ARTICLE IN PRESS

Telematics and Informatics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

A trustworthy system for secure access to patient centric sensitive information

Yachana, Navroop Kaur*, Sandeep K. Sood

Guru Nanak Dev University Regional Campus Gurdaspur, India

ARTICLE INFO

Keywords:
Patient Centric Big Medical Data
Information security
Privacy
Trust
Access Control

ABSTRACT

Smart technological innovations in healthcare are continuously generating digitized medical information about each patient, leading to the creation of Patient Centric Big Medical Data (PCBMD). Rapid adoption of PCBMD in healthcare ushers at the cost of its security and privacy concerns. Current methods focus on identifying authorized users who can access PCBMD but they barely identify the insider attackers. Alternatively, these methods do not prevent information leak by authorized users. Working towards this direction, this paper proposes a Trust based Access Control (TAC) system which not only identifies authorized users for PCBMD but also defends Sensitive Personal Information (SPI) of a patient from insider attacks. The proposed method calculates the trust value of each user by considering various quantitative parameters. Based on the calculated trust values, access rights are granted to each user such that SPI can be accessed by only highly trustworthy users. To implement access rights securely, a privacy scheme is also proposed. The experimental results show that the proposed security system can be efficiently used to protect the SPI of patients.

1. Introduction

A colossal data set of digitized and accurate medical data of patients, persistently generated by the technological enhancements and smart innovations in medical field, is referred as Patient Centric Big Medical Data (PCBMD). Many useful predictions and information can be drawn from PCBMD with the application of effective mining techniques. For example, a doctor can get information about the medical condition of a patient and recommend better medications by comparing it with the other similar types of patients around the world. Similarly, an insurance company can devise more effective policies based on the available data. Although mining PCBMD leads to useful results, however, it paves its way to various security and privacy loopholes. For example, disclosure of sensitive patient's health data can result in its illegal use, thereby, putting millions of patient records at risk. Therefore, security and privacy of PCBMD is one of the major challenges in medical data mining.

The existing methods (Liu et al., 2011; Zhou et al., 2013) identify the authorized users who are allowed to have full access of data in PCBMD. Nevertheless, such systems fail to identify the insider attackers, leading to a problem of information leak by authorized users. Working towards this direction, the proposed system initially divides the information in PCBMD into two granular levels, namely, Sensitive Personal Information (SPI) and Non-Sensitive Personal Information (NSPI). SPI corresponds to that data which can potentially identify a particular individual. Alternatively, it is the information which, when disclosed, can result in privacy breach to the user. SPI includes: person's name, address, all elements of dates directly related to patient (e.g. D.O.B, admission date, discharge date, date of death), medical record number, medical conditions, health plan number, biometric identifiers (e.g. finger and voice

* Corresponding author.
E-mail address: navonline98@gmail.com (N. Kaur).

http://dx.doi.org/10.1016/j.tele.2017.09.008
Received 29 September 2016; Received in revised form 13 August 2017; Accepted 26 September 2017
0736-5853/ © 2017 Published by Elsevier Ltd.

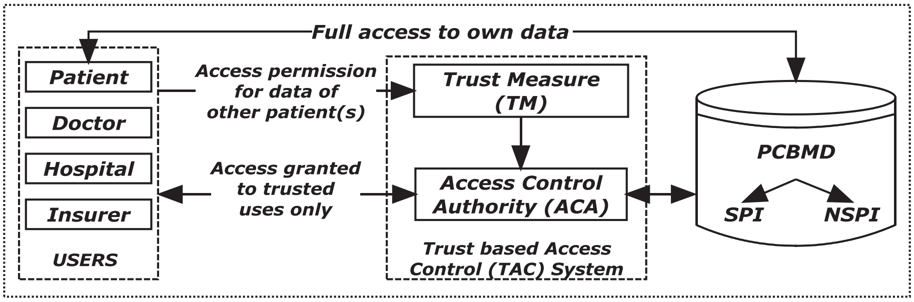

Full access to own data
Patient
Doctor
Hospital
Insurer
USERS
Access permission for data of other patient(s)
Access granted to trusted uses only
Trust Measure (TM)
Access Control Authority (ACA)
Trust based Access Control (TAC) System
PCBMD
SPI
NSPI

Fig. 1. Overview of proposed system.

prints), unique identifiers such as passport or SSN, personally identifiable financial information, photographs etc. On the other hand, information other than SPI is considered as NSPI. Few examples of NSPI are symptoms of particular disease, prescribed medications, medical test results, prevention measures, location of infected area with virus (like Ebola virus in South Africa), virus alerts, etc. Correspondingly, users are also divided into granular levels based upon their trustworthiness. Highly trustworthy users are allowed to have access of SPI while the users with lower trust values are allowed to have access of only NSPI.

The overview of proposed system is shown in Fig. 1. Here, the system identifies various users such as patients, doctors, hospitals, and insurance companies. A patient has full access of his/her own data. On the other hand, the access to other users' data is controlled by the proposed TAC system. TAC, initially, calculates the trust values of users using Trust Measure (TM) module (as described in Section 3.1). Thereafter, the measured trust value is passed to Access Control Authority (ACA) which provides access of PCBMD to only trustworthy users (Trojer et al., 2014) (as described in Section 3.2). The basic idea of the proposed system is to decrease the trust values of inside attackers and thereby, secure SPI of patients that can be leaked by authorized but malicious users. Thus, the proposed system imposes access control restrictions that assure confidentiality of PCBMD.

The novelty of the work presented here lies in the fact that it not only identifies the authorized users, but also controls access to SPI using trust values. In other words, the proposed system provides access of SPI only to highly trustworthy authorized users. For providing secure access to PCBMD based on user's accurate trust value, a privacy based scheme is proposed which is an essential step of ACA and is discussed later in Section 3.2.3. The proposed scheme conveys the sensitive information only to an intended recipient. It also provides the capability of shielding against information disclosure by authorized users. Moreover, as described in Section 3, TM prevents users to provide fake feedback about their rivals, resulting in improved feedback reliability. Thus, TAC is a secure system with an efficient trust evaluation mechanism described as TM module and a protected ACA communication for providing security and privacy to PCBMD.

The rest of the paper is structured as follows: Section 2 presents a survey on different trust and access control systems; Section 3 proposes a reliable trust based access control system for security of patient's PCBMD; Section 4 presents an experimental evaluation of proposed system. It also gives comparative analysis of proposed system with other existing systems; and finally Section 5 concludes the paper.

2. Related work

Research in the field of secure healthcare data management is gaining attention day-by-day. Patient–centric system primarily considers the needs of patient; however, they invite various security breaches. The current studies in patient-centric healthcare information system point towards various challenges such as security, privacy, reliability, availability, integrity, confidentiality, access control, etc. (Shahzad et al., 2016) of patient data. The related work is presented in the following two significant subsections: related work on trust and on access control.

