International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue-01
January 2018

Secure and Efficient Protocols for Providing Personalized Privacy Protection over High-Dimensional Healthcare Data

Ch. Sowjanya & V. Sridhar Reddy
[1]M. Tech Student, Department of CSE, Vignana Bharathi Institute of Technology, Village Aushapur, Mandal Ghatkesar, District RangaReddy, Telangana, India
[2]Associate Professor, Department of CSE, Vignana Bharathi Institute of Technology, Village Aushapur, Mandal Ghatkesar, District RangaReddy, Telangana, India

Abstract— *According to the modern rule released by Health and Human Services (HHS), healthcare information will be outsourced to cloud computing services for medical studies. A significant concern concerning outsourcing attention information is its associated privacy problems. However, previous solutions have centered on cryptographically techniques that introduce important value once applied to attention information with high-dimensional sensitive attributes. To deal with these challenges, we tend to propose a privacy-preserving framework to transit insensitive knowledge to commercial public cloud and therefore the rest to trustworthy personal cloud. Under the framework, we tend to design two protocols to produce customized privacy protections and defend against potential collusion between the general public cloud service supplier and therefore the information users. We tend to derive obvious privacy guarantees and finite information distortion to validate the projected protocols. Intensive experiments over real-world datasets are conducted to demonstrate that the projected protocols maintain high usability and scale well to massive datasets.*

Index Terms-- *Healthcare Data, Hybrid Cloud, Privacy Preserving, Outsourced Computing;*

I. INTRODUCTION

Cloud computing is getting more and more widespread as it can give low-priced and on demand use of huge storage and processing resources. As the volume of data grows, also increasing is the Total price of ownership which includes storage infrastructure value, management value and human administration cost. Therefore in cloud storage systems, reducing the amount of data that want to be transferred, stored, and managed becomes a crucial. Cloud computing is an emerging technology where the cloud service providers (CSP) offers to provide an efficient data storage facilities to shown in below figure1. CSP is the authority for controlling the information that is stored in the cloud system. The CSP can undertake the authority to look over the data items without the permission of data owner. However, the data that are stored in the cloud service providers are stored in the plain text format which may be known to the service providers. Hence there is a possibility of leakage in these sensitive information. The threats to these leakage of the data are classifies into two categories such as internal threats and external threats. External threats might be due to the outside hackers performing attacks by finding the network vulnerabilities for

accessing the data about the data owner. Internal attacks are done by the inner intruders who are responsible for the protection of the database information where the data owner's sensitive information are stored in plaintext format. Gaining access to healthcare information may be an important demand for medical practitioners and pharmaceutical researchers to check characteristics of diseases. In recent years, the proliferation of cloud computing services allows hospitals and establishments to transit their healthcare information to the cloud that provides present information access and on-demand high quality services at a low price. The U.S. Department of Health and Human Services (HHS) released the Omnibus Rule, which defines cloud service providers (CSPs) as business associates for healthcare information.

Figure 1: Architecture of Cloud Computing

Currently, many CSPs, as well as Box, Microsoft, Verizon and dell, have declared their support for this business associate agreement. Despite the advantages of healthcare cloud services, the associated privacy problems are widely involved by individuals and governments. Privacy risks rise once outsourcing personal healthcare records to cloud because of the sensitive nature of health data and also the social and legal implications for its revealing. A natural methodology is to encode healthcare information before transiting them to cloud. However, process encrypted information are not economical and is restricted to specific operations, and therefore is not appropriate for healthcare information with versatile usages. An alternate resolution is applying existing privacy-preserving data publishing (PPDP) techniques, like partition-based anonymization, and differential privacy, to the outsourced healthcare information. However, as we tend to show below, once the following practical necessities are thought-about, the present works are not applicable within the context of healthcare information outsourcing. the most contributions of this paper are threefold. First, we tend to propose a privacy-preserving framework for high dimensional healthcare data outsourcing. To the most effective of our data, this can be the primary framework considering high dimensional sensitive attributes and personalized privacy requirements over different attributes. Second, through formal analytic study, we tend to derive demonstrable privacy guarantees and finite data distortion achieved by the planned framework. We tend to show that the projected framework will defend against the collusion between the public cloud and also the DUs whereas still retaining high usability. Finally, for the primary time, we tend to conduct experiments on real-world healthcare datasets with high-dimensional sensitive attributes to validate the projected framework.

