A Knowledge-Constrained Role-Based Access Control model for protecting patient privacy in hospital information systems

Article

Accepted Version

A Knowledge-Constrained Access Control Model for Protecting Patient Privacy in Hospital Information Systems

Runtong Zhang, *Senior Member*, *IEEE*, Donghua Chen, Xiaopu Shang, *Member*, *IEEE,* Xiaomin Zhu, and Kecheng Liu

***Abstract*—Current access control mechanisms of the hospital information system can hardly identify the real access intention of system users. A relaxed access control increases the risk of compromise of patient privacy. To reduce unnecessary access of patient information by hospital staff, this paper proposes a Knowledge-Constrained Role-Based Access Control (KC-RBAC) model in which a variety of medical domain knowledge is considered in access control. Based on the proposed Purpose Tree and knowledge-involved algorithms, the model can dynamically define the boundary of access to the patient information according to the context, which helps protect patient privacy by controlling access. Compared with the Role-Based Access Control model, KC-RBAC can effectively protect patient information according to the results of the experiments.**

***Index Terms*—Privacy, knowledge, access control, information systems, hospital**

I. Introduction

The hospital information systems, including Hospital Information System (HIS) [1], Laboratory Information Management System (LIS) [2], Picture Archiving and Communication System (PACS) [3], Radiology Information System (RIS) [4] and Electronic Medical Record System (EMRS) [5], have been adopted in many hospitals to meet the needs of different users. Data in these systems which are associated with the personal information of the patients are crucial resources for the users. For example, physicians can carry out diagnosis and treatment for patients according to their personal information and medical history; and medical researchers can review and study the cases. As a consequence, to make the information and data accessible to approporiate users is essential for the design of such information systems. However, the information in medical and health data in the hospital information systems is extremely personal and private; therefore access to which should be strictly controlled to protect patient pravicy and avoid information leak. Protecting the information in such systems is an important issue. To protect patient privacy, from the legal and technical levels, many studies [6–7] have been conducted in the area of information security. However, the existing access control models have not given full consideration on domain knowledge such as the hospital processs and clinical pathways. Following a preliminary version of this work [8], this paper studies the privacy-protection mechanism for information systems in the hospital and puts forward a Knowledge-Constrained Role-Based Access Control (KC-RBAC) model, based on the Role-Based Access Control (RBAC) model.

The rest of this paper is organized as follows. Section II reviews related RBAC work and patient privacy. Section III describes our research motivation and the formal model description of KC-RBAC, where Knowledge Module (KM) and Purpose Tree (PT) are introduced. Section IV gives the three algorithms to identify real user purposes, determine user permissions, and protect patient information using medical knowledge. Experiments have been conducted to verify the performance of KC-RBAC model and are discussed in Section V. This paper ends with conclusions and perspectives.

II. Related Work

A widely adopted model, the RBAC model, is proposed to simplify the access management of information systems [9]. In RBAC, system users are distinguished by roles according to the access policies, which enables the system to assign a new system user with the corresponding access rights automatically rather than case by case manually. Hung et al. [10] propose a privacy-based framework to tackle the need in e-Healthcare services to protect health information. The access purpose is also considered in the design of access control mechanism, which allows privacy officers to specify what data should not be used for certain purposes [11]. With the use of big data, the knowledge from the large volume of medical data can help improve data management with security and performance constraints [12]. The personal health record system offers new opportunities for personalized healthcare management because patients normally worry about their personal information being used inappropriately [13]. Facing the rapid growth of users and information, Hsu et al. [14] propose the role-based access control model to deal with authorization in the healthcare

Manuscript received November 29, 2016; revised January 19, 2017; accepted April 13, 2017. This work was supported by a key project of National Natural Science Foundation of China under grant number 71532002, the Fundamental Research Funds for the Central Universities under grant number 2016YJS057, and Beijing Logistics Informatics Research Base.

R. Zhang, D. Chen, and X. Shang are with the Department of Information Management, School of Economics and Management, Beijing Jiaotong University, Beijing, China (e-mail: rtzhang@bjtu.edu.cn; 15113181@bjtu.edu.cn; sxp@bjtu.edu.cn). X. Shang is the corresponding author of this paper.

X. Zhu is with the School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing, China (e-mail: xmzhu@bjtu.edu.cn). X. Zhu is the corresponding author of this paper.

K. Liu is with Henley Business School, University of Reading, Reading, United Kingdom (email: k.liu@reading.ac.uk)

systems. To capture domain knowledge in a formal language, such as ontology modeling and semantic modeling, to enable automatic performance of access control, is necessary [15]. Gritzalis et al. [16] conduct a study on risk assessment to protect patient privacy in a shared care platform for the treatment of patients suffering from beta thalassemia. Røstad et al. [17] believe that the main problem facing the busy clinicans is to avoid being exposed with both irrelevant and relevant information at the same time. Hence, most studies focusing on the privacy protection in hospital information systems are based on the RBAC.

Another research mainstream focuses on the improvement of application systems using semantic approaches and medical domain knowledge [18, 19]. In the personal health record and health knowledge sharing system design, Lee et al. [20] use the Integral Healthcare Enterprise-Cross Enterprise Document Sharing (IHE-XDS) and the W3C Web Ontology Language (OWL) to maintain the Personal Health Records (PHRs) and collate the useful health Web resources related to the personal diseases. Making better use of existing data in the electronic health records to identify eligible subjects can improve efficiency and quality of medical care [21]. Discovering the knowledge from the medical records may support medical personnel in making clinical decisions and also help improve personalized medicine and care [22]. Considering the difficulties of ensuring the security and appropriate use of patient health information contained in medical records, an explanation-based auditing system (EBAS) [23] was proposed to distinguish the clinical or operational reasons and present metrics to determine which hospital employees are responsible for treating a given diagnosis.

