A Privacy-aware Graph-based Access Control System for the Healthcare Domain

Yuan Tian[1], Biao Song[1], M.Mehedi.Hassan[2] and Eui-Nam Huh[1*]
[1] KyungHee University, Dept. Computer Engineering,
Internet Computing & Network Security Lab. Room No.322.
1Seocheon-dong, Giheun-gu,Yongin-si, Gyeonggi-do 446-701, Republic of Korea.
[2]King Saud University, Riyadh, College of Computer and Information Sciences,
[e-mail: (ytian, bsong, johnhuh)@khu.ac.kr,seyam27@hotmail.com]
*Corresponding author: Eui-Nam Huh

Received May 14, 2012; revised August 13, 2012; accepted September 13, 2012; published October 29, 2012

Abstract

The growing concern for the protection of personal information has made it critical to implement effective technologies for privacy and data management. By observing the limitations of existing approaches, we found that there is an urgent need for a flexible, privacy-aware system that is able to meet the privacy preservation needs at both the role levels and the personal levels. We proposed a conceptual system that considered these two requirements: a graph-based, access control model to safeguard patient privacy. We present a case study of the healthcare field in this paper. While our model was tested in the field of healthcare, it is generic and can be adapted to use in other fields. The proof-of-concept demos were also provided with the aim of valuating the efficacy of our system. In the end, based on the hospital scenarios, we present the experimental results to demonstrate the performance of our system, and we also compared those results to existing privacy-aware systems. As a result, we ensured a high quality of medical care service by preserving patient privacy.

Keywords: Healthcare information systems, privacy preservation, access control, privacy preference, purpose, privacy data graph

A preliminary version of this paper [34][36] appeared in MoMM2009 and MASS '09, International Conference on 20-22 Sept. 2009. This version reveals our early thoughts about the privacy data graph. However, from the definition to the mathematical expression, the privacy data graphs presented in previous versions are quite different from the one in this paper. In this paper, we focused on a privacy data graph considering role hierarchy. With the concept of role hierarchy and constraints, the size of the privacy data graph was minimized to reduce the system overhead in the running phase.

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2012-(H0301-12-2001). The corresponding author is Eui-Nam Huh.

http://dx.doi.org/10.3837/tiis.2012.10.016

1. Introduction

Privacy has been acknowledged as a vital building block in the medical domain from both moral and legal concerns [1][6][13][18]. On the one hand, virtually everyone agrees that one's medical information should not be open to inspection by any interested person [12][18]. Electronic health record (EHR) [15][16][54] contains a great deal of sensitive information such as psychiatric care, substance abuse, genetic predispositions to diseases, abortions, and so on. Unauthorized disclosure of this information may cause serious consequences including social embarrassment or prejudice, difficulties in obtaining or continuing insurance contracts and loans, or limits on a person's ability to obtain and maintain employment [11]. On the other hand, most also agree that some privacy sacrifices in the medical domain are necessary for scientific advancement, comprehensive public health surveillance and protection, and other social goals [18]. Physicians need and expect to access a patient's full medical records in order to help diagnose diseases correctly and to design effective treatment plans that take into account many complicating factors [44].

It is not surprising that a great deal of attention has been paid to finding an appropriate trade-off between maintaining patient privacy and providing better quality healthcare. Since the electronic sharing of patients' personal health information requires their informed consent, healthcare information networks need an access control model that can enforce individual access policies tailored to specific circumstances [14][44]. After studying the existing approaches, we found that the most common practice in preserving privacy at the role levels is the Role-Based Access Control (RBAC) model [5][17][21][27][29][30][31]. RBAC is a method for managing privacy data that allows permissions to be set based on limited roles when there are a large number of users and therefore allows for the user data to be centrally managed. RBAC models are widely used in the healthcare field. Generally, two major roles are commonly considered: patient as the data subject and hospital staff as the data requester. In addition, various ontology methods are used to classify those roles into many mutually exclusive sets [10], which range from clinic staff to healthcare provider. It has been proven that the use of this traditional role level privacy preservation is necessary to either facilitate management or to avoid medical accidents.

Other researchers have focused on privacy preservation at the personal level and have proposed anonymity models [26][32][38][39] to protect personal privacy information. Anonymity is defined as "the state of being not identifiable within a set of subjects" [38]. If the data cannot be traced back to an individual, then the collection and usage of the data does not usually pose a threat to the individual's privacy. The size of the set of subjects influences the strength of the anonymity; the larger the set, the stronger the anonymity. Anonymity allows for viable alternatives to the collection of personal data as well as the legal collection of certain data without requiring user consent [39]. An additional type of data service model for privacy protection [22] at the personal level was proposed in order to minimize the need for disclosure of personal privacy data for customers (or for patients at a hospital).

By observing the existing approaches, we found that the preservation of privacy through anonymity or by restricting access to data is not a new problem. Many studies have been conducted in either the RBAC domain or in the field of anonymity systems; however, none of these studies have provided solutions that combine the two models in order to simultaneously fulfill the requirements of both role level privacy preservation and privacy protection at the personal level.

Therefore, we took these two requirements into consideration in this study and developed a conceptual system: a graph-based, access control model to safeguard patient privacy. We propose a privacy data graph in this study in order to enable both RBAC and personal level access control. As some vulnerability may be created after certain types of sensitive data are disclosed, the proposed privacy data graph can help to prevent data breaches caused by access to this sensitive data. We focused on the role hierarchy in RBAC which helps to simplify the privacy data graph by reducing its nodes and edges. In addition, the notion of purpose was added to specify the intended usage of the data, which establishes the actual boundaries of data processing [4][8]. The management of patient information can be subtly transferred from the role level to the personal level using the privacy data graph that we have proposed. Patients can express their privacy concerns and customize their privacy preferences after the general role design and management have been completed; as a result, we ensured appropriate access of healthcare data to allow complete and high quality care by the providers. The design approach including full detailed views of the proposed system can be found in Section 3 and Section 4 of this paper.

In this paper, we describe two typical hospital scenarios in order to illustrate to what kinds of applications the proposed system is applicable and how they will work. In our scenarios, we discussed two typical roles: the patient as the data subject and the physician as the data requester. The user could take corresponding roles according to his responsibility. We first discussed background information and other relevant studies (Section 2) and then presented a component-based view of the proposed system and the design approach (Section 3). The process of preserving data privacy in our system is explained by three different scenarios (Section 4). We examined various cases in the experiments to evaluate our system performance, and the experimental results show that our system significantly reduces the storage space and execution time compared with existing privacy-aware systems (Section 5). In the end, we summarized the contributions of this paper (Section 6).

2. Related Work

This research is related to many areas of the privacy protection field. The solution introduced in this paper aims at dealing with two open issues. First, inproving the flexibility, scalability, and heterogeneity of the overall infrastructure where the data transactions occor. Second, integrate privacy into the development process and infrastructure employed for data transactions. Therefore, this section will briefly review the related issues that motivated our work.

Different technologies have been proposed to preserve data privacy. RBAC [17][23][24][25][27][43][20] is the primary concept we considered throughout this study. RBAC was proposed by Ferraiolo et al. [27] in 1992 to ensure that certain data or resources could only be accessed by authorized users. The RBAC models are widely used in the healthcare field, because the typical hospital roles (patient, doctor, nurse, etc.) can be used to describe the hospital scenario quite clearly. In many RBAC models, role hierarchy [17][27] defines the inheritance relationship among roles, and it is always applied in order to simplify the administration tasks. The junior roles [27] in the role structure inherit all of the permissions of the senior roles.

An example of role hierarchy we used in this paper is illustrated in **Fig.1,** which provides a view of the different roles involved in the overall hospital environment and how they are categorized in the hierarchy architecture [51][37]. For example, the roles of "*Attending*

Doctor", "*Resident Doctor*", and "*Intern*" may be categorized under the role of "*Physician*", and thus they will inherit all of the permissions from the "*Physician*" role.

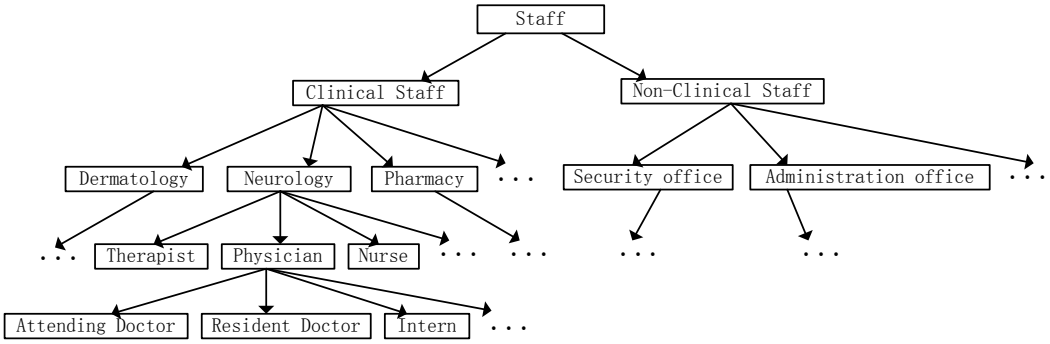

Fig. 1. Role Hierarchy

The patient data we used in the system is from an EHR database. An EHR [16][19][46] has the ability to generate a complete record of a clinical patient encounter, including evidence-based decision support, quality management, and outcomes. Also, the EHR is well categorized through medical ontologies [3], which mainly focused on the representation and (re)organization of medical terminologies, including disambiguation of polysemous terms and organization of very large corpora.