2.1. *Trust*

Jia et al. (2012) stated that trust mechanism is a crucial factor in the terms of information security. Authors presented a trust based model using peer communications. Thereafter, based upon the proposed model, they introduced a power-law division of neighbors. Saleem et al. (2015) proposed an encryption algorithm that guarantee privacy of sensitive information while transferring information from source to destination. The authors posited that the proposed algorithm enhanced the ability to take accurate decisions. Boukerche and Ren (2009) provided a trust assessment method with reliable multicast approach to evaluate the performance of respective nodes. The proposed method allowed only trustworthy nodes to participate in the data transmission by restricting the participation of malicious nodes. The authors elicited that the proposed method enhanced the overall security of system. van Deursen et al. (2008) proposed trust based method in healthcare field known as *"Hedaquin"*. The proposed method specified the quality of healthcare data in a personal health record based on the trust of user as well as information provided by different equipments. Liu et al. (2011) presented a trust based system to identify malicious nodes in advance. The intersection time for separating malicious nodes is diminished by merging intimacy values with biased estimation. The authors in Refs. Wang et al. (2015) and Abdel et al. (2015) worked on an approach for calculating trust for web services. In 2016, Manaman et al. (2016) proposed

N–gram learning approach for calculating online trust of organizations. The authors presented the fact that most companies use social media such as Facebook, and Twitter to increase their trust. Thus, the authors used data from social media (such as tweets) for measuring company's trust. Wu et al. (2013) discussed a method for measuring trust of a service. Kraounakis et al. (2015) worked on a trust formation system in ubiquitous systems. The work of Liu et al. (2015) provided online ranking system for object quality and user trust based on their activity level in the system via iterative algorithm. Zhu et al. (2015) discussed authenticated trust calculation and management system for an integrated cloud-sensor network.

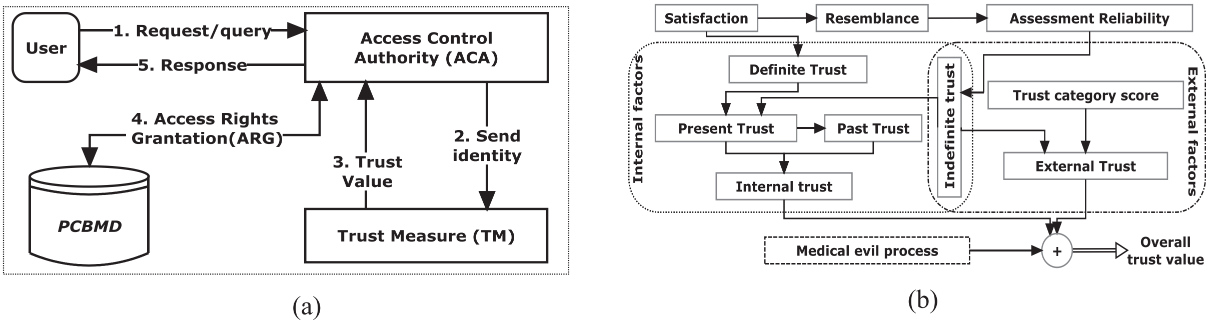
2.2. Access control (AC)

Hauer (2015) worked in the area of information security and discussed that AC to sensitive data plays a significant role in data security. This fact is further supported by Hu et al. (2014) by eliciting that AC is a critical security factor. The authors proposed a distributed AC model for managing sensitive information and assets in complex big data systems. Saleem et al. (2015) designed a machine-to-machine low cost and secure method for e-Healthcare society. The proposed method is aimed to improve the quality of patient's care by reducing the stress level of various users. Moreover, the authors posited the requirement of sharing patient's health information with outsiders. For securely sharing patient's health information, the proposed method used random key management method and modified Kerberos algorithm. Zhou et al. (2013) designed a Role-Based Encryption (RBE) AC system. On encoded information, RBE authorizes access strategies and ensures information security in a cloud reposition system. RBE utilized cryptographic methods where the proprietor of the information is permitted to encipher it so that only a legitimate user can decrypt it. Hu et al. (2010) designed public key encryption algorithm for gathering trustworthy information from implantable medical appliances via sensor supported transmissions. Dong et al. (2015) proposed a system for trustworthy sharing and protection of sensitive information on big data stage by using re-encryption algorithm. Chen et al. (2014) proposed granular AC model having multi-labels in healthcare that contributes to adaptable big data security preservation. Li et al. (2013) proposed patient-driven system which provides secure AC by encrypting each patient's personal health record information with the help of attribute-based encryption techniques. Yu et al. (2010) provided AC mechanism for reliable and extensible data in cloud computing. Hong et al. (2015) used resource-set tree based key production method for providing AC which reduces authentication keys in cloud system. Yang et al. (2014) provided an efficient AC scheme with different dynamic strategy upgrading methods for big data in cloud. Khan et al. (2014) developed a protected mobile cloud based medicinal services system using wireless body area networks. Li et al. (2015) proposed fine grained AC method with adequate attribute revocation and strategy upgrading capabilities in smart grid.

All the preceding works explained distinct approaches to determine a person to whom access of medical data can be given. However, to the best of our knowledge, these approaches lack in three aspects. Firstly, none of the approaches prevent untrustworthy authorized users from accessing sensitive information of patients. Secondly, none of the authors discussed the method of defense against insider attackers who can leak the information as authorized users. Thirdly, all the above mentioned approaches gave access of whole data set to authorized users rather than requisite data. In order to address such lacunae, the proposed method divides patient's data into sensitive and non-sensitive information so that access to SPI can be controlled at a finer granular level. Thus, the proposed system provides the approach for innovative creation of secure and more proficient system.

3. Proposed system

Fig. 2(a) depicts the proposed system of TAC in healthcare. Here, access to PCBMD has been divided into a three main components, namely User, Trust Measure (TM) and Access Control Authority (ACA). User is the person who wants to have access of patient-centric data from PCBMD. As stated earlier, user can be a patient, doctor, hospital, and/or an insurance company. In the proposed system, initially, user sends a request/query to access PCBMD to ACA. On receiving the user's request/query, ACA sends requester's identity to the TM component. TM calculates the trust value of requestor by considering some quantitative factors and sends the calculated value back to ACA. ACA then compares the request/query with the computed trust value of user. Later, it filters the queries according to the trust value of users and then provides access permissions accordingly. The complete description and structure of TM and ACA is given in the following subsections.



(a)

(b)

Fig. 2. (a) Proposed system and (b) Calculating trust value of each user.

