International Journal of Latest Trends in Engineering and Technology
Special Issue April-2018, pp. 052-056
e-ISSN:2278-621X

DECISION SUPPORT SYSTEM FOR DISEASE DIAGNOSIS USING CLASSIFICATION IN BIG DATA

Arun Manicka Raja M[1], Umamaheswari G[2], Valarmathi M[3], Sowndharya S[4]

Abstract -"Big Data" consists of very huge capacities of heterogeneous data that is being engendered, often, at high speeds. These data sets cannot be managed and processed using outdated data management tools and applications at hand. Big Data requires the use of a new set of tools, applications and frameworks to procedure and manage the data. Key enablers for the growth of "Big Data" are: Increase of storage capacities ◦ Increase of processing power ◦ Availability of data. Cloud computing is the result of the evolution and adoption of existing technologies and paradigms. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in privately owned on Personal Health Record (PHR). Cloud computing poses privacy concerns because the service provider can access the data that is in the cloud at any time. Implementing the cloud solution involves significant changes in the processes being followed by the healthcare providers. Healthcare facilities have been dependent on legacy systems and processes many of which are outdated and are not efficient. Sharing the medical records of individuals among healthcare providers and researchers around the world can accelerate advances in medical research. While the idea seems increasingly practical due to cloud data services, maintaining patient privacy is of paramount importance. Several variants of popular cryptosystems are also discussed, including the necessary modifications to security assumptions.
Keywords: Hadoop, Cloud computing, PHRs, HIPAA, PBE, DPA, Homomorphic Encryption, e-health, Privacy Preserving.

1. INTRODUCTION

As the world is turning deeper into the "Digital Age," we're witnessing an explosive growth in the volume, velocity, variety, veracity, and value (the 5Vs) of data produced over the Internet. Volume: Many factors contribute towards increasing Volume streaming data and data collected from sensors etc., Variety: Today data comes in all types of formats emails, video, audio, and transactions etc., Velocity: This means how fast the data is being produced and how fast the data needs to be processed to meet the demand. Variability: Along with the Velocity, the data flows can be highly inconsistent with periodic peaks. Value: Value of the data also needs to be considered when the data is coming from multiple sources. The data must be linked, matched, cleansed and transformed into required formats before actual processing

Cloud computing is an emerging trend in software field as it focuses on shared resources. Cloud computing purely relies on virtualization, which increases the infrastructure deployment in reduced cost and IT operations are scampered up. Moving to cloud in Health Care is the major challenge rather than other fields because of the above stated reason. So, we are proposing architecture to overcome the pitfalls in the cloud by moving to cipher cloud. The paper is structured as follows. First, we briefly introduce basics of cloud computing with its deployment models and services rendered by it. Section 2 deals with the related work from the literature. A motivation behind cloud computing in Healthcare is also provided. In Section 3, we discuss about the proposed work in which PHR (Personal Health Record) architecture is combined with inter cloud models to enhance data security and reliability. Section 4 presents the methodology to achieve the above. Section 5 draws some concluding remarks and some ideas for future work.

2. RELATED WORK

Patients and health organizations take compensations of the new technology by improving patient's quality of service through a distributed high-integrated platform, coordinating of medical process as well as reducing IT infrastructure investment or maintenance costs which leads to a better healthcare environment. Concentrating on the Global Market for Cloud Computing in Healthcare, IT industries invest heavily to build infrastructure for cloud to support it and help organizations take benefit from it. The rate of increase in adopting cloud is directly proportional to the rate of achieving greater efficiencies. This results in providing extraordinary sharing capabilities between the healthcare organizations and patients alike. The Challenges of

[1] Assistant Professor, Department of Computer Science and Engineering, R M K College of Engineering and Technology, Chennai, Tamil Nadu, India
[2] U·G Student, Department of Computer Science and Engineering, R M K College of Engineering and Technology, Chennai, Tamil Nadu, India
[3] U·G Student, Department of Computer Science and Engineering, R M K College of Engineering and Technology, Chennai, Tamil Nadu, India
[4] U·G Student, Department of Computer Science and Engineering, R M K College of Engineering and Technology, Chennai, Tamil Nadu, India

