International Journal of Engineering & Technology, 7 (4.10) (2018) 504-507

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Survey on Securing and Optimizing Health Care Bigdata

K.Kavitha [1]*, D.Anuradha [2], P.Pandian [3]

[1,3] Faculty of School of Advanced Science, VIT, Vellore campus-632 014
[2] Faculty of Computer Science and Engineering, VIT, Vellore campus-632 014
**Corresponding author E-mail: kavinphd@gmail.com*

Abstract

Huge amount of health care data are available online to improve the overall performance of health care system. Since this huge health care Big-data is valuable and sensitive, it requires safety. In this paper we analyze numerous ways in which the health care Big-data can be protected. In recent days many augmented security algorithm that are suitable for Big-data have emerged like, El-Gamal, Triple-DES, and Homomorphic algorithms. Also authentication and access control can be implemented over Big-data using Role-Based Access Control (RBAC) and Attribute-Based Access Control (ABAC) schemes.

Along with security to Big-data we try to evolve the ways in which the valuable Big-data can be optimized to improve the Big-data analysis. Mathematical optimization techniques such as simple and multi-purpose optimization and simulation are employed in Big-data to maximize the patient satisfaction and usage of doctor's consulting facility. And also, to minimize the cost spent by patient and energy wasted.

***Keywords**: Big-data; Healthcare; Information security; Optimization, Multi-objective optimization; Simulation.*

1. Introduction

Nowadays, healthcare environment is extensively more explosive and complex than in years past. Healthcare developments are progressively essential in healthcare innovativeness and health care groups are making gradually more volumes of high velocity data in an immeasurable plan. These huge volumes of data need a sophisticated environment to collect, store, process and retrieve it in an efficient way. This special environment is called Big-data. In the recent years medication has become available online with the support of Big-data. Digitizing the health care data is the trend of new era of health care industry. This ended up in what is called the big-data, which is describing as a huge data collection with different structures and variable rate of growth. This Big-data plays an important role in almost all business intelligence. The decisions made by analyzing the Big-data are very vital for the business growth in corporate and patient's welfare in hospitals. Doctors do their consultation over internet by diagnosing the health compliant with the data available on the fly. In most of the situations either the doctor or the patient could not travel physically to the hospital. This will reduce the cost of healthcare for the hospital and also for patients [15].

The Australian government spends about 130 million dollars for healthcare alone [12]. Big-data was used extensively in controlling and preventing Ebola virus in Africa by isolating the places where the people movement is excessive and identifying best place to set up treatment places [10, 21, 23]. The applications that are available on smart phones can collect the data and analyze to forecast the health complications well in advance [21, 23]. Huge amounts of clinical data can be analyzed to find out the best subject for clinical trials. Regular monitoring, collection and analysis of medical data of a patient at high risk can reduce the preventable deaths. The medical data collected over a long period of time can be used to diagnose the side effects of new medicines. Medical history of members in a family can help in DNA analysis to figure out the diseases due to heredity [21, 23]. By integrating different big data technologies in healthcare the performance and throughput of the health care information system can be faster and safer.

The Indiana Health Information Exchange in United States is a non-profit charity organization, which provides a secure and reliable technology network of health care information connecting about 90 hospitals, community health centers, rehabilitation units and other healthcare solution providers located in Indiana. This system allows medical information to be submitted from any location which is connected and the same information can be accessed from any location in the network [13]. To utilize this facility the sensitive health data of the patients should be secured from data leakage and damage. Since the data stored are real, latest and sensitive data they must be secured efficiently. Any breaches in the Big-data store can lead to drastic effects such as wrong prediction of disease, wrong prescription of medicines, false forecast of health risks and irrelevant treatment, etc. The primary privacy issue with Big-data is Insider attack. Since the insiders are given the complete access to the storage they can easily steel the data. Also some cloud service providers may steel and sell the data to make money [3, 5]. Hence the data used for analysis must be secured in an efficient way.

A frame work which provides a big data driven model for the optimization of healthcare processes is discussed in [31]. An enlightened optimize model of access controller in big data using assurance interval and digital signatures discussed in [2]. A survey of cost effective big data in healthcare applications [32] which considerably reduce the cost incurred in health care systems. Improving the health care systems performance by simulation optimization [11] is described well in this paper. A multi-objective demonstrating in health care facilities is a well-organized way to optimize the specified data. Multiple optimal solutions taking into account weights if considered objective function [18].