ARTICLE IN PRESS

3.1. Trust Measure (TM)

Trust, in the context of security in healthcare, is defined as a global evaluation of character or trustworthiness for a particular user. It is generally measured by those who have experienced direct or indirect interaction with the users whose trust is being measured and is based on feedback, past behavior, observation, etc. In the proposed method, TM component provides an effective path to measure the trust value of users.

TM component calculates the trust of each user by using three quantitative factors namely, medical evil process, internal factors and external factors. Medical evil process is the record of misconduct or mishandling cases of particular user. For example, a doctor involved in illegal medical practices/carelessness/negligence has a record of misconduct cases. The sensitive patient-centric personal information must not be disclosed to such users. Hence, medical evil process forms a significant factor to calculate trust. On the other hand, internal factors calculate trust of a user from the feedback provided by those who have direct interaction with the user. When there is no such direct interaction, then feedback from trusted third parties (such as friends, relatives, or others having direct interaction with the user) is considered. In such a case, external factors are also considered.

Fig. 2(b) depicts the process of calculating trust of each user by using various quantitative factors. All these factors are explained in detail in the subsections ahead. The arrows in Fig. 2(b), shows the dependence of one factor on another. For example, satisfaction is used in the calculation of both resemblance and definite trust. Similarly, present trust is determined using definite and indefinite trust. Furthermore, as stated earlier, there are three main factors namely, medical evil process, internal factors and external factors. These factors are shown by dotted rectangles in Fig. 2(b). Internal factors include definite trust, indefinite trust, present trust, and past trust. Using these, an internal trust is calculated. Similarly, external trust is calculated using indefinite trust and trust category score as external factors. It can be noted that indefinite trust is included in both internal and external factors. Later, an overall trust value is calculated using medical evil process, internal trust and external trust. In the following subsections, these factors as well as overall trust value are described in detail along with their mathematical foundations.

Apart from medical evil process, internal and external factors, there are various other qualitative factors such as payment options, physical appearance of the user, etc. These qualitative factors are not considered here because nothing can be interpreted from these factors about the trust of user in the current scenario.

3.1.1. Medical evil process evaluation

As stated earlier, medical evil process defines misconduct cases by users. It plays an important role in calculating trust of a user who request access permission for particular information related to patients. In order to calculate medical evil rate, a process called trust deduction is used. In this process, initially, each user is assigned "K" points of trust. As the medical evil cases against the user increases, points of trust decrease. Weighing factor (π) is used to determine the loss in trust points of each user.

Let $MR_{user}(i)$ be the i^{th} medical misconduct record (Alhaqbani and Fidge, 2010) of a '*user*', then trust of '*user*', denoted by RP_{user}, is given by Eq. (1).

$$RP_{user} = 1 - \frac{\sum_{i=1}^{n} \pi(i) MR_{user}(i)}{K}; \tag{1}$$

Using trust of user, the medical evil process evaluation rate (EP_{user}) is calculated using Eq. (2).

$$EP_{user} = \begin{cases} RP_{user}; & where\ RP_{user} \geqslant 0 \\ 0; & Otherwise \end{cases} \tag{2}$$

3.1.2. Satisfaction, resemblance and assessment reliability

As healthcare industry shifts towards patient-centric models, satisfaction, resemblance and assessment reliability are essential and commonly used factors for measuring the trust of different users. As shown in Fig. 2(b), these factors are required to calculate internal and external factors. Therefore, this sub-section provides a brief explanation of satisfaction, resemblance and assessment reliability. Later, internal and external factors are discussed. Furthermore, in the sub-sections that follow, it adopts feedback based system and two notations 'x' and 'y' are taken into consideration. Here, 'x' is the evaluator who has direct or indirect interaction with the user. For simplicity, 'y' denotes the user who is requesting access to data.

3.1.2.1. Satisfaction. As the name implies, satisfaction measures how much satisfied evaluator x is from the services provided by user y. It is a feedback based factor where evaluator rates the user based upon its personal experience with the user. Intuitively, more is the satisfaction of evaluator x on user 'y', higher is the trust. Therefore, it is an important factor for trust measurement.

Let $Sf^{t-1}(x,y)$ denote the amount of satisfaction 'x' has on the facilities provided by 'y' at (t − 1)th time interval (Das and Islam, 2012) and Sf_{prst} be the most recent satisfaction of 'x' on 'y'. The predicted value of satisfaction at time $t+1$, denoted by $Sf^{t+1}(x,y)$ is given by Eq. (3).

$$\mathrm{Sf}^{t+1}(x,y) = \begin{cases} 0, & if\ \ t = 0 \\ \beta \times \mathrm{Sf}_{prst} + (1-\beta) \times \mathrm{Sf}^{t-1}(x,y); \quad 0 \leqslant \beta \leqslant 1, & otherwise \end{cases} \tag{3}$$

Here, Sf_{prst} is given by Eq. (4) and $Sf^{t-1}(x,y)$ is given by exponential average of previous values of satisfaction as given by Eq. (3). On the other hand, β is the relative weight which has great impact on the calculation of $Sf^{t+1}(x,y)$ as given by Eq. (5). Therefore, its value is selected carefully such that $\mathrm{Sf}_{\mathrm{prst}}$ is given higher weight than $Sf^{t-1}(x,y)$ in order to give higher weightage to the most recent information.

$$Sf_{prst} = \begin{cases} 0, & \text{if } x \text{ is completely unsatisfied from } y \\ 1, & \text{if } x \text{ is completely satisfied from } y \\ \in (0,1), & \text{otherwise} \end{cases} \tag{4}$$

$$\text{if } \beta = \begin{cases} 0, & \text{then } Sf^{t+1}(x,y) = Sf^{t-1}(x,y) \\ 1, & \text{then } Sf^{t+1}(x,y) = Sf_{prst} \end{cases} \tag{5}$$

3.1.2.2. *Resemblance*.

This factor measures the degree up-to which feedback given by two different evaluators x and x_1 about the same user ‘y’ is similar (Das and Islam, 2012). More is the resemblance between the ratings of two evaluators, more accurate is the rating. Higher accuracy of rating leads to higher precision of trust calculation. Therefore, resemblance plays an important role in determining the accuracy of trust calculation.

