# Dual Locks : Partial sharing of health documents in Cloud

Mahmood Ahmad[1], Zeeshan Pervez[2], and Sungyoung Lee[1]

[1] Department of Computer Engineering, Kyung Hee University South Korea
rayemahmood,sylee@oslab.khu.ac.kr
[2] School of Computing, University of the West of Scotland, Paisley, PA1 2BE, UK
zeeshan.pervez@uws.ac.uk

**Abstract.** While working with the sensitive data e.g, related to health, there is a barrier of mistrust while selecting cloud services. To overcome this barrier various standards of cryptosystem are used like encrypted outsourcing, attribute based encryption and oblivious access policies. The default access model of authorization on encrypted data gives full access permission to its user. To narrow down the access scope as a subset on given authorization is a non-trivial task. To design such systems multiple encryption and decryption keys, data partitioning or attribute based encryption are few available options. These techniques involve extra computation cost and complex issue of key management. In this paper we have proposed a framework to restrict authorization on encrypted data with selective access. The underlying model is independent from complex issue of key management. The proposed model also avoids one dimension of side channel attacks on secure data and that is to learn from the patterns of encrypted traffic. Our experimental results show that selective authorization based on proposed model is compute efficient and create random pattern for user access even for similar queries.

**Keywords:** health data, data sharing, Cloud Computing, security and privacy

## 1 Introduction

The enormous volume of data in todays era of digitization is ready to be explored and shared by scientists, research institutes and enterprise organizations for enhanced knowledge exploration. Ease of accessing the internet, widespread of e-applications, user awareness, obvious benefits of digitized world on humanity are main reasons for this data proliferation. With same trend in health care domain, services like Personal Health Record (PHR) allow its users to create, manage and share its medical records with entities like physicians, friends and family members [8]. Due to the management and maintenance cost of systems like PHR, they are outsourced to third parties or cloud infrastructure [13],[10]. Besides these potential advantage of cloud infrastructure, health data requires optimal level of security too. From the consumers' and data sharing perspective,

these concerns remain the primary inhibitor for adoption of cloud computing services[2]. Besides PHR,hospital information management systems (HIMS)and electronic medical records (EMR) require even greater level of security due to the larger volume and variety of data. At a fine grained level the personal health information (PHI) or the electronic health records (EHR) are required to be protected in terms of their storage and utilization. Other than threat of mistrust from third party or cloud service provider, the inside rouge user pose an equal threat for the privacy breach of same information. A similar example is give in [8] where a user took the PHI data home of 26.5 million users without prior permission of his employer. With all these concerns encrypted storage of sensitive data is highly recommended while availing service of cloud infrastructure [3],[7].Encrypted storage maximizes the privacy aspect of data however; at the cost of lower scope of its utility. Encrypted data requires decryption keys by authorized user to avoid its unsolicited disclosure. To minimize the usage of network bandwidth and to avail optimal computation powers of cloud infrastructure various searchable encryption SE schemes have been proposed[1],[4]. With SE, if Alice is an authorized user, she does not need to download the entire data from cloud to her machine for decryption and then searching, rather, her encrypted query can be evaluated efficiently with the help of available SE schemes.The SE schemes allow users to access entire data for which authorization has been granted by the owner.

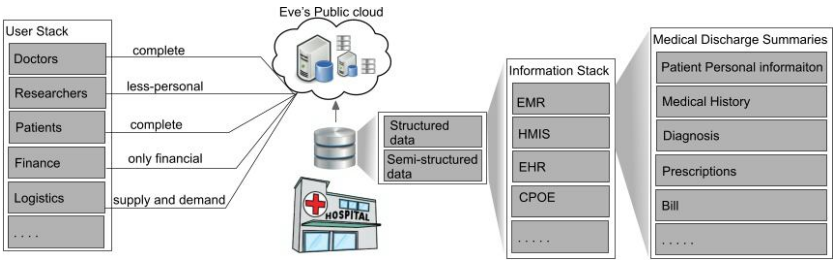


**Fig. 1.** Overall architecture

Another challenge beyond authorized access on sensitive data is to restrict authorization for selective segments of information which comes under the category of fine grained access control over encrypted data [5]. Consider an example where a National Hospital (NH) is using HIMS and its IT infrastructure is utilizing services of public cloud owned by Eve. To narrow down our example, we will consider the scenario for medical discharge summaries. In medical discharge summaries various sections like medical history, prescription, allergies, diagnosis and accounts information is entered and kept as record.These documents hold value for various analytical processing e.g, frequently prescribed medicines, prevalent diseases or the financial statements. The overall architecture is shown in Figure 1. To perform any sort of analysis, it is required that personal informa-