II. RELATED WORK

Recently deferential privacy has gained considerable attention as a substitute for partition-based approaches. Varied approaches are planned for implementing deferential privacy in information publication. many works try and handle the multi-dimension issue in

deferential privacy present a general framework to unharnessed multi-dimensional information cubes by optimally choosing a part of an information cube for publication. Pang et al introduce deferentially private indices to reduce errors on multi-dimensional datasets. However, the multi-dimensional issues discussed in deferential privacy are usually restricted to be less than twenty, whereas datasets with higher dimension are not covered. Besides, none of those approaches are designed for information outsourcing with collusion resistance. Wong et al. assume that adversaries recognize the anonymization algorithm. This additional data will facilitate adversaries breach the privacy, which is named minimalists attack. The notion of customized privacy is projected by Xiao and to allow each individual to specify his/her own privacy preference on one sensitive attribute. This model makes assumptions that the sensitive attribute includes a taxonomy tree and every individual specifies a guarding node within the taxonomy tree as his/her privacy preference. The privacy is desecrated if the reasoning confidence on any sensitive value within the subtree of individual's guarding node is higher than the pre-defined threshold. Xiao and Tao's approach specifies different privacy level on one sensitive attribute, whereas our work considers individuals' privacy preferences at attribute level. Nevertheless, none of those approaches are designed for high-dimensional information outsourcing with collusion resistance, in which case they lack the thought of multiple sensitive attributes and customized privacy issues at the attribute level. Privacy-preserving information outsourcing in the main adopt encoding techniques to protect sensitive information Yuan and Yu encode the biometric information before outsourcing it to the cloud, which might perform kNN search within the encrypted information. Li et al. leverage ranked predicate encoding to establish a scalable framework for approved non-public keyword search on cloud information. Cao et al alter privacy-preserving multi-keyword hierarchic search over encrypted cloud information. However, these solutions are limited to specific operations that are not appropriate for tending information outsourcing that supports a spread of queries. Besides, encoding ends up in massive overhead once responsive queries. Another complete of privacy-preserving approaches is PPDP techniques. Basically, the works on privacy protection in information publication is divided into two categories; partition based mostly approaches and differential privacy. Several partition based mostly privacy models are projected to tackle totally different privacy issues. K-anonymity is developed to prevent adversaries with Quash-Identifier (QI) information from re-identifying an individual with a chance higher than 1k. Fragmentation is used into break sensitive associations among attributes.

III. FRAME WORK

In this paper we tend to initial propose a privacy-preserving framework for high dimensional healthcare information outsourcing. To the most effective of our information, this can be the primary framework considering high dimensional sensitive attributes and customized privacy needs over different attributes. Second, through formal analytic study, we tend to derive demonstrable privacy guarantees and delimited

information distortion achieved by the planned framework. We tend to show that the planned framework will defend against the collusion between the public cloud and therefore the DUs whereas still retentive high usability. Finally, for the primary time, we have a tendency to conduct experiments on real-world healthcare datasets with high-dimensional sensitive attributes to validate the planned framework. Hybrid cloud may be a new framework planned for secure cloud computing. Sedic-modifies Map-Reduce s files system to move sanitized information to the public cloud and keep sensitive information on the personal cloud. Privacy-aware information retrieval on hybrid cloud is investigated in several from these works, we tend to think about the salient options of real health care information, and supply privacy protection against collusion between the general public CSP and therefore the DUs. We tend to think about the situation wherever a hospital has to transit its health care information to the cloud to produce ubiquitous information access and services at low value. To produce privacy protection of individuals' information, the hospital outsources the health care information to a hybrid cloud, which consists of a personal cloud that keeps sensitive information inside the hospital and a public business cloud that handles the remainder of the dataset. The figure illustrates the healthcare outsourcing design, wherever an information holder (e.g., a hospital) outsources a tending dataset owing to the hybrid cloud, and approved DUs (e.g., medical practitioners and pharmaceutical researchers) to achieve access to the tending information from the cloud for medical information analysis. Specially, the information holder initial splits the initial dataset into insensitive and sensitive elements, and outsources them to the personal cloud and public cloud, severally.