The degree of resemblance can be computed by initially determining the variation in feedback given by a set of evaluators who ever had interacted with the same user ‘y’. Let $SA = \{x_i\}$ be the set of evaluators such that $x_i \neq x$ and x_i had interaction with ‘y’. Then variation $(Var^{t+1}(x,x_i))$ of feedback of x_i from that of x at time $t + 1$ is given by Eq. (6).

$$Var^{t+1}(x,x_i) = \begin{cases} \sqrt{\sum_{x_i \in SA} \frac{[Sf^{t+1}(x,y) - Sf^{t+1}(x_i,y)]^2}{|SA|}}, & \text{if } |SA| > 0 \\ 0, & \text{if } |SA| = 0 \end{cases} \tag{6}$$

Using $(Var^{t+1}(x,x_i))$, the resemblance $(Res^{t+1}(x,x_i))$ of feedback given by two evaluators at time $t + 1$ is given by Eq. (7). Here, ∇ is fluctuation constant which gives the upper limit of allowed variation. The value of variation less than or equal to fluctuation constant (∇) implies that the feedback of two evaluators is more alike. In such a case, ‘x’ is rewarded. On the other hand, ‘x’ is punished if variation is greater than ∇. In Eq. (7), R and P denote reward and punishment component respectively to update the resemblance. Depending on the system, both of these components can be changed dynamically. Since, creating trust is tougher than losing it, therefore greater value is allocated to punishment component as compared to reward i.e. $0 < \mathrm{R} < \mathrm{P} < 1$. Furthermore, the punishment component prevents the evaluators from providing a fake feedback of their rivals. On the other hand, reward component encourages evaluators to provide an accurate rating of ‘y’. Hence, reward and punishment components result in improved feedback reliability.

$$Res^{t+1}(x,x_i) = \begin{cases} Res^{t}(x,x_i) + R*(1-Res^{t}(x,x_i)), & \text{if } Var^{t+1}(x,x_i) \leqslant \nabla \\ Res^{t}(x,x_i) - P*Res^{t}(x,x_i), & \text{if } Var^{t+1}(x,x_i) > \nabla \end{cases} \tag{7}$$

From Eq. (7), it can be observed that as variation increases beyond ∇, resemblance decreases.

3.1.2.3. *Assessment reliability*.

It is used to measure accuracy or reliability of the feedback provided by ‘x’ for user ‘y’ (Das and Islam, 2012). During trust calculation, feedback provided by users with higher reliability is more trustworthy than the one with lower reliability. It is a direct logarithmic function of resemblance. More the resemblance between two feedbacks, the higher is assessment reliability and vice versa. Assessment reliability at (t + 1)th time interval, denoted by $AR^{t+1}(x,y)$, is by Eq. (8).

$$AR^{t+1}(x,y) = \begin{cases} 1 - \frac{\ln(Res^{t+1}(x,x_i))}{-4.606}, & \text{if } Res^{t+1}(x,x_i) > 0.01 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

3.1.3. *Internal factors*

As stated earlier, internal factors include definite trust, indefinite trust, present trust, past trust and overall internal trust. These are explained here along with their mathematical foundations.

3.1.3.1. *Definite trust*.

Definite trust $(Def^{t+1}(x,y))$ measures the value of trust obtained by evaluator ‘x’ on user ‘y’ from its own personal experience (Das and Islam, 2012). Better are the services provided by user ‘y’ to evaluator ‘x’, higher is the value of $Def^{t+1}(x,y)$. Definite trust is given by Eq. (9).

$$Def^{t+1}(x,y) = (Sf^{t+1}(x,y)) \tag{9}$$

3.1.3.2. *Indefinite trust*.

Indefinite trust $(Indef^{t+1}(x,y))$ measures the value of indirect trust by ‘x’ on ‘y’ (Das and Islam, 2012). Here ‘x’ has no personal experience or little direct experience with ‘y’ and hence the value of trust is based on the experience of some other recommending agent ‘r’. Eq. (10) represents indefinite trust. Here, RA denotes the set of recommending agents who have ever communicated with the target agent ‘y’.

$$Indef^{t+1}(x,y) = \begin{cases} \frac{\sum_{r \in RA-\{x\}} AR^{t+1}(x,r) \times Def^{t+1}(r,y)}{\sum_{r \in RA-\{x\}} AR^{t+1}(x,r)}, & if\ |RA{-}\{x\}| > 0 \\ 0, & if\ |RA{-}\{x\}| = 0 \end{cases} \tag{10}$$

3.1.3.3. Present trust. Present trust is defined as present attitude of ‘y’ with ‘x’ and with the recommending agent ‘r’ (Das and Islam, 2012). Therefore, it is combination of definite trust and indefinite trust. Present trust ($Pre^{t+1}(x,y)$) is given by Eq. (11).

$$Pre^{t+1}(x,y) = \mu_p \times Def^{t+1}(x,y) + (1{-}\mu_p) \times Indef^{t+1}(x,y) \tag{11}$$

$$\text{where } \mathrm{\mu_p} = \frac{D^{t+1}(x,y)}{D^{t+1}(x,y) + I^{t+1}(x,y)} \tag{12}$$

In Eq. (12), $D^{t+1}(x,y)$ is number of direct interactions that ‘x’ has with ‘y’ and ‘$I^{t+1}(x,y)$’ is number of mean indirect interactions that ‘r’ has with ‘y’ at time t + 1. $I^{t+1}(x,y)$ is given by Eq. (13).

$$I^{t+1}(x,y) = \begin{cases} \frac{\sum_{r \in RA-\{x\}} AR^{t+1}(x,r) \times D^{t+1}(r,y)}{|R-\{x\}|}, & if\ |R{-}\{x\}| > 0 \\ 0, & otherwise \end{cases} \tag{13}$$

3.1.3.4. Past trust. Past trust is defined as a trust which can be computed from past happening. With the passage of time, present trust changes to past trust. Past trust ($Pst^{t+1}(x,y)$) (Das and Islam, 2012) is given below in Eq. (14).

$$Pst^{t+1}(x,y) = \frac{\partial \times Pst^{t}(x,y) + Pre^{t+1}(x,y)}{2} \tag{14}$$

where $\partial (0 \leqslant \partial \leqslant 1)$ is a neglecting factor. Its purpose is to forget past malicious actions of the user and replace it with the most recent behavior for large number of activities. Such a mechanism allows the user to be considered good if his/her behavior has actually changed from past.

3.1.3.5. Internal trust. Internal trust, T^{Int}_{user}, depends on both present trust and past trust (Das and Islam, 2012). It determines expected behavior of the user in future and is given by Eq. (15). Here, γ is the relative weight whose value is selected carefully such that $\mathrm{Pre}^{t+1}(x,y)$ is given higher weight than $\mathrm{Pst}^{t+1}(x,y)$ in order to give higher weightage to the most recent information.

$$T^{Int}_{user} = \begin{cases} \gamma \mathrm{Pre}^{t+1}(\mathrm{x,y}) + (1{-}\gamma)\mathrm{Pst}^{t+1}(\mathrm{x,y}), & if\ \ either\ \ \mathrm{Pre}^{t+1}(\mathrm{x,y})\ \text{and/or}\ \mathrm{Pst}^{t+1}(\mathrm{x,y}) \\ 0, & if\ \ neither\ \ \mathrm{Pre}^{t+1}(\mathrm{x,y})\ \text{nor}\ \mathrm{Pst}^{t+1}(\mathrm{x,y}) \end{cases} \tag{15}$$

3.1.4. External factors

Indefinite trust and trust category score are two factors which contribute to external trust calculation. Indefinite trust is included in both internal as well as external factors and is already explained in Section 3.1.3. Here, trust category score and external trust are explained in detail.

