ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

IOT Based Healthcare and Medical Knowledge Extraction System for Medical Big Data

Kunal Sharad Raghorte[1], Simpi[1], A.U. Gahankari[2], Gopal Sakarkar[2,]

MCA Student, Department of Computer Application, G.H Raisoni College of Engineering, Nagpur, India[1]

Assistant Professor, Department of Computer Application, G.H Raisoni College of Engineering, Nagpur, India[2]

ABSTRACT: With development of cloudlet and wearable technology, it is necessary to provide best medical data sharing over internet. As we know, medical data sharing is critical and challenging issue because medical data contains patient's sensitive information. The functions of the cloudlet include privacy protection and intrusion detection. In this system, medical data sharing is done in energy efficient fashion using Number Theory Research Unit (NTRU). NTRU provide computationally fast and efficient method to implement public key cryptography. Protection against malicious attacks is achieved through a Collaborative Intrusion Detection System (CIDS).

Along with growth in technology, young generation always prefer to search health related information, doctor's suggestion on any health related problem on web through internet. Tens of millions of health related queries are searched every day. E.g. medhelp.com, xywy.com. This new type of healthcare service brings opportunities and challenges to doctors, patients and service providers. To overcome these challenges, Medical Knowledge Extraction (MKE) is proposed that extract medical knowledge from noisy Q-A website and estimate the trustworthiness of answers with doctor's expertise using truth discovery method.

KEYWORDS: Privacy protection, Collaborative Intrusion Detection System (CIDS), Healthcare, Number Theory Research Unit (NTRU), Medical Knowledge Extraction (MKE), Truth discovery.

I. INTRODUCTION

Human services social stage, for example, Patients-Like Me, can acquire data from other comparable patients through information partaking regarding client's own particular discoveries.

In spite of the fact that sharing medicinal information on the interpersonal organization is valuable to the two patients and specialists, the delicate information may be spilled or stolen, which causes protection and security issues without productive assurance for the common information.

1.1 MOTIVATION

Traditional healthcare system require the delivery of medical data to the cloud which involves user's private information which is sensitive and that delivery causes communication energy consumption. As we know sharing this medical data on social network is useful for both patients and doctors. The patient's sensitive data might be stolen or leaked which causes privacy and security problem.

In order to provide better medical data sharing in an energy efficient fashion and secure whole healthcare system from malicious attacks and provide the most reliable answer from noisy Q-A pairs without any supervision, a cloudlet based healthcare and medical knowledge extraction system is proposed. The proposed system also aims to protect the user's medical data stored in remote cloud of hospital.

DOI: 10.15680/IJIRCCE.2018.0602060

1.2 BACKGROUND

Along with the development of clouds, cloudlet technology and wearable technology, it is necessary to provide better medical data sharing over the internet .In traditional healthcare system, the medical data which involves user's sensitive information was delivered to the remote cloud which causes communication energy consumption. In order to reduce this communication energy consumption, a novel healthcare system is proposed by using the flexibility of cloudlet. Privacy protection and intrusion detection are the two main functions of the cloudlet. Large amount of patients and doctors are involved in the medical crowd sourced question answering website in recent years. Extraction of medical knowledge from the noisy question-answers pair and filter out unrelated or incorrect information are challenging issues. To overcome these issues, Medical Knowledge Extraction system is proposed that can automatically provide high quality knowledge triples and estimate the expertise for the doctors.

1.3 NEED

It is necessary to provide privacy protection to the user's body information and medical records stored in remote cloud of hospital. Intrusion avoidance is necessary to protect the whole healthcare system from malicious attacks. Also to provide more reliable answers to the patient's health related questions.

II. STATE OF THE ART

Y. Shi, S. Abhilash, and K. Hwang [1]. We have specified a sequence of authentication, authorization, and encryption protocols for securing communications among mobile devices, cloudlet servers, and distance clouds. Securing mobile cloud services is the major barrier to the integration of BTOD (bring your own devices) and BYOC (bring your own cloud) in our daily applications. We use the cloudlet mesh to perform collaborative intrusion detection among multiple cloudlets. Network attacks are a serious matter that confronts both cloud providers and massive number of mobile users who access distance clouds in our daily-life operations. We extend their work to support security functionalities in offloading the distance clouds.

M. S. Hossain [2]. The proposed approach uses Gaussian mixture modeling for localization and is shown to outperform other similar methods in terms of error estimation. The design and development of such systems requires access to substantial sensor and user contextual data that are stored in cyberspace. We will conduct more workload measurements to record the resource utilization of CPU, memory, storage, and network bandwidth. This enables a range of emerging applications or systems such as patient or health monitoring, which require patient locations to be tracked.