tion of patient should not be disclosed to anyone else other than the clinicians. Information on drug usage is required to be shared with the supply department and account section need to look into the billing section of this summary. A single document now has multiple users with definite restrictions as imposed by the data access policy of hospital.This issue can be handled by disintegrating the clinical discharge summary document into various sections(bins) and then assign each bin to its intended recipient. This approach might work for trivial access structure ignoring the additional leakage of information which is explained through an example.
Alice and Bob are two researchers in the hospital having access on the disease and drug sections. Recently it is reported that diabetic patients aged 40 and above have serious reaction with certain drug usage. Alice and Bob investigate the information with repetitive queries. Although the information is encrypted in the cloud, yet similar queries encrypted with same keys can reveal a common pattern on Eves' cloud. These common pattern in shape of similar queries or replies might help the curiosity of Eves by learning beyond required. To overcome, both Alice and Bob may require different keys, however, it will make the overall key management a complex process and process of segregating the segments(bins) useless. Considering all these concerns as motivation to our proposed solution we have formalized a *Dual-Lock* mechanism to overcome the aforesaid limitations. The proposed methodology is analogous to real world scenario of banking lockers. The bank locker can be opened by two keys, where one key is held with the bank and other is kept with the consumer. The unlocking process is possible only when bank administration and owner has right pair of keys. The rest of the paper is structured as below. Section 2 is about related work. Main idea and proposed methodology is given in Section 3. Section 4 covers the evaluation and results. Conclusion is given in Section 5.

## 2 Related Work

The trend in sharing health data between its stakeholder is not as much progressive as in other disciplines due to the sensitivity and security issues related with that [12]. This issue persist despite knowing the fact that its sharing can greatly helps researchers to minimize the rate of illness and can save human lives. The direct concern of security with health data is usually dealt with encrypted storage [3],[7] where it can be shared with authorized parties conveniently. Health data sharing take place between person to person e.g., PHR systems, between hospitals through HIMS and standards like HL7. Personal health record (PHR) is an emerging patient-centric model of health information exchange, which is often outsourced on computer clouds. However, there have been wide privacy concerns as personal health information could be exposed to those third party servers and to unauthorized parties. For PHR system security Attribute based encryption (ABE) is proposed [9]. In this system the concept of multi authority ABE has been introduced. The proposed idea mainly relies on keys which is a costly operation. A system that works with multiple keys becomes cumbersome

for consumer in terms of remembering and managing these keys. In another technique [6] that uses the ABE. After highlighting the importance of data and security concerns while storing it in the untrusted domain of cloud computing, they emphasis is to protect data and authorize access is permitted only if the patient attributes meet the ABE construction.

Besides protecting data from unauthorized access, inferred knowledge is another challenge while dealing with the sensitive data. To protect inferred knowledge also known as additional slip away of information techniques like k- anonymization [14] and l-diversity [11] with their various variations are used. The purpose of all these techniques is to protect data either from unauthorized access or avoiding the additional leakage of information. In between these two concerns we have proposed a new methodology with least instrumentation to provide further selective access on authorized data.

## 3 Main Idea

The motivation behind proposed idea is to exploit optimal resources in cloud environment with flexible, controlled and trace-free recourse sharing. With optimal resources utilization we mean that for every request call, a constant operation will output the required result. The controlled resource sharing will ensure that a user does not learn anything for which it is not authorized. The proposed algorithm also gives no clue to infer any additional knowledge by an honest but curious cloud service provider from user request logs. The proposed methodology also protects the pattern discovery of network traffic if intercepted by an eavesdropper or malicious user. The same strength is also effective for inside intruders.

### 3.1 Notations and Assumptions

The notations used in proposed system is given in table 1.

**Table 1.** Notations used in the descriptive detail of *Dual Locks*

| Notation | Description |
|---|---|
| $D = \{f_1, f_2, .., f_n\}$ | $D$ is a data containing set of files |
| $f_i = s_1, s_2, s_3, ..., s_m$ | Each file consist of m set of sections/attributes |
| $P(x, y)$ | A polynomial P defined over roots $x$ and $y$ |
| $\Delta_{y_{1...n}}$ | Each $\Delta$ uniquely identified a section $s_i \in f_i$ |
| $\ell\{\gamma, \delta\}$ | To unlock each section of $f \in D$, $\ell$ holds one and only one combination of $\gamma$ and $\delta$ for a particular section $s$. Formal proof of this concept is given in next section |
| $ER_P(x)$ | Evaluation result of a polynomial P with root $x$ |