3.1.4.1. Trust category score. Trust category of users is determined by the patients by rating them as bad, fair, average, good and best. Let R^{t+1}_{user} be a number such that it is equal to 0, 1, 2, 3, or 4 corresponding to the rating bad, fair, average, good and best respectively. Using R^{t+1}_{user}, trust category score of user (CS^{t+1}_{user}) can be computed by using Eq. (16).

$$CS^{t+1}_{user} = \frac{R^{t+1}_{user} + \frac{2}{y}}{2 + \sum_{l=1}^{y} R^{t+1}_{l}} \tag{16}$$

Eq. (16) does not provide single valued result. By applying Eq. (17), normalized result of trust category score of user (CS^{t+1}_{user}) can be computed.

$$CS^{t+1}_{user} = \frac{1}{2}\left(\sum_{l=1}^{y} v(l) CS^{t+1}_{user}\right) + \frac{1}{2};\ where\ \ v(l) = \frac{2l{-}2}{y{-}1}{-}1 \tag{17}$$

3.1.4.2. External trust. The external trust, T^{Ext}_{user}, is calculated using indefinite trust and trust category score and is given by Eq. (18).

$$T^{Ext}_{user} = Indef^{t+1}(x,y) \times CS^{t+1}_{user} \tag{18}$$

3.1.5. Overall trust value

Based on the above calculated factors, the overall trust value (van Deursen et al., 2008) of user can be calculated by Eq. (19).

$$Trust\ \ value = \mu_r(\gamma_1 T^{Int}_{user} + \gamma_2 T^{Ext}_{user}) + \mu_e EP_{user} \tag{19}$$

$$\text{where } \mu_r + \mu_e = 1; \text{ and } \gamma_1 + \gamma_2 = 1$$

It can be observed that the calculated value always lies in the range of [0–1]. Here, if computed trust value = 0 then user is "totally untrustworthy". On the other hand if trust value = 1, then the user is said to be "fully trusted". In other words, higher is the trust value, more trustworthy is the user. Based upon this trustworthiness, access permissions are granted by Access Control Authority (ACA) as explained in detail in the following subsection.

3.2. Access control Authority (ACA)

As stated earlier, the goal of ACA is to provide access to sensitive information to only trustworthy users. To achieve this goal, it works in three steps. The first step identifies the trustworthiness level of the user who submitted the query. In the second step, it identifies the access level of the incoming query. The last step compares the trust level of user with access level of query. If the user is trustworthy enough, access is granted to him/her, else it is denied. For providing secure access, Personal Information Privacy Preservation Scheme (PIPPS) is proposed. All these steps are explained in detail in the sub-sections that follow.

3.2.1. Identifying trust level of the user

For identifying the trust level of a user, ACA fetches trust value of the user from TM as shown in Fig. 2(a). If the trust value lies in the range [0–0.4], then the user is said to be at Trust Level 1 (TL1). On the other hand, the value that lies in the range (0.4–0.7] is said to be at Trust Level 2 (TL2) while the value that that lies in the range (0.7–1.0] is said to be at Trust Level 3 (TL3). Therefore, three levels, TL1, TL2, and TL3 are defined such that TL1 < TL2 < TL3 and each user belongs to one of the three levels.

3.2.2. Identifying access level of query

The system identifies two types of queries namely read, and write/update queries. Read queries allows the user to read SPI or NSPI or both. On the other hand, user can write new information for a particular patient or can update invalid existing information to valid information using write/update query. If the incoming query is a read query which reads only NSPI, it is said to be at Access Level 1 (AL1). On the other hand, if read query involves SPI, then it is said to be at Access Level 2 (AL2). Furthermore, every write query (which writes NSPI or SPI) is said to be at Access Level 3. Therefore, AL1 is the lowest access level while AL3 is the highest access level. Alternatively, the access rights of queries at different access levels for SPI and NSPI are summarized in Table 1 and Table 2 respectively. In order to implement these access rights securely, a scheme known as Personal Information Privacy Preservation Scheme (PIPPS) is designed. Complete description of PIPPS is given in Section 3.2.3.

3.2.3. Personal information privacy preservation scheme (PIPPS)

In PIPPS, access rights are implemented securely between two parties using the concept of a key allocation method. This method allocates a key to end users who want to transmit data while restricting others to know the key. For allocating keys to different users, a center known as Key Allocation Center (KAC) is used. It allocates two types of keys. One is session key which encrypts communication between end users temporarily. The other one is known as master key which is unique and helps to transmit session keys in encrypted form.

In PIPPS, a user at trust level '*i*' initially constructs its private specific key pool key_i such that it consists of an initial key, P_i, and $i-1$ master keys as given by Eq. (20).

$$\boldsymbol{key_i} = \{\boldsymbol{P_i}, \boldsymbol{key_i^{i-1}}, \boldsymbol{key_i^{i-2}}, \ldots, \boldsymbol{key_i^{1}}\} \tag{20}$$

The master keys are obtained from initial key $\boldsymbol{P_i}$ by using Eq. (21).

$$\boldsymbol{key_i^j} = \boldsymbol{H}(\boldsymbol{P_i} \| \boldsymbol{j}) \ \boldsymbol{where} \ (1 \leqslant \mathbf{j} < \boldsymbol{i}) \tag{21}$$

Here, $\boldsymbol{H}()$ is a secure hash function such as SHA-256. The initial keys are used for granting access to user for the queries at same level while the master keys are used for queries at lower level. For example, let there are three users at trust levels TL1, TL2 and TL3 respectively. The key pools for users are constructed using Eq. (20) such that

$$key_1 = \{P_1\} \quad key_2 = \{P_2, key_2^1\} \quad \text{and} \quad key_3 = \{P_3, key_3^1, key_3^2\}$$

Here, the user at trust level TL1 has only one key i.e. initial key P_1. This key is used for securely accessing the information at access level AL1. Since there is no other key, therefore, user at trust level TL1 cannot access information at access levels AL2 and AL3. Similarly, user at trust level TL2 has two keys. Here again, the initial key, P_2, is used to access information at same level i.e. AL2. On

Table 1
Access rights for SPI.

Query	Access Levels		
	Low (AL1)	Medium (AL2)	High (AL3)
Read	No	Yes	Yes
Write/Update	No	No	Yes

Table 2
Access rights for NSPI.