A. Sajid and H. Abbas [3]. The system is privacy assured where cloud sees neither the original samples nor underlying data. It handles well sparse and general data, and data tampered with noise. We have proposed a privacy-aware cloud assisted healthcare monitoring system via compressive sensing. The random mapping based protection ensures no sensitive samples would leave the sensor in unprotected form. Wireless sensors are being increasingly used to monitor/collect information in healthcare medical systems. Despite the increasing popularity, how to effectively process the ever-growing healthcare data and simultaneously protect data privacy, while maintaining low overhead at sensors, remains challenging.

R. Mitchell and I.-R. Chen [4]. We exhibit that our interruption location system can viably exchange false positives off for high discovery likelihood to adapt to more complex and concealed aggressors to help ultra protected and secure MCPS applications. For security basic MCPSs, having the capacity to recognize aggressors while restricting the false alert likelihood to ensure the welfare of patients is of most extreme significance. We intend to break down the overheads of our discovery strategies, for example, the different separation based techniques in correlation with contemporary methodologies. We propose and analyze a behavior-rule specification-based technique for intrusion detection of medical devices embedded in a medical cyber physical system (MCPS) in which the patient's safety is of the utmost importance.

M. Quwaider and Y. Jararweh [5]. The proposed work additionally endeavors to limit the conclusion to-end parcel delay by picking powerfully a neighbor cloudlet, with the goal that the general deferral is limited. The objective was goal to limit end-to-end bundle cost by progressively picking information gathering to the cloud utilizing cloudlet based framework. Execution of the proposed framework was assessed by means of expanded rendition of CloudSim test system. The gigantic measure of information gathered by BAN hubs requests versatile, on-request, effective, and secures capacity and handling foundation.

J. Zhao, L. Wang, J. Tao, J. Chen, W. Sun, R. Ranjan, J. Kolodziej, A. Streit, and D. Georgakopoulos [6]. We describe an EHR security reference model for managing security issues in healthcare clouds, which highlights three important core components in securing an EHR cloud. The goal of this research is to advance the Map Reduce framework for large-scale distributed computing across multiple data centers with multiple clusters. The designed security framework has the ability to prevent the most common attacks, such as MITM attack, replay attack, and delay attack, and ensures a secure communication of GHadoop over public networks. The Map Reduce tasks are firstly scheduled among the clusters using Hadoop’s data-aware scheduling policy and then among compute nodes use the existing cluster scheduler on the target clusters.

N. Cao, C. Wang, M. Li, K. Ren, and W. Lou [7]. We first offer a basic idea for the multi keyword ranked search over encrypted cloud data (MRSE) based on effective comparison measure of coordinate matching. We have adopted an efficient strategy to researching security models and security prerequisites for medicinal services application mists. We have talked about vital ideas identified with EHR sharing and incorporation in human services mists and investigated the emerging security and protection issues in access and administration of EHRs. The far reaching utilization of electronic wellbeing record (EHR), building a protected HER sharing condition has pulled in a ton of consideration in both human services industry and scholarly group.

H. Mohamed, L. Adil, T. Saida, and M. Hicham [8]. We propose a community oriented model comprises of the Intrusion Detection and Prevention System capacities based appropriated IDS and IPS, with the utilization of a half and half discovery strategy for tending to the issues of assaults experienced, particularly conveyed assaults, for example, port examining assaults and disseminated inside set up inside a Cloud Computing condition by clients qualified for get to, including the mix of the Signature Apriori Algorithm for creating new assault marks whose goal is to build up the working of our security framework to have the capacity to identify and piece different sorts of assaults and interruptions. Security arrangements are not yet adjusted to this new idea. Without a doubt, in such a situation, the more clients and ways, the more prominent the interruption is viable. We additionally join the mark apriori calculation to improve and refresh our database mark to dissect and analyze data got. Cloud Computing has emerged as a model to process large volumetric data. They add that Cloud Computing deals with different fundamentals like virtualization management, fault tolerance and load balancing.

R. Zhang [9]. We depict an EHR security reference demonstrate for overseeing security issues in medicinal services mists, which features three vital center segments in securing an EHR cloud. We have adopted a precise strategy to researching security models and security prerequisites for medicinal services application mists. We have examined critical ideas identified with EHR sharing and combination in social insurance mists and broke down the emerging security and protection issues in access and administration of EHRs. The widespread use of electronic health record (EHR), building a secure EHR sharing environment has attracted a lot of attention in both healthcare industry and academic community.