2. Big Data Security in Health Care Page Style

Data encryption schemes are useful to avoid possibilities of breaches such as packet sniffing and theft of data. Health care systems have may have to face attacks like distributed denial of service (DDOS), malicious software attacks, stealing data by insiders and outsiders [17]. Multiple biometric values are used to create keys for encryption to implement randomness [33]. With the Map-Reduce framework, the malicious activities of some nodes can be identified by the auditor nodes, which perform monitoring and accountability tests (A Tests) [30, 26]. The data model used in the scheme described in [5] collects the data from mobile phones, sensor networks as well as from social media. Symmetric key encryption is used to protect the data. A public key cryptosystem called ElGamal encryption algorithm is used for encrypting the symmetric key. This is simple, efficient and less-overhead public key encryption algorithm. The algorithm choice for encrypting the data depends on the sensitivity and complexity of the data. A mobile phone app is used to receive health care information in text, number and image formats from the user. The collected data are encrypted using the symmetric key and sent to the Cloud Service Provider (CSP). The CSP and the mobile app are assumed to be honest and not leaking the data out. The patient have the authority to decide whether his/her medical information to be shared to the respective doctor or not. Once the data are shared the corresponding doctor can access the data in social network and can decrypt the content. Whenever the patient wishes to revoke the permission of the doctor to access his/her data, the patient has the option to revoke the permission. Since the data is first encrypted and uploaded in the cloud, there is no possibility of Insider’s attack. Even if the mobile phone is stolen also, no one else can use the app to encrypt the data and upload the same to cloud without knowing the symmetric key. This work demonstrates the usefulness and effectiveness of Cloud in secured sharing of sensitive medical data patients.

Many of the security algorithms perform the user authentication at the entry level, mainly to prevent Man-in-the-Middle attacks. The Secure Socket Layer (SSL) and Transport Layer Security (TLS) are the security protocols which implement data protection over the Internet. Sensitive information in the store can be monitored using an algorithm called Blue Eye. This algorithm provides data security and relationship management between original data and replicated data. It also does user authentication for allowing users to read and modify critical data [25].

Searching is made efficient using the comparison of cosine similarity of any two vectors. But it requires the data to be decrypted for finding the cosine value of it [20]. Hence the special type of encryption algorithms namely, Homomorphic algorithms such as Pallier encryption, which does not require decryption. But the homomorphic algorithms are basically time-consuming algorithms.

3. Privacy Preserving in Big Data

De-identification is a classical privacy preserving technique used in data mining. When compared to the aggregation and operations on the encrypted data, the de-identification method makes the data mining activities more simple and effective. Still the hackers can get outside information to perform re-identification, which is a security breach in de-identification method [27].

The Hybrid Execution (HybrEx) is a model for private and public cloud functionality. HybrEx model uses public cloud if there is no privacy preserving requirements. The private cloud will be used by the model if the organization requires privacy to be implemented over the sensitive data being handled. [16, 24]. Once authenticated, the users of the data can enter a Big-data management system but their access will still be influenced by a set of rules of control policy which is usually based on the privilege and right of each doctor authorized by data owner or a trusted third party.

Role-Based Access Control (RBAC) and Attribute-Based Access Control (ABAC) are the most primitive and famous models for e-Patient data [28, 22, 9]. Different authorization models are analyzed in [1]. The Role Based Access Control model is widely used. But it not flexible to fine-grained access policies. That is if the patient wants only his/her data would be available within a time slot or a specific set of doctors could access the data, RBAC will not be useful [29, 7, 19, 8, 4]. In this scenario, ABAC can be chosen which provides the patients more control over the access control management. But the implementations of ABAC are monolithic and not suitable for distributed environment. It was found that an open authorization framework called SAFAX was best suited to the Big-data environment. This framework was developed by the Eindhoven University of Technology, which provides facilities to setup authorization solutions [14]. The modules of the SAFAX framework are available as web services, which provide flexibility and scalability. And hence, web service components of this framework can be used by any application irrespective of their location and without implementation of the app.

4. Data Optimization

The past few years, healthcare systems all over the world are enduring necessary transfiguration as they transfer from a volume-based to a value-based healthcare delivery model in an exertion to meet increasing demand for care and improve healthcare quality while limiting costs. Healthcare is a rich data field. When we collect more and more data, it will increase the demand for big data analytics. When we optimize the patient care in the growth of Big-Data in Healthcare, the main objective for every healthcare laborers is to provide the best quality care to all of their patients. Healthcare laborers can construct new, more productive processes to increase patient satisfaction through the strategic use of new mobile technologies, hospital wireless networks and latest improvements in big-data and analytics. Optimized health care system provides maximum performance with no lost resources. First we find patients having lot of risk and confirm they get the treatment they want. Second we improve processes to avoid maximum numbers of days a patient will stay in a hospital. Optimized healthcare system provides increase the satisfaction level of patients while minimizing the expenditure.

