Volume 5, Issue 3, March 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Survey of Security of Medical Intermediate Data Set using Cloud Computing

Varsha R
MTech(CSE) , BMSCE,
Bangalore, India

Indiramma M
Prof. Dept. of CSE, BMSCE,
Bangalore, India

Abstract:- Cloud computing is an emerging field in the development of business and organizational environment. Cloud services can help the healthcare industry to access and manage health records effectively in order to provide better patient care. A well implemented cloud storage system allows hospitals to process tasks effectively and more quickly, without causing a failure in the performance. Cloud computing has proven to be extremely beneficial and also cost effective for patients and healthcare providers. One of the biggest concerns with moving healthcare data to the cloud is security. Nowadays many organizations are very much interested in storing and retrieving their important and large amount of data into cloud with encryption and decryption techniques are used to preserve the intermediate data sets in cloud. Intermediate data sets are stored into cloud and retrieved from multiple parties. By encrypting all intermediate datasets, we can ensure cost-effectiveness and time consuming properties.

Keywords: Anonymisation, Cloud Computing, Data Storage Protection, Privacy Preserving, Intermediate Datasets, Proxy Re-encryption, Optimized Balanced Scheduling, Privacy Upper Bound Heuristic Value.

I. INTRODUCTION

Healthcare is important as everyone needs it several points in their daily life. Healthcare is becoming important as new technologies are being incorporated into the existing infrastructure. In recent forefront of new technologies, the patient records are being put in electronic format (EMR) enabling patients to access their records via the Internet and also cloud computing environment. The remote patient is monitoring with more feasibility at anytime, anywhere and any place also. The combination of these technologies will improve the quality of health care by making it more personalized and reducing costs and medical errors. To make it more beneficial and acceptable, associated privacy and security issues need to be analyzed.

Healthcare identifies important factors to consider when moving the data and applications to cloud computing. For healthcare organizations to maintain higher levels of flexibility and security, cloud-based solutions are becoming an increasingly support on-demand provisioning infrastructures. Several concerns have held healthcare organizations back from moving ahead toward full-scale cloud migrations. Healthcare organizations may have security concerns due to the growth of privacy and security breaches.

In cloud computing, considering the security threats there may be chances of easily hacking of the intermediate medical dataset in cloud. Remote backups of the data should be created regardless of whether the Cloud Service Provider (CSP) is already providing backup service for the data. It’s better to have multiple data backups when the need for data restoration arises.

The limitation of existing work is to achieve privacy-preserving cost. Preserving the privacy of intermediate datasets becomes a challenging problem because adversaries may recover privacy-sensitive information by analyzing multiple intermediate datasets. Many people are dependent on healthcare organizations, emergency workers, and hospitals to provide medical information on time for their critical need and also to provide ongoing treatment to saves lives. To provide optimal service, doctors and other medical professionals need timely access to patient data.

II. RELATED WORK

According to the work presented by Tejaswi Turla et.al., shows that an approach is utilized that recognizes which division of intermediate data set requires to be encrypted and which does not, in order to conserve the privacy preserving cost. A tree structure is shaped from the generation relationships of intermediate data sets to analyze privacy broadcast among data sets. The problem of conserving privacy preserving cost as a constrained optimization problem which is delivered by decomposing the privacy leakage constraints. A heuristic value & PP (privacy preserving) algorithm has originated for reducing the privacy preserving cost. The strategy of heuristic search can also be effective for stochastic shortest-path problem and in general MDPs. The strategy is to solve the problem only for states that are reachable from the start state by following an optimal policy. Privacy-preserving cost of intermediate data sets stems from frequent en/decryption with charged cloud accommodations. Cloud resource vendors have established sundry pricing models to fortify the pay-as-you-go model, e.g., Amazon Web Accommodations pricing model. Practically en/decryption needs computation potency, data storage, and other cloud accommodations. This cumulated price is denoted as PR. PR signifies the overhead of en/decryption on per GB data per finishing [01].