K. Hung, Y. Zhang, and B. Tai [10]. As an essential piece of this framework, a cuffless BP meter has been created and tried on 30 subjects in a sum of 71 trials over a time of five months. Utilization of portable correspondence is never again restricted to communication. New interests and requests are remote information and interactive media administrations, as 3G telephones are accessible. The world’s maturing populace and pervasiveness of ceaseless infections have prompt popularity for tele-home human services, in which imperative signs observing is basic.

III. PROBLEM STATEMENT

To design, develop and implement a Cloudlet based healthcare and Medical Knowledge Extraction system in order to provide privacy protection to the user's body data, intrusion avoidance with high intrusion detection rate and more reliable answer to the patient's question.

The key generation scheme is used to generate the private and public key pair. The process begins by choosing two small polynomials f and g, where small is defined as having coefficients much smaller than the large modulo p and modulo q.

The user must compute the inverse of f modulo q and the inverse of f modulo p such that
f * fq = 1 (mod q) and f *
fp = 1 (mod p).
The inverse of f is calculated both modulo p and modulo q, generating,
fp = f-1 (mod p) and fq = f-1 (mod q).
The values of f and fp are retained as the private key pair and the public key h is calculated using p, fq and g. The public key is as follows:
h = pfq * g (mod q).... (1)

NTRU encryption
The encryption process starts by generating a polynomial message m whose coefficients lie in an interval of length q, which is normally cantered around zero. A small random blinding polynomial, r, is then generated and used to obscure the message [8]. The final encryption uses m, r and the public key h to generate e, the encrypted message that is as follows:
e = r * h + m (mod q).... (2)

NTRU decryption
The decryption process first uses the private key f to calculate:
a = f * e (mod q).... (3)
The coefficients of a must be chosen in the proper interval of length q to ensure the highest probability that the decryption process will be successful. Once the coefficients of a are chosen on the proper interval, a is reduced modulo p and the second private key is used to compute:
b = a (mod p)..... (4)
c = fp * b (mod p)..... (5)
If decryption has successfully completed, then the polynomial c will be equal to the original message.

Set of medical questions Q and their corresponding answers,

$$\{x_q^d\}q€Q, d€D$$

An external entity dictionary with entity types, and real- value vector representations of entities.
The basic idea of Cloudlet based healthcare and medical knowledge extraction system is to provide privacy protection and intrusion avoidance for cloudlet based medical data sharing over internet. This system also provides most trustworthy answers to the patients. The system is built up by utilizing the flexibility of cloudlet.

IV. PROPOSED SYSTEM

The proposed system has following stages:
Data collection by wearable device: The body data of user collected by wearable devices are protected by NTRU method in order to provide protection before it is transmitted to the cloudlet. The encrypted data will then stored in nearby cloudlet through cellular network or Wi-Fi.

Collaborative Intrusion Detection System (CIDS): CIDS is designed among M different Intrusion Detection System (IDS) in order to get high detection rate. The M IDS are assumed to detect independently. Before transmitting data to remote cloud, CIDS based on cloudlet mesh is designed to complete the intrusion detection task. O nce a malicious attack is detected, CIDS will fire an alarm or block the visit.

Medical data privacy protection in remote cloud: Remote cloud contain data generated from patients treated in hospital. Data in remote cloud is in Electronic Medical Records EMR) form. In this stage EMR are divided into 3 classes:
1. EID (Explicit Identifier) - EID are the properties which can identify users apparently. Ex Name, Phone no, Email, Home Address.
2. QID (Quasi-Identifier) – QID are the properties which can identify users approximately i.e. date of birth, gender.
3. MI (Medical Information) – Some clinical manifestation and disease type.
To protect privacy of data and make it convenient for doctors to access data from remote cloud, EID and QID are encrypted.

Disease prediction model: As doctor has access to remote cloud, a disease prediction model is built based on decision tree. The prediction will reported to users and provide doctors on demand.

Medical Knowledge Extraction System (MKE): MKE extract knowledge triples < question, diagnosis, trustworthiness degree > from Q-A pair. Entity extraction is applied on Q-A pair that gives entities from noisy Q-A pair. On these extracted entities (i.e.< question, diagnosis, source >) truth discovery method is applied. The goal of truth discovery is to resolve the conflicts and find truth i.e. most trustworthy answer for each question by estimating doctor’s reliability i.e. doctor expertise.