Within health informatics, an ontology is a formal description of a health-related domain. Physicians developed their own specialized languages and lexicons to help them efficiently store and communicate general medical knowledge and patient-related information. Such terminologies, optimized for human processing, are characterized by a significant amount of implicit knowledge. Moreover, ontology-based applications have also been built in the field of medical natural language processing to solve problems in the field of medical terminology, including disambiguation of polysemous terms and organization of very large corpora.

We designed a medical ontology which is grounded in clinical scenario referred [3][2], where each data is already well categorized. The use of ontology is mainly focused on the representation of medical terminologies, by which physicians are able to store and communicate general medical knowledge and patient-related information efficiently.

3. The Proposed System

We proposed a privacy-aware system which could be used in various situations through the switching of data and the controlling of data access. In this paper, the user can be any one who is involved in the hospital and healthcare centers, i.e., patient, staff or non-clinical staff. These situations could include use of EMR (Electronic Medical Record) or EHR databases, use by hospitals or medical research institutes and use by third party privacy service providers. As illustrated in **Fig. 2**, patients' personal data was stored in EMR or EHR databases from which hospitals or healthcare centers could download information to their local databases in a timely manner. The reverse arrowhead in the figure signifies the case data upload from the local database to the EMR or EHR. Any data access happening between the users is organized by our system according to the hospital regulations and the preferences of the data subject. We designed a medical ontology which is grounded in clinical scenario referred [3][2], where each set of data is already well categorized. The use of ontology is mainly focused on the

representation of medical terminologies, by which physicians are able to efficiently store and communicate general medical knowledge and patient-related information.

The foundation of our proposed system is a local database that is used to store the patients' privacy data as well as a data hierarchy applied by most hospitals and healthcare centers in order to facilitate data management. Our novel *privacy data graph* was established based on this database and data hierarchy in order to safeguard patients' sensitive data and to enable regular data access policies. To support personalized medical services, the regular data access policies could be converted to personal data access policies when the anonymity requirement for personal data is taken into consideration in the conversion algorithm.

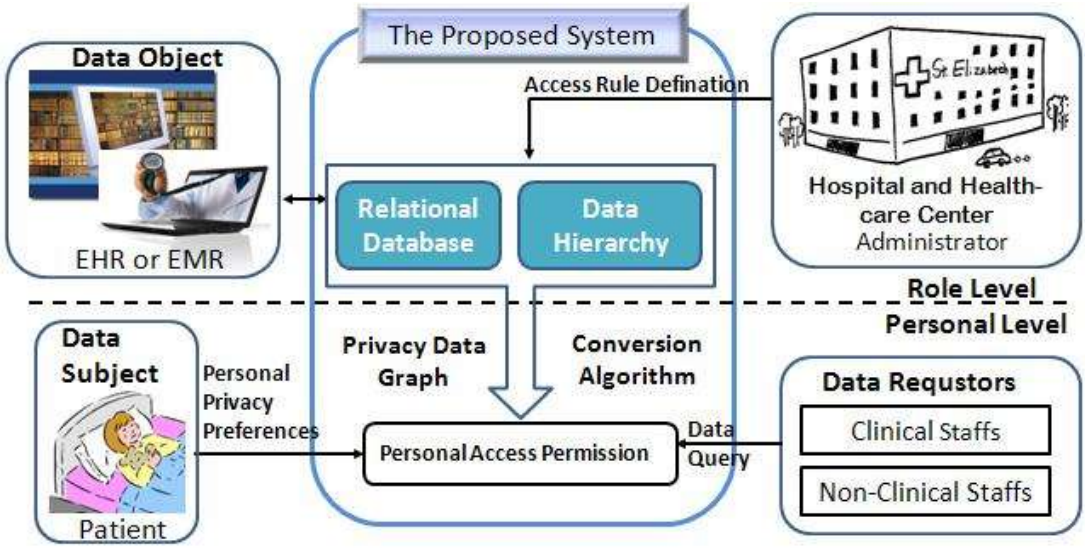

Fig. 2. General Application Scenario

3.1 A Motivating Example

In the healthcare field, relational databases are currently the predominant choice for the storage of medical records. The database includes privacy-aware technology which combines data and its associated usage practices into a single unit. The creation of this single unit simplifies the complex process in which the collected data is brought into compliance with the declared privacy practices [39]. In addition, data hierarchy is used to manage the data in a hierarchical form because the record is always a collection of related fields [5][21]. These methods focus on both the data itself and the relationships among the data.

- Relational Database

In order to clearly illustrate our proposed system, we used the following motivating example. We assumed that the patients' data was exported from an EHR or EMR system and stored in a local database. For the sake of simplicity, only a small number of datasets are listed to illustrate the function of the proposed system, as shown in **Table 1**.

One dataset contained multiple tabulations, and those tabulations could be linked by certain constraints. In the following table, the term ``ID'' signifies the tabular data in our execution example. The ``ID'' is defined as the only identifying information for each patient, and it is the primary key for each data set. For example, if ``ID'' is linked to ``Cranial Nerve Symptoms'' and ``Cranial Nerve Symptoms'' is linked to ``Parkinson's Diagnosis,'' as in dataset 7, then it is possible to determine the relationship between ``ID'' and ``Parkinson's Diagnosis.''

Table 1. Database Information

Dataset	Primary Key	Field
Dataset1	ID	Name
Dataset2	ID	Email
Dataset3	ID	Mobile Number
Dataset4	ID	Telephone Number
Dataset5	ID	Address
Dataset6	ID	Payment
Dataset7	ID	Cranial Nerve Symptoms, Parkinson's Diagnosis
Dataset8	ID	Cranial Nerve Symptoms, Mental Disorder
Dataset9	ID	Cranial Nerve Medication, Parkinson's Diagnosis
Dataset10	ID	Cranial Nerve Medication, Mental Disorder
Dataset11	ID	Bone Symptoms, Rickets
Dataset12	ID	Bone Medication, Rickets

- Data Hierarchy

The data in the database can be organized in a tree structure, where each node represents a piece of data and each edge between the data represents a hierarchical relation. A node in the higher level contains the information of the nodes beneath it, which means that access to the junior nodes should be allowed if the access permission has been given to their senior node. In this study, we referred to the node that was associated with the specific data as the data element and to the node for accessing the data as the high level data hierarchy. An example was applied to this hierarchy in **Fig. 3** to illustrate how a patient's data is organized using the tree structure. For example, if the high level data hierarchy ``Identity'' could be accessed, then its subordinate data elements ``Name'' and ``ID'' could also be accessed.

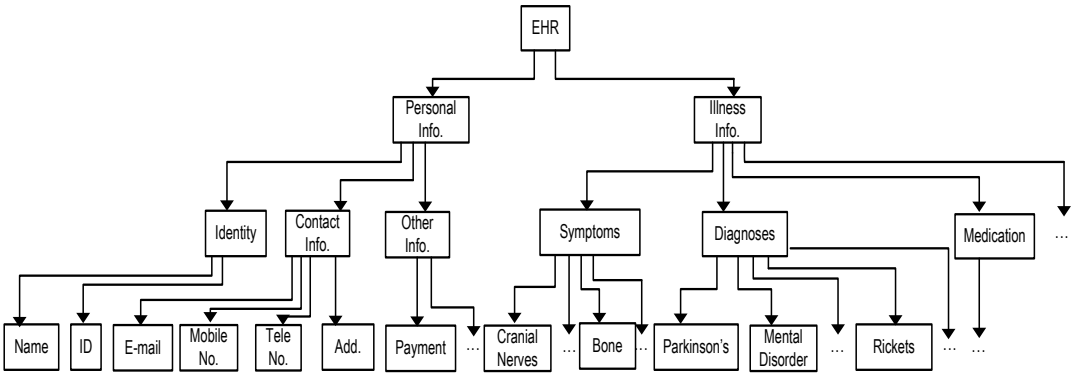

Fig. 3. Tree Architecture in the Database

3.2 Details of the Privacy Data Graph

In order to design a privacy-aware system that was compatible with both of the methods noted above, we used the *privacy data graph* proposed in our previous studies [33][34][36], using an access policy that took into consideration both the data and the relationships among the data. The access policy of each role was defined by system administrators. An example of the proposed privacy data graph for the role of Patient is shown in **Fig. 4**. A tree architecture, like that discussed in Section 3.1, is presented in this graph. In addition, the data fields which appear in the database tables are represented by nodes and are connected using directed edges.