It is also widely argued that, the knowledge in medical domains should be seriously considered when using RBAC model in HIS. Some knowledge like the relationship between different diseases is given in some international coding standards, such as ICD-10 (International Classification of Diseases 10th edition) and SNOMED CT (Systematized Nomenclature of Medicine-Clinical Terms), and other biomedical domain ontologies. Beimel et al. [24] present the SitBAC knowledge framework to infer new knowledge based on the incoming data access request from the realization process. For the pervasive healthcare systems, Li et al. [25] propose an authorization model that supports specifying and enforcing authorizations in a flexible and efficient way to conceptualize the data and explicitly express the relationship among concepts and instances. Considering that personal healthcare applications range over many disciplines, Blobel et al. [26] introduce the care paradigms and discuss the requirements to meet the business objectives. In order to acquire better accuracy and coverage, Gordona et al. [27] propose a cost-effective semi-automated method for generating a userful knowledge compendium with minimal reliance on domain experts. To some extend, these studies are to make use of medical domain ontology to improve the performance of RBAC.

A comparison of the main access control models reviewed above is given in Table I from the aspects of role, purpose, knowledge, confidentiality, integrity, and flexibility. For the patient privacy protection in hospital information systems, current access control models still need further improvement in reducing access rights by combining the medical domain knowledge and the existing RBAC model.

TABLE I
COMPARISON OF THE EXISTING ACCESS CONTROL MODELS

Models	RBAC [9]	PE-RBAC [10]	P-RBAC [11]	OBAM [26]
Role	YES	YES	YES	YES
Purpose	NO	NO	YES	NO
Knowledge	NO	NO	NO	YES
Confidentiality	NO	YES	NO	NO
Integrity	YES	NO	NO	NO
Flexibility	NO	NO	NO	YES

III. A KNOWLEDGE-CONSTRAINED ACCESS CONTROL MODEL

A. *Motivation and General Idea*

RBAC and its relevant models can restrict access rights according to the role of the users, but this is based on their specialties. Theoretically, some employees may have rights to access all the medical records of the patients even if their jobs have no relationship with patient information at that time. This brings potential risks of patient privacy disclosure. Thusly, to make the users' access right more accurate, we take the medical domain knowledge and context information into consideration to formulate a novel knowledge-constrained access control model—KC-RBAC.

Figure 1 gives an example of the traditional RBAC model, where A_i (i = 1, 2, 3, …, n) represents the i^{th} independent access right. In this case, ROLE 1 possesses several access rights which are different from those of ROLE 2.

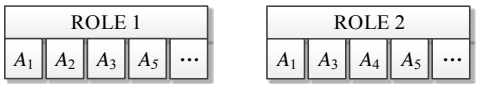

Fig. 1. Difference of access rights between two different roles in the traditional RBAC model

Following the case of Figure 1, taking knowledge and context into consideration, another case is given in Figure 2.

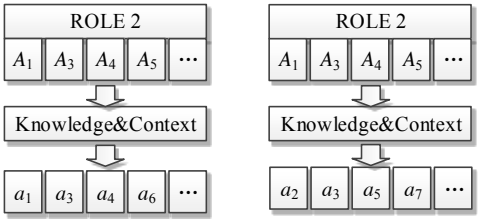

Fig. 2. Difference of access rights for the same role when considering knowledge and context

In this case, the access rights (A_i, i=1,2,…, m) of the same role (ROLE 2) are redefined as a_j (j=1,2,…, n) by introducing the knowledge and context factors. It brings a better control on accessing data and information. Even for the same role, the acquired patient information is different because the knowledge has a dynamic effect on those user permissions. Moreover, this mechanism has no impact on the origin system business logic. The context in Figure 2 is usually associated with the real access purposes of system users. In the hospital process, user intention is recognized as the original purposes that the user intends to do something. It differs from the predefined user purposes of visiting the systems. In this paper, user intentions are usually obscure and hard to be predicted, whereas user purposes are very clear and predefined in the design of system. Figure 3 illustrates the application of KC-RBAC in real life.

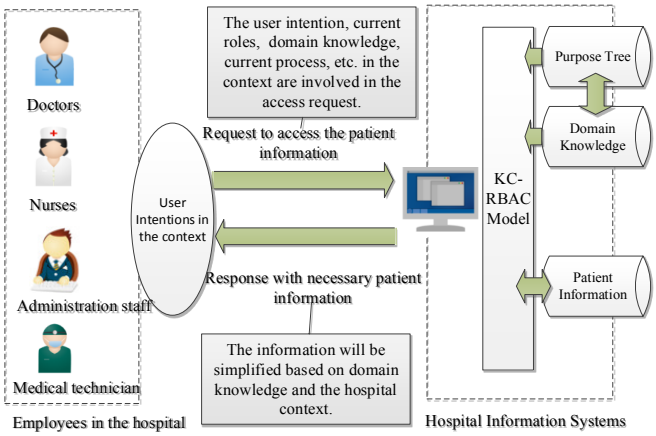

Fig. 3. Illustration of application with KC-RBAC in real life

The rest of the paper explores the models and algorithms to implement our motivation and idea.