Cloud Computing in Health Care is mostly due to two important concerns associated with security and interoperability. Health Insurance Portability and Accountability Act (HIPAA) when moving health record to the cloud. The healthcare data comprises of sensitive information yet migration of the medical records to the cloud (third-party) may be trusted. To prevent uncovering of information to unauthorized persons, security activities such as imposing access controls, providing authentication, checking authorization can be done. Patient centric framework and mechanisms for data access control to PHRs stored in semi-trusted servers. Attribute-Based Encryption (ABE) technique was used to encrypt each patient's PHR file. It exploits multi-authority ABE for the privacy of patient's by dynamic modification of access policies. Revoking of access policy is not possible at all the moments and the attributes which were known to the user leads to privacy concern. Some applications analyzed in this paper were primarily inspired from. The work exhibits slow running time. The work on the other hand has considerably faster running times but at the expense of having an incomplete implementation of the HE schemes that needs to provide the client with information about the depth of the computation made in order to correctly decrypt the result.

3. PROPOSED WORK

In the proposed work, for ensuring security, encryption tools are used along with attributes of user. Usage of tools overcomes the complexity in key management in PBE schemes. So, we are proposing architecture by combining the concepts of homomorphic encryption and PBE in which the PHR encrypted by encryption tool is again encrypted by using Parameter Based Encryption (PBE).

- The intention of the mission is to ensure the correct diagnosis of any illness with the help of decision support system.
- The decision support system is used for implementing the healthcare with the use of software.
- Hadoop is used to classify and predict the disease of the patient based upon the symptoms.
- Patient's Health records (PHR's) are maintained in the public cloud where each and every patient is provided with an ID.
- Since the PHR's contain the sensitive information the records are encrypted using the Homomorphic Based Encryption (HBE).
- The goals are: Ease of retrieval/collection of the specific information, less time consumption, cost effective, scalable, Fault tolerant and increase in security.

4. METHODOLOGY

4.1 Homomorphic Encryption:

Homomorphic encryption is a form of encryption that allows computations to be carried out on cipher text, thus generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext.

4.2 Personal Health Record

A personal health record is a health record where health data and information related to the care of a patient is maintained by the patient. [1] This stands in contrast to the more widely used electronic medical record, which is functioned by institutions (such as hospitals) and contains data entered by clinicians or billing data to support insurance claims. The purpose of a PHR is to provide a complete and accurate precipitate of an individual's medical history which is accessible online. The health data on a PHR might include patient-reported outcome data, lab results, data from devices such as wireless electronic weighing scales or collected passively from a smartphone.

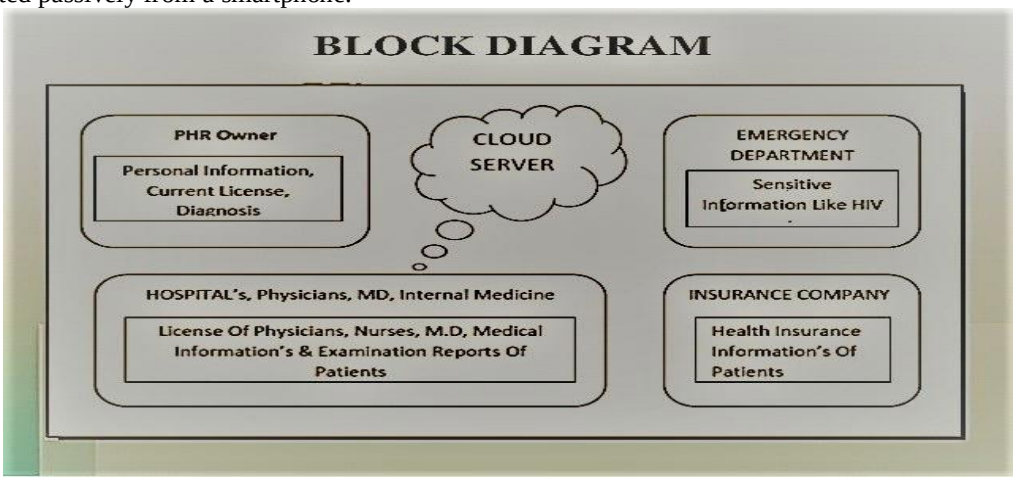

Figure 1. Proposed Architecture Of Phr

4.3 Key Management And Module Description:

The overarching theme of a PHR involves a patient centric tool that is controlled for the most part, by the individual. It should be immediately available electronically, and able to link to other systems, either in a "drag-drop" or "drag-drop"