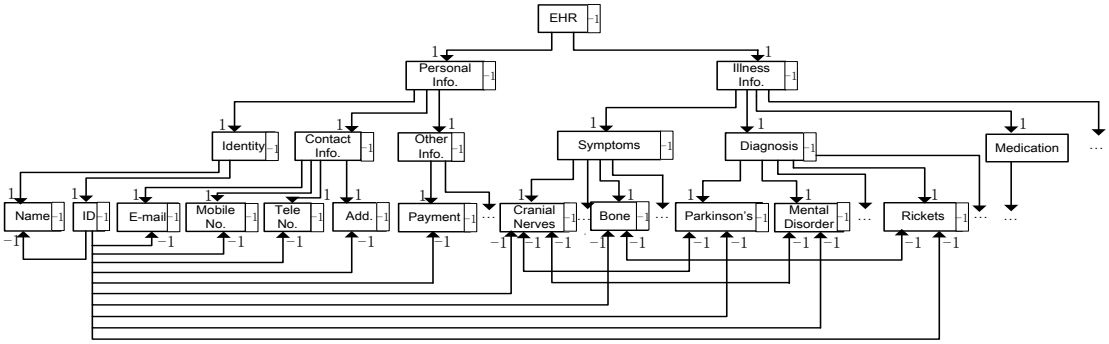

Fig. 4. Privacy Data Graph

Some basic concepts of the *privacy data graph* are formally expressed in order to establish a common terminology. Two values were associated with the node in the proposed graph to indicate the different categories of data access. The node with value "1" could be accessed, whereas access was forbidden to the node with value "-1." As we noted above, the node for accessing data was referred to as the high level data hierarchy, and the node with the specific data was called the data element in the privacy data graph. However, the data element did not signify a single data set; it represented a collection of data from the same category for every user in the same role. For example, if the data element "Name" is set as "1," then each patient's name is accessible, which is supported in the database by the query "Select 'Name' from "Dataset 1."

The linkage between the nodes was referred to as ``relationship" in the *privacy data graph*. Each relationship was accompanied by either "1," "0" or "-1," which represented the associated access control as well as its anonymity situation. There are two broad types of relationships:

1. The relationship in the tree hierarchy:

As we noted above, access to the junior nodes should be allowed if access permission has been given to their senior node. We used the value "1" to denote this type of relationship.

2. The relationships in the database:

- From "primary key" to "field":

In this case, the values of "1" and "-1" were used to represent the permission and prohibition of the associated access, respectively. For example, in the relationship between "ID" and "Name" in **Fig. 8**, "1" indicates that the query "Select `Name' by `ID'" is allowed to be executed and, on the basis of this data request, the data element ``Name" can be matched using the data element "ID." Since the ID was the primary key in the database, we considered it as a unique characteristic in our example. Therefore, a one-to-one relationship existed between each ID to Name pair, and the data could be re-identified even though the data element "ID" represented a set of IDs from users with the role of "Patient" [26][35]. For example, as shown in **Table 2**, only one corresponding Name could be found for each ID.

Table 2. From "Primary Key" to "Field"

ID	Name
10001	Ada Wang
10002	Jason A.Landay
10003	Carmen Etheridge
10004	Shoaib Siddiqui
…	…

Anonymous data for each patient is sacrificed under this kind of data access method. A value of "-1" denoted that the data query was forbidden from the primary key to the field.

In this example, the values ``1" and ``0" were used to represent the permission and prohibition of associated access, respectively. The meaning of ``1" for this example is the same as that noted above. Even though the value ``0" indicates that the access query was allowed to be executed, this type of data request does not begin at the primary key, so only many-to-many linkages could exist among the data elements. For instance, if we executed the data query ``Select `Cranial Nerve Symptoms' by `Mental Disorder' from dataset 7, we can see that in **Table 3**, the data in the data element ``Cranial Nerve Symptoms" (abbreviated CNS) does not match the data in the data element ``Mental Disorder" (abbreviated MD), because ambiguous linking protects the anonymity of the data. Therefore, we could not speculate any specific Mental Disorder data by querying the Cranial Nerve Symptoms data. Based on the personal anonymity requirement, further review of the anonymity situation is required under those circumstances noted above.

Table 3. From "Field" to "Field"

ID	Cranial Nerve Symptoms (CNS)	Mental Disorder (MD)
10001	CNS1	MD3
10002	CNS1	MD2
10003	CNS3	MD1
10004	CNS2	MD1

It should be noted that the proposed *privacy data graph* is able to handle the role level access control and can also properly address the anonymity problem. This feature allows our system to simultaneously support RBAC and allows users to control their privacy preferences.

3.3 Basic Concepts of the Privacy Data Graph

3.3.1 Definitions

Prior to presenting the proposed system, in order to establish a common ground of concepts, some basic definitions are necessary to be expressed formally.

Definition 1. (Role) Let R be the set of roles in our system, denoted by *R* where *ri* *R* denotes any role.

Definition 2. (Privacy data graph) Let $D_i = (V, A)$ be the privacy data graph of r_i.

A privacy data graph comprises a set V of vertices together with a set A of arcs (directed edges). In D_i, let V be the set of privacy data elements of r_i. $v_j \in V$ indicates any data element of r_i. Let A be the set of relationships between the data elements of r_i. An arc $a = (v_x, v_y)$ is considered to be directed from v_x to v_y, v_y is called the head of a and v_x is called the tail of a. The a indicates the relationship between two data elements v_x and v_y.

Suppose r_i equals "patient", we give an example of privacy data graph D_i as shown in **Fig.5**. A data element of r_i can be "Name" or "ID" which is represented as a vertex v. An arc a in A indicates a relationship between two patients' data elements like the relationship between "ID" and "Name".

Fig.6 shows two data tables that contain part of patients' privacy data. In table1 of **Fig.6**,

suppose that v_x = "Name" and v_y ="ID" represent patients' names and IDs. $a=(v_x, v_y)$ denotes given any patient's name, it is possible to get the corresponding ID in table1.

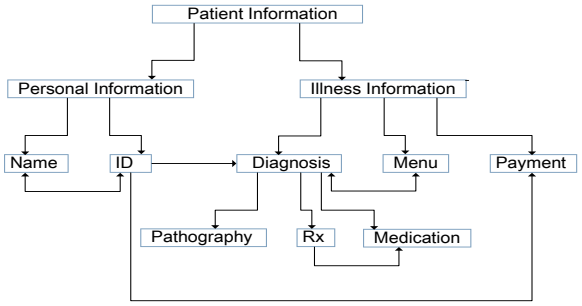

Fig.5. The privacy data graph of patient

Table1

Name	ID
L.James	01
D.Wade	02
M.Yao	03
...	...

Table2

ID	Diagnosis
01	Diabetes
02	Ulcer
03	Tuberculosis
...	...

Fig.6. Part of patients' privacy data

Definition 3.(Access Policy of Privacy Data Elements & Connection) Let $w(v_j)$ be the weight of v_j and $w(a_j)$ be the weight of a_j. In privacy graph, $w(v_i)$ and $w(a_j)$ are the access policy of v_i and a_j, respectively.

Both of $w(v_j)$ and $w(a_j)$ can be chosen from {-1, 0, 1}. $w(v_j)$ = -1 denotes that the data elements indicated by v_i cannot be accessed. $w(v_j)$ = 0 denotes the data elements indicated by v_i can not be accessed now but it can be modified as -1 or 1. $w(v_j)$ =1 means the data elements indicated by v_i can be accessed. For $w(a_j)$, the value defines whether the relationship indicated by a_j is available or not.

In Fig.2, suppose an arc $a=(v_x,v_y)$ where v_x= "Name" and v_y ="ID". A query *"SELECT Name FROM table1"* is submitted to our system. If $w(v_x)$ =-1 or $w(v_x)$ =0, this query will be rejected. If $w(v_x)$ =1, this query will be accepted. Then another query *"SELECT ID FROM table1 ORDER BY Name"* is submitted to our system. If $w(a)$ = -1 or $w(a)$ = 0, this query will be rejected. If $w(a)$ =1, this query will be accepted.

Assume $w(v_x)=1$, $a=(v_x,v_y)$ and $w(a)=1$, we can get $w(v_y)=1$. Since the data indicated by v_x and the relationship between v_x and v_y are disclosed, so the data indicated by v_y is also disclosed at the same time. However, suppose that $w(v_x)=1$, $a=(v_x,v_y)$ and $w(v_y)=1$, we can

not get $w(a)=1$ because the connections of personal data elements on v_x and v_y cannot be found.