B. Model Description

Based on the RBAC model, medical domain knowledge and user purpose are involved in the new model.

1) General Model

KC-RBAC is used to define the mapping relationships between roles and permissions by referring to the medical knowledge and the PT with the knowledge inferences and semantic analysis to obtain more precise control of permission. PT serves as a tree hierarchy that describes the purpose of the use of the hospital information systems. Though the role hierarchy in hospitals is static, permissions for a role, as aforementioned, are dynamic and determined by the PT of the user. Following the basic structure of the traditional RBAC model, our proposed KC-RBAC model is shown in Figure 4.

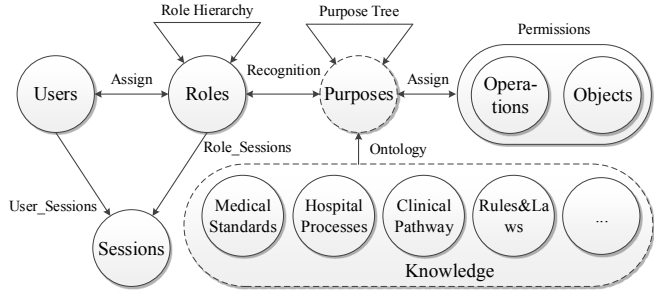

Fig. 4. Knowledge-Constrained Role-Based Access Control Model (KC-RBAC), which follows RABC model

Different from the RBAC model, the knowledge module in Figure 4 is introduced to identify the real access purposes of the users. The access purpose in this paper is a concept that represents the real needs of the system users in a hospital. For example, if a doctor wants to offer medical advices to his or her patients, the system should provide the doctor only the necessary access rights to the relevant patient information. However, the key to identify what the doctor actually needs depends on the KC-RBAC model which analyzes the user intention based on the knowledge. The knoweldge relevant to the scope of the patient information acquired includes the hospital proceses, roles, as well as the domain knowledge in the medical field. Thus, the access control model can restrict the hospital employees' access to other information which is unnecessary to the process of medical diagnosis and treatment. Such an approach will protect patient privacy as far as possible.

2) Knowledge Module

The Knowledge Module is the key component in KC-RBAC to determine the access rights of system users to protect patient privacy. Those basic knowledge, such as ICD-10 and SNOMED CT, provides the ability of knowledge inference in the KC-RBAC.

In RBAC, the role hierarchy is preliminary knowledge in medical information systems. Many different roles with different purposes exist in the hospital because of the different processes in the medical departments. Thus, as a key part of the knowledge-based system, the role hierarchy is very important. The medical institutions normally have many employees, so the rights of the role in the system are hard to distinguish even by using contexts. To give strict protection for patient privacy in information systems in the hospital, the permissions of the role hierarchy in Figure 4 should be divided into two types, namely, public and private permissions. The logic of the knowledge module is shown as Figure 5. Figure 5 shows that the key part of the Knowledge Module is the knowledge reasoning engine. It provides the user with patient information that a user actually needs based on the calculation of medical knowledge. Except for the aforementioned basic knowledge, the main knowledge sources in medical information systems are the knowledge contained in the patients medical records corresponding to the knowledge bases such as the SNOMED CT; the domain query language of which is SNOMED CT Expression Constraint Language (SNOMED ECL) to support the search for knowledge in massive knowledge. The number of medical records in the hospital increases rapidly, and the invisible domain knowledge in those records are valuable. Thus the medical records are useful to be reused in our proposed model.

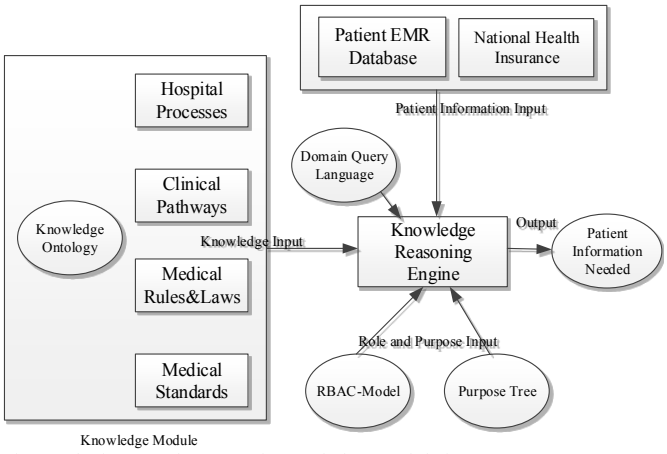

Fig. 5. The input and output of Knowledge Module in KC-RBAC

3) Purpose Tree Module

The Purpose Tree (PT) Moduel is another key compnent that illustrates the user purposes of vising the system in tree hierarchical structure. In the different processes of the hospital, physicians and other hospital staffs may request for different information of the patient in HIS, which we call the motivation of visiting as visiting purpose. In order to describe the visiting purpose, PT is introduced in this model. Table II gives a typical PT of HIS. The Level-1 column suggests several general purposes that may appear in the HIS. Based on this, more specific system access purposes are introduced in Level-2 and Level-3. The relationships of these purposes can be graphitized in a tree hierarchy with three levels.

