Journal of Web Engineering, Vol. 14, No.3&4 (2015) 346-360


# CREDENTIAL PURPOSE-BASED ACCESS CONTROL FOR PERSONAL DATA PROTECTION

NORJIHAN ABDUL GHANI
*Faculty of Computer Science and Information Technology,*
*University of Malaya, 50603 Kuala Lumpur, Malaysia*
*norjihan@um.edu.my*

HARIHODIN SELAMAT ZAILANI MOHAMED SIDEK
*Advanced Informatics School, Technology, University of Malaysia, Jalan Semarak,*
*54100 Kuala Lumpur, Malaysia*
*harihodin@ic.utm.my, zailani@ic.utm.my*

Received April 20, 2014
Revised April 6, 2015

Web-based applications enable users to carry out their business transactions virtually at any time and place. They require users to disclose almost all their personal information. Organizations on the other hand will collect, process, and store a huge amount of this information, which results in a greater risk of information disclosure. Enforcing personal information protection in databases requires controlled access to systems and resources and is only granted to authorized users. Previous research on purpose-based access control does not fully support personal data protection, especially users' rights and less user participation towards their personal data once it is released via web applications. This paper formulates a solution to control access while ensuring that personal data is protected and that users have full control over their own data. This model, which implements two-phase security involving user authentication using personal credential and data authorization based on purpose, is presented. The purpose of this model is to protect personal information that has been collected via web-based applications by using data access control.

*Key words*: Wireless computing, specification, requirements, acceptance criteria
*Communicated by*: B. White & M. Gaedke

## 1 Introduction

The existence and the rapid development of the World Wide Web (WWW) on the Internet have literally transformed people's lives in recent years. These applications provide the capabilities to collect and store many types of personal information related to individuals in the course of the business activities. Personal information is collected, stored and used in various types of information systems, therefore, privacy protection for these types of information, especially in this environment, is a major concern. Enforcing data protection therefore requires that every access to a database must be controlled by ensuring that only authorized access can take place. This can be done through the process of access

control. Access control is the process of mediating every request to resources and data maintained by a system, and determining whether the request should be granted or denied [6]. A fundamental component in enforcing privacy and data protection represented by the access control is to control all access to a system and ensure that all and only authorized access can take place [15]. Due to the importance and crucial needs of privacy protection, there is a demand for privacy aware access control in protecting personal data stored in a database.

The problem arises regarding the expansion of the user population in these types of application. Web-based applications operate in a more complex and open environment system where traditional access control mechanisms based on user or login name and password for qualifying the subjects are no longer appropriate [7]. To support this, Bertino and Sandhu in [7] suggested using a flexible user specification and scalable access control mechanisms through which user authorization is based on user attributes (e.g. user credential). In recent years, much work has been conducted to introduce a privacy aware access control in order to support the data privacy requirements. Most of the work showed the importance of purpose as a basic requirement for developing an access control for specifying a privacy policy. The notion of purpose was introduced in [1] with the concept of the Hippocratic Database. Kabir and Wang in [16] pointed out that the first reasons that privacy protection cannot be easily achieved by traditional access control is that traditional access control, such as RBAC, focuses on which user is performing which action on which data object, whereas a reliable access control for protecting personal data concerns which data object is used for what purpose. Instead of purpose, the data subject participation principle also becomes an important factor. Users should be provided with the right to their own personal data stored in databases. Right should enable the owner (also known as data subject) of the data to access their own information regardless of what purpose [2]. They should be able to access and participate in whatever data belongs to them stored in a database. In [5], Bertino *et al.* presented an access control as one of the available approaches in protecting personal data stored in the databases. According to Chauduri *et al.* in [12], through access control, the system can restrict the access to authorized users only and can guarantee the protection of the data object.

The main contribution of this paper is to propose a Credential Purpose-Based Access control (CrePBAC) model. It presents an appropriate mechanism in controlling the access in order to protect personal data stored in an open database from unauthorized access via web-based applications. Various aspects of data security and privacy with special emphasis on mechanisms based on access control in protecting personal data stored in open databases are considered. The rest of the paper is organized as follows. Section 2 briefly overviews some related works in this area. Section 3 presents the proposed CrePBAC model, which also overviews the basic components involved in the CrePBAC model, meanwhile Section 4 presents the CrePBAC model. Section 5 discusses how access decisions are determined and the implementation of the CrePBAC. We compare our proposed model with the closest previous models in Section 6. Section 7 concludes the paper.

## 2 Related Works

This work is related to several topics in the area of privacy and security of data management, especially in managing and protecting personal data. We also exploit the tremendous work carried out for data subject participation and purpose issues, which mainly focuses on the secure management of data.