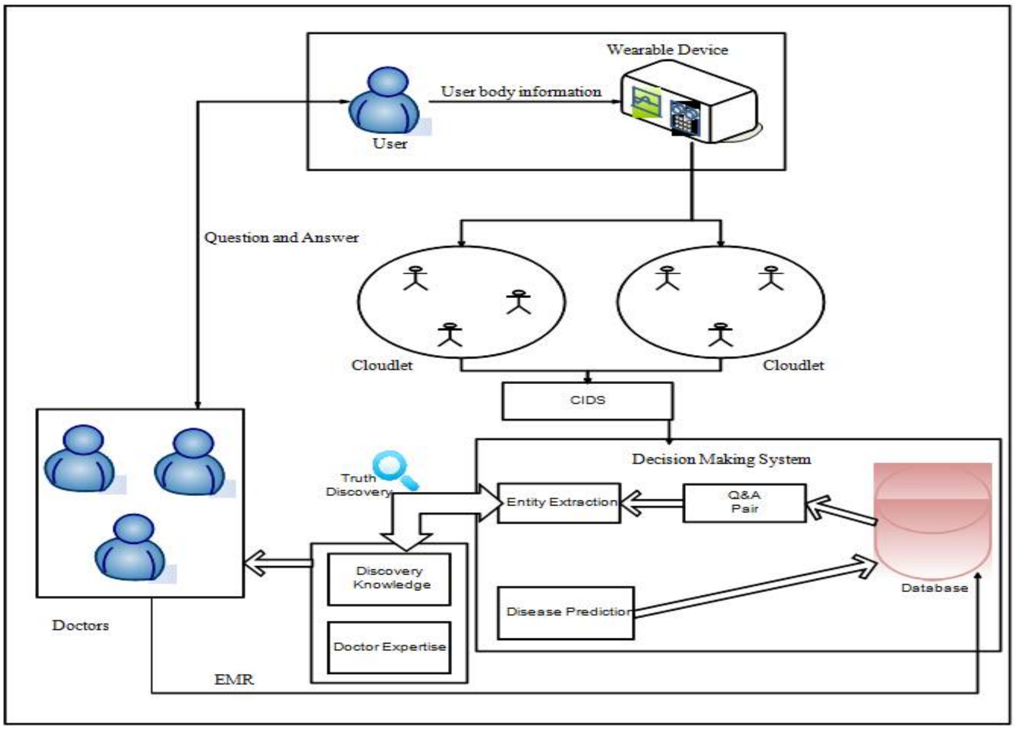

Figure: 1. Proposed architecture diagram.

V. EXPERIMENTAL RESULT

The body information gathered by wearable device is transmitted to the adjacent cloudlet. That information is additionally conveyed to the remote cloud where specialists can get to for disease finding. In the main stage, user's vital signs gathered by wearable gadgets are conveyed to gateway of cloudlet. In this stage, information security is the primary concern. In the second stage, client's information will be additionally conveyed toward remote cloud through cloudlets. A cloudlet is framed by a specific number of cell phones whose proprietors may require as well as offer some particular information substance. In this manner, both security insurance and information sharing are considered in this stage.

	1	2	3	4	5	6	7
Doctor 1	0.2	0.25	0.05	0.1	0.3	0.15	0.27
Doctor 2	0.22	0.17	0.31	0.37	0.7	0.16	0.23
Doctor 3	0.2	0.43	0.05	0.05	0.1	0.09	0.07
Doctor 4	0.05	0.17	0.2	0.25	0.21	0.05	0.17
Doctor 5	0.23	0.45	0.05	0.15	0.18	0.8	0.2
Doctor 6	0.26	0.3	0.08	0.23	0.43	0.23	0.45