The work carried out by A. Jeeva et.al., shows that an approach that identifies which part of intermediate dataset needs to be encrypted and which part does not need to be encrypted, to save the privacy preserving cost. A tree structure from the generation relationships of intermediate dataset has been modeled to analyze privacy propagation among datasets. The problem of saving privacy preserving cost as a constrained optimization problem which is addressed by decomposing the privacy leakage constraints. An algorithm has been designed accordingly for this purpose. Evaluated results on real-world datasets and larger extensive datasets have demonstrated the cost of preserving privacy in cloud and can be reduced significantly over existing ones, where all the datasets are encrypted. Intermediate dataset management is becoming an important research area. Privacy preserving for intermediate datasets is one of important yet challenging research issues, and needs intensive investigation. Encrypt all such data using a single key, and the key to be shared with all users of the service. Unfortunately, this has the problem that a malicious or compromised cloud could obtain access to the encryption key, or by compromising or clouding with an existing user [03].

The work carried out by Mr. C. Radhakrishnan et.al., tells that an upper bound constraint based approach is used where data set needs to be encoded, in order to reduce the privacy preserving cost so we investigate privacy aware efficient scheduling of intermediate data sets in cloud by taking privacy preserving as a metric together with other metrics such as storage and computation. Optimized balance scheduling strategies are expected to be developed towards overall highly efficient privacy aware data sets are scheduling and mainly in the overall time reduction and data delivery overhead is reduced by the load balancing based scheduling mechanism. Dynamic privacy preservation model is supported by the system and along with that a high security provisioning is done with the help of full suppression, semi suppression and Value Generalization Hierarchy Protocol. The problem of managing the intermediate data which is generated during dataflow computation deserves deeper study as a first-class problem. There are two major requirements that any effective intermediate storage system needs to satisfy: first is the availability of intermediate data, and second is the minimal interference on foreground network traffic which is generated by the dataflow computation [04].

According to the author Mr. C. Radhakrishnan et.al., point of view in cloud computing saving privacy-preserving cost is the major constrained problem. This proposed work decomposes the problem into simple small problem and it will determine which part of intermediate data set needs to be encrypted and it also saves the privacy preserving cost. An algorithm for this is designed accordingly to obtain a global privacy preserving solution. Finally this algorithm is constructed and based on the result of the algorithm from which the data sets are to be encrypted and determined. From execution, general data set illustrates that the privacy-preserving cost of intermediate data set can be much more reduced over existing ones of all data sets which are encrypted. Privacy aware efficient scheduling of intermediate data set in cloud by taking privacy preserving as a metric together with other metrics such as storage and computation [05].

The work represented by the author A. Balachandra Rao et.al., suggested that the design an upper bound on privacy measure of stored datasets. Privacy protective for intermediate information set is one among necessary nevertheless difficult analysis problems and desires intensive analysis and investigation. By this a heuristic privacy leakage constraint based approach to identify which processing datasets need to be encrypted and which do not. Based on the privacy requirements produced by the data holders, an optimal heuristic privacy leakage control approach can be implemented in order to minimize the privacy preserving cost of the processed datasets in cloud data [06].

According to the work proposed by Gurudayal Singh Bhandari et.al., explained that the security plays a vital during the transmission of data from the sender to the receiver in any environment. The challenge in privacy preserving Back-Propagation Neural Network Learning is avoiding the attack of personal data privacy. Due to the enlargement of distributed computing environment. In such distributed scenarios, privacy concerns often become a big concern. Secure computation provides a solution to this problem. With the invention of new technologies and computing, it has been more convenient than ever for users across the Internet, who may not even know each other whether it is data mining, or in any networks that resolving privacy problems has become very important [11].

The work presented by the author Mr. R. Sathish et.al., describes that the shared data values are maintained under third party cloud data centers. Cloud modes are used to store the processed data values. Privacy leakage upper bound constraint model is used to protect the intermediate data values. Dynamic privacy management and scheduling mechanism are integrated to improve the data sharing with security. Privacy preserving cost is reduced by the joint verification mechanism. Based on the scheduling mechanism, the data delivery overhead is reduced. Dynamic privacy preservation model is supported by the system [12].

The author Dong Yuan et.al., explains that many scientific workflows are data intensive where a large volume of intermediate data is generated during their execution. Intermediate data needs to be stored for sharing or reuse purpose but they are selected and stored according to the system storage capacity which is determined manually. More intermediate data can be stored in scientific cloud workflows based on a pay-for use model. In this proposed work, the author builds an Intermediate data Dependency Graph (IDG) from the data provenances in scientific workflows. Based on the IDG, the author develops a novel intermediate data storage strategy that can reduce the cost of the scientific cloud workflow system by automatically storing the most appropriate intermediate datasets in the cloud storage [15].

