

Conditional Purpose Based Access Control Model for Privacy Protection

Md Enamul Kabir

Hua Wang

Department of Mathematics and Computing
University of Southern Queensland,
Toowoomba, Queensland 4350, Australia,
Email: {kabir, wang}@usq.edu.au

Abstract

This paper presents a model for privacy preserving access control which is based on variety of purposes. Conditional purpose is applied along with allowed purpose and prohibited purpose in the model. It allows users using some data for certain purpose with conditions. The structure of conditional purpose based access control model is defined and investigated through a practical paradigm with access purpose and intended purpose. An algorithm is developed to achieve the compliance computation between access purposes and intended purposes. According to this model, more information from data providers can be extracted while at the same time assuring privacy that maximizes the usability of consumers' data. This model extends traditional access control models to a further coverage of privacy preserving in data mining atmosphere. Its interior is a new structure for managing collected data in an effective and trustworthy way. This structure helps enterprises to circulate clear privacy promise, to collect and manage user preferences and consent. The implementation of the idea in the paper shows the flexibility of the model, and finally we provide comparisons of our work to other related work.

Keywords: access control, access purpose, intended purpose, conditional intended purpose, prohibited intended purpose.

1 Introduction and Motivation

Privacy¹ preservation of individuals is a challenging problem in the data-mining environment. Enterprises collect customer's personal identification information along with other attributes during any kinds of marketing systems. It is a natural expectation that the enterprise will use this information for various purposes, this leads to concern that the personal data may be misused. Many enterprises collect, store and use huge amount of personal information. A study conducted by the Federal Trade Commission has shown that 97 percent of websites were collecting at least one type of identifying information such as name, e-mail address, or postal address of customers (Federal Trade Commission 2000). Privacy preservation in data-mining environment has become a great concern both for enterprises and individuals.

Copyright ©2009, Australian Computer Society, Inc. This paper appeared at the 20th Australasian Database Conference (ADC 2009), Wellington, New Zealand. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 92, Athman Bouguettaya and Xuemin Lin, Ed. Reproduction for academic, not-for profit purposes permitted provided this text is included.

¹Privacy is defined as the right of an individual to decide when, how, and to what extent he/she would like to share his/her information

As individuals are more concern about their privacy, they are becoming more reluctant to carry out their businesses and transactions online, and many organizations are losing a considerable amount of potential profits (Forrester Research 2001). The research shown that on-line commerce was reduced by US\$15 billion in 2001 due to individual privacy concerns. These reactions from individuals imitate an altering awareness about how data is managed. Therefore without a clear compromising between individuals and enterprises, data quality and data privacy cannot be achieved and many organizations are seriously thinking about privacy issues of consumers. By demonstrating good privacy practices, many businesses are now trying to build up solid trust to customers, thereby attracting more customers (Baker & Peter 2003). Considering the privacy of customers, enterprise has to develop a secure privacy policy to remove the fear of customers. Thus in an internal management system, a reliable, efficient, effective and secure privacy policy should be established depending on customer's requirements.

A lot of work has been developed in order to protect the privacy of individuals and showed that the notion of purpose should be used as the basis for access control for specifying a privacy policy (Agrawal et al. 2002, Powers et al. 2002, LeFevre et al. 2004, Agrawal et al. 2005, Byun et al. 2005, 2008). According to Yang et al. (2007), a privacy policy ensures that data can only be used for its intended purpose (intended usage of data), and an access purpose (intention for accessing data objects) is compliant with the data's intended purpose. During the last few years, rapid technological developments especially in the field of information technology directed most attention and energy to the privacy protection of Internet users. Unless customers' data is suitably protected, individuals' privacy can be breached revealing their personal information. On the other hand, these collected data sets are the most important tools for a wide range of studies. Again the data that is more protected usually loses data quality. So it is necessary to come up a point where both data quality and data privacy are achieved. Although a significant number of works has been developed in this area (OASIS, Agrawal et al. 2002, Powers et al. 2002, LeFevre et al. 2004, Agrawal et al. 2005, Byun et al. 2005, 2008), research has yet to be done in order to remove the dilemma between data quality and data privacy.