TABLE II
THE PURPOSE TREE FOR HIS

	Level 1	Level 2	Level 3
HIS PT	Decision Making	Analysis	Development, Judgment
		Planning	Strategy
	Management	Flow	Human, Resource
		Quality	Service, Feedback
		Statistics	Organization
		Monitor	Video, Pictures, Report
	Medical Care	Advise	
		Charge	Outpatient, Inpatient
		Inspector	Physical Examination
		Procedure	Check In/Out, Transfer
		Processing	Clinical, Care, Healthcare
		System	X-rays
		Management	Medical Insurance, Case
	Scientific Research	Medical Research	Data Analysis, Data Mining
		Case Studies	Internship, Training

However, the operations of different hospitals vary from each other, and additional work is thusly necessary in the development of the system.

4) Model Formulation

Based on the description of the KC-RBAC model, we give specific formal definitions for KC-RBAC in the following.

Definition I (Symbols and Module Relationships in KC-RBAC):

- U, R, RH, P, S, PM, I, and K represent users, roles, role hierarchy, purposes, sessions, permissions, information, and knowledge, respectively
- U-R, U-S, S-R, R-P, and P-P represent the relationships of user–role, user–session, session–role, role–purpose, and purpose–permission respectively
- U–$R \subseteq U \times R$, U–$S \subseteq U \times S$, S–$R \subseteq S \times R$, $I \subseteq K$, where $\times$ means a many-to-many relationship and $I_u = \{i | i \in I, \text{Owner}(i) \in U\}$
- $RH \subseteq R \times R$, $\text{PT} \subseteq P \times P$, where the role hierarchy is defined based on the RBAC model and PT is a tree structure that shows the top–down relationship of the purposes
- $K = \{k_d \cup k_r | k_d \in \text{domain ontology}, k_r \in \text{rules}\}$. K will update itself as time passes
- $PM = R - P \subseteq R \times K \times P$. The number of PM for R is restricted by K and P, $R \times K = \{rk | rk \subseteq K, rk = (r, k, p), rk \in \text{rules}\}$, $R \times K \times P = \{p | \{p\} \subseteq \text{PT}_u, p \in R \times K\}$. $\text{PT}_u = \text{PT} \times U$, PT_u means the purposes of the current users.
- I_u (R, K, P, and PM) is the expected output of the model. Outside the system, the input of the model is the set of purpose (PT_u). To sum up, $\text{PT}_u \xrightarrow{(K,R)} I_u$.

The above definition gives the relationship between modules in the model. The module of ROLES, SESSIONS, and PERMISSIONS still follows the rules from the traditional RBAC model. Based on Definition I, the following definition gives clearer description on the structure of the proposed PT module.

Definition II (PT):

- P, PT, UI, and OU (purpose, Purpose Tree, user intention, and outside user, respectively)
- $P_u = f_p(\text{UI}, \text{PT}) = \{p | p \in \text{PT}, u \text{ is } p\text{'s owner}\}$, $\text{UI} = f_u$ (R, OU, and PM). F_p is a function to obtain what purpose they want based on the determination of UI and PT. f_u is a function to help inference for user intentions.
- P_u is the result of the total calculation of PT for user u.

Definition III (Knowledge Ontology):

- AT, DE, RE, and P (attributes, description, relationship, and purpose, respectively)
- K = (AT, DE, RE, and P), $P \in \text{PT}$ is the last part of P, which is the key part to associate knowledge and purpose in this paper. The other parts are traditional relationship of domain ontology.
- $K_u = K_R \cap K_{PM} \cap K_P$, K_R is a set of knowledge constrained by ROLES, K_{PM} is a set of knowledge constrained by permission, and K_p is all knowledge with purpose p.

In Definition II, UI indicates the set of purposes such as hospital employees. OU represents those people who want to protect their privacy as much as possible (such as the patients). Here, we define a function f_p as

$$f_p(UI, PT) = \{p \mid diff(p, iu) > e\} \quad p, ui \in \text{PT,UI}, 0 \le e \le 1 \tag{1}$$

$$diff(x, y) = \begin{cases} e = sim(x, y) & x \cong y \\ 0 & x \neq y \end{cases} \tag{2}$$

In Equation (1), the result of f_p is defined as a set in which the elements are determined by function $diff(p, iu)$ as in Equation (2). The function *diff* aims to calculate the similarity between user purpose and user intention. The function $sim(x, y)$ calculates the similarity between x and y. All the results of the calculation of function *diff* should be treated as a set of purposes that the users will own. The *UI* can be calculated by

$$UI = \{ui \mid ui \in OU, ui \in R \cap PM\} \tag{3}$$

Thusly, according to the Definition II, the purpose set of specific users can be calculated. As aforementioned, the ontology of medical domain knowledge, such as SNOMED CT, can help the system find the most knowledgeable relationships in patient information for better privacy protection in Definition III.

IV. ALGORITHMS AND EXPERIMENTATION

The formal model description above outlines the basic components of KC-RBAC and their relationships. Three methods in the KC-RBAC are proposed: Purpose Identification Algorithm (PRA), Permission Determination Algorithm (PDA), and Knowledge-involved Information Filtration Algorithm (KIFA).

A. Purpose Identification Algorithm (PRA)

PRA is an algorithm to identify the actual purposes of visiting the hospital information systems based on the intentions before using the systems. Algorithm I gives the step-by-step description.