The concept of the Hippocratic Database, which was introduced in [1], proposes that the databases should include privacy protection as a central concern. The notion of purpose has been introduced in Hippocratic Database and much work has been done to extend this work. The Hippocratic Database includes privacy policies and authorization associated with each attribute and each user's purpose [2]. Strawman architecture is proposed, in which access control is based on purpose, and privacy metadata is used that consists of two tables that are referred to as privacy policies and privacy authorizations tables. In [19], LeFevre *et al.* presented an approach to enforce the privacy policy at the database level. Query modification was used as a way to implement this approach.

Inspired by the concept of purpose, as introduced in the Hippocratic Database, many researchers have extended the access control based on purpose, such as [9], [16], [18], [20], [21] and [23]. Most of these works focused on the concept of the purpose in determining which data are accessed by users. In [10], Byun and Li introduced access control based on purpose in which an appropriate metadata-model must be developed in order to support such privacy protection access control. In their work [9], Byun *et al.* sought compliance between the intended purpose defined for data and the access purpose requested by the user at the runtime. Unfortunately, the work of Byun et al. is only preliminary and still requires other researchers to extend their work. In order to fulfil the privacy protection for users, every data access must obey the privacy policies on which users have conditionally or unconditionally agreed [10]. This approach provides more sophisticated concepts of the purposes, which are organized in a hierarchy.

Kabir and Wang in [17] extended the access control based on purpose with Conditional Purpose Based Access Control. This model introduced a variety of purposes in which conditional purpose is applied together along with allowed purpose and prohibited purpose in determining the access permission. This model extended the work done by [1], [9] and [10]. The work by Kabir and Wang in [7] identified that a key feature of this model compared to the Basic PBAC is that it supports conditional purpose and prohibited purpose, thus allowing users to specify the data that should be used conditionally or should not be used for a set of purposes. A conditional purpose is introduced in addition to explicit prohibitions that make data providers more flexible in giving information. Kabir and Bertino in [18] extended their work adding the RBAC in their access control model. It presented a CPBAC [16] and injected it with RBAC, which is referred to as a RPAC model. This model enables organizations to operate as a reliable keeper of their customers' data. The model is useful for internal access control within an organization as well as for information sharing between organizations, as many systems are already using RBAC mechanisms for the management of access permission.

Purpose-Aware Role-Based Access Control (PURBAC), which was introduced in [20], was an extended version of the role-based access control model to capture the privacy requirements of an organization. This access control extends the RBAC with purposes as a central entity in RBAC where the assigning of permission to roles is based on the purpose related to privacy policies. However, this model assigns a purpose as a separate entity in defining the permission. Sun and Wang proposed the Purpose Based Usage Access Control Model. The authorization rule permits or denies the access of a subject to an object based on subject and object attributes. The key feature of this approach is that an access decision is not only based on decision factors, such as authorizations, obligations and conditions, but also the continuity properties (ongoing authorization).

All of these works proposed different approaches to protect the privacy of individuals through different models. However, it does not consider the type of operating environment, whereas our aim is to protect the personal information that is being collected and disclosed via the open environment. Although purpose-based access control models focused on the concept of purpose, which is closely related to ours, there are some differences to our approach. While we are considering the security and privacy of personal data that is operating in an open environment, we have to take into account several circumstances of this environment. We believe that purpose alone is not suitable in today's era of open environment. Enforcing the privacy of personal information requires that every access through an open system must be controlled, and that only authorized access can take place. This work was motivated by the large number of users who are trying to access the personal information via online applications.

## 3 Proposed CrePBAC

This section presents a two-phase security of access control model for the purpose of protecting the personal data stored in open databases. Section 3.1 identifies five requirements for the development the CrePBAC in protecting the personal data stored in open databases. Six basic components of the CrePBAC model have been derived from these requirements and the explanations of the components are also given in Section 3.2.

### *3.1. Requirements for the Development of Credential Purpose-Based Access Control*

Demands for personal data protection technology are stronger than before. It is crucial that the collected data, especially personal data, must be enforced with privacy policies with the information systems managing them. The emphasis in this access control has been primarily on determining, specifying, maintaining and enforcing policies for controlling access to the personal data inside these databases. This type of access control is needed to ensure the protection of personal data from being accessed by unauthorized users, which requires the development of access control. To further develop this access control system, a set of requirements of the CrePBAC model has been outlined.

*R#1: Access Control based on Purpose*

In most recent research concerning the protection of personal data and its privacy, the notion of purpose has been widely used as the base for controlling access in achieving the personal data protection [9], [16] and [20]. All data collected must be enforced with the purpose of the collection and for what purpose it will be used. In addition, every access request sent by users must be accompanied with the purpose for which the data is accessed.