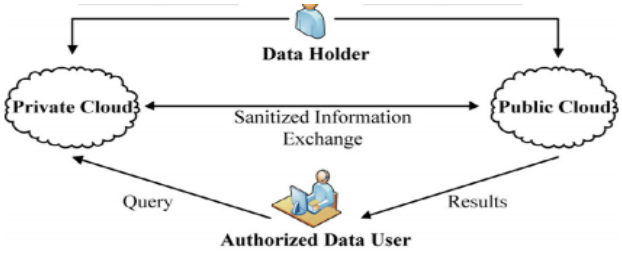


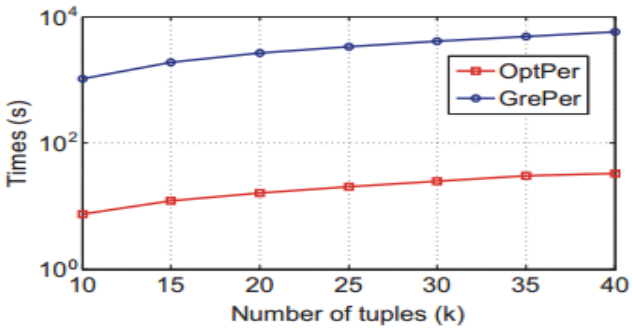
Figure 2: Architecture of Healthcare Hybrid Cloud

Then, to provide privacy protection, sequences of operation are comprehensive at the private and public clouds to sanitize the dataset before it will be accessed by DUs. When sanitization, licensed DUs will post queries on the cloud for information analysis the authorized DUs post queries to the private cloud, and therefore the private cloud communicates with the public cloud to generate results. To preserve individual's privacy, the data shared with the public cloud should be rigorously sanitized. However, according to the notion of minimalist in anonymization, anonymization mechanisms aim to realize privacy guarantee with minimal information distortion, and this settled try provides a loophole for attacks. Thus, the anonymized results for queries could leak personal data to the DUs. To thwart such privacy breach, differential privacy protection is required to disarrange the anonymization results. The intuition of differential privacy is that the removal or addition of one record does not considerably have an effect on the result of any analysis. The core plan of the privacy-preserving outsourcing framework is to share partition strategy between clouds to derive alter information whereas keeping sensitive information on the personal cloud. to supply customized protection on

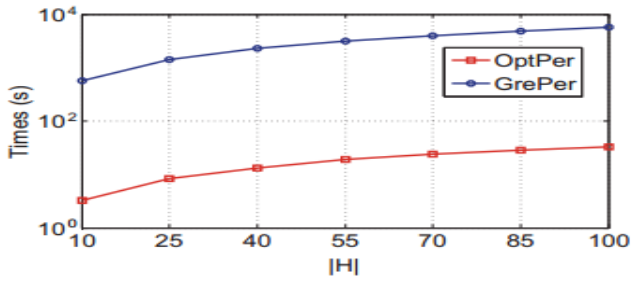
sensitive information, the dataset is split into multiple partitions and generalization operations are applied on personal sensitive attributes. Partition data is shared between clouds to derive the best partition strategy. The protocols are built supported two elements. The primary part is perfect partitioning, that aims to seek out a partition set which will satisfy customized privacy needs with minimal information distortion. The second part is privacy budget allocation, which optimally allocates totally different fractions of randomness to every partition operation in order that the ultimate partition set is differentially personal whereas the information distortion is decreased. The remainder of this section elaborates the two elements and therefore the sanitization protocols. We tend to compare the performance of our protocols with ancient anonymization and differential privacy approaches. For honest comparison, we tend to extend these approaches to be applicable to high-dimensional dataset. The anonymization approach, denoted a MulAnony, first decomposes the dataset into disjoint partial datasets that every partial dataset contains the tuples with an equivalent privacy requirement, and then applies anonymization on every partial dataset independently.