### 3.2 Methodology

Sara is a security expert and looks after the automation process for the National Hospital. The encrypted Data $D$ of hospital is outsourced to a public cloud owned by Eve. As a public cloud owner, Eve is considered as trusted but curious. With trust we mean that data storage and policy of sharing this data with authorized users is rightfully executed by Eve however for its curiosity Eve tries to learn beyond obvious and permitted information disclosure. Authorized users on this data holds a valid key to access the information. Recently medical discharge summaries have also been uploaded on the cloud. There is a range of user groups within the hospital that will be accessing the information on these documents. To avoid unnecessary disclosure of information it is required that an employee working in logistic department should only be able to view relevant information related to logistics. Similarly people working in research department has nothing to do with supply and demand issues. To achieve fine grained level of access either multiple keys per group are required to be generated or data has to be categorized in sub categories. To achieve it a dual lock mechanism is followed that fits into the infrastructure already running. The fist locking component is constructed using the unique identifier $\delta$ for each section $s$ of a file with in $D$. This unique identifier is used to construct a polynomial $P(x, y)$ where $x$ and $y$ are two large random numbers such that $|x - y| = \delta$. The second locking component is constructed using the same mechanism and is handed over to authorized user which we call as $\gamma$. The value of $\gamma$ is send with user request to Eve Cloud. Eve then calculates the composite polynomial using $\delta$ and $\gamma$. This composite polynomial is then evaluated on $x$ and $y$ pair that were used in constructing $\delta$ and $\gamma$. This operation will end up in $ER_p(r1)$, $ER_p(r2)$, $ER_p(r3)$ and $ER_p(r4)$. The figure 2 shows how these values are dealt with XOR and NOR logical gates. The output result of gate operation is then multiplied with the user public key and reply is encrypted with that key. In case of valid request the output will remain decipherable by the user secret key and in case of invalid request the output will not be recoverable.

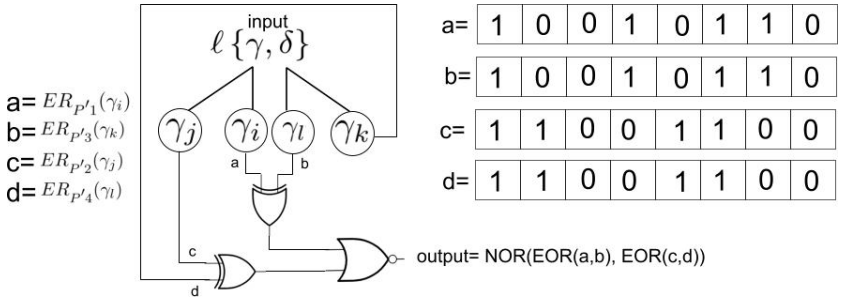


**Fig. 2.** Internal working of Dual Locks operations

### 3.3 Proof of valid roots for composite polynomial

In this section we will prove that a composite polynomial constructed with a finite set of roots satisfies the basic principle of proposed methodology. We well show that only complete and exact participation of roots can result in required output as manipulated with logical gates.
Let $\gamma_{1\ldots n}$ be the set of valid roots. $\gamma_i$,$\gamma_j$,$\gamma_k$,$\gamma_l$ have been used to construct two polynomials $P_1(\gamma_i, \gamma_j)$and $P_2(\gamma_k, \gamma_l)$ such that $\gamma_i < \gamma_j < \gamma_k < \gamma_l$. Also $\gamma_j - \gamma_j = \gamma_l - \gamma_k = \Gamma$, where $\Gamma$ is an integer value. These polynomials are then added together resulting in $P'$. This newly constructed polynomial $P'$ is then evaluated as $ER_{P'1}(\gamma_i)$ $ER_{P'2}(\gamma_j)$, $ER_{P'3}(\gamma_k)$ and $ER_{P'4}(\gamma_l)$. The proposed methodology works only when equation 1 is satisfied.

$$E\,R_{P'1} = ER_{P'3}, ER_{P'2} = ER_{P'4}\} \tag{1}$$

Let us consider that there exist another root value $\gamma_x|\gamma_x \notin \{\gamma_i, \gamma_j, \gamma_k, \gamma_l\}$ for which equality of equation 1 still holds. But while constructing the $P_1$ and $P_2$, $\gamma_x$ has not been used therefore if equality holds for equation 1, that means $\gamma_x$ is used while constructing the polynomial and it is equal to at least one of $\{\gamma_i, \gamma_j, \gamma_k, \gamma_l\}$.

## 4 Evaluation and Results

The signature for Encrypted data varies either with different key or variant input plain text. In our experimental evaluation we have used it without using any encryption technique. The output pattern revealed by our proposed methodology is random and trace free even without encryption to find out the pre encryption resistance for pattern trace. After applying encryption on these inputs will generate different output with similar keys.