Algorithm I PRA
Input: Purpose Tree *PT*, User Intention *UI*

Output: a subset of *PT* determined by input *UI*

1) Let p_1, p_2, ..., p_n as input purposes from user's intention (*UI*)
2) Let *r* as the list result of <purpose, knowledge> that users have
3) Let p_t as the subset of Purpose Tree that shows the relationship of purposes the users have
4) Let K_p as the knowledge (concepts or relationships) of the specific purpose of *PT*
5) for each p_i in *UI* where i=1,2,..., n do
6) if(*PT*.hasNode(p_i) and $K(p_i) \in K_p$)
7) *r*.push(p_i,$k_p(p_i)$);
8) for each a_i in *r* where i=1,2,3...,n-1 do
9) for each b_j in *r* where j=1,2,3...,n do
10) if((a_i not equals b_j) and ($diff(a_i, b_j) < e$)){
11) if(a_i.isParent(b_j))
12) p_t.push($a_i \rightarrow b_j$);
13) else
14) p_t.push($b_j \rightarrow a_i$);
15) }
16) return p_t;

To evaluate the performance of Algorithm I, it is necessary to compare the difference between a set of predicted purposes and a set of original purposes by assigning weights to the purposes and knowledge. The weight of a purpose is an integer, ranging from 1 to 10. A larger value means a more important purpose. The difference can be calculated by Equition (2). Additionally, the aggregate weight of UI (user intentions) and UP (recognized purposes) can be calculated as

$$W_i = a\sum_{p\in U_i} W_p + b\sum_{p_k\in PT, k\in K} W_k \tag{4}$$

$$W_p = a\sum_{p\in U_i} W_p + b\sum_{p_k\in PT, k\in K} W_k - c\sum_{p,k} W_r \tag{5}$$

where W_i stands for the weight of UI, W_p stands for the weight of the user intention, and W_k means the weight of knowledge associated with the purpose p. The constant a and b are set as 0.2 and 0.8 respectively. In Equation (5), W_p stands for the weight value of UP, and W_k stands for the knowledge associated with purpose p. If there is a strong relationship between W_k and p, W_r, as a coordinator, is able to enhance its effect in the expecting results. The constant c is set as 0.2. Thusly, a weighted PT is built to quantify the model.

B. Permission Determination Algorithm (PDA)

Algorithm I returns a subset of the PT, which represents the actual purposes of a system user. Given that the system modules in the hospital are associated with specific access rights for a role, the following algorithm (Algorithm II) is presented to summarize all the permissions based on their actual purposes determined by Algorithm I.

Algorithm II PDA
Input: List<Purpose> *P*, Role *R*, PurposeTree *PT*, List<Permission> *PM*
Output: List<Permission>

1) Let p_1, p_2, ..., p_n as the Purpose Tree calculated by FindCurrentUserPurpose() in Algorithm I
2) Let U_c as the current user in HIS
3) Let r_1, r_2,..., r_n as elements of Role *R*
4) Let *PT* as Total Purpose Tree
5) Let pm_1, pm_2,..., pm_n as elements of List<Permission>
6) Let upm_1, upm_2,..,upm_n as elements of List<Permission> that *r* has
7) Let PM_r as List<Permission> that user have
8) for each pm_i in *PM* where i=1,2,...,n do
9) if($pm_i \in ump$ and (*r* isRoleOf U_c)){
10) for each p_j in *P* where j=1,2,...,n do
11) if($p_j \rightarrow PM_i$){
12) PM_r.push(pm_i);
13) }
14) }
15) return PM_r

Algorithm II enables the KC-RBAC model to cluster those access rights based on the result of Algorithm I because the subset of PT may contain the duplicate access rights generated in Algorithm I.

C. Knowledge-Involved Information Filtration Algorithm

The outputs of Algorithm I consist of the purposes and their associate medical standard knowledge including ICD-10 and SNOMED CT, which contains massive medical concepts and relationships. Moreover, the purposes are associated with specific permissions (access rights) in Algorithm II. To integrate their relationships, Algorithm III, also called Knowledge-Involved Information Filtration Algorithm (KIFA), is presented below.

Algorithm III KIFA
Input: List<Purpose> *P*, Knowledge *K*, Roles *R*, Permission *PM*, Information *I*
Output: I_u as user information

1) Let k_1, k_2, k_3..., k_n as elements of Knowledge in libraries (*K*)
2) Let r_1, r_2, r_3..., r_n as elements of Knowledge that matches the purposes (K_i)
3) Let i_1, i_2, ...,i_n as elements of necessary information(*i*)
4) Let I_1, I_2, ..., i_n as elements of total information(*I*)
5) for each k_i where i=1, 2,..., n in *K* do
6) for each p_j where j=1, 2,..., n in *P* do
7) if(k_i.attribute(*P*) EQ p_j){
8) if($p_j \in (R \cup PM)$){
9) *r*.push(k_i)
10) }
11) }
12) for each i_i where i=0, 1, 2..., n do
13) for each k_j where j=01,2... n do
14) if(I_i hasknowledge(k_j)){
15) *i*.push(I_i);
16) }
return *i*;

By using medical domain knowledge-driven libraries, based on the actual purposes of requesting of information, the KC-RBAC model can finally extract the necessary data from the system when hospital employees request patient information.

V. RESULTS AND DISCUSSION

Based on the proposed model and algorithms, this section summarizes the results of our experiments. More importantly, we report how the model works in the existing hospital information systems in real cases.

A. Case Studies of KC-RBAC in the Hospital

To evaluate the performance of KC-RBAC, we established a prototype hospital information system with real medical records of cancer patients (8,654 copies of records) to simulate the real process of querying the patient information in a hospital. These data contain four kinds of illness information, including esophageal cancer, stomach cancer, kidney cancer, bowel cancer, and leukemia (2,033 records). Each record contains 27 data fields, 11 of which are patient privacy data fields.