Definition 4. (Connectivity between Data Elements) Let $k(v_x,v_y)$ be the connectivity between the data elements indicated by v_x to v_y

Note "$k(v_x,v_y)=1$" means that there exists at least one connected path from v_x to v_y. Note "$k(v_x, v_y)=0$" means the connectivity from v_x to v_y does not exist. A $p(v_x,v_y)$ for $k(v_x,v_y)=1$ contains several arcs $\{a_1, a_2, ..., a_m\}$ where for each $a_{j=1...m} \in p$, $w(a_j)=1$. The head of each a_j is the tail of a_{j+1} except the tail of a_1 is v_x and the head of a_m is v_y. For instance, in **Fig.5** we assume $a_1=(v_x,v_y)$ and $a_2=(v_y,v_z)$ where v_x = "Name", v_y ="ID" and v_z ="Diagnosis". If $w(a_1)=1$ and $w(a_2)=1$, we say $k(v_x,v_z)=1$. Otherwise, $k(v_x,v_z)=0$.

The connectivity $k(v_x,v_z)=1$ between two data elements represents a chain of data disclosure from v_x to v_z. In **Fig.5**, suppose $a_1=(v_x, v_y)$ and $a_2=(v_y, v_z)$ where v_x = "Name", v_y = "ID" and v_z = "Diagnosis". $k(v_x,v_z)=1$ denotes it is possible to get any patient's diagnosis from our system if patients' names are released.

Definition 5.(Access Policy Graph) Given a pair of roles $< r_i, r_j >$, two directed graphs D_{ij} and D_{ji} are defined by the administrator of r_i and r_j respectively. D_{ij} denotes r_i's privacy data access policy for r_j and D_{ji} denotes r_j's privacy data access policy for r_i. D_{ij} and D_i have the same nodes and arcs. D_{ij} is different with D_i in terms of the weights of nodes and arcs are modified.

Definition 6.(Personal Privacy Preference Policy) For each user taking a specific role, we define a set *PV*s to store personal privacy preferences for different roles. Let PV_i be the user's personal privacy preference against r_i. Also, $pv_j \in PV_i$ indicates any personal data element.

Definition 7.(Access Policy of Personal Data Elements) Let $w(pv_j)$ be the weight of pv_j which indicates whether pv_j is disclosed or not. In PV_i, the $w(pv_j)$ is defined as the access policy of pv_j.

The $w(pv_j)=-1$ is defined by system administrator or role administrator as the personal privacy data element indicated by pv_j can't be accessed. Individual user can not modify $w(pv_j)$ in this case. The $w(pv_j)=0$ means the data element indicated by pv_j can't be accessed now but the situation can be modified(to 1 or 0) by individual user. The $w(pv_j)$= 1 means the data element indicated by pv_j can be accessed. If the $w(pv_j)$=1 was modified from the $w(pv_j)$=0, again, it can be modified to 0 by the user.

3.3.2 Usages of the Privacy Data Graph

In this section, we present the usages of our privacy data graph. Broadly, it can be summarized into three categories.

- Remove nodes and edges.

This function could be used in the system design phase by system administrators to define the access field for a specific role by removing unrelated information. Firstly, the administrator should declare which data elements are not allowed to be accessed according to the role's capabilities and delete those data elements as well as the related linkages (relationships). Secondly, if there exists a high level data hierarchy node which does not contain any junior data elements, remove that node and its related linkages. Finally, a new privacy data graph for a certain role would be generated from the original graph.

- Modify the values on nodes and edges.

This function is used to set the role level access policy. As we mentioned before, the values on nodes and edges denote the access permission or prohibition to the node and the edge. The system administrator can set those values and modify them in the system design phase.

- Associated disclosure searching function.

A searching algorithm is applied in this function to check whether any node has connections with the recently disclosed node. If a node is allowed to access (with value ``1'') and search another node through the linkage between them (the edge with value ``1''), then the other node is also disclosed, and the value on that node should be changed to ``1.'' This algorithm should be repeated until there is no further node disclosed.

Detailed examples of the use of our *privacy data graph* can be found in Section 4.1 and Section 4.2 of this paper.

4. Research Examples & Methods

In this section, we describe the two scenarios we used to help illustrate what kinds of applications we wanted to support and roughly how they would work in our proposed system.

4.1 Scenario 1 - Dr. Lee's Rights

Dr. Lee is an intern who works in the department of neurology in hospital A. This hospital recently applied a privacy-aware system to assist in the safeguarding of patient data. As a primary-care doctor in the infectious disease department, Dr. Lee is only allowed to access the limited amount of information that is related to his field. For the purpose of helping the doctor to diagnose patient conditions, the system administrators defined a uniform access rule to indicate what kind of information could be obtained by each doctor in a specific department. Dr. Lee can access related patient information for diagnosis unless the data is very sensitive, in which case he may need the patient's further consent to access it.

4.1.1 Role Hierarchy

In order to ease the management of patient information and to minimize the storage space required, the role hierarchy was predefined by the system administrator. According to the medical ontology [3], in **Fig. 7**, the illness information was categorized into several aspects which includes symptoms, diagnosis, medication and other information. For simplicity, we only enumerated a few items under each category, where some were related to neurology (e.g., cranial nerves symptoms) while others were not (e.g., bone symptoms). The scope of the patient information that the intern in the neurology department may need is marked by shading. If the value of every data element and the relationship between the data in the Patient EHR was set at the default of ``-1,'' then access to that information would initially be forbidden. The system administrator would set a series of access purposes by modifying the initial value.

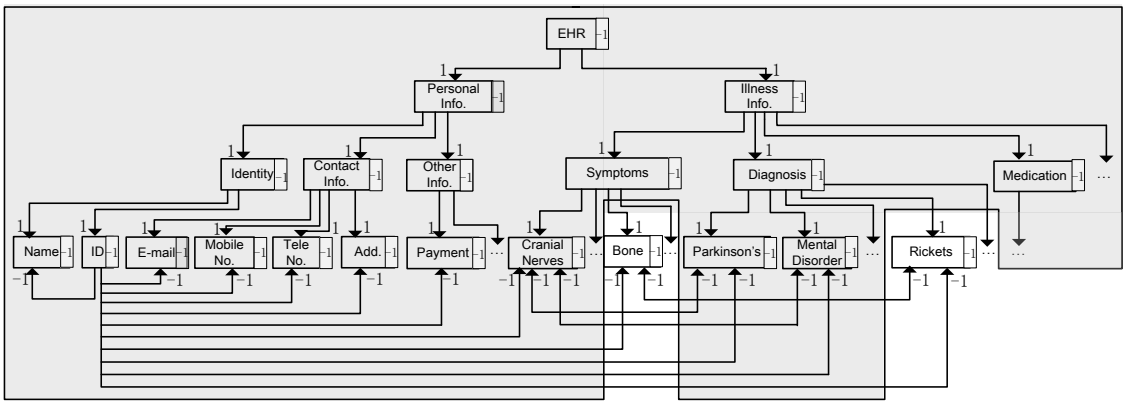

Fig. 7. Patient EHR (The shaded area is that allowed to be accessed by the intern in neurology department)

4.1.2 Access Purpose

In the next step of this process, the system administrator defined a set of access purposes. The access purposes in **Table 4** were used by the intern from the neurology department in order to request *patient* information.

Table 4. Access Requests from the Intern in the Neurology Department

Element	Revealed Data Element	Revealed Linkage
1.Cranial Nerve Symptoms Analysis	"Cranial Nerve Symptoms"	-
2.Parkinson's Analysis	"Parkinson's Diagnosis"	-
3.Mental Disorder Analysis	``Mental Disorder Diagnosis"	-
4.Checking Names	-	(``ID", ``Name")
5. Checking Payments	-	(``ID", ``Payment")
6. Checking Cranial Nerve Symptoms	-	(``ID", ``Cranial Nerve Symptoms")
7. Checking Parkinson's Diagnosis	-	(``ID", ``Parkinson's Diagnosis")
8. Cranial Nerve Symptoms-Parkinson's Analysis	``Cranial Nerve Symptoms" ``Parkinson's Diagnosis"	(``Cranial Nerve Symptoms", ``Parkinson's Diagnosis") (``Parkinson's Diagnosis", ``Cranial Nerve Symptoms")
9.Cranial Nerve Symptoms-Mental Disorder Analysis	``Cranial Nerve Symptoms" ``Mental Disorder Diagnosis"	(``Cranial Nerve Symptoms", "Mental Disorder Diagnosis") ("Mental Disorder Diagnosis", ``Cranial Nerve Symptoms")

Dr. Lee was able to access all of the information, which included both the data elements (i.e., ``Cranial Nerve Symptoms") and the linkages between the data elements (i.e., ``(`ID', `Name')"). It should be noted here that ``Cranial Nerve Symptoms" is a collection of the Cranial Nerve Symptoms data from every individual with the role of *patient*, and the doctor usually requests this access purpose for medical research. In addition, the set of linkages from ``ID" to ``Name" is between every patient's ID to their name. In this example, Dr. Lee could obtain a specific patient's name by executing ``Select `Name' by `ID'" if that patient's ID was given, because a one-to-one match existed between ``ID" and ``Name." However, he could not perform an analysis of the linkage between the Cranial Nerve Symptoms and Mental Disorder data, even though both of the data elements and linkages were provided, because a plurality of

the same Cranial Nerve Symptoms and Mental Disorder data could exist, and thus the data could not be matched.