method. The PHR is intended to provide functionality to help an individual maintain a longitudinal view of his or her health history, and may be comprised of information from a plethora of sources—i.e., from providers and health plans, as well as from the individual. Data collected by the system is administrative and/or clinical, and the tool may provide access to a wealth of forms (advance directives) and advice (diet, exercise, disease management). A PHR would help the individual collect behavioural health, public health, patient entered and patient accessed data (including medical monitoring devices), medication information, care management plans and the like, and could be connected to providers, laboratories, pharmacies, nursing homes, hospitals and other institutions and clinical resources.

At its heart, the PHR should endow with the ability for the individual to capture and maintain demographic, insurance coverage, and provider information. It should also provide the ability to capture health history in the form of a health summary, problems, conditions, symptoms, allergies, medications, laboratory and other test results, inoculations and encounters. Additionally, personal care planning features such as advance directives and care plans should be available. The system must be secure and have appropriate identity and access management capabilities, and use standard nomenclature, coding and data exchange standards for consistency and interoperability. A host of optional features have been addressed over the course of this initiative, including secure messaging, graphing for test results, patient education, guideline-based reminders, appointment scheduling and reminders, drug-drug interfaces, formulary management, health care cost comparisons, document storage and clinical trial eligibility. The effective use of a PHR is a key point for improving healthcare in terms of self-management, patient-provider communication and quality outcomes.

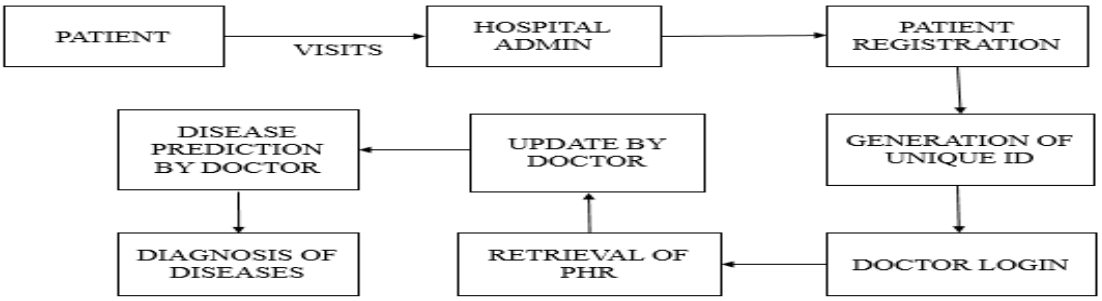

Figure 2. Data Flow Model

The modules are explained here:

- *Admin:*

The role of admin is to maintain all PHR's of the patient on the regular basis. Hospital admins are responsible for the day-to-day operation of a hospital, clinic, manages care organization or public health agency. To coordinate the actions of all the departments and ensure they function as one, hospital admin's must hold a wide set of skills and knowledge.

- *Doctor:*

Hospital doctors examine, diagnose and treat patients who have been referred to the hospital by GPs and other health professionals. They apply their medical knowledge and skills to the diagnosis, prevention and management of disease. In relation to the proposal, the doctor is provided with the login details where he/she can personally use their account for diagnosing the patients.

- *Encryption:*

The translation of data into a secret code. Encryption is the most effective way to achieve data security. To read an encrypted file, you must have access to a secret key or password that enables you to decrypt it. Unencrypted data is called plain text; encrypted data is referred to as cipher text. Competing algorithms were to be judged on their ability to resist attack, as compared to other submitted ciphers, though security strength was to be considered the most important factor in the competition.

- *Pre-processing:*

In this module, the doctor diagnoses the existing patients where they are informed to take the prescribed test. When the patient visits the doctor again he/she just provide the unique ID by which the doctor gets the information about the patient and the input values are provided from the test reports.

- *Disease prediction:*

It might have happened so many times that you or someone yours need doctors help immediately, but they are not available due to some reason. The Disease Prediction application(DPA) is an end user support. Here, we propose a web application that allows users to get instant guidance on their disease through an intelligent system. The application is fed with various details and the disease associated with those details. The application allows user to share their symptoms related issues. It then processes user specific details to check for various illness that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's details. Based on result, the can contact doctor accordingly for further treatment. The system allows user to view doctor's details too. So, it is very useful in case of emergency.