Patient personal data, such as the names, identification

numbers, addresses, and so on, are very sensitive for both patients and medical employees. The proposed PT, KM, and RH are introduced independently in this system. Two test indicators are used to evaluate the results of the experiments, including the number of medical records accessed (MRA) and the amount of medical knowledge (MK) as shown in Table III. In this simulation, we assume that each medical record is a 1*n matrix, each cell in the matrix can be seen as the basic unit of patient information. At the same time, each cell can be mapped to a set of medical domain knowledge. Thusly, the MRA is measured based on the number of medical records multiplied by the filtering ratio of the purpose, knowledge, role and knowledge-role. And the MK is measured based on corresponding MRA multiplied by the number of concepts and rules of domain knowedge in each cell. These relationships and equations are listed in Table III. A series of access control models are separately implemented upon the system to examine patient privacy protection performance of the different models. In order to illustrate how the model can protect patient privacy, two cases are given to discuss and evaluate the advantages of KC-RBAC.

TABLE III
THE FORMAL EQUATIONS TO CALCULATE MRA AND MK OF THE DIFFERENT MODELS IN THE EXPERIMENTS

Models	**Components**	**MRA**	**MK**
Open-AC	-	N (all records)	MRA*1
RBAC	Role	$N * r(\%)$	MRA*K_0
P-RBAC	Purpose	$N *r(\%)*p(\%)$	MRA*K_p
K-RBAC	Knowledge	$N*r(\%)*k(\%)$	MRA*K_r
KC-RBAC	Role, Purpose, Knowledge	$N*r(\%)*k(\%)*p(\%)$	MRA*(K_0+K_p+K_r)

(N, number of medical records in the system; r(%), percentage of the role-based records; p(%), percentage of the purpose-based records; k(%), percentage of the knowledge-based records; K_0, K_p, and K_r, the purpose-, role-, and context-lead knowledge)

Case 1: The matron (head nurse) specifies that a nurse is responsible for a patient, and the KC-RBAC-based system can automatically filter and provide information about the care but doesn't disclose unnecessary information of the patient.

In Case 1, the nurses follow hospital policies to take care of patients, which means that they do not have to know all of the patient information. The KC-RBAC model in this case firstly determined the preliminary permission of the nurses based on their roles and context of hospital policies and associated the nurses to their patients. Thusly, the nurses know only the necessary patient information. Role, context of policies, and permissions are considered to protect patient privacy in the nursing process. When applying KC-RABC model in the context of Case 1, Algorithm I can filter out the unreasonable user intentions and only give necessary access rights to user according to the using context. Figure 6 denotes the difference of patient information leaks between two identification algorithms by using the intention-based and purpose-based rules respectively. In the Intention-based Identification algorithm, we simulated the process of visting HIS without consideration on whether their visiting purposes are necessary for the users in some context (e. g. a nurse on duty tries to search all records for specific patients in the system). In contrast, for the Purpose-based Identification algorithm, the user intentions are double checked by Algorithm I and mapped to a set of predefined purposes in PT. Based on this, in the simulation, we generate a random array as the knowledge set, and then assume some mapping relationships between the generated knowledge and patient information (data cells in the 1*n matrix). Then we run ten rounds of comparative experiments. The simulation results are shown as Figure 6.

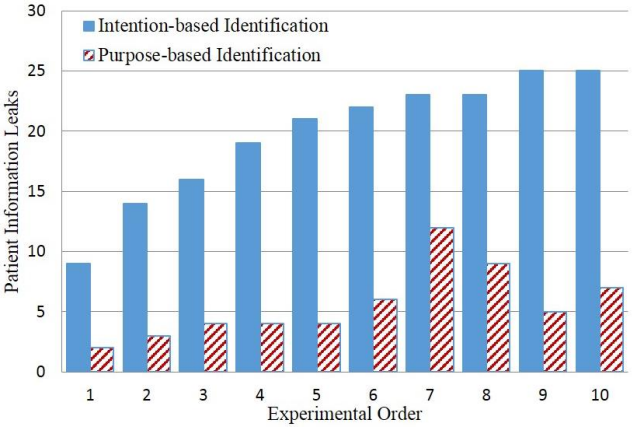

Fig. 6. The results of the comparison between the Intetion-based Identification test and the Purpose-based Identification test.

In Figure 6, the horizontal axis represents the number of rounds and the vertical axis represents the estimated values of patient information leaks in the simulation. The larger the value of vertical axis, the more information or access rights the user acquire. The values of the Intention-based Identification in vertical axis are higher than those of the Purpose-based Identification in each round of experiments. This indicates that only the Intention-based Identification is hard to reduce the information exposed to system users if no further procedures are adopted to identify whether the intentions are neccessay.

Case 2: A doctor wants to do scientific research based on the electronic medical records in hospital information systems. Given that he/she may have a specific scientific research purpose, the KC-RBAC-based system automatically filters out all personal information and only displays purpose-related information.

In this case, the information that the doctor acquired in Case 2 contains not only the related medical records but also the knowledge associated with the related medical domains, which help the doctors do their research. The medical domain is involved in the response of the KC-RABC model to provide doctors with more medical reference information. Given that the purpose of the doctor is scientific research, the patient information will be hidden, reducing the risk of privacy leaking. A series of experiments with several different access control models were simulated according to Table III. One experiment is to compare the difference of the MRA indicator in several models when querying the medical records. Another one is to compare the difference of the MK indicator in different models when querying medical records for research. The results of the experiments are shown in Figure 7.