Next, we described the access purposes noted above by modifying the penalty value in the *privacy data graph*. In **Fig. 8**, the data elements ``Cranial Nerve Symptoms" and ``Parkinson's Diagnosis" were revealed due to the previously permitted access requests that were set at a value of ``1." The linkages between ``ID" and ``Name," ``ID" and ``Payment," ``ID" and ``Cranial Nerve Symptoms," and ``ID" and ``Parkinson's Diagnosis" were also revealed. As we discussed in Section 3.1.1, if the linkages in the database were searched from ``Primary Key" to ``Field," then the value should be set as ``1," and if the relationships were searched from ``Field" to ``Field," then a value of ``0" should be used to denote those linkages. The access rules are illustrated below.

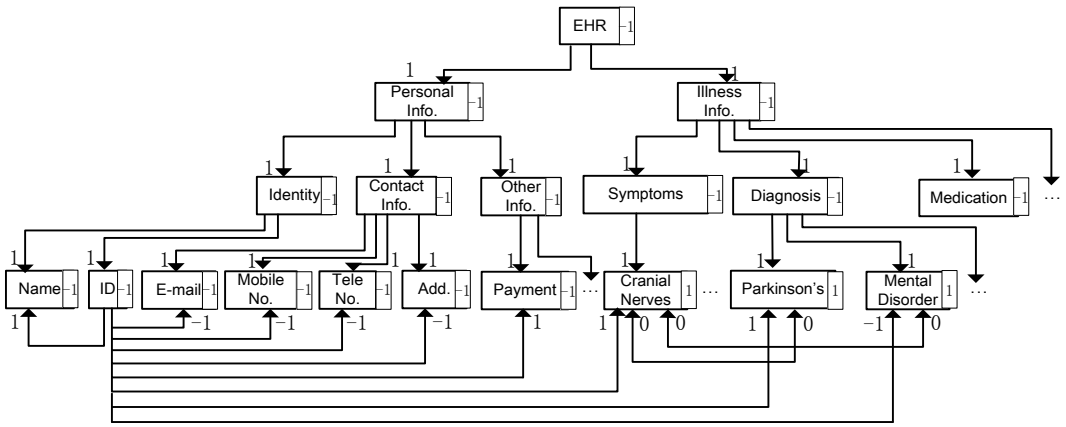

Fig. 8. *Patient* Access Rule

4.2. Scenario 2 - Alice's request

Alice went to hospital A to get a checkup. Accordingly, the hospital received her EHR and uploaded or updated it as was appropriate. Dr. Lee was the doctor appointed to preside over Alice's treatment. Since the treatment process was a one-to-one relationship between Alice and Dr. Lee and the only way to check a specific patient's information was by using that patient's ID, Alice had to give Dr. Lee access to her ID. This does not mean that Alice gave up her right for privacy. On the contrary, she could easily determine to what extent her information was disclosed to Dr. Lee by using the privacy-aware system.

4.2.1 Personal Analysis

Personal analysis is a process that transfers a role level access policy to a personal level access policy. The access rule for Patient is a centralized control for a set of patient data or relationships among those data. However, after Alice gave Dr. Lee permission to access her ID, a *privacy data graph* with her personal access rule was generated. The data element ``Name" in the graph was Alice's name, and the linkage from ``ID" to ``Name" indicated the permission to link Alice's ID to her name.

With the exception of the value of her exposed ID being set at ``1," the values of all of her other data were set at the default of ``-1." The value of the linkages does not change because the same access rule is maintained with the personal level access policy. A searching algorithm (presented in the Appendix) was executed to determine what data would be disclosed in this step. First, the algorithm checks those linkages that are associated with ``ID." Additional data elements could be disclosed if the linkages from ``ID" to the other data

elements were revealed (with a penalty value of ``1"). For example, Dr. Lee could easily gain access to Alice's Mental Disorder data if he was allowed to access her Mental Disorder from her ID linkage. If the linkages from ``ID" to the other data elements were not revealed (with a penalty value of ``-1"), then the values of those other data elements remained ``-1." Then, the searching algorithm repeated the previous step to determine whether the other data elements had connections with the recently disclosed data elements.

We must also discuss what happens when the linkage has a value of ``0." We have already noted that, in this case, the relationships between the data elements can be ambiguous because a plurality of the same data elements may exist. Alice could declare her anonymity requirements by defining a value, and if the value of a data element exceeded this pre-determined value, then that set of data could not be traced back to the individual and vice versa. Sample Patient Cranial Nerve Symptoms and Mental Disorder data are presented in **Table 5**.

Table 5. A database example

ID	Cranial Nerve Symptoms	Mental Disorder (MD)
10001	CNS1	MD3
10002	CNS1	MD2
10003	CNS3	MD1
10004	CNS2	MD1
10005	CNS3	MD2
10006	CNS2	MD3
10007	CNS1	MD3
10008	CNS3	MD2

Alice's ID is ``10003," and she defined her anonymity requirement as ``3." Alice's corresponding Cranial Nerve Symptoms is CNS3, as is noted in the table below. Since there are three CNS3 values listed in this table, Alice's anonymity requirement is satisfied. Under these conditions, we believe that Alice's Mental Disorder data could not be linked back to her Mental Disorder and would be modified to be ``-1." **Figure 9** depicts Alice's personal access rule.

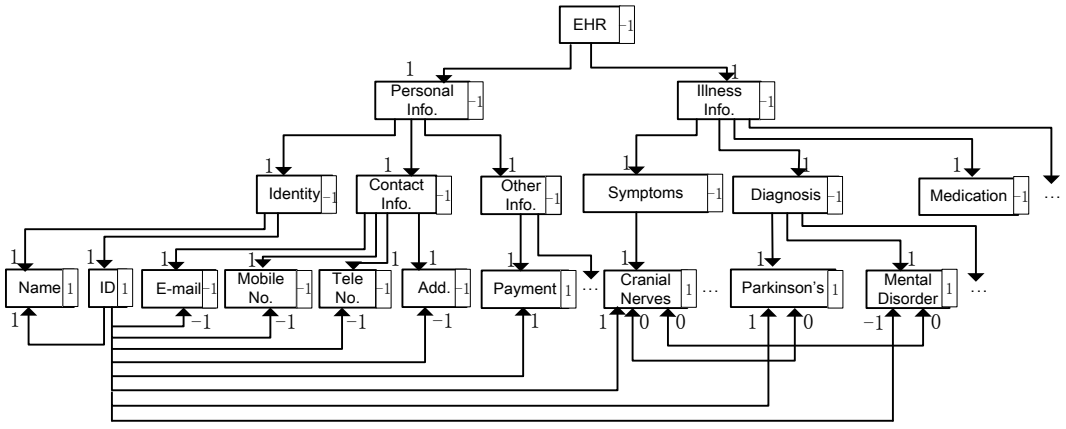

Fig. 9. Alice's Personal Access Rule

Table 6 lists the data for Alice's default personal access permissions after applying the algorithm discussed above.

Table 6. Default Personal Access Permission

Personal Privacy Data	**EHR**	**Personal Info.**	**Illness Info.**	**Identity**	**Contact Info.**	**Other Info.**
Access Permission	-1	-1	-1	-1	-1	-1
Personal Privacy Data	**Symptoms**	**Diagnosis**	**Name**	**ID**	**E-mail**	**Mobile No.**
Access Permission	1	1	-1	-1	-1	-1
Personal Privacy Data	**Tele.No.**	**Address**	**Payment**	**Cranial Nerves**	**Parkinson's**	**Mental Disorder**
Access Permission	-1	-1	1	1	1	1

In order to access specific patient information, the doctor could access the patient's information through the personal requests that were defined previously by the system administrator. In our example, six personal information requests were defined and are listed in **Table 7**. The doctor is allowed to access the first four sets of requested information as those requests have been permitted in the default personal setting, while he still needs to get the patient's permission for the other requests. This enables the patient to be aware of what information is disclosed to his/her doctor. One particular request is the ``Contact,'' in which the requested personal information is alternative.