Query	Access Levels		
	Low (AL1)	Medium (AL2)	High (AL3)
Read	Yes	Yes	Yes
Write/Update	No	No	Yes

the other hand, master key, key_2^1, is used to access information at lower access level AL1. In other words, the user at trust level TL2 can access information at access levels AL1 and AL2. On the similar grounds, the user at trust level TL3 can access information at access levels AL1, AL2, and AL3.

From the above example, it can be observed that PIPPS ensures that the user is allowed to access only that information which matches the trust level of the user. Therefore, data will be delivered to only an appropriate user without uncovering any patient centric personal information, thereby securing PCBMD.

Furthermore, for establishing a secure connection between a user “d” and ACA “e”, pair-wise keys are created. There are three cases for pair-wise key generation as described below.

Case 1: $RL_{user} = AL_{query}$ i.e. trust level of user is equal to access level of query.

When both “d” and “e” are associated to same level “b”, then pair-wise keys are obtained as follows.

a) “d” sends request to KAC for a session key. The message includes the identity of “d” and “e” and nonce. Nonce is a random number and varies with each request of user. It defends against masquerade attack.
b) KAC responds back by sending session key encrypted with master key that KAC also shares with “e”. Master key is sent to “e” in order to set up connection and confirm “d’s” identity. Thus only “d” can successfully read the message. The master key can be computed using a pseudo-random function f such that $key_b^e = f_{P_b}(d)$
c) “d” broadcast request message including its ID and trust level i.e. “b” and also nonce for secure communication and authentication with “e” such that $d \rightarrow^* : d, b, nonce$. Thus the protected exchange begins with “d” and “e”.
d) In response, “e” send nonce to “d” including its ID + MAC such that $e \rightarrow d$: $eMAC(key_b^e$,d|e),nonce.
e) After this, “d” can access the query by sending its ID, level of query, nonce as well as MAC of “e”. If nonce and MAC sent by “d” matches with that of “e” then “e” can verify that the request is valid. And then pair-wise key is generated.
f) Both “d” and “e” enumerates a pair-wise key such that $key_b^{de} = f_{key_b^e}(d) = f_{key_b^e}(e)$

Case 2: $RL_{user} > AL_{query}$ i.e. trust level of user is greater than access level of query.

When “d” and “e” are associated to different levels such that “d” is at level “a” and “e” is at level “c” ($1 \leqslant$ c$< a$) then pair-wise keys are obtained as:

$$\boldsymbol{key_{ac}^{de} = f_{key_{ac}^{e}}(d) \text{ where } key_{ac}^{e} = f_{key_{c}^{a}}(e)} \tag{22}$$

Case 3: $RL_{user} < AL_{query}$ i.e. trust level of user is less than access level of query

As stated earlier, there is no master key for accessing information at higher level, therefore, in this case, simply deny the access request of user.

Hence, PIPPS implements the access rights for SPI and NSPI by isolating the personal information into different levels as per safety specifications. It assures that user is only authorized to have access permissions to patient centric personal information corresponding with its appropriate trust level. This process boosts the capacity of shielding against inner attacks, particularly the one initiated by unknown malicious user.

4. Experimental evaluation

4.1. Experimental setup

In order to evaluate the proposed system experimentally, a medical database (UCI Machine Learning Repository, 2017) is taken. This data set consists of clinical care data of one million patients collected over the duration of ten years (1998–2008) from 130 hospitals in US. It consists of 55 attributes out of which 12 attributes corresponds to SPI and the rest corresponds to NSPI. This database is bootstrapped (Bao et al., 2012) to one thousand billion patients by using correlation co-efficient as variant and [0.5512 0.9447] as confidence interval. This bootstrapped data is stored at i2.xlarge storage optimized instance of Amazon EC2 (EC2 Instance Types – Amazon Web Services, 2017). A java based implementation is used to test the proposed system by sending queries (Cassavia et al., 2016) from one hundred virtual users to the bootstrapped database. A random trust value is assigned to each of the hundred virtual users which act as their actual trust value. These allocated values are used for measuring the accuracy rate of TM component of proposed system.

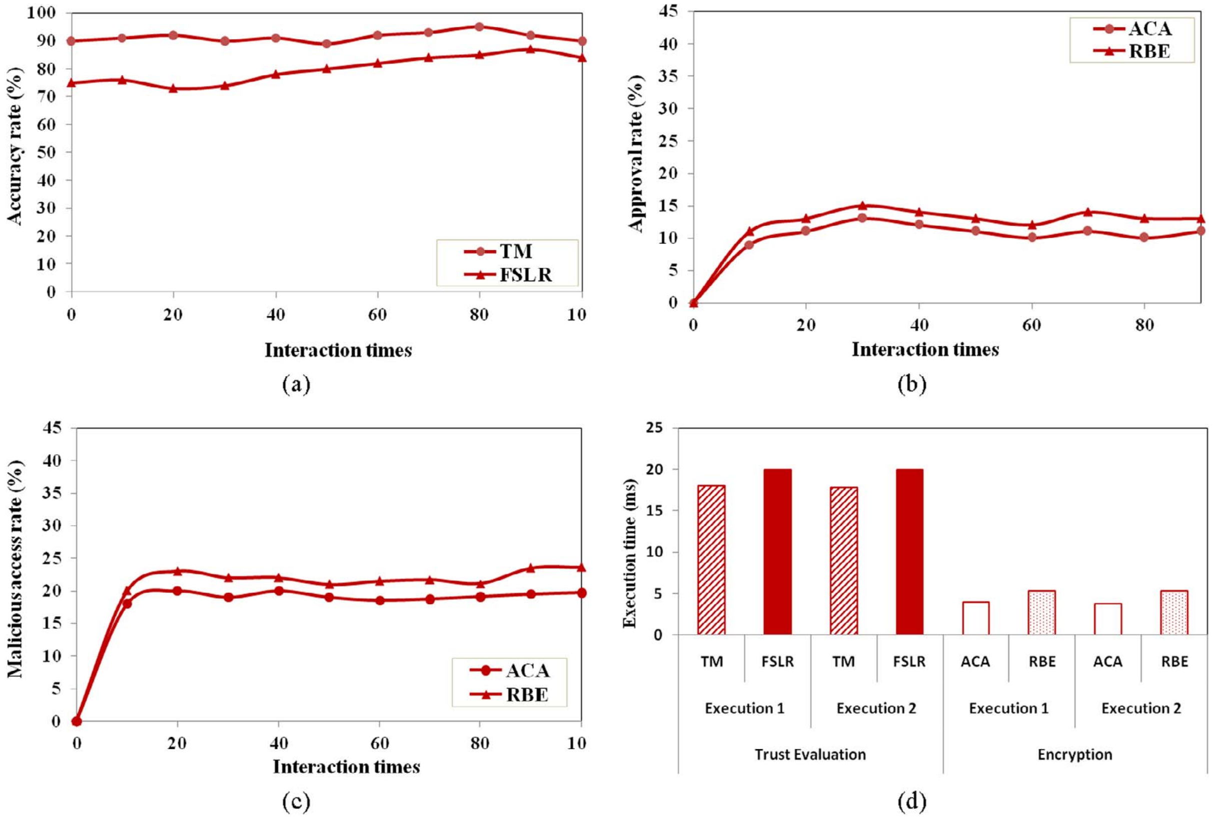


Fig. 3. (a)-(d): Experimental results: (a) Accuracy Rate. (b) Approval rate. (c) Malicious access rate. (d) Execution time for trust evaluation and encryption.