*R#2: Flexible User Specification based on Credential*

Dagdee and Vijaywargiya, in [14], stated that the use of a traditional identity mechanism should be replaced with a more flexible user specification. In addition, Bertino and Sandhu in [7] proposed that the use of user attributes is more appropriate in authenticating the user rather than using an identity mechanism, such as login and user names. This research identifies the use of a flexible user specification based on the user's credentials when authenticating the user.

*R#3: Support for Rich Privacy Related Metadata*

The research should be able to provide a comprehensive and accurate privacy-related metadata. According to [4], metadata represent the core of access control mechanisms specifically tailored towards policy. The privacy-specific metadata should be associated with the data, stored in the database together with the data, and sent with the data whenever the data flow to the other parties in the system. HDB, which has been introduced in [1], is an example of a database system that implements the privacy-related metadata.

*R#4: Support for Users' Rights*

According to Barker in [3], users should have the rights to their own personal data and they should at least know for what purpose their data will be used. The requirement defines the ability to support data subject to the rights to their personal data stored in the open databases.

*R#5: Support for Open Databases in Open Environments*

In this research, personal data in open databases are accessible from web-based applications. As a result, an access control mechanism is required to control access towards personal data stored in open databases, which usually operates in an open environment, such as web-based applications.

These five requirements are important in designing and implementing the CrePBAC access control system. The following section continues the discussion regarding CrePBAC components in which these five requirements are taken into account.

### *3.2. Specification of Access Control Model Components*

This section explains in detail the components of the CrePBAC model. The discussion starts with the introduction of six components involved in the CrePBAC model.

a. Users
   Users refer to an individual who owns the data, discloses it through web-based applications and stores it in the organization's databases. Protecting the personal data requires the user to have more control and rights towards their personal information on web-based applications.
b. Personal Credentials

   The specification of authentication is not only based on user identity but also on the user characteristics, in which each user is associated with one or more credentials. According to Camenisch *et al.* in [10], one of the credential features is when it can be used as proof of ownership, by binding a credential to its legitimate owner authentication information can be attached and evaluated to the credentials.

**Definition 1:** *Personal credential: Any personal attribute belonging to a specific person. A set of personal credentials denoted as PC with* $pc_1, pc_2, \cdots, pc_n$ *is an attribute of PC, which is represented as:* $\{pc_1, pc_2, \cdots, pc_n\} \in PC$.

*Credential-types of Personal Credential*

Personal credentials are organized by type, in which it identifies the properties of the personal credential [26]. The credential-type (CT) is used for a better and easier specification in which personal credentials with similar structures are grouped together. In [10], Camenisch *et al.* specified that credentials must be of a certain type that determines the attributes contained in the credential. Then, the policy specifies the credential type that must be used to satisfy the policy for authenticating users.

Figure 1 shows an example of the credential-type hierarchy that exists for the *customer*, which is classified in two categories – non-personal and personal. Non-personal is a credential-type that cannot be used to identify a person while personal is a credential-type that can be used to identify a person.

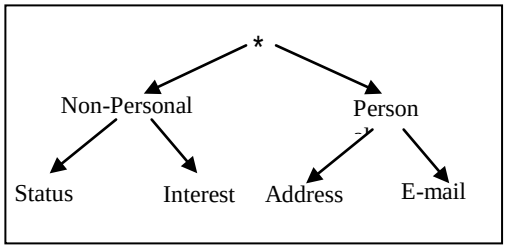


Figure 1 Credential-type Hierarchy for *customer*

A hierarchy H of credential types is a subcredential of $(CT, <_{sc})$, where *CT* is a subcredential of all types and $<_{sc}$ is a subcredential of *CT*. Two credential types $ct_1, ct_2 \in CT$, where $ct_2$ is a subcredential of $ct_1$ if and only if $ct_2 < CTct_1$. Figure 1 shows that address is a subcredential of Personal. A hierarchy *H* has a unique root, which is denoted as *, hence $ct_2 <_{sc} *$ for each $ct_1 \in CT$.

*Personal Credential Properties*

Personal credential *myICnum* and *myEmail* are instances of *ICNum* and *E-mail*. Each credential instance (*CI*) is characterized by a unique identifier, credential owner, credential type and a set of properties $\{pc_1, pc_2, \cdots, pc_n\}$. It can be denoted as:

$$(c\_id,\ c\_owner, \{pc_1, pc_2, \cdots, pc_n\}, c\_type)$$

Each credential property (*CP*) is characterized by attribute names and its value. It states that $(att1 : val1,\ att2 : val2, \cdots, attn : val\,n)$, where $att1, att2, \cdots, attn \in A(c\_id)$ are the names of the attributes of *CP* and $(val1,\ val2, \cdots val\,n) \in V$ are their values.

c. Purpose

In web-based applications, personal data are collected for a specific usage purpose. Since an organization's privacy policy mainly concerns which data object is used for which purpose, the access requests are made for specific purposes. The purpose describes the reasons for collection and data access [9]. It represents how personal data will be used by the user. In [1], the notion of purpose is defined as a basic concept upon which decisions to access personal information are made. For example, in online shopping web-based applications, *customer's address* is used for

*Delivery* and *Marketing* purposes. Purposes naturally have a hierarchical relationship among them, such as generalization and specialization relationships, which we refer to as a purpose tree.