Table 7. Personal Information Access Requests from the Primary-care Doctor

Purpose Description	**Revealed Personal Data**
1. Checking Name	Name
2. Checking Payment	Payment
3. Checking Cranial Nerve Symptoms	Cranial Nerve Symptoms
4. Checking Parkinson's Diagnosis	Parkinson's Diagnosis
5. Checking Mental Disorder Diagnosis	Mental Disorder Diagnosis
6. Contact	Mobile Num./Address/Tele. Num./E-mail

5. Results & Analysis

The problem of the extra computational cost introduced by adopting privacy control always needs to be considered when designing a privacy-aware system, especially when the system contains EHR, which contain a large amount of important medical information and data for patients.

In this section, we present the experimental results for the performance of our system based on the hospital scenarios described above, and we also compared those results to existing privacy-aware systems. We presented the comparison to see how our system performs when the proposed *privacy data graph* was applied to different privacy-aware RBAC models as the number of role, patient and data elements increased. We also compare the proposed privacy data graph and the existing privacy-aware RBAC model with regard to potential data disclosure detection.

5.1 Performance Evaluation of the Proposed *privacy data graph* Using Existing Privacy-aware RBAC Models

Traditional RBAC technology aims at providing an efficient data and privacy management solution for the healthcare domain; however, it has no flexibility in personal data access control. In our system, the proposed *privacy data graph* is based on privacy-aware RBAC to define privacy policy where a users' data access is permitted or denied. In this experiment, we applied our proposed *privacy data graph* to two typical privacy-aware RBAC models, Core PRBAC (CPRBAC) and Hierarchical PRBAC (HPRBAC), which are commonly used in healthcare information systems [20][43][44][50]. CPRBAC is the base model that provides the basic components, while HPRBAC enhances CPRBAC with hierarchical organization. We conducted three experiments to evaluate which model is more compatible with the proposed *privacy data graph*. In order to easily illustrate our results, we used CPRBACS to denote our system using the CPRBAC model, while HPRBACS was used to denote the system using the HPRBAC model. The variables and parameters used for this group of experiments are listed in **Table 8**.

Table 8. Workload parameters in the experiments

Variable	Range	Parameter	Value
Number of Patients N_P	*1-1000*	Patients per doctor	*5*
Number of Data Elements N_E	*1-50*	Level of hierarchy	$Log_3 N_R$
Number of Roles N_R	*1-9*	Reduction of data element per hierarchy level	*1/4*
-	-	Number of doctors	N_P /5
-	-	Number of role level purposes	N_E/5

Our first experiment was conducted in order to measure the performances of the systems when the number of patients increased.

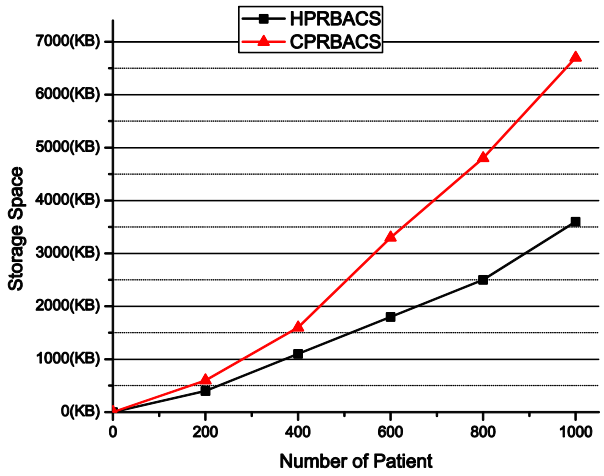

Fig. 10. The Performance of HPRBACS Compared to CPRBACS without Taking Role Hierarchy into Consideration (Impact of the Number of Patients

During this experiment, we set fixed values of N_R = 9, N_E = 50 and varied the number of patients from 10 to 10,000. The results of this experiment are plotted in **Fig. 10**, where it is

easy to see that, as additional patients appeared in the system, the storage space of HPRBACS increased much less than that of CPRBACS without taking the role hierarchy into consideration. For example, the CPRBACS used *3300KB* with 600 patients, while the HPRBACS only used *1800KB*. Therefore, the HPRBACS reduced the storage space by almost *45%* compared to that of the CPRBACS without taking the role hierarchy into consideration. This resulted from the fact that HPRBACS does not have to maintain an individual doctor's access permission for every patient's data because some of the unrelated data are removed during the hierarchy process.Then we set fixed values of N_P = 100, N_E = 50 and varied the number of roles N_R from 1 to 9. The simulation results are presented in **Fig. 11**.

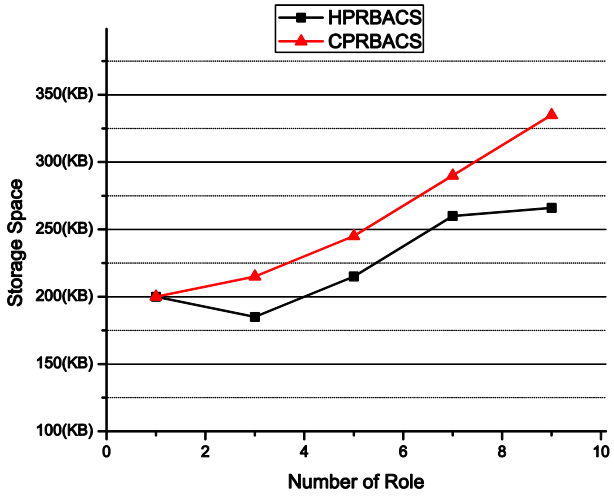

Fig. 11. The Performance of HPRBACS Compared to CPRBACS without Taking Role Hierarchy into Consideration (Impact of the Number of Roles)

In **Fig. 11**, we observed that, by applying role hierarchy, the HPRBACS performed better than did the CPRBACS without taking role hierarchy into consideration as the number of roles increased.

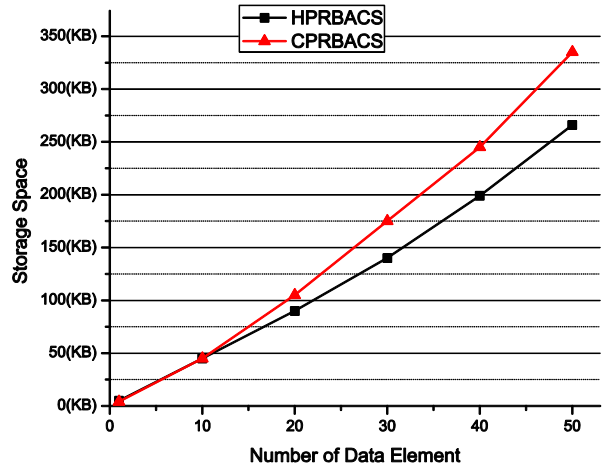

Fig. 12. The Performance of HPRBACS Compared to that of CPRBACS without Taking Role Hierarchy into Consideration (Impact of the Number of Data Elements)

For example, with three roles, the storage space of our system was *175KB*, which was a reduction of almost *22%* compared to that of the CPRBACS. However, the role hierarchy is not equal to Log_3N_R in reality; therefore, this simulation was conducted to illustrate the benefits of applying role hierarchy. In the following experiment, we set fixed values of N_R *=9*, N_P *=100* and varied the number of data elements from 1 to 50. The simulation results are plotted in **Fig. 12**. From **Fig. 12**, we can conclude that, when the number of data elements increased, the HPRBACS still performed better than did the CPRBACS without taking role hierarchy into consideration. For example, the storage spaces of CPRBACS and HPRBACS were *175KB* and *140KB*, respectively, when there were 30 data elements in the systems.

5.2 The Comparisons of the Proposed Privacy Data Graph and the Existing Privacy-aware RBAC Model with Respect to Potential Data Disclosure Detection

Comparing to [20][43], our approach provides more comprehensive privacy protection by considering both role level data access and personal level data access. In the following two simulations, we randomly generate patients' database according to **Table 1** and the anonymity requirement for each patient ranging from 3 to 9. All patients' medical information can be accessed at role level, but cannot be related with any patient's identity. We randomly generate five groups of personal level data queries where the number of data queries in each group ranges from 100 to 900. Each query requests a specific patient's medical information by using patient's ID information. Because our focus is on the capability of detecting potential data disclosure, we choose the most important aspect related to that objective, i.e., the total amount of personal data which has been related to the identity of corresponding patient. In our first simulation, we vary the number of personal level data queries. The results are presented in **Fig.13**.