5. CLOUD COMPUTING AND ITS BENEFITS:

5.1 Improved medical research:

Much in the way big data is making it possible for doctors to treat their patients in improved ways, that the cloud makes it possible via storing and sharing data to speed up the research process. With the ability to collect outside data from multiple fields, data analysts can use the cloud to pool this data and abbreviate it into better results, allowing the medical professionals to get a clearer and more advanced image of the subjects they are researching. These kinds of advances are the kind that cure diseases and improve the kind of care being given.

5.2 Remote Patient Care:

With new mobile devices being made to monitor a patient's ailment, and even applications on your smartphone making it possible to keep your doctor up-to-date on your condition or get remote consolation, cloud is making it possible for you to get high quality care without ever treading into a hospital. If you're unable to get to the hospital, don't want to spend the money, or dislike visiting doctors altogether, you can use new devices powered by the cloud to transmit your condition or ask for advice from a doctor on standby, which will allow both patients and medical professionals to catch dangers early.

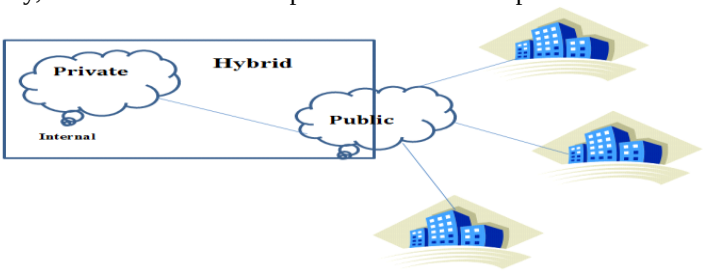

Figure 3. Public Cloud

The cloud services are exposed to the public and can be used by anyone. Virtualization is typically used to build the cloud services that are offered to the public. An example of a public cloud is Amazon Web Services (AWS).

6. SECURITY ISSUES IN CLOUD COMPUTING:

We understand that the Cloud Computing is lacking security, confidentiality and visibility. To Provide Infrastructure (IaaS), Platform Service (PaaS) or Software (SaaS) as a Service is not sufficient if the Cloud provider does not guaranty a better security and confidentiality of customer's data. By tenacity, we consider a Cloud Computing as any conduct or storage of personal or professional information which are realized outside the concerned structure to secure the. Secure storage and treatment of data requires using a modern aspect of cryptography that Secure Cloud Computing through Homomorphic Encryption has the criteria for treatment such as, the necessary time to respond to any request sent from the client and the size of an encrypted data which will be stored on the Cloud server. Transfer the processing of your data to a third party; it is also transferring some of the responsibility associated with their security and compliances. The advantages of cloud computing include reduced costs, easy maintenance and re- provisioning of resources, and thereby increased profits. Our proposal is to encrypt data before sending it to the cloud provider, but to execute the calculations the data should be decrypted every time we need to work on it. Until now it was impossible to encrypt data and to trust a third party to keep them safe and able to perform distant calculations on them. So to allow the Cloud provider to perform the operations on encrypted data without decrypting them requires using the cryptosystems based on Homomorphic Encryption [2].

7. CONCLUSION AND FUTURE ENHANCEMENT:

The Security of Cloud Computing based on fully Homomorphic Encryption is a new concept of security which empowers to provide the results of intentions on encrypted data without knowing the raw entries on which the design was carried out respecting the confidentiality of data. The terabytes of patient health records are maintained in database which help clinicians to predict the correct diagnosis of any illness of the patient by the process of decision support system. The main contribution of this work lies in the fact that Disease Prediction application(DPA) is able to assist in the quick diagnosis of rare diseases using a small amount of patient information. We plan to enlarge DPA to other diseases using the following strategy. Initially, we will examine which datasets are accessible that associate diseases with their symptoms. Furthermore, we will select those datasets that include groups of diseases related with their patients' symptoms. To conclude, we will apply strategies similar to those offered in the up-to-date paper and test how well these methods scale to other sets of diseases.