The work proposed by Ms. Apeksha Sakhare explained that the anonymization is a viable technique to secure cloud computing. It omits the misuse of sensitive data, but it cannot be concluded as a complete solution to preserve confidentiality. Lots of techniques for anonymization have been implemented but still there is a fear of security breach. Research for anonymization and de-anonymization is in process. Anonymisation techniques which are currently safe may fail in future. In future, the privacy preserving in cloud needs many efforts [16].

According to the work represented by Ravindra Suresh Kamble et.al., have described an approach that identifies which part of intermediate data sets needs to be encrypted while the rest does not.. The author suggests that the problem of saving privacy-preserving cost as a constrained optimization problem which is addressed by decomposing the privacy leakage constraints [09].

III. SECURITY MODELS FOR HEALTHCARE

In this paper we describe an EHR security reference model for managing security issues in healthcare clouds. Currently, each provider has its own database for Electronic Medical Record's (EMR). The electronic records sharing between different EMR systems are called Electronic Heath Records (EHRs). EMR is a legal record of what happened to the patient their encounter at Care Delivery Organization (CDO) across inpatient and outpatient environments and its owned by the CDO. EMR is created, used and maintained by the healthcare organization to document, monitor and manage healthcare delivery with the organization.

EHR is a subset of EMR record maintained by the CDO and created and owned by the patient. An EHR typically has patient input and can be shared across multiple CDOs within community, region, or state. [13] Based on ISO/TS 18308 [1] standard, the primary purpose of EHR is to provide a documented record of healthcare which supports present and future healthcare received by the patient from same or other clinicians. This documentation provides a mean of communication between the clinicians and the patients. [14]

IV. BASIC SECURITY CONCEPTS OF HEALTHCARE CLOUD

Common security issues of healthcare cloud are ownership of information, authenticity, and authentication, non-repudiating the information, patient permission, uprightness and confidentiality of data.

A. Ownership of Information:

Generator of the information is known as the ownership of information. Ownership of information is necessary to protect against unauthorized access or misuse of patient's medical information [15].

B. Authenticity and Authentication:

Authenticity refers to the truthfulness of the origins, commitments and intentions. Authentication is the act of confirming the claims made by the person are true and correct. Authentication of information poses problems like man-in-middle attack and is often implemented with authenticating identity [15].

C. Non-Repudiating the Information:

Non-Repudiating implies that one party of a transaction cannot deny having received a transaction nor can the other party deny having sent a transaction. EMR uses technology such as digital signatures and encryption to establish the truthfulness of the original information. [15]

D. Patient Permission:

Patient can allow or deny sharing their information with other healthcare practitioners. To implement patient consent in healthcare patient may grant rights to the users on the basis of the role held by the user. [15]

E. Uprightness and Confidentiality of Data:

Uprightness in healthcare means that data has not been accessed by unauthorized users. Confidentiality means that data is been made available only to the persons who have access for it. [15]

F. Availability:

Availability means that the information should be available when it is required in the desired format for the future purpose. [15]

V. CLOUD CHALLENGES IN HEALTHCARE

Some of the challenges faced by the healthcare providers as they are moving to the cloud are:

A. Privacy Challenges:

Putting personal health information into the third party raises a concern where patient privacy laws are concerned. There may be a possibility that patient data might be lost or misused while moving towards the cloud. Violation of patient data carries heavy fines including recovery of costs. A solution for this might be using private cloud model where only the users who have access to the patient data can view it. [16]

B. Security Challenges:

A cloud provider would have many security experts developing security aspects so that the information regarding health will be well safe-guarded to ensure that the information is safe. [16]

VI. BENEFITS OF CLOUD ADOPTION IN HEALTHCARE

A. Collaboration with Patients:

Patient's records are made available anywhere, anytime for healthcare professionals for diagnosis and allowing physicians to access critical data and adjust their diagnosis based on their informed decisions.

B. Mobility:

Each mobile app is backed up by a cloud infrastructure. By storing data in the cloud, healthcare service providers enable faster access to information anywhere and anytime.

C. Decreased Costs:

There is no need for health care institutions and doctors to invest in hardware infrastructure and maintenance because these are already taken care by the cloud computing providers.

D. Speed :

Cloud-based tools can upgrade and improve their features faster and less expensively with no service interruption. Cloud services enable faster access to important information for health service providers and their patients.