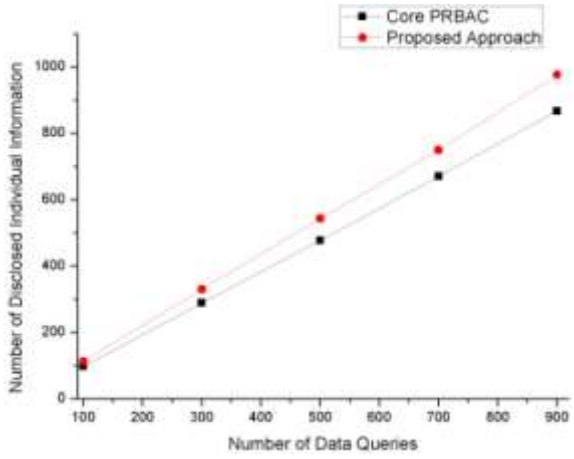

Fig. 13 Capability of Detecting Potential Data Disclosure with Varied Number of Data Queries

It can be seen from **Fig. 13** that our proposed approach is capable of detecting potential data disclosure. Although the core PRBAC model provides personal level data management as well, there is no potential data disclosure detection algorithm in the core PRBAC model. Thus, only the requested personal data is considered as the disclosed individual information. On the other hand, our proposed approach considers the data linkages. Thus, our privacy data graph addresses about 10% of individual information as potential data disclosure.

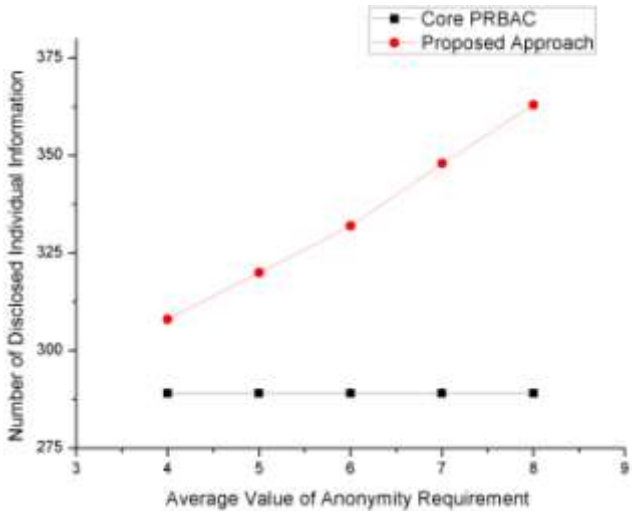

Fig. 14 Capability of Detecting Potential Data Disclosure with Varied Value of Anonymity Requirement

In our second simulation, we fix the number of data query as 300, and vary the average anonymity requirement. The results can be found in **Fig. 14**. The results we observe from **Fig.14** show that the average has effect in our approach. Larger amount of potential data disclosure can be detected with the increase of average anonymity value. The core PRBAC does not take the anonymity into consideration, which makes constant number of disclosure individual information.

6. Conclusions

The privacy and management of sensitive data protection are very critical issues at the current time because people are paying closer attention to their personal information. In this paper, we presented a privacy-aware system for the healthcare domain that facilitated the management of patient data and safeguarded patient privacy. The contributions of our findings to the field of privacy preservation can be summarized by answering the following questions:

- Can the *privacy data graph* simultaneously support RBAC and allow users to set their own privacy preferences?

An algorithm that allows the role level rules to be converted into personal level rules must be developed in order to simultaneously support RBAC and personal level privacy preferences. We developed a corresponding conversion algorithm based on the *privacy data graph* that can easily achieve this goal.

- Can the proposed system achieve privacy preservation by restricting access to data in order to meet anonymity requirements?

Access policies, especially personal level access policies, should differ, because anonymity requirements may differ from one scenario to another. We established a system that can be used to express and enforce various access policies based on different anonymity requirements rather than a system that could only be used with a particular anonymity technology or access policy.

- Can a privacy-aware system improve privacy protection and reduce storage overhead?

Since we used a *privacy data graph* to represent the data elements and linkages, potential

data disclosure can be detected and presented to subjects. In addition, the system provides an interface for individual users so that they can express their personal privacy preferences in a simple way without violating any role level rules. Therefore, the storage overhead can be reduced due to the simple preference expressions of the individual users.

In conclusion, our system supports RBAC as well as personal level access control based on the proposed *privacy data graph*. In addition, the notion of purpose was added in order to specify the intended data usage and to allow users to set personal privacy preferences based on their intended purpose. In the end, we provided simulations and comparative results which support the claims made in this paper.

References

[1] Gerardo Canfora, Elisa Costante, Igino Pennino, and Corrado Aaron Visaggio, "A three-layered model to implement data privacy policies", *Computer Standards & Interfaces, Elsevier Science Publisher*, pp.398-409, 2008.

[2] Berkeley biological open source project, http://www.berkeleybop.org/ (Accessed: 2 Sep, 2010)

[3] Ontology of Clinical Research (OCRe), http://bioportal.bioontology.org/ (Accessed: 10 Sep, 2010).

[4] Paolo Guarda and Nicola Zannone, "Towards the development of privacy-aware systems", *Information and Software Technology, Butterworth-Heinemann*, pp.337-350. 2009.

[5] JiWon Byun, Elisa Bertino, and Ninghui Li, "Purpose based access control of complex data for privacy protection", *Symposium on Access Control Models and Technologies, ACM*, pp.102-110, 2005.

[6] Stefan Sackmann, Jens Struker, and Rafael Accorsi, "Personalization in privacy-aware highly dynamic systems", *Communications of the ACM*, pp.32-38, 2006.

[7] E.Bertino, J.-W.Byun and N.Li, "Privacy-preserving database systems", *in: FOSAD*, vol.3655, pp.178-206, 2005.

[8] A. Tumer, A. Dogac and H. Toroslu, "A Semantic based privacy framework for web services", in *Proc. of ESSW'03*, 2003.

[9] Peter Bodorik, Dawn Jutla and Mike Xuehai Wang, "Consistent privacy preferences (CPP): model, semantics, and properties", *Symposium on Applied Computing, ACM New York*, pp.2368-2375, 2008.

[10] D.M. Eyers, J. Bacon and K. Moody, "OASIS role-based access control for electronic health records," *Software, IEE Proceedings*, vol. 153, issue. 1, pp.16-23, 2006.

[11] Thomas C. Rindfleisch, "Privacy, Information Technology, and Health Care", *Commun. ACM*, vol. 40, no. 8, pp. 92-100, 1997.

[12] Divakaran Liginlal, Inkook Sim and Lara Khansa, "How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management", *Computers & Security*, vol.28, pp.215-228.

[13] Westin A.F., Privacy and Freedom, Atheneum, Newyork, 1967.

[14] Calvin S.Powers, Paul Ashley and Matthias Schunter, "Privacy Promises, Access Control, and Privacy Management," *IEEE Computer Society*, pp13, 2002.

[15] Maxwell J.Mehlman, J.D., "Emerging Issues: The Privacy of Medical Records," Article (CrossRef Link)

[16] Electronic medical record http://en.wikipedia.org/wiki/Electronic_medical_record#Privacy (Accessed: 3 Sep,2009).

[17] Sandhu, R., Coyne, E.J., Feinstein H.L. and Youman, C.E. (August 1996), Role-Based Access Control Models, *IEEE Computer*, vol. 29, no. 2, pp. 38-47.

[18] James M. Humber and Robert F. Almeder, "Privacy and health care(Biomedical Ethics Reviews)", *Humana Press*, 2001.

[19] Janet Colwell, EHR era ushers in stricter privacy, security, from the April ACP Internist, copyright at 2010 by the American College of PhysiciansApril, Article (CrossRef Link) (Accessed: 23 May, 2010).

[20] Q. He., "Privacy enforcement with an extended role-based access control model. NCSU Computer Science Technical Report" TR-2003-09, Feb 2003.

[21] Ji-Won Byun, Ninghui Li, "Purpose Based Access Control for Privacy Protection in Relational Database," Springer Berlin/Heidelberg, vol. 17, no. 4, pp.603-619, Jul 2008,.

[22] Fabio Massacci, John P Mylopoulos and Nicola Zannone, "Hierarchical hippocratic databases with mnimal disclosure for virtual organizations," Springer-Verlag New York, Inc. Secaucus, NJ, USA, pp.370-387.

[23] Georgios V. Lioudakis, Eleftherios A. Koutsoloukas, Nikolaos L. Dellas, Nikolaos Tselikas, Sofia Kapellaki,

George N. Prezerakos, Dimitra I. Kaklamani, and Iakovos S. Venieris, "A middleware architecture for privacy protection," *Computer Networks:The International Journal of Computer and Telecommunications Networking*, Elsevier North-Holland, Inc, pp.4679-4696, 2007

[24] M.Hilty, D.A.Basin, A.Pretschner, On obligations, in *Proc. of ESORICS'05*, vol. 3679, pp.98-117, 2005.