*Purpose and Purpose Tree*

For preserving the privacy of users, each and every data access must adhere to the privacy policies on which the users have agreed. Data access requests by the users are made for a specific data usage purpose or purposes. This represents how the data is going to be used by the user itself.

**Definition 2:** (see [9]) Purpose and Purpose Tree. *A purpose is defined as a reason for data collection or data access. A set of purposes (P) is organized in a tree structure, referred to as a purpose tree (PT). The purpose directly dictates how accesses to data objects should be controlled*.

Each node represents a purpose in P and each edge represents a hierarchical relation between two purposes. For example, data usage purpose *Tele-marketing* is a specialization of purpose *Marketing*, as shown in Figure 2.

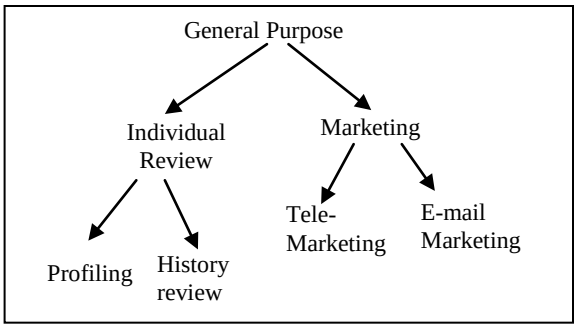


Figure 2 An Example of a Purpose Tree

There are two types of purpose: access purpose and data usage purpose. The privacy policy is to ensure that data can only be accessed for its data usage purpose, and the access purpose should be in compliance with the data usage purpose.

**Definition 3:** (see [9]) Access Purpose. *An access purpose is defined as a purpose for accessing data, which is determined or validated by the system when data access is requested.*

**Definition 4:** (see [9]) Data usage purpose. *A data usage purpose is defined as specified usages for which data objects are accessed. That is, the purpose is associated with the data and thus regulates data access as data usage purpose.*

Thus, any access decision is made based on the relationship between the access purpose and the data usage purpose. The data usage purpose can be defined as allowable usage purpose and prohibited usage purpose [9]. The allowable data usage purpose is the access request granted for a particular purpose while the prohibiting usage purpose is the access request strictly not granted for a particular usage purpose. That is, an access request is granted if the access purpose is entailed by the allowable usage purpose but not entailed with the prohibited usage purpose.

Let *PT* be a purpose tree with *P* as a set of the purposes in *PT*. We denote *AP* for access purpose and *UP* for data usage purpose and *UP* is a tuple $\langle aup, pup \rangle$, where $aup, pup \subseteq P$ are two sets of purposes. The *aup* refers to the allowable data usage purposes and *pup* represents prohibited data usage purposes. An access is only allowed if $P = aup$ and the access is denied if $AP = pup$.

d. Actions
In [8], Braghin *et al.* defined actions as rights that users can perform when accessing personal data through web-based applications. However, this paper only considers two actions: *select* and *update*.

e. Objects
The data represents the information referring to users that can be processed by the system [7]. The objects involved in this research are personal data stored in the databases.

f. Access decision evaluation

Access decision evaluation is based on the relationship between a-two phase security: user authentication and data authorization. User authentication is an association between *user* $\rightarrow$ *personal credential* while authorization data is an association between *access purpose* $\rightarrow$ *usage purpose*. The notation $\rightarrow$ explains, for example the personal credential $pc_1$ belongs to user $u_1$. It refers to the required credential that should be provided by the user before being verified as an authenticated user.

## 4 CrePBAC Model

CrePBAC is designed to satisfy the need for simplifying the access control management and directly presenting access control policies. The key concepts of the CrePBAC access control model introduced in this paper are:

a. *personal credential*, which represents personal attributes that belong to data users, and
b. *purpose*, which represents the reason for the data being accessed or used.