VII. CONCLUSION

This paper surveys which part of intermediate data sets needs to be encrypted while the rest does not, in order to save the privacy preserving cost. This paper also provides a cloud service that can help healthcare industries to access and manage health records effectively in order to provide better patient care.

ACKNOWLEDGMENT

The work reported in paper is supported by the college through the TECHNICAL EDUCATION QUALITY IMPROVEMENT PROGRAMME [TEQIP-II] of the MHRD, Government of India Autonomous under VTU

REFERENCES

[1] Tejaswi Turla, S. Siva Skandha and G. Ravi Kumar, “*EnsuringSecurity for Intermediate Datasets in Cloud Using Upper BoundHeuristic Value And PP Algorithm*,” International Journal of Professional Engineering Studies Volume 4, Issue 2, 2014.

[2] A. Jeeva, “*Enabling Cost-Effective Privacy Preserving of Intermediate Sensitive Data Sets in Cloud Using Proxy Re-Encryption Technique*,” International Journal of Innovative Research in Computer and Communication Engineering Volume 2, Special Issue 1, 2014.

[3] D. Tejaswini and C. Rajendra, “*Privacy Aware Efficient Data Scheduling of Intermediate Data Sets in Cloud*,” International Journal of Advanced Engineering and Global Technology Volume 2, Issue 09, 2014.

[4] Mr. C.Radhakrishnan and Ms. C.P.Darani, “*Privacy Preserving of Intermediate Data Sets in Cloud*,” International Journal of Engineering Research and Development Volume 10, Issue 5,2014

[5] A. Balachandra Rao and Dr. Sai Satyanarayana Reddy, “*A Heuristic Privacy Leakage Control Approach for Profitable Privacy Protection of Processing Datasets in Cloud*,” International Journal of Emerging Technology and Advanced Engineering Volume 4, Issue 7, 2014.

[6] Ravindra Suresh Kamble and Sheikh Gouse, “*Cost-Effective Privacy Preserving of Intermediate Data Sets in Cloud using Privacy Leakage Upper Bound Constraint-Based Approach*,” International Journal of Science and Research (IJSR) Volume 3, Issue 8, 2014.

[7] Gurudayal Singh Bhandari and Abhishek Chauhan, “*A Survey on Privacy Preservation in Cloud Computing*,” International Journal Of Emerging Technology and Advanced Engineering Volume 4, Issue 7, 2014.

[8] Mr. R. Sathish and Mr. T. Sathish kumar, “*Privacy Preserved Data Scheduling for Cloud Data Services*,” International Journal Of Innovative Research in Computer and Communication Engineering Volume 2, Special Issue 1, 2014.

[9] Dong Yuan, Yun Yang, Xiao Liu and Jinjun Chen, “*A Cost- Effective Strategy for Intermediate Data Storage in Scientific Cloud Workflow Systems*,” Swinburne University of Technology Hawthorn, Melbourne, Australia.

[10] Ms. Apeksha Sakhare and Ms. Swati Ganar, “*Anonymization: A Method To Protect Sensitive Data In Cloud*,” International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013.

[11] Ali Mirarab, Najmeh Ghasemi Fard and Mahboubeh Shamsi, “*A cloud solution for medical image processing*,” Ali Mirarab etInt. Journal of Engineering Research and Applications 2248-9622, Volume 4, Issue 7, Version 3, 2014.

[12] Fatma E.Z and A. Elgamal, “*Secure Medical Images Sharing over Cloud Computing environment*,” (IJACSA) International Journal of Advanced Computer Science and Applications, Volume 4, 2013.

[13] D. Garets and M. Davis, “*A HIMSS Analytics White Paper. Electronic Medical Records vs. Electronic Health Records: Yes*,” There Is a Difference. January 26, 2006.

[14] “*Electronic Medical Record/Electronic Health Record Use by Office-based Physicians: United States,*” 2008 and Preliminary 2009

[15] H. Linden, D. Kalra, A. Hasman, J. Talmon, “*Inter-organization future proof HER systems-A review of the security and privacy related issues*,”. International Journal of Medical Informatics, 78(2009), 141-160.

[16] S. Ramgovind, M. M. Eloff, E. Smith. “*The Management of Security in Cloud Computing*” In PROC 2010 IEEE International Conference on Cloud Computing 2010.