[25] David F. Ferraiolo, Ravi S. Sandhu, Serban Gavrila, D. Richard Kuhn, and Ramaswamy Chandramouli. "Proposed NIST standard for role-based access control," *ACM Transactions on Information and Systems Security*, vol. 4, no. 3, pp. 224-274.

[26] L.Sweeney, "k-anonymity: a model for protecting privacy, International Journal on Uncertainty," *Fuzziness and Knowledge-based Systems*, vol. 10, no. 5, pp.557-570, 2002.

[27] D.Ferraiolo and R.Kuhn, "Role-Based Access Controls," in *Proc. 15th NIAR-NCSC Nat'l Computer Security Conf., Nat'l Inst.Standards and Technology*, pp.554-563, 1992.

[28] Bresciani, P., Giorgini, P., Giunchiglia, F., Mylopoulos, J., Perini, A., "TROPOS: An agent-oriented software development methodology." *JAAMAS*, vol. 8, no.3, pp.203-236, 2004

[29] Mor Peleg, Dizza Beimel, Dov Dori and Yaron Denekamp, "Situation-Based Access Control: Privacy management via modeling of patient data access scenarios", *Journal of Biomedical Informatics*, vol. 41, pp. 1028-1040, 2008.

[30] Elisa Bertino, "RBAC models-concepts and trends", *Computers \& Security*, vol. 22, Issue. 6, pp.511-514, Sep 2003,.

[31] R. Sandhu, ``Role Hierarchies and Constraints for Lattice-based Access Controls", in E. Bertino, H. Kurth, G. Martella, and E. Montolivo Eds., *Computer Security- Esorics'96*, LNCS N.1146, pp.65-79.

[32] Stephen S. Yau, Yin Yin, "a privacy preserving repository for data integration across data sharing services," *IEEE Transactions on Services Computing*, vol. 1, Issue 3, pp.130-140, Jul 2008.

[33] Yuan Tian, Biao Song, Eui-Nam Huh, "Relationship based privacy management for ubiquitous society," *ICCSA*, vol. 1, pp.853-867, 2009:.

[34] Yuan Tian, Biao Song, Eui-Nam Huh, "A novel graph-based privacy policy management system, management and service science," *International Conference*, pp.1-4, 2009Article (CrossRef Link)

[35] Josep Domingo-Ferrer, Yucel Saygin, "ecent progress in database privacy", *Date \& Knowledge Engineering* vol. 68, 1157-1159, 2009.

[36] Yuan Tian, Biao Song, Eui-Nam Huh, "A Purpose-based Privacy-aware System using Privacy Data Graph"

[37] Ontologycomparision, Article (CrossRef Link) (Accessed: 2 Sep, 2010)

[38] Andreas Pfitzmann and Marit Koehntopp. Anonymity, unobservability, and pseudonymity -a proposal for terminology. In Hannes Federrath, editor, Proceedings Workshop on Design, Issues in Anonymity and Unobservability, volume LNCS 2009. Springer Verlag, 2001.

[39] Marc Langheinrich, Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems, Lecture Notes In Computer Science; Vol. 2201, Proceedings of the 3rd international conference on Ubiquitous Computing, Atlanta, Georgia, USA, pp.273-291.

[40] Fair Information Practices, http://whatis.techtarget.com/definition/0,sid9_gci213501,00.html (Accessed: 3 Dec,2009).

[41] US Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computers and the Rights of Citizens, Chapter IV: Recommended Safeguards for Administrative Personal Data Systems (1973).

[42] A Review of the Fair Information Principles: The Foundation of Privacy Public PolicyArticle (CrossRef Link), l (Accessed: 3 Dec,2009).

[43] Lorenzo D. Martino, Qun Ni, Dan Lin and Elisa Bertino. "Multi-domain and Privacy-aware Role Based Access Control in eHealth." *In the International Conference on Pervasive Computing Technologies for Healthcare*, Jan 2008.

[44] Reid, Jason F. and Cheong, Ian and Henricksen, Matthew P. and Smith, Jason "A Novel Use of RBAC to Protect Privacy in Distributed Health Care Information System"s. *in 8th Australasian Conference on Information Security and Privacy*, Jul Wollongong..

[45] COPPA Safe Harbors discussed, Cybertelecom Federal Internet Law \& Policy - an Educational Project. Krohn & Moss Consumer Law Center, Article (CrossRef Link) (Accessed: 22 Dec,2009).

[46] EHR, Article (CrossRef Link) (Accessed: 3 Mar, 2011)

[47] Lorrie Faith Cranor, Praveen Guduru, Manjula Arjula, User Interfaces for Privacy Agents, ACM Transactions on Computer-Human Interaction, Vol.13, No.2, June 2006, pp. 135-178.

[48] Scott Lederer, Jennifer Mankoff, Anind K. Dey, Who wants to know what when? privacy preference determinants in ubiquitous computing, Conference on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA, 2003, pp. 724 - 725.

[49] Robert W. Proctor, Kim-Phuong L. Vu, and M. Athar Ali, Usability of User Agents for Privacy-Preference Specification, Human Inerface, Part 2, HCII2007, LNCS 4558, Springer Berlin/Heidelberg, 2007, pp.

766-776.
[50] Qun Ni, Elisa Bertino and Jorge Lobo. Privacy-aware RBAC - Leveraging RBAC for Privacy, IEEE Security \& Privacy Magazine, Volume 7, Number 4, pp. 35-43, July/August 2009.
[51] Semantic Web Applications in Neuromedicine, Article (CrossRef Link).

Appendix

In this algorithm, the input contains a privacy graph D and one or many disclosed data element stored in V. Graph traversal process is used in this algorithm. Searching starts the vertices representing disclosed data elements. The output of this algorithm contains a set of data elements which will be potentially disclosed. Pseudo code and annotation is clearly presented as follow.

```
1. procedure personal preference converting algorithm( D, V )
2. Create set C = φ
3.              ▷ C stores potentially disclosed data elements
4. foreach v_i in D              ▷ Initialize D
5. {
6.    w(v_i) = 0
7. }
8. foreach v_i in V              ▷ Mark disclosed data elements
9. {
10.   find corresponding w(v_i) in D
11.   w(v_i) = 1
12.}
13.repeat :
14.   foreach w(v_i) = 1
15.   {
16.      foreach a_j = (v_i, v_k) & w(v_k) = 0 & w(a_j) == 1
17.      {
18.         C = C ∪ v_k
19.         w(v_k) = 1          ▷ v_k will be potentially disclosed
20.      }
21.      foreach a_j = (v_i, v_k) & w(v_k) = 0 & w(a_j) == 0
22.      {
23.         if (the anonimity requirement of a_j is not filled)
24.         {
25.            C = C ∪ v_k
26.            w(v_k) = 1        ▷ v_k will be potentially disclosed
27.         }
28.      }
29.   }
30.until : C will not be changed
31.return C
32.end procedure
```

Yuan Tian received her B.S. from the Department of Computer Sciences, Hebei Normal University in 2008. Since 2008, she has been working on her Master's degree and currently is a Master's candidate in the Department of Computer Engineering at Kyung Hee University, Korea. Her research interests include privacy, security, grid and cloud computing.

Biao Song received his B.S. from the Department of Computer Sciences, Hebei Normal University in

2008. He is currently working toward his Ph.D. degree in the Department of Computer Engineering at Kyung Hee University, Korea. His research interests include task co-allocation, dynamic collaboration in cloud computing, remote display protocol in thin client environments and IPTV service delivery over virtual networks.

Mohammad Mehedi Hassan received his B.Sc. degree in Computer Science and Informaion Technology from the Islamic University of Technology, Dhaka, Bangladesh in 2003. He received his Ph.D. degree in Computer Engineering from Kyung Hee University, South Korea in 2010. He was a Research Professor at the Computer Engineering Department, Kyung Hee University, South Korea from March, 2011 to October, 2011. Currently he is with King Saud University, Kingdom of Saudi Arabia as an Assistant Professor and is the Chair of Pervasive and Mobile Computing, in the College of Computer and Information Science. His current research interests are cloud computing, data intensive computing, media clouds, mobile clouds, game theory, dynamic VM resource allocation, IPTV, virtual networks, sensor networks and publish/subscribe systems.

Eui-Nam Huh earned his B.S. degree from Busan National University in Korea, his Master's degree in Computer Science from the University of Texas, USA in 1995 and his Ph.D. degree from Ohio University, USA in 2002. He also served the WPDRTS/IPDPS community as the program chair in 2003. He has been an editor of the Journal of Korean Society for Internet Information and been a Korea Grid Standard group chair since 2002. He was also an Assistant Professor in Seoul Women's University, South Korea. He is now with Kyung Hee University, South Korea as a Professor in the Dept. of Computer Engineering. His areas of research interest are cloud computing, ubiquitous computing, high performance networks, sensor networks, distributed real rime systems, grids, and network security.