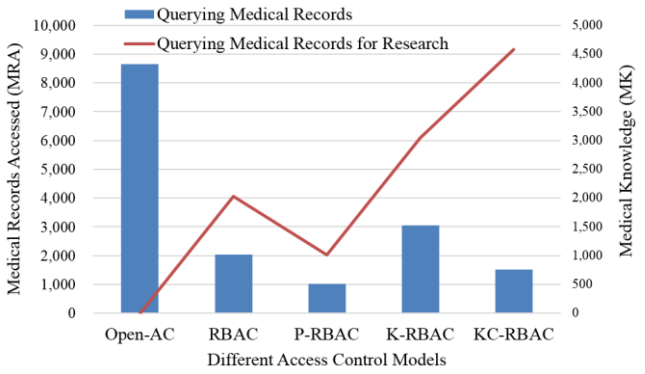

Fig. 7. Performance evaluation of protecting patient privacy for different access models in the prototype system

In Figure 7, Open-Access Control model (Open-AC), RBAC, Purpose-RBAC (P-RBAC), Knowledge-RBAC (K-RBAC), and KC-RABC are evaluated based on the prototype information systems in a hospital. According to the medical records accessed by the user (MRA indictor), the Open-AC model caused a large risk of privacy disclosure, whereas KC-RBAC has better performance in comparison with Open-AC, RBAC, and K-RABC. The P-RBAC is the model that only considers the purpose of using the system, which is not suitable for hospital information system because the considered purposes could be illegal and irrelevant to the hospital process. According to the evaluation of the amount of medical knowledge provided by the system (MK indicator) in Figure 7, the KC-RBAC provided the best support of knowledge-based purposes in Case 2, which shows another advantage of the KC-RBAC that is able to improve the hospital information systems.

B. Discussion of the Model's Application

The KC-RBAC model is proposed to protect patient privacy in the medical information systems of a hospital. The existing RBAC models provide the users with a role-based access control model with static restrictions. By contrast, the proposed KC-RBAC can identify the real purposes of system users by referring to the biomedical domain knowledge. Different system users in the hospital information systems have different roles in a variety of hospital processes. However, even for users with the same role in the system, their intentions to use the information systems may be different and vary in different contexts. As shown in Figure 8, in a HIS's deployed RBAC model, the system user can access all information and data authorized to his or her role. While in a KC-RBAC system, the accessing rights of a user additionally are restricted by using context which is supported by knowledge.

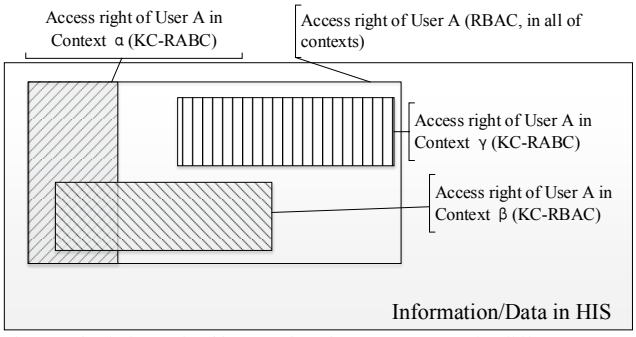

Fig. 8. The information/data explosed to system user in different contexts when adopting KC-RBAC and RBAC.

The KC-RBAC model should be deployed at the fundamental level of the information system of the hospital because the patient information is widely existing in different sub-systems. All the access requests on patient information by hospital employees will be reorganized and modified by the KC-RBAC model to protect patient privacy. However, the KC-RBAC model is designed for the access control in those hospitals with complex processes and organizational structure, and as a consequence, it is not suitable for individual doctors.

The input of Knowledge Module in the KC-RBAC is dynamic, while the role hierarchy of RBAC model is static. For example, the adopted knowledge includes ICD-10 and SNOMED CT, the structure of which supports the expansion of medical knowledge in the future without changing our model. The user intentions for using the system are associated with the purpose hierarchy, and also mapped to the medical domain knowledge that includes the semantic maps of medical concepts. This flexible knowledge organization allows possible knowledge inference to our Knowledge Module.

VI. Conclusion

In a traditional RBAC model, one role is always mapped to a fixed set of access rights, which means the user with a specific role can access to corresponding modules of the systems in any context. This kind of access control mechanism lacks flexibility. In this paper, we propose a KC-RBAC model which has a more precise access control according to the user's context. The identification of visiting purpose is realized based on the PT which indicates the relationships of purposes in differnet context. Compared with current RBAC model and its improvement models, KC-RBAC can effectively zoom out the scope of data and information that a system user can access. The reduction of patient information range accessed by the system user objectively enhances the privacy protection for patients. Based on results of the experiments, the proposed model is shown to be highly flexible to protect patient privacy in HIS compared with other access control mechanisms. Although the model is special for HIS, it also can be used for privacy protection in other information systems if corresponding context knowledge can be described and encoded in the knowledge module.

References

[1] V. P. Aggelidis and P. D. Chatzoglou, "Methods for evaluating hospital information systems: a literature review," *EuroMed J. Business*, vol. 3, pp. 99-118, 2008.

[2] S. K. Dubey, A. Anand and H. Jangala, "Laboratory information and management system: A tool to increase laboratory productivity," *Clinical Research & Regulatory Affairs*, vol. 29, no. 2, pp. 45-56, 2012.