Many privacy policy access control models have been proposed in order to protect the privacy of consumers. Byun et al. (2005, 2008) pointed out that privacy protection cannot be easily achieved by traditional access control models as it focuses on which user is performing which action on which data

Table 1: Hypothetical data base illustrating AIP and PIP

name	age	address	income	name _{ip}	age _{ip}	address _{ip}	income _{ip}
Alice	35	21, West St., TBA, QLD 4350	35000	$\langle\{G\}, \{\Phi\}\rangle$	$\langle\{\Phi\}, \{G\}\rangle$	$\langle\{G\}, \{A, S\}\rangle$	$\langle\{G\}, \{M\}\rangle$
Bob	29	45, Fay CT., TBA, QLD 4350	23000	$\langle\{G\}, \{\Phi\}\rangle$	$\langle\{G\}, \{M\}\rangle$	$\langle\{G\}, \{A, S\}\rangle$	$\langle\{G\}, \{A, M\}\rangle$
Ron	56	20, Anita Dr., TBA, QLD 4350	56000	$\langle\{G\}, \{\Phi\}\rangle$	$\langle\{G\}, \{M\}\rangle$	$\langle\{G\}, \{A, S\}\rangle$	$\langle\{G\}, \{A\}\rangle$
Jak	48	25, Wuth St., TBA, QLD 4350	48000	$\langle\{G\}, \{\Phi\}\rangle$	$\langle\{G\}, \{M\}\rangle$	$\langle\{G\}, \{A\}\rangle$	$\langle\{G\}, \{A, M\}\rangle$

G={General purpose}, A={Admin purpose}, S={Shipping purpose}, P={Purchase purpose},
M={Marketing purpose}, ip={Intended purpose}=(AIP, PIP)

object. But a reliable privacy policy are concerned with which data object is used for which purpose. They also suggested that the notion of purpose must play a major role in access control models and that an appropriate metadata model must be developed to support such privacy centric access control models in order to protect data privacy. An approach is developed that is based on intended purposes, which specify the intended usage of data, and access purpose, which specify the purposes for which a given data element is accessed. Usually, during the data collection procedure customers are informed about the purposes of enterprises. Customers then decide whether their information could be used or not for a certain purpose. That means data providers are given an option of their data with certain purposes. If an individual mentions that his/her data could not be used for a certain purpose, then his/her information is not accessible for the purpose. Usually data providers are reluctant to use any part of their information for any purposes and so there is a possibility of losing information. But more information can be extracted from data providers by providing more possible options of using their information. It is possible to protect the privacy of individuals in this model, but there is a shortcoming of information loss. An intended purpose is divided (IP) into two parts: Allowed Intended Purposes (AIP) (explicitly allows to access the data for the particular purpose) and Prohibited Intended Purpose (PIP) (data access for particular purposes are never allowed). In order to recognize the model clearly, suppose that a company uses consumers' data for the purpose of General, Admin, Marketing and Shipping and consider the hypothetical database in Table 1.

In Table 1, the value of Alice's attribute income_{ip} is $\langle\{G\}, \{M\}\rangle$, which means that Alice income could be used for General purpose but strictly prohibited to use for Marketing purpose. If we take a query

```
SELECT name
FROM Table 1
FOR Marketing Purpose
```

it gives the name of Alice, Bob, Ron, Jak and if we have a query

```
SELECT name, age
FROM Table 1
FOR Marketing Purpose
```

it returns nothing because prohibited intended purposes override the allowed intended purposes. This model protects privacy of consumers as it considers customers' requirements but it occurs more information loss. So a natural question arise

“ Is it possible to extract information from PIP at least conditionally?”

The answer of this question is achieved in this article by adding a new term conditional purpose in the intended purpose. In order to extract more data

and protect data privacy, conditional purpose plays a role in access control models. In this paper, we address this goal by presenting a model of purpose management, which is a fundamental building block on which conditional purpose based access control can be developed. Our proposed model is based on access purpose and intended purpose. Both access purposes and intended purposes are specified with respect to a hierarchical structure that organizes a set of purposes for a given enterprise. A key feature of our proposed model is that it supports conditional purpose and prohibited purpose, thus allowing users to specify that data should be used conditionally or should not be used for a set of purpose.

Observing these challenges and the satisfaction of both enterprises and customers, we need a better model to extract more information from customers with privacy guarantees. To overcome this challenge, we propose a new access control called conditional based access control model. In the access control model it is enable to extract information from PIP by giving conditions called Conditional Intended Purpose (CIP). Our proposed model is helpful for enterprises to establish an ideal privacy policy and to manage data in a sensitive, effective and trustworthy way. It also helps policy makers and the experts in the data-mining environment.