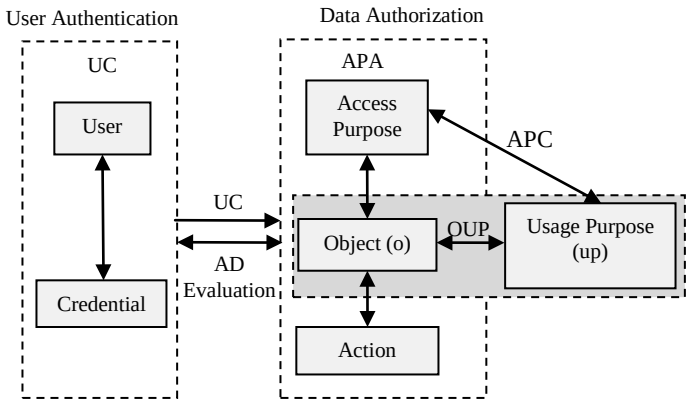


Figure 3 CrePBAC Access Control Model

Like other PBAC, the CrePBAC model also uses purposes in extending the model. In CrePBAC, users who are successfully authenticated through personal credential are not guaranteed access to the personal data. In addition, gaining the access purpose permission does not mean that the users are directly performing operations on request objects. This is because all personal data are dynamically bound with the respective usage purposes according to the equivalent privacy metadata. Figure 3 shows the CrePBAC access control model with six components as well as the interaction between these components, and explains how the access decision is evaluated and determined.

Based on the CrePBAC access control model above, definitions of its associated components have been formalized. The definitions refer to the components of the model and how it is used in acquiring the final access decision.

a. User, Credential, Object, Purpose, Action and Access Decision represent the set of user, personal credential, object, purpose, operation and access decision evaluation.
b. $UP = \{(aup, aup) \mid aup, pup \subseteq P\}$ is the set object's usage purpose, where *aup* indicates the object's allowable usage purpose and *pup* represents the object's prohibited usage purpose.
c. $OUP = \{(o, up) \mid o \in Object, up \in UP\}$ is the set of data object with its usage purpose.
d. $UC = \{(u, c) \mid u \in User, c \in Credential\}$ is the set of users with its credential.
e. $APA = \{(o, a, ap) \mid o \in Object, a \in Action, ap \in Purpose\}$ is the set of access purpose allowable for an action on data object.
f. Access Purpose Compliance is defined as a mapping between access purpose and usage purpose for the data object, i.e. $APC \subseteq AP \times UP$.
g. Access Decision $AD \subseteq UC \times APA$ is a many to many mapping between the User with their Credential and their access allowable purpose. It determines that the action that a certain user (with trusted credential) performs on an object is based on a certain access purpose

Besides the formal definition above, we can also define the set of functions to facilitate the CrePBAC model as below:

a. User_Credential_Compliance $(u, c)$ = TRUE if $u \in$ User and $c \in$ Credential
b. Access_Purpose_Compliance $AP \times UP \rightarrow \{\text{TRUE, FALSE}\}$ is used to determine the compliance between access purpose and data object's usage purpose.
   Purpose_compliance $(aup, pup)$=TRUE if $ap \in aup$ and $ap \notin pup$ and $AP \rightarrow_{PT} UP$.
c. Access_purpose_authorization $UC \rightarrow APA$ is a mapping of user (with authenticated credential) onto access purpose allowable
   access_purpose_authorization $(uc) = \{apa \in APA \mid uc, apa \in AD\}$.
d. Access_ decision $(uc, apa)$ = TRUE if $AD(uc \in UC \wedge apa \in APA)$.

*4.1. User Authentication and Data Authorization*

This model introduces a two-phase security whereby the users are required to fulfil both phases; user authentication $(UC)$ and data authorization $(APA)$ before access is given to that user. We define function for user authentication, as follows:

Let $(C)$ is a set of *Credential* and $(U)$ is a set of *User* where:

$$UC = \{(u,c) \mid u \in User,\ c \in Credential\}$$

$(C)$ is compliance with $(U)$ if:

$$assigned\_credential(c) = \{c \in C \mid \langle u,c \rangle UC\} \text{ is TRUE}$$

If the result is TRUE, it indicates that the users who sent the access request are authenticated and it may proceed to the second phase of the security mechanism, data authorization.

Data authorization phase is only applicable for authenticated users. Before granting the user with the authority to access personal data stored in open databases, first it must accomplish the access purpose authorization.

Access purpose authorization, $APA$, is only authorized for users that are verified as authenticated users. Furthermore, an access purpose must be in compliance with the allowable usage purpose, but not with the prohibited usage purpose. The access is granted if and only if both conditions are fulfilled. The purpose compliance between $ap$ and $up$ is represented as:

$$\text{Access\_Purpose\_Compliance}(aup, pup) = \text{TRUE if } ap \in aup \text{ and } ap \notin pup \text{ and } AP \rightarrow_{PT} UP\ .$$

$ap$ is compliant to $up$ if the following conditions are satisfied:

i. i. $ap \in aup$

ii. ii. $ap \notin pup$

$AP \rightarrow_{PT} UP$ means condition $_{PT}UP$ is a necessary conditions for $ap$. Otherwise, $ap$ is not compliance with $up$. For example, if $up$ = ({Individual review},{Purchase}). If $ap$ = Purchase_History, then $AP \rightarrow_{PT} UP$. But if the $ap$ = Tele-Marketing, then $AP \rightarrow_{PT} UP$.