8. REFERENCES:

[1] M.Mozaffari-KermaniS.Sur-Kolay, A. Raghunathan, and N. K. Jha, "Systematic poisoning attacks on and defences for machine learning in healthcare on health prediction," IEEE J. Biomedical and Health Informatics, vol. 19, no. 6, pp. 1893–1905, 2015.

[2] M. Shoaib, N. K. Jha, and N. Verma, "Signal processing with direct computations on compressively sensed data," IEEE Trans. Very Large Scale Integr. Syst., vol. 23, no. 1, pp. 30–43, 2015.

[3] G. N. Forrest, T. C. Van Schooneveld, R. Kullar, L. T. Schulz, P. Duong, and M. Postelnick, "Use of electronic health records and clinical decision support systems for antimicrobial stewardship," Clinical Infectious Diseases, vol. 59, pp. 122–133, 2014.

[4] D. Suryakumar, A. H. Sung, and Q. Liu, "Influence of machine learning vs. ranking algorithm on the critical dimension," Int. J. Future Computer and Communication, vol. 2, no. 3, pp. 215–220,2013.

[5] M. Li, S. Yu, K. Ren, and W. Lou, "Securing personal healthrecords in cloud computing: Patient-centric and fine-grained data access control in multi - owner settings," in SecureComm'10, Sept. 2010, pp. 89–106. J. Baek, R. Safavi-Naini, and W. Susilo, "Public Key Encryption with Keyword Search Revisited," in Computational Scienceand Its Applications – ICCSA 2008, ser. Lecture Notes in Computer Science, O. Gervasi, B. Murgante, A. Lagan`a,D. Taniar, Y. Mun, and M. Gavrilova, Eds. Springer Berlin Heidelberg, 2008, vol. 5072, pp. 1249–1259. [Online]. Available:dx.doi.org/10.1007/978-3-540-69839-5 96

[6] D. Boneh, C. Gentry, S. Halevi, F. Wang, and D. Wu,"Private Database Queries Using Somewhat Homomorphic Encryption," in Applied Cryptography and Network Security, selector Notes in Computer Science, M. Jacobson, M. Locasto,P. Mohassel, and R. Safavi-Naini, Eds. Springer BerlinHeidelberg, 2013, vol. 7954, pp. 102–118. [Online]. Available:dx.doi.org/10.1007/978-3-642-38980-1 7

[7] D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Ros¸u, andM. Steiner, "Highly-Scalable Searchable Symmetric Encryption with Support for Boolean Queries," in Advances in Cryptology– CRYPTO 2013, ser. Lecture Notes in Computer Science. Canetti and J. Garay, Eds. Springer Berlin Heidelberg, 2013,vol. 8042, pp. 353–373. [Online]. Available: dx.doi.org/10.1007/978-3-642-40041-4 20

[8] S.Vidya, K.Vani, D.Kavinpriya," Secured Personal Health Records Transactions using Homomorphic Encryption in Cloud Computing", IJERT, December 2012.

[9] B. Hayes, "Alice and Bob in Cipherspace," 2012, vol. 100, no. 5, pp. 362–367.

[10] Aderemi A. Atayero*, OluwaseyiFeyisetan, "Security Issues in Cloud Computing: "The Potentials of Homomorphic Encryption", Journal of Emerging Trends in Computing and Information Sciences, VOL. 2, NO. 10, pp. 546-552, October 2011

[11] V.Lyubashevsky, C. Peikert, and O. Regev, "On Ideal Lattices and Learning with Errors over Rings," in Advances in Cryptology – EUROCRYPT 2010, ser. Lecture Notes in Computer Science, H. Gilbert, Ed. Springer Berlin Heidelberg, 2010, vol. 6110, pp. 1–23. [Online]. Available: dx.doi.org/10.1007/978-3-642-13190-5 1.

[12] R. Lindner and C. Peikert, "Better Key Sizes (and Attacks) for LWE-based Encryption," in Proceedings of the 11th International Conference on Topics in Cryptology: CT-RSA 2011, ser. CT-RSA'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 319–339. [Online]. Available: http://dl.acm.org/citation.cfm?id=1964621.1964651C. A. Melchor, "High-Speed Single-Database PIR Implementation," in the 8th Privacy Enhancing Technologies Symposium (PETS 2008), 2008, pp. 1–12.