IV. EXPERIMENTAL RESULTS

In our experiments, as expected, the running time of each algorithm grows linearly with the number of tuples, and GrePer runs much faster than OptPer in the same settings. The below chart describe that Running time vs. number of tuples (in thousands)



In below chart we run both protocols on datasets with different dimensions of sensitive attribute. The results demonstrate that both OptPer and GrePer achieve linear complexity with respect to the dimension of sensitive attribute. The chart describe that Running time vs. the Dimension of sensitive attribute



V. CONCLUSION

This paper studied the problem of privacy-preserving health care information outsourcing. A framework based on hybrid cloud was projected to produce customized privacy protection over high-dimensional health care information. Under the framework, we tend to devised two sanitization protocols to anonymize the knowledge set on the personal and public clouds supported randomized data partitioning. The protocols are proven to be immune to collusion between the public Cloud Service Provider and also the data users. Analytical results are derived to verify the usability and efficiency of the protocols. Experiments on real-life datasets validate the prevalence of our approaches over range of

baseline techniques.

REFERENCES

[1] Fung BC, Wang K, Chen R, Yu PS. Privacy preserving data publishing: A survey of recent developments. ACM Comput Survey (CSUR). 2010 Jun 1; 42(4):14.

[2] Tsai J, Bond G. A comparison of electronic records to paper records in mental health centers. Int J Qual Health C. 2008 Apr 1; 20(2):136.

[3] Kierkegaard P. Electronic health record: Wiring Europe's healthcare. CLSR. 2011 Sep 30; 27(5):503-15.

[4] Qiao Y, Asan O, Montague E. Factors associated with patient trust in electronic health records used in primary care settings. Health Policy Technol. 2015 Dec 31; 4(4):357-63.

[5] Tamersoy A, Loukides G, Nergiz ME, Saygin Y, Malin B. Anonymization of longitudinal electronic medical records. IEEE T INF Techno B. 2012 May; 16(3):413-23.

[6] Gostin LO, Levit LA, Nass SJ, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research: 2009 Feb 24; Washington: National Academies Press.

[7] Sellapan P, Ng YH. A Tool for Healthcare Information Integration. JICT. 2006; 5:29-44.

[8] Lin YM, Zakariah MI, Mohamed A. Data leakage in ICT outsourcing: risks and countermeasures JICT.2010; 9: 87-109.

[9] Dubovitskaya A, Rove V, Varian M, Abider K, Schumacher MI. A Cloud-Based eHealth Architecture for Privacy Preserving Data Integration. Proceedings of the ICT Systems Security and Privacy Protection; 2015 May 26; Springer; 2015.

[10] Malin B, Karp D, Scheuermann RH. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J Invest Med. 2010 Jan 1; 58(1):11-8.

[11] Utility-aware anonymization of diagnosis codes. IEEE J Biomed Health Inform. 2013 Jan; 17(1):60-70.

[12] El Emam K. Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PloS One. 2011 Dec 2; 6(12):e28071.

[13] Daglish D, Archer N. Electronic personal health record systems: A brief review of privacy, security, and architectural issues. Proceedings of the World Congress on Privacy, Security, Trust and the Management of e-Business; 2009 Aug Delta Brunswick, Canada. IEEE; 2009. p. 110-20.

[14] Appari A, Johnson ME. Information security and privacy in healthcare: current state of research. Int Journal Internet Enterprise Manag. 2010 Jan 1; 6(4):279-314.