The access purpose authorization is made based on the relationship between access purpose and its object (with usage purpose) and operation towards the data. It must be fulfilled before the access is granted. The access purpose authorization refers to

$$APA = \{(o,a,ap) \mid o \in Object,\ a \in Action,\ ap \in Purpose\}$$

is the set of access purposes allowable for operation on the data object (with its usage purpose).

*4.2. CrePBAC Access Decision Determination*

The access purpose authorization then checks the relationship between the UC and APA, which we denote as:

$$\text{access\_purpose\_authorization}\ (uc) = \{apa \in APA | (uc, apa) \in AD\}.$$

Access decision (AD) is the relationship between authenticated users with its data access purpose authorization:

$$\text{access\_ decision}\ (uc, apa) = \text{TRUE if}\ AD(uc \in UC \wedge apa \in APA).$$

From the above function, the access decision must fulfil both the user authentication and the authorization data. It is defined that an access decision is granted if it successfully satisfies both phases in the CrePBAC.

## 5 CrePBAC Access Decision

The CrePBAC model is implemented as a two-phase security model, comprising two levels of security; authenticating user and authorizing data access. It was implemented using a query modification algorithm, as discussed in the next section.

*5.1. CrePBAC Query Modification Algorithm*

The use of a query modification in implementing the CrePBAC mechanism is important in order to protect the personal data from unauthorized access. In this CrePBAC implementation, the query modification approach is adopted when users used to rewrite queries so that the database only returns the personal data for which the user is authorized [19]. Our CrePBAC query modification algorithm reflects the CrePBAC model discussed in the previous section. The query modification must be in compliance with all the conditions before the access is granted. This algorithm filters out any personal data, and then the decision will be made whether the access is allowed or prohibited with respect to the purpose of access.

CrePBAC query modification algorithm is outlined in Figure 4. The query modification must be in compliance with two-phases before the access is granted. The CrePBAC query modification algorithm is implemented in two levels:

i. User authentication: identifying and verifying users before authenticating to the second level, that is
ii. Data authorization: check and give the authorization irrespective of whether or not the personal data's request for a specific action is in compliance with the access purpose.

The CrePBAC query modification algorithm, which starts at line 4 to line 8, illustrates the first phase of the model; user authentication $(UC)$ as outlined in CrePBAC model in Figure 3. The user $(U)$ must enter the required credential $(C)$ and the credential must be in compliance with the credential $(U)$. If both credentials are in compliance with the user, the user is known as an authenticated user and is able to proceed to the second phase, data authorization. Line 10 to line 15 show how the data authorization phase checks whether the authenticated user should be given access

to data or not, depending on the first condition, which is stated in line 10. As explained in Figure 3, the access purpose must be in compliance with the usage purpose as:

Purpose_compliance $(aup, pup)$=TRUE if $ap \in aup$ and $ap \notin pup$ and $AP \rightarrow_{PT} UP$ .

```
       Input : Username and password
// Verify the compliance between username and password
2      If failed, ACCESS DENIED
3      endif
// Verify the User Authentication
4      Input : User u need credential c
5      If UA(UC(u, c) = False
// The process of user authentication is not successful
6      ACCESS DENIED
7      Else if
8      If UA(UC(u, c) = True
// go to the next step
9      Input : User tagged the access purpose ap
// Verify the compliance between ap and up
10     If ap ∈ aip and ap ∉ pip then
// go to the next step
11     Else
12     ACCESS DENIED
// verify the Data Authorization
13     Let R be the relation
14     Let A_i be the attributes of relation R
15     Let O be the operation
16     If DA(Ai, R, O) = False
17     ACCESS DENIED
18     Else
19     ACCESS GRANTED
// Return the query result
Output : List of attributes
```

user authentication

data authorization

Figure 4 CrePBAC Query Modification Algorithm

Then, once both purposes are in compliance with each other, it will then check for data authorization, as shown in line 13 to line 19. The algorithm checks whether the authenticated user $(UC)$ is trying to access an object with an allowable action, as shown in Figure 3, which can be achieved through:

$$APA = \{(o, a, ap) \mid o \in Object, a \in Action, ap \in Purpose\}$$

Later, an access decision as to whether to grant or deny is decided at this level.

## 6 Comparisons

There are some related works on privacy protection access control models discussed in the literature. The closest works related to this article are [1], [9], [13] and [14]. This section will provide a comparison between these models and our proposed approach.