[3] R. Wetering, R. Batenburg, J. Versendaal, R. Lederman and L. Firth, "A balanced evaluation perspective: picture archiving and communication system impacts on hospital workflow," *J. Digital Imaging*, vol. 19, pp. 10-17, 2006.

[4] J. J. Nance, C. Meenan and P. G. Nagy, "The future of the radiology information system," *American J. Roentgenology*, vol. 200, pp. 1064-1070, 2013.

[5] T. Kuroda, H. Sasaki, T. Suenaga, Y. Masuda, Y. Yasumuro and K. Hori, *et al.*, "Embedded ubiquitous services on hospital information systems," *IEEE Trans. Informat. Technology in Biomed.*, vol. 16, pp. 1216-1223, 2012.

[6] E. V. Eikey, A. R. Murphy, M. C. Reddy and H. Xu, "Designing for privacy management in hospitals: Understanding the gap between user activities and IT staff's understandings," *Int. J. Med. Informat.*, vol. 84, pp. 1065-1075, 2015.

[7] H. Zhang, S. Mehotra, D. Liebovitz, C. A. Gunter and B. Malin, "Mining Deviations from Patient Care Pathways via Electronic

Medical Record System Audits," ACM Trans. Management Informat. Syst., vol. 4, A. 17, 2013.

[8] R. Zhang, D. Chen, and X. Shang, "Privacy preserving for patients' information: a knowledge-constrained access control model for hospital information systems," In Proc. IEEE INDIN 2016, Poitiers, France, 2016, pp. 921-926.

[9] R. S. Sandhu, E. J. Coyne, H. L. Feinstein and C. E. Youman, "Role- based access control models," IEEE Computer, vol. 29, pp. 38-47, 1996.

[10] P. Hung, "Towards a privacy: access control model for e-healthcare service", 3rd Conf. Privacy, Security and Trust, New Brunswick, Canada, October 12-14, 2005.

[11] J. Byun and N. Li, "Purpose based access control for privacy protection in relational database systems," Int. J. Very Large Data Bases, vol. 17, pp. 603-619, 2008.

[12] M. Viceconti, P. Hunter and R. Hose, "Big data, big knowledge: big data for personalized healthcare," IEEE J. Biomed. and Health Informat., vol. 19, pp. 1209-1215, 2015.

[13] J. Li, "Ensuring privacy in a personal health record system," Computer, vol. 48, pp. 24-31, 2015.

[14] W. Hsu and J. Pan, "The secure authorization model for healthcare information system," J. Med. Syst., vol. 37, pp. 1-5, 2013.

[15] A. V. Deokar and O. F. EI-Gayar, "On semantic annotation of decision models," Informat. Syst. and e-Business Management, vol. 11, pp. 93- 117, 2012.

[16] S. Gritzalis, C. Lambrinoudakis, D. Lekkas and S. Deftereos, "Technical guidelines for enhancing privacy and data protection in modern electronic medical environments," IEEE Trans. Information Technology in Biomed., vol. 9, no. 3, pp. 413-423, 2005.

[17] L. Røstad and O. Nytrø, "Towards dynamic access control for healthcare information systems," Studies in Health Technology & Informat., vol. 136, pp. 703-8, 2008.

[18] V. P. Gurupur, S. C. Suh, R. R Selvaggi, P. R. Karla, J. S. Nair and S. Ajit, "An approach for building a personal health information system using conceptual domain knowledge," J. Med. Syst., vol. 36, no. 6, pp. 3685-3693, 2012.

[19] W. A. Khan, A. M. Khattak, M. Hussain, M. B. Amin, M. Afzal and C. Nugent, et al., "An adaptive semantic based mediation system for data interoperability among health information systems," J. Med. Syst., vol. 38, no. 8, pp. 1-18, 2014.

[20] L. Lee, Y. Chou, E. Huang and D. Liou, "Design of a personal health record and health knowledge sharing system using IHE-XDS and OWL," J. Med. Syst., vol. 37, no. 2, pp. 1-12, 2013.

[21] M. B. Ateya, B. C. Delaney and S. M. Speedie, "The value of structured data elements from electronic health records for identifying subjects for primary care clinical trials," BMC Med. Informat. & Decision-making, vol. 16, no. 1, 2015.

[22] N. W. Changa, H. J. Dai, J. Jonnagaddala, C. W. Chen, R. T. Han Tsai and W. L. Hsu, "A context-aware approach for progression tracking of medical concepts in electronic medical records," J. Biomed. Informat., vol. 59, pp. S150-S157, 2015.

[23] D. Fabbri and K. LeFevre, "Explaining accesses to electronic medical records using diagnosis information," J. American Med. Informat. Association, vol. 20, no. 1, pp. 52-60, 2013.

[24] D. Beimel and M. Peleg. "Using OWL and SWRL to represent and reason with situation-based access control policies," Data & Knowledge Eng., vol. 70, no. 6, pp. 596-615, 2011.

[25] Z. Li, C. H. Chu and W. Yao, "A semantic authorization model for pervasive healthcare," Research Collection School of Informat. Syst., vol. 38, pp. 76-87, 2014.

[26] Bernd Blobel, "Ontology driven health information systems architectures enable pHealth for empowered patients," Int. J. Med. Informat., vol. 80, no. 2, pp. 17-25, 2010.

[27] C. L. Gordona and C. Weng, "Combining expert knowledge and knowledge automatically acquired from electronic data sources for continued ontology evaluation and improvement," J. Biomed. Informat., vol. 57, Part C, pp. 42-52, 2015.