. Unlike the outcome of other disclosure protection techniques that involve condensation, data scrambling and swapping, or adding noise, all records within a k-anonymized dataset remain truthful. De-identifying data through common formulations of k-anonymity is unfortunately NP-hard if one wishes to guarantee an optimal anonymization. A practical method for determining an optimal k-anonymization of a given dataset is proposed. An optimal anonymization is one which perturbs the input dataset as little as is necessary to achieve k-anonymity, where "as little as is necessary" is typically quantified by a given cost metric. Several different cost metrics have been proposed, though most aim in one way or another to minimize the amount of information loss resulting from the generalization and suppression operations that are applied to produce the transformed dataset.

The ability to compute optimal anonymizations is more definitively investigate the impacts of various coding techniques and problem variations on anonymization quality. It also allows to better quantify the effectiveness of stochastic or other non-optimal methods.The experiments to illustrate the feasibility of this approach is performed. To demonstrate that despite the problem's inherent hardness, provably optimal k-anonymizations can be obtained for real census data under two representative cost metrics in most cases within only a few seconds or minutes. Some parameter settings (specifically, very small values of) remain challenging; but even under these conditions, the

algorithm can still be used to produce good solutions very quickly, and constantly improving solutions throughout its execution. This anytime quality allows it to be used to obtain good anonymizations even when an optimal anonymization is out of reach.

VII. Existing System

Several security schemes for data sharing on untrusted servers have been proposed. In these approaches, data owners store the encrypted data files in untrusted storage and distribute the corresponding decryption keys only to authorized users. Thus, unauthorized users as well as storage servers cannot learn the content of the data files because they have no knowledge of the decryption keys. Designing an efficient and secure data sharing scheme for groups in the cloud is not an easy task. Identity privacy is one of the most significant obstacles for the wide deployment of cloud computing. Without the guarantee of identity privacy, users may be unwilling to join in cloud computing systems because their real identities could be easily disclosed to cloud providers and attackers. On the other hand, unconditional identity privacy may incur the abuse of privacy. For example, a misbehaved staff can deceive others in the company by sharing false files without being traceable. Therefore, traceability, which enables the group manager (e.g., a company manager) to reveal the real identity of a user, is also highly desirable. it is highly recommended that any member in a group should be able to fully enjoy the data storing and sharing services provided by the cloud, which is defined as the multiple-owner manner. Compared with the single-owner manner, where only the group manager can store and modify data in the cloud, the multiple-owner manner is more flexible in practical applications. More concretely, each user in the group is able to not only read data, but also modify his/her part of data in the entire data file shared. Last but not least, groups are normally dynamic in practice, e.g., new staff participation and current employee revocation in a company. The changes of membership make secure data sharing extremely difficult.

- Designing an efficient and secure data sharing scheme for groups in the cloud is not an easy task.
- Groups are normally dynamic in practice. These changes of membership make secure data sharing extremely difficult.
- In existing approaches, data owners store the encrypted data files in untrusted storage and distribute the corresponding decryption keys only to authorized users. Thus, unauthorized users as well as storage servers cannot learn the content of the data files because they have no knowledge of the decryption keys.

VIII. Proposed System

This paper proposes a secure multi-owner data sharing scheme. It implies that any user in the group can securely share data with others by the untrusted cloud. Our proposed scheme is able to support dynamic groups efficiently. Specifically, new granted users can directly decrypt data files uploaded before their participation without contacting with data owners. User revocation can be easily achieved through a novel revocation list without updating the secret keys of the remaining users. The size and computation overhead of encryption are constant and independent with the number of revoked users. We provide secure and privacy-preserving access control to users, which guarantees any member in a group to anonymously utilize the cloud resource. Moreover, the real identities of data owners can be revealed by the group manager when disputes occur.

- Secure multi-owner data sharing scheme in healthcare record.
- Any user in the group can securely share data with others by the untrusted cloud.
- Provide secure and privacy-preserving access control to users, which guarantee any member in a group to anonymously utilize the cloud resource.