Recently, in [8], Byun *et al.* introduced a purpose based access control, which focused on how to determine the purpose for which certain data are accessed by a given user. While [18] introduced an authorization based on role and purpose. Their proposed solution relies on the well-known control RBAC model, as well as the notion of conditional role, which is based on the notions of the role attribute and system attribute. It supports data access control based on the purpose information. Besides this work, more and more research works have been done proposing a purpose as a base for access control, as discussed in Section 2. However, this work substantially differs from those proposals. Within the context of web-based systems, the limitation of this approach is when it only provides the authorization based on purpose, and, specifically, on roles. Firstly, their approach is based on the notion of the purpose and the role. On the other hand, this approach provides the concept of a personal credential. In this research, the personal credential is used to authenticate the user before authorizing them to access the personal data. In web-based applications, the use of a personal credential is important rather than an identity mechanism, such as username and password. This is necessary to support the features of the web application itself.

Previous work on HDB, which was introduced by Agrawal *et al.* in [1], was designed with privacy metadata that is stored in a database. Privacy metadata consists of two tables – privacy policy and privacy authorizations – stored in two different tables. Hippocratic Databases extend the architecture of standard DBMSs with components that ensure personal data is handled in compliance with its associated privacy definitions. Our approach differs from these HDBs when we apply the Federated Database concept in designing and implementing the HDB. In our approach, HDB acts as a filter database in which the process of sending, checking and verifying the privacy is based on personal credential and purpose, and this authorization, which is based on user parameters, happens transparently. Compared with HDB, besides privacy metadata, we also have credential metadata, so that the privacy checking is not only based on purpose, but also personal credential. HDB implements the privacy checking by giving the user privileges to access personal data by authorization based on purpose; however, our approach has two-phase security, authentication and authorization. HDBs have been proposed as an answer to the privacy requirement and personal data protection by introducing privacy-metadata based on purposes where it defines which data object is used for which purpose. However, this approach enhances the previous metadata by proposing the use of credential metadata together with privacy-metadata based on purpose. The two types of metadata in our approach are used to define which data object is used for which purpose and which data object is allowed for the data subject. As mentioned earlier, our approach of HDB implementation is two phases, which includes two types of metadata:

iii. Credential metadata: credential metadata are used for user authentication
iv. Privacy metadata: privacy metadata store the privacy related metadata and are used for defining the authorization for access.

Kabir et al., in [18], presented a CPBAC [16] and injected it with RBAC, which referred to a RPAC model that enables enterprise to operate as reliable keepers of their customers' data. The model is useful for internal access control within an organization as well as for information sharing between organizations, as many systems are already using RBAC mechanisms for the management of access permission. However, we strongly believe that the use of role is not suitable enough compared to personal credential. In contrast, our proposed model in this paper uses personal credential in giving the

authorization. Again, the proposed model is illustrated with a personal credential to achieve the compliance computation between access purpose and intended purpose.

Credential-based access control (CBAC) in open environment was proposed in [13]. According to them, in an open and dynamic scenario, parties may be unknown to each other and the traditional separation between authentication and access control cannot be applied anymore [15]. This model provides a more flexible user specification, such as user credentials to define access control policy and anyone who possesses the desired credentials is granted access to shared data source. This model does not require central control and allows users to specify their own trust specification. It uses various types of credential, such as identity credential, attribute credential and standard credential. In the proposed system, credentials are used to define access control policy and anyone who possesses the desired credentials is granted access to the shared data resource. However, in our approach, we present that the purpose to access an object is also an important criteria that needs to be considered before granting access to the user, especially in an open environment. We believe that purpose is an important component that needs to be considered as discussed in [9].

## 7 Conclusions and Future Work

Protecting the personal information privacy is important in today's environment. The credential purpose based access control model or CrePBAC is an access control model that is designed based on five requirements that have been identified when proposing an access control to support personal data protection. Access control based on purpose, which is specifically designed to support personal data protection must also consider user's participation in respect of their own data. Hence, instead of purpose, the use of personal credential and the data subject participation is also important in developing the privacy-aware access control.

The CrePBAC model has been successfully implemented using HDB technology via web-based applications. An enhancement of the HDB was done by introducing the credential metadata and data subject's right inside the HDB. Every query submitted must go through two phases of security: user authentication and authorization data. The limitation of this work is when an added security always has a negative effect on performance, it is important to ensure that the system remains reasonably fast. However, performance issues are not within the scope of this research. The basic concept has been explained in this paper. Our future work includes completing the credential purpose based access control system by designing the access control policies and implementing its mechanisms. To improve our current implementation, we plan to extend the work by considering the performance aspect. As the performance is an important factor in database access, this aspect must be taken into account in future works.