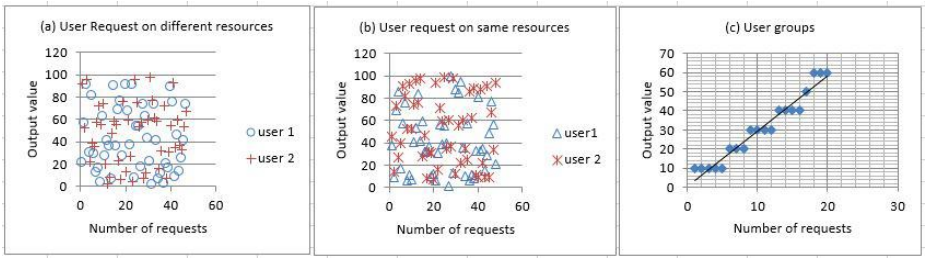


**Fig. 3.** Pattern analysis of user queries with different scenarios

Fig 3(a) shows the pattern output analysis of two users who are accessing different resources [3]. In Fig 3(b) two users are accessing the similar resource again and again but both figures Fig 3(a) and Fig 3(b) are hard to distinguish. In Fig 3(c) we showed the output patterns without using the mechanism of dual locks. In this figure, Fig 3(c), if users are requesting the same resource they will end up in forming a cluster of similar queries.

## 5 Conclusion

In this paper we have proposed a framework that is used to narrow down user authorization for selective attributes only. The goal of proposed framework is two fold. First, It avoids using complex management of encryption keys and is usable in existing systems where encrypted access is required for selective authorization. Second, For every user request, irrespective from similar query or users from same groups, the output pattern always end up in random pattern. This random pattern helps to avoid additional leakage of information especially while availing services of public cloud.

## Acknowledgment

This research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2014-(H0301-14-1003)

## References

1. Michel Abdalla, Mihir Bellare, Dario Catalano, Eike Kiltz, Tadayoshi Kohno, Tanja Lange, John Malone-Lee, Gregory Neven, Pascal Paillier, and Haixia Shi. Searchable encryption revisited: Consistency properties, relation to anonymous ibe, and extensions. In *Advances in Cryptology–CRYPTO 2005*, pages 205–222. Springer, 2005.
2. Deyan Chen and Hong Zhao. Data security and privacy protection issues in cloud computing. In *Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on*, volume 1, pages 647–651. IEEE, 2012.
3. Richard Chow, Philippe Golle, Markus Jakobsson, Elaine Shi, Jessica Staddon, Ryusuke Masuoka, and Jesus Molina. Controlling data in the cloud: outsourcing computation without outsourcing control. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, pages 85–90. ACM, 2009.
4. Reza Curtmola, Juan Garay, Seny Kamara, and Rafail Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. In *Proceedings of the 13th ACM conference on Computer and communications security*, pages 79–88. ACM, 2006.

[3] The results in figure 3 are normalized to lower scale as shown on y-axis whereas x-axis represents the number of user requests

5. Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attribute-based encryption for fine-grained access control of encrypted data. In *Proceedings of the 13th ACM conference on Computer and communications security*, pages 89–98. ACM, 2006.
6. Luan Ibraimi, Muhammad Asim, and Milan Petkovic. Secure management of personal health records by applying attribute-based encryption. In *Wearable Micro and Nano Technologies for Personalized Health (pHealth), 2009 6th International Workshop on*, pages 71–74. IEEE, 2009.
7. Seny Kamara and Kristin Lauter. Cryptographic cloud storage. In *Financial Cryptography and Data Security*, pages 136–149. Springer, 2010.
8. Ming Li, Shucheng Yu, Yao Zheng, Kui Ren, and Wenjing Lou. Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. *Parallel and Distributed Systems, IEEE Transactions on*, 24(1):131–143, 2013.
9. Ming Li, Shucheng Yu, Yao Zheng, Kui Ren, and Wenjing Lou. Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. *Parallel and Distributed Systems, IEEE Transactions on*, 24(1):131–143, 2013.
10. Hans Löhr, Ahmad-Reza Sadeghi, and Marcel Winandy. Securing the e-health cloud. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pages 220–229. ACM, 2010.
11. Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.
12. Elizabeth Pisani and Carla AbouZahr. Sharing health data: good intentions are not enough. *Bulletin of the World Health Organization*, 88(6):462–466, 2010.
13. Robert Steinbrook. Personally controlled online health data-the next big thing in medical care? *New England Journal of Medicine*, 358(16):1653, 2008.
14. Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.