IX. Process Model

Every hospital databases has their own login page. The census database has the separate login to view the reports. The hospital databases are integrated. The databases are interconnected to get the valid reports for the classification. Every hospital can make their websites here to get interconnected with the census database to get the real and valid classification reports. Also tracking the patients is

very easy. The patient can register to this site and can get the security key. The security key only gives the permission to view the patient histories and the treatments and medicines which given already. With the use of username and password, the user only can view the personal information. So the security key is very important. The key is uniquely provided by the administrator.

The census database is classified by the methods, k-anonymity and the l-diversity. From these methods, the quantity of the patients who all affected by and with the disease information can be obtained. The authorized person only able to view the reports by the integrated databases. The classification divided by disease, pin code, age based. So the aggregation made the valid report for the future proposals. This proposal only shows the quantity of the diseases by hierarchy. It will not show any personal information about the patients. The security made this into effective and Data integration architecture.

X. Conclusion

In this paper, we design a secure data sharing scheme, Mona, for dynamic groups in an untrusted cloud. In Mona, a user is able to share data with others in the group without revealing identity privacy to the cloud. Additionally, Mona supports efficient user revocation and new user joining. More specially, efficient user revocation can be achieved through a public revocation list without updating the private keys of the remaining users, and new users can directly decrypt files stored in the cloud before their participation. Moreover, the storage overhead and the encryption computation cost are constant.

References

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

[2] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. Int'l Conf. Financial Cryptography and Data Security (FC), pp. 136-149, Jan. 2010.

[3] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing," Proc. IEEE INFOCOM, pp. 534-542, 2010.

[4] M. Kallahalla, E. Riedel, R. Swaminathan, Q. Wang, and K. Fu, "Plutus: Scalable Secure File Sharing on Untrusted Storage," Proc. USENIX Conf. File and Storage Technologies, pp. 29-42, 2003.

[5] E. Goh, H. Shacham, N. Modadugu, and D. Boneh, "Sirius: Securing Remote Untrusted Storage," Proc. Network and Distributed Systems Security Symp. (NDSS), pp. 131-145, 2003.

[6] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved Proxy Re-Encryption Schemes with Applications to Secure Distributed Storage," Proc. Network and Distributed Systems Security Symp. (NDSS), pp. 29-43, 2005.

[7] R. Lu, X. Lin, X. Liang, and X. Shen, "Secure Provenance: The Essential of Bread and Butter of Data Forensics in Cloud Computing," Proc. ACM Symp. Information, Computer and Comm. Security, pp. 282-292, 2010.

[8] B. Waters, "Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization," Proc. Int'l Conf. Practice and Theory in Public Key Cryptography Conf. Public Key Cryptography, http://eprint.iacr.org/2008/290.pdf, 2008.

[9] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data," Proc. ACM Conf. Computer and Comm. Security (CCS),pp. 89-98, 2006.

[10] D. Naor, M. Naor, and J.B. Lotspiech, "Revocation and Tracing Schemes for Stateless Receivers," Proc. Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO), pp. 41-62, 2001.

[11] D. Boneh and M. Franklin, "Identity-Based Encryption from the Weil Pairing," Proc. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO), pp. 213-229, 2001.

[12] D. Boneh, X. Boyen, and H. Shacham, "Short Group Signature," Proc. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO), pp. 41-55, 2004.

[13] D. Boneh, X. Boyen, and E. Goh, "Hierarchical Identity Based Encryption with Constant Size Ciphertext," Proc. Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), pp. 440-456, 2005.

[14] C. Delerablee, P. Paillier, and D. Pointcheval, "Fully Collusion Secure Dynamic Broadcast Encryption with Constant-Size Ciphertexts or Decryption Keys," Proc. First Int'l Conf. Pairing-Based Cryptography, pp. 39-59, 2007.