Figure: 2. Doctor Expertise

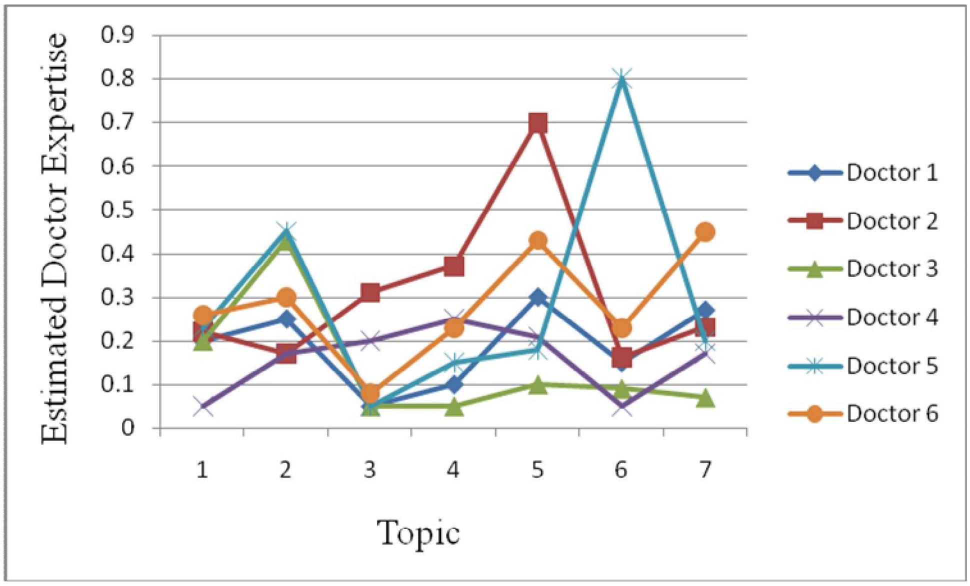

Figure: 3. Observe Doctor Expertise over Topics

Figure 2 clearly show that the estimated doctor expertise scores are quite different over topics. Means doctor cannot be experts in all topics, some doctors are experts in some topic and other also. For example, doctor 1 might be an expert on topic 5, but not on other topics. Doctor 5 has a high expertise score on topic 7, while his expertise scores on other topics are quite low. These observations confirm the intuition about the necessity of fine-grained doctor expertise estimation.

VI. CONCLUSION

In this project, we investigated the problem of privacy protection and sharing large medical data in cloudlets and the remote cloud. We developed a system which does not allow users to transmit data to the remote cloud in consideration of secure collection of data, as well as low communication cost. However, it does allow users to transmit data to a cloudlet. Firstly, we can utilize wearable devices to collect user's data, and in order to protect users privacy, we use NTRU mechanism to make sure the transmission of user's data to cloudlet. Secondly, for privacy-preserving of remote cloud data, we partition the data stored in the remote cloud and encrypt the data in different ways, so as to not just ensure data protection but also accelerate the efficacy of transmission. Finally, we propose collaborative IDS based on cloudlet mesh to protect the whole system. Medical Knowledge Extraction (MKE) systems that can automatically provide high quality knowledge triples extracted from the noisy question-answer pairs, and at the same time, estimate expertise for the doctors who give answers on these Q& A websites.

REFERENCES

[1] A. Sajid and H. Abbas, "Data privacy in cloud-assisted healthcare systems: State of the art and future challenges," Journal of Medical Systems, vol. 40, no. 6, pp. 1-16, 2016.

[2] R. Mitchell and I.-R. Chen, "Behavior rule specification-based intrusion detection for safety critical medical cyber physical systems," Dependable and Secure Computing, IEEE Transactions on, vol. 12, no. 1, pp. 16-30, 2015.

[3] Y. Shi, S. Abhilash, and K. Hwang, "Cloudlet mesh for securing mobile clouds from intrusions and network attacks," in The Third IEEE International Conference on Mobile Cloud Computing, Services, and Engineering,(Mobile Cloud 2015). IEEE, 2015.

[4] M. Quwaider and Y. Jararweh, " Cloudlet-based efficient data collection in wireless body area networks," Simulation Modelling Practice and Theory, vol. 50, pp. 57-71, 2015.

[5] M. S. Hossain, "Cloud-supported cyber-physical localization framework for patients monitoring," 2015.

[6] J. Zhao, L. Wang, J. Tao, J. Chen, W. Sun, R. Ranjan, J. KoÅC´ odziej, A. Streit, and D. Georgakopoulos, "A security framework in g-hadoop for big data computing across distributed cloud data centres," Journal of Computer and System Sciences, vol. 80, no. 5, pp. 994-1007, 2014.

[7] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 1, pp. 222-233, 2014.

[8] H. Mohamed, L. Adil, T. Saida, and M. Hicham," A collaborative intrusion detection and prevention system in cloud computing," in AFRICON, 2013. IEEE, 2013, pp. 1-5.

[9] R. Zhang and L. Liu," Security models and requirements for healthcare application clouds," in Cloud Computing (CLOUD), 2010 IEEE 3rdInternational Conference on. IEEE, 2010, pp. 268-275.

[10] K. Hung, Y. Zhang, and B. Tai, "Wearable medical devices for telehome healthcare," in Engineering in Medicine and Biology Society, 2004. IEMBS 04.26th Annual International Conference of the IEEE, vol. 2. IEEE, 2004, pp. 5384-5387.