4.2. Experimental results

Since TM and ACA forms two crucial components of the proposed system, therefore, their performance is measured separately. Initially, the experiment is started with 10 interactions. Later, experiment is repeated for ten times such that ten interactions are incremented every time. The performance is measured every time. For measuring performance, the accuracy rate of calculating trust of a user by TM is compared with that of a similar approach called Familiarity and Subjective Logic Based Reputation system (FSLR) (Liu et al., 2011). The results of the comparison are shown in Fig. 3(a). Similarly, the approval rate and malicious access rate of ACA is compared with Role-Based Encryption (RBE) (Zhou et al., 2013) model. The results of the comparison are shown in Fig. 3(b) and (c). Moreover, the execution time of TM and ACA are compared with FSLR and RBE respectively and the results are shown in Fig. 3(d).

4.3. Discussion of results

4.3.1. Accuracy rate of trust

Accuracy rate is defined as the ratio of estimated trust value to the actual trust value. The results shown in Fig. 3(a) show that the accuracy rate is higher in case of TM as compared to FSLR. Furthermore, the accuracy rate remains almost constant for TM while it increases slowly for FSLR. This is due to the fact that TM can efficiently reward users for providing honest recommendations and punish users for incorrect recommendations. On the other hand, FSLR does not consider reward and punishment factors and cannot avoid inaccurate recommendations. Thus, accuracy rate of TM is better than FSLR.

4.3.2. Approval rate

It is defined as the amount of the effectively acknowledged access requirements dissatisfying the safety demand to the overall amount of access requirements. As stated earlier, the approval rate of ACA is compared with that of RBE. It can be observed from Fig. 3(b) that approval rate of ACA is lower as compared to that of RBE. This is due to the fact that RBE does not taken into account the capacity of shielding against information disclosure attack, particularly the one initiated by unknown malicious agent. On the other hand, ACA uses trust level of users providing secure access to PCBMD. In addition, PIPPS has the capacity to prevent untrustworthy users from accessing SPI of the patients. Therefore, less number of users are allowed to access PCBMD, thereby lowering the approval rate of ACA.

4.3.3. Malicious access rate

It is defined as ratio of successfully accessing the unauthorized personal information in overall interactions. Due to inherent property of PIPPS to shield against unauthorized access to personal information of a patient, the malicious access rate is lower for ACA as shown in Fig. 3(c). On the other hand, RBE does not differentiate between SPI and NSPI and grants access to whole of the patient's information to an authorized user. In other words, when a legitimate authorized user deceives, then user accesses all secret keys and personal information. It can likewise get access to the unapproved assets by acquiring different roles and significant secret keys through dispatching the data disclosure attack. This leads to information leaks by authorized users, thereby increasing malicious access rate.

4.3.4. Execution time

For effectively comparing the execution time, the proposed system is executed twice. The execution time for TM and ACA are measured for both the executions. In addition, the execution time for both FSLR and RBE are measured and compared with that of proposed method as shown in Fig. 3(d).

For trust evaluation, execution time of TM is compared with that of FSLR. The results in Fig. 3(d) show that the execution time of TM is slightly less than that of FSLR in execution 1 as well as in execution 2. Moreover, TM takes little less time in execution 2 (17.8 ms) as compared to that of execution 1 (18 ms). This is due to the fact that the results obtained in execution 1 act as input for calculating the results in execution 2. Thus previous result enhances the results of the next execution and slightly lowers the execution time.

Similarly for encryption, execution time of ACA is compared with RBE. The results in Fig. 3(d) show that the execution time of ACA is slightly less than that of RBE in execution 1 as well as in execution 2. Moreover, ACA has lower execution time for execution 2 (3.8 ms) as compared to execution 1 (4 ms). This is due to the fact that the initialization of keys is done in execution 1 only. On the other hand, execution 2 calculates pair-wise keys only, which slightly lowers execution time. On the other hand, execution time for RBE is same in both the cases since it executes whole of the algorithm in execution 2 also.

Clearly, in execution 1, the overall execution time of two operations is as: proposed system (18 ms for TM and 4 ms for ACA which is equal to 22 ms) which is less than that of the existing systems (20 ms for FSLR and 5.3 ms for RBE which is equal to 25.3 ms). Similarly, in execution 2, overall time in proposed system (17.8 + 3.8 = 21.6 ms) is also less than the existing systems (20 + 5.3 = 25.3 ms).

From the above results, it can be concluded that the proposed system is better than the existing systems in terms of security and privacy of patient centric medical data. Moreover, execution time of proposed system is lower than the existing systems.

5. Conclusion

In this paper, a trust based access control system is proposed which provides secure and authorized access rights to Patient Centric Big Medical Data. Patient-centric healthcare information systems confront various challenges such as security, privacy, reliability, availability, integrity, confidentiality, access control of patient data. The proposed system presents an approach to address security and privacy of patient-centric big medical data. It calculates accurate trust value of various users and provides secure access to PCBMD accordingly. It prevents users to provide fake feedback about their rivals resulting in improved feedback reliability. It can efficiently reward users for providing honest recommendations and punish users for incorrect recommendations. For providing secure access to PCBMD based on user's accurate trust value, a secure privacy scheme (PIPPS) is also proposed. PIPPS has the capacity to prevent untrustworthy users from accessing sensitive information of patients. Thus, proposed scheme provides sensitive information only to an intended recipient without uncovering any personal data. The detailed simulation testing and comparative analysis have proved that the proposed system is secure, reliable and efficient as compared to other existing systems.

References

Abdel, O., Bentahar, J., Otrok, H., Mourad, A., 2015. A survey on trust and reputation models for web services: single, composite, and communities. Decis. Support Syst. 74, 121–134.

Alhaqbani, B., Fidge, C.J., A time-varient medical data trustworthiness assessment model. In: 11th International Conference on Ethical Issues Secur. Monit. Trends Glob. Healthc. Technol. Adv., pp. 130–150, 2010.