## References

1. Agrawal, R., Kiernan, J. and Srikant, R. (2002). *Hippocratic Database.* Proceedings of the 28th International Conference on Very Large Data Bases, 143-154.
2. Al-Fedaghi, S. (2007). Beyond Purpose-Based Privacy Access Control. 18th *Australasian Database Conference (ADC 2007), Ballarat, Australia. Conferences in Research and Practice in Information Technology*. 63.
3. Barker, S. (2010). Personalizing Access Control by Generalizing Access Control. *Proceedings of the Seventh ACM Symposium on Access Control Models and Technologies.* 149-158.

4. Bertino, E. (2005). Purpose Based Access Control for Privacy Protection in Database Systems. *Database Systems for Advanced Applications. Lecture Notes in Computer Science*, 3453, 1003-1007.
5. Bertino, E., Byun, J. W. and Li, N. (2005). Privacy-Preserving Database Systems. *Foundations of Security Analysis and Design III, Lecture Notes in Computer Science*. 3655: 178-206.
6. Bertino, E., Ghinita, G. and Kamra, A. (2011). Access Control for Databases: Concepts and Systems, *Foundations and Trends in Databases*. 3(1-2): 1-148.
7. Bertino, E. and Sandhu, R. (2005). Database Security-Concepts, Approaches, and Challenges. *IEEE Transactions on Dependable and Secure Computing*. 2(1): 2-19.
8. Braghin, S., Coen-Porisini, A., Colombo, P., Sicari, S. and Trombetta, A. (2008). Introducing Privacy in a Hospital Information System. *Proceedings of The 4th International Workshop on Software Engineering for Secure Systems*. 9-16.
9. Byun, J. W., Bertino, E. and Li, N. (2005). Purpose Based Access Control of Complex Data for Privacy Protection. *Proceedings of 10th ACM Symposium on Access Control Models and Technologies.* 102-110.
10. Byun, J.-W. and Li, N. (2008). Purpose Based Access Control for Privacy Protection in Relational Database Systems. *The International Journal on Very Large Data Bases*, 17(4), 603 - 619.
11. Camenisch, J., Modersheim, S., & Neven, G. (2009). Credential-Based Access Control Extensions to XACML, www.w3.org/2009/policyws/papers/Neven.pdf.
12. Chauduri, S., Kaushik, R., and Ramamurthy, R. (2011). Database Access Control & Privacy: Is There A Common Ground. *Proceedings of the 5th BiennialConference on Innovative Data Systems Research*. January 9-12. Asilomar, California, USA, 2010. 96-103.
13. Dagdee, N., and Vijaywargiya, R. (2009a). Access Control Methodology for Sharing of Open and Domain Confined Data using Standard Credentials. *International Journal on Computer Science and Engineering*. 1(3), 148-155.
14. Dagdee, N. and Vijaywargiya, R. (2009b). Credential Based Hybrid Access Control Methodology for Shared Electronic Health Records. *International Conference on Information Management and Engineering*. 3-5 April. S.D. Bansal Coll. of Technol., Indore. 624-628.
15. Di Vimercati, S. D. C., Foresti, S. and Samarati, P. Authorization and Access Control. In: Petkovic, M and Jonker, W. *Security, Privacy, and Trust in Modern Data Management.* Berlin/DE. Springer-Verlag Berlin and Heidelberg GmbH & Co. KG. 39; 2010.
16. Kabir, M. E., and Wang, H. (2009). Conditional Purpose Based Access Control Model for Privacy Protection. *Proc. 20th Australasian Database Conference*. 92, 135-142.
17. Kabir, M. E., Bertino, E. (2011). A Conditional Purpose Based Access Control Model with Dynamic Roles for Privacy Protection. Expert Systems with Applications. 38(2011), 1482-1485.
18. Kabir, M. E., Wang, H. and Bertino, E. (2012). A Role-involved Purpose-based Access Control Model. Information System Frontiers. 14, 809-822).
19. LeFevre, K., Agrawal, R., Ercegovac, V. and Ramakrishnan, R. (2004). Limiting Disclosure in Hippocratic Databases. *Proceedings of the Thirtieth International Conference on Very Large Data Bases*. 30, 108-119.
20. Masoumzadeh, A. and Joshi, J.B.D. (2008) PuRBAC: Purpose-Aware Role-Based Access Control. *Proceedings of the OTM 2008 Confederated International Conferences*. 1104-1121.
21. Peng, H., Gu, J., & Ye, X. 2008. Dynamic Purpose-Based Access Control. *Proceedings of the International Symposium on Parallel and Distributed Processing with Applications 08.* 695-700.
22. Stoupa, K., Simeoforidis, Z., and Vakali, A. (2006). Credential-Based Policies Management in an Access Control Framework Protecting XML Resources. , *Lecture Notes in Computer Science.* 4263, 603-612.
23. Sun, L. and Wang, H. (2010). Dynamic Purpose Based Usage Access Control. *World Academy of Science, Engineering and Technology 2010*. 619-624.