Simulation modelling is the process of creating and analyzing a digital prototype of a physical model to predict its performance in the real world. Simulation models diverge significantly in their breadth, deepness and practicality. It is used to address a wide-ranging of medical and economic decisions in healthcare occur in real life. In health care, simulation model needs modelling all the necessary features of human anatomy, physiology, pathology and response to medical treatment. Since scheduling is an important component of the happening, exhibition, development, organization and the model should be continuous in results of diseases. Presently, information are seized, kept and investigated in order to excerpt visions after the fact. A lot of profits should be recognized if data could be endlessly examined in real-time. It supports the acceptance of fast achievement. The objective of optimization will vary depends on schedule. The claim of optimization provides the stability between the calculating productivity and the computational expenditure for different figuring loads.

Now a days, Researchers targets utilization of huge volumes of medical data while joining multimodal data from distinct sources. Multi-objective method established on mixed integer linear programming to sustenance the action in healthcare facilities. It achieves an effective feasible operational plan for each resource on a day-to-day basis and also provide the patients and the caretaker fulfilment while controlling expenditures and concerning patient's favorites. Big data takes increase in time leads to increase in cost into a traditional comparative database for analysis. In recent years optimized big data provides reduce the waiting time and reduce the cost and also fulfilment of patient satisfaction. It conveys stable workload and minimal waiting time of caretakers and maximal patient satisfaction and minimal roaming time. A multi-objective modelling for health care system can be profitable for enhanced decision making.

Our objective is to determine a technique for each resource that fulfils several criteria: Minimization of the sum of the transport times, maximization of the balancing of the route dependency costs. For each route, calculate the sum of the level of dependence of all the visited patients. The objective is to divide up high dependent patients between the resources. Maximization of the degree of proficiency, each resource is more effective when he understands an activity for which he has a great level of competence.

5. Conclusion

Huge scope is available in the fields of health data research, knowledge discovery, clinical data analysis, and individual health management. Optimization and security of the health data are the two main issues that require more innovations. Security to Big-data can be implemented in different ways. Encrypting the data and access control are the techniques used. Depending upon the Big-data application, either encryption or access control will be used. In some systems both encryption and access control can also be used. Worth full researches had been made on the application encryption algorithms like, Homomorphic, ElGamal, and Triple-DES. User authentication can also use encryption. Not all the end users need to be given full access to the data. Restriction in data access will enforce access control over the Big-data. Resourcefully working with optimize and simulate the healthcare data sources provides immediate returns in terms of patient results and minimize the expenditures. It reduces the number of unnecessary hospitalizations. More difficult data provides in healthcare as a result leading to more chances for big data analytics.

6. Future Scope

The category of Homomorphic algorithms allows the users of the Big-data to work on the encrypted data. But formulating fully homomorphic algorithm for the target Big-data requires complex computation. The uncontrolled deviations and uncertain nature of the input data becomes more sensitive and influence more on the output of any Big-data analytical model. This fact can be used to alert the patient in advance for medical assistance.

References

[1] AlexandruSoceanu, AlexandruEgner, TraianMuntean (2015), Managing the Privacy and Security of eHealth Data. 20th International Conference on Control Systems and Science, 978-1-4799-1780-8/15 $31.00 © IEEE.

[2] AmineRahmani, Abdelmalek Amine and Mohamed Reda Hamou (2015), A mathematical model of access control in big data using confidence interval and digital signature. Computer Science and Information Technology, 183-198.

[3] Arthur C (2011), PlayStation Network: hackers claim to have 2.2m credit cards. The Guardian Technology Blog. http://www.guardian.co.uk/technology/blog/2011/apr/29/playstationnetwork-hackers-credit-cards.

[4] Balana, open source XACML 3.0 implementation http://xacmlinfo.org/category/balana.

[5] Brosette SE, Spragre AP, Jones WT, Moser SA (2000),A data mining system for infection control surveillance. Methods Inf Med, 303-310.

[6] DananThilakanathan, Yu Zhao, Shiping Chen, Surya Nepal, Rafael CalvoandAbelardo Pardo A (2014), Protecting and Analysing Health Care Data on Cloud",978-1-4799-8085-7/14 $31.00 ©IEEE

[7] Dolski S, Huonder F and Oberholzer S (2007), HERAS-AF: XACML 2.0 Implementation, Technical Report, University of Applied Sciences Rapperswil.