Bao, X., Bahl, P., Kansal, A., Chu, D., Choudhury, R.R., Wolman, A., 2012. Helping mobile apps bootstrap with fewer users. In: Proc. 2012 ACM Conf. Ubiquitous Comput. – UbiComp '12, p. 491.

Boukerche, A., Ren, Y., 2009. A secure mobile healthcare system using trust-based multicast scheme. IEEE J. Sel. Areas Commun. 27 (4), 387–399.

Hu, V.C., Grance, T., Ferraiolo, D.F., Kuhn, D.R., 2014. An access control scheme for big data processing, COLLABORATECOM, pp. 1–7.

Cassavia, N., Ciampi, M., De Pietro, G., Masciari, E., 2016. A big data approach for querying data in EHR systems. In: Proc. 20th Int. Database Eng. Appl. Symp. – IDEAS '16, pp. 212–217.

Chen, H., Bhargava, B., Zhongchuan, F., 2014. Multilabels-based scalable access control for big data applications. IEEE Cloud Comput. 1 (3), 65–71.

Das, A., Islam, M.M., 2012. SecuredTrust : a dynamic trust computation model for secured communication in multiagent systems. IEEE Trans. DEPENDABLE Secur. Comput. 9 (2), 261–274.

Dong, X., Li, R., He, H., Zhou, W., Xue, Z., Wu, H., 2015. Secure sensitive data sharing on a big data platform. Tsinghua Sci. Technol. 20 (1), 72–80.

EC2 Instance Types – Amazon Web Services (AWS). [Online]. Available: https://aws.amazon.com/ec2/instance-types/. [Accessed: 09-Jan-2017].

Hauer, B., 2015. Data and information leakage prevention within the scope of information security. IEEE Access 3, 2554–2565.

Hong, S., Kim, H., Kim, T., Chang, J., 2015. A user access control scheme for reducing authentication keys in cloud systems. Int. J. Secur. Appl. 9 (4), 217–228.

Hu, F., Hao, Q., Lukowiak, M., Sun, Q., Wilhelm, K., Radziszowski, S., Wu, Y., 2010. Trustworthy data collection from implantable medical devices via high-speed security implementation based on IEEE 1363. IEEE Trans. Inf Technol. Biomed. 14 (6), 1397–1404.

Jia, C., Xie, L., Gan, X., Liu, W., Han, Z., 2012. A trust and reputation model considering overall peer consulting distribution, IEEE Trans. Syst. MAN, Cybern. PART A

Syst. HUMANS, 42(1), pp. 164–177.

Khan, F.A., Ali, A., Abbas, H., Haldar, N.A.H., 2014. A cloud-based healthcare framework for security and patients' data privacy using wireless body area networks. Procedia Comput. Sci. 34, 511–517.

Kraounakis, S., Demetropoulos, I.N., Michalas, A., Obaidat, M.S., Sarigiannidis, P.G., Louta, M.D., Member, S., 2015. A robust reputation-based computational model for trust establishment in pervasive systems. IEEE Syst. J. 9 (3), 878–891.

Li, M., Yu, S., Zheng, Y., Ren, K., Lou, W., 2013. Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. IEEE Trans. Parallel Distrib. Syst. 24 (1), 131–143.

Li, H., Liu, D., Alharbi, K., Zhang, S., Lin, X., 2015. Enabling fine-grained access control with efficient attribute revocation and policy updating in smart grid. KSII Trans. INTERNET Inf. Syst. 9 (4), 1404–1423.

Liu, Y., Li, K., Jin, Y., Zhang, Y., Qu, W., 2011. A novel reputation computation model based on subjective logic for mobile ad hoc networks. Future Gener. Comput. Syst. 27, 547–554.

Liu, X., Guo, Q., Hou, L., Cheng, C., Liu, J., 2015. Ranking online quality and reputation via the user activity. Phys. A 436, 629–636.

Manaman, H., Jamali, S., Aleahmad, A., 2016. Online reputation measurement of companies based on user-generated content in online social networks. Comput. Human Behav. 54, 94–100.

Saleem, K., Derhab, A., Al-muhtadi, J., Shahzad, B., 2015. Human-oriented design of secure machine-to-machine communication system for e-healthcare society. Comput. Human Behav. 51, 977–985.

Saleem, K., Derhab, A., Al-muhtadi, J., Shahzad, B., Orgun, M.A., 2015. Secure transfer of environmental data to enhance human decision accuracy. Comput. Human Behav. 51, 632–639.

Shahzad, B., Orgun, M.A., Thuemmler, C., 2016. Fundamental issues in mobile healthcare information systems, Mob. Inf. Syst., vol. 2016, Article ID 6504641, 2 pages. doi: 10.1155/2016/6504641.

Trojer, T., Katt, B., Özata, T., Breu, R., Mangesius, P., Schabetsberger, T., 2014. Factors of access control management in electronic healthcare: The patients' perspective, 47th Hawaii Int. Conf. Syst. Sci., pp. 2967–2976.

UCI Machine Learning Repository: Diabetes 130-US hospitals for years 1999-2008 Data Set. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008. [Accessed: 07-Jan-2017].

van Deursen, T., Koster, P., Petković, M., 2008. Hedaquin: a reputation-based health data quality indicator. Electron. Notes Theor. Comput. Sci. 197 (2), 159–167.

Wang, S., Zheng, Z., Wu, Z., Lyu, M.R., Yang, F., 2015. Reputation measurement and malicious feedback rating prevention in web service recommendation systems. IEEE Trans. Serv. Comput. 8 (5), 755–767.

Wu, Y., Yan, C., Ding, Z., Liu, G., Wang, P., Jiang, C., 2013. A novel method for calculating service reputation. IEEE Trans. Autom. Sci. Eng. 10 (3), 634–642.

Yang, K., Jia, X., Ren, K., 2014. Secure and verifiable policy update outsourcing for big data access control in the cloud. IEEE Trans. Parallel Distrib. Syst. 9, 1–11.

Yu, S., Wang, C., Ren, K., Lou, W., 2010. Achieving secure, scalable, and fine-grained data access control in cloud computing. IEEE INFOCOM. http://dx.doi.org/10.1109/INFCOM.2010.5462174.

Zhou, L., Varadharajan, V., Hitchens, M., 2013. Achieving secure role-based access control on encrypted data in cloud storage. IEEE Trans. Inf. Forensics Secur. 8 (12), 1947–1960.

Zhu, C., Nicanfar, H., Leung, V.C.M., Yang, L.T., 2015. An authenticated trust and reputation calculation and management system for cloud and sensor networks integration. IEEE Trans. Inf. Forensics Secur. 10 (1), 118–131.