[8] Enterprise-Java-XACML, http://code.google.com/p/enterprise-javaxacml

[9] Hagner M (2007), Security infrastructure and national patent summary. In Tromso Telemedicine and eHealth Conference.

[10] Halamka JD (2015), Using Big Data to Make Wiser Medical Decisions, Harvard Business Review,[Online]. Available: https://hbr.org/2015/12/using-big-data-to-makewiser-medical-decisions.

[11] Hamid Reza Feili (2013), Improving the health care systems performance by simulation optimization. Journal of Mathematics and Computer Science, 7. 73-79.

[12] Healthcare costs rise to $130bn, or $5800 per Australian: report. http://www.news.com.au/lifestyle/health/health-spending-reaches-130b-report/story-fneuz9ev-1226481443042

[13] Indiana Health Information Exchange (2016), http://www.ihie.org

[14] Kaluvuri S, Egner A, Hartog J and Zannone N (2015), SAFAX – An Extensible Authorization Service for Cloud Environments, Frontiers in Computer and Network Security.

[15] Karvelas P (2014), Australias mental health system must become more efficient. The Australian. Source: http://www.theaustralian.com.au/nationalaffairs/policy/australiasmental-health-system-must-become-more-efficient/story-fn59nokw-1226850819260.

[16] Ko SY, Jeon K, Morales R (2011), The HybrEx model for confidentiality and privacy in cloud computing. In: 3rd USENIX workshop on hot topics in cloud computing, HotCloud'11, Portland.

[17] KupwadePatil H and Seshadri R (2014), Big Data Security and Privacy Issues in Healthcare, IEEE International Congress on Big Data, 762–765.

[18] Laila En-nahli, Hamid Allaoui and IssamNouaouri (2015), A multi-objective modelling to human resource assignment and routing problem for home health care services. Science Direct, 698-703.

[19] Liu A, Chen F, Hwang J and Xie T (2011), Designing fast and scalable XACML policy evaluation engines. IEEE Trans. Computers, 60(12) 1802-1817.

[20] Liu S (2013), Exploring the Future of Computing. IT Professional, vol. 15(1), 2–3.

[21] Marr B (2015), How Big Data Is Changing Healthcare. Forbes, [Online]. Available: http://onforb.es/1bfRQ0b.

[22] Mohan A, Blough DM (2010), An Attribute-Based Authorization Policy Framework with Dynamic Conflict Resolution. Proceedings of the 9th Symposium on Identity and Trust on the Internet.

[23] Patrizio A (2007), Salesforce.com Scrambles To Halt Phishing Attacks. Internet News.com. http://www.internetnews.com/ent-news/article.php.

[24] Priyank J, Manasi G and Nilay K (2016), Big data privacy: a technological perspective and review. In Journal of Big Data.

[25] RuiZhangand Ling Liu (2010), Security Models and Requirements for Healthcare Application Clouds. IEEE 3rd International Conference on Cloud Computing.,

[26] Samanthula BK, Elmehdwi Y and Jiang W (2015), K-nearest neighbour classification over semantically secure encrypted relational data. IEEE Transactions on Knowledge and Data Engineering, vol. 27(5), 1261–1273.

[27] Sathya S, Sethukarasi T (2016), Efficient privacy preservation technique for healthcare records using big data. International Conference On Information Communication And Embedded System (ICICES) 978-1-5090-2552-7.

[28] Science Applications International Corporation (2004) (SAIC). Role-Based Access Control (RBAC) Role Engineering Process Version 3.0.

[29] Sun'sXCAMLImplementation,http://sunxacml.sourceforge.net/

[30] Ulusoy H, Kantarcioglu M, Pattuk E and Kagal L (2015), Accountable MR: Toward accountable MapReduce systems. IEEE International Conference on Big Data, 451–460.

[31] VassilikiKoufi, Flora Malamateniou and George Vassilacopoulos (2015), A big data driven model for the optimization of healthcare processes, European Federation for Medical Informatics, 697-701.

[32] Vishnu S Basuthkar, ChetanaSrinivas (2016), A survey of cost effective big data in healthcare applications. International Journal of Computer Applications, 23-27.

[33] Waqar A, Raza A, Abbas H and Khurram Khan M (2013), A framework for preservation of cloud users data privacy using dynamic reconstruction of metadata. Journal of Network and Computer Applications, vol. 36(1), 235–248.