The reminder of this paper is organized as follows. We present a brief overview of privacy related technologies in Section 2. Since purpose is used as the basis of access control, a brief description of the notion of purpose is described in Section 3. In Section 4 we present comprehensive descriptions of our proposed access control model. Section 5 is devoted to compliance check and access control using query modification. We compare our proposed model with the most recent access control models in Section 6. Concluding remarks are included in Section 7.

2 Related Work

This work is related to several topics in the area of privacy preservation in data mining atmosphere.

The most notable technique to protect privacy is the W3C's Platform for Privacy Preferences (P3P) that formally specify privacy policy by service providers (Marchiori 2002). P3P provides a way for a web site to encode its data collection in a machine-readable format known as a P3P policy, which can be compared against a user's privacy preferences Yang et al. (2007). Byun et al. (2008) pointed out that P3P does not provide any functionality to keep promises in the internal privacy practice of enterprise. Thus it can be said that a striking privacy policy with inadequate enforcement mechanism may place the organizations at risk of reputation damage. The concept of Hippocratic database introduced by Agrawal et al. (2002) that amalgamates privacy protection in relational database system. A Hippocratic database includes privacy policies and authorizations that associate with each attribute and each user

the usage purpose(s) (Al-Fedaghi 2007). Agrawal et al. (2002) presented a privacy preserving database architecture called Strawman which was based the access control on the notion of purposes, and opened up database-level researchers of privacy protection technologies. After that, purpose based access control introduced by Byun et al. (2005, 2008) and Yang et al. (2007), fine grained access control introduced by Rizvi et al. (2005) and Agrawal et al. (2005) are widely used access control models for privacy protection. In IT system the proposed Enterprise Privacy Authorization Language (EPAL) of IBM is a language for writing enterprise privacy policies to run data handling practices. An EPAL policy defines hierarchies of data-categories, user-categories, and purpose (Byun et al. 2008). A set of actions, obligations, and conditions are also defined by an EPAL policy.

A lot of works (Bertino et al. 1996, Denning et al. 1988, Sandhu & Chen 1991, 1998) provide many valuable insights for designing a fine-grained secure data model. In a multilevel relational database system, every piece of information is classified into a security level, and every user is assigned a security clearance (Byun et al. 2008). LeFevre et al. (2004) proposed an approach to enforcing privacy policy in database setting. This work focus on ensuring limited data disclosure, based on the premise that data providers have control over who is allowed to see their personal data and for what purpose. They introduced two models of cell-level limited disclosure enforcement and suggest an implementation based on query modification techniques.

Byun et al. (2008) present a comprehensive approach for privacy preserving access control model. In their access control model multiple purposes to be associated with each data elements and also support explicit prohibitions. This model is based on the notion of purpose as it plays a central role and is the basic concept on which access decisions are made. According to Byun et al. (2008) a purpose describes the reason(s) for data collection and data access, access purpose is intension for accessing data objects and intended purpose is the specified usages for which the data objects are collected. Massacci et al. (2005) pointed out that most privacy-aware technologies use purpose as a central concept around which privacy protection is built.

All of these works proposed different approaches to protect the privacy of individuals through different models without being considering to extract more information. Our aim is to preserve privacy of individuals as well as extracting more information. With this aim in this paper we propose a model that has significantly improved the work of Byun et al. (2008). It has improved in three different remarkable ways. First, we introduce conditional purpose in addition to explicit prohibitions that make data providers more flexible to give information. Second, the enterprise can publish an ideal privacy policy to manage data in a sensitive, effective and trustworthy way, and third it reduces the information loss as it shows that we can extract more information from data providers.

3 Purpose, Access Purpose and Intended Purpose

Data is collected for certain purpose. For instance, a nation wide demographic survey in Australia, data may be collected to know the socioeconomic and demographic characteristics of all Australians. Each

Table 2: Predetermined Intended Purposes

	Group 1	Group 2	Group 3
Name	$\langle\{G\}, \{T\}, \{\Phi\}\rangle$	$\langle\{G\}, \{\Phi\}, \{T\}\rangle$	$\langle\{G\}, \{\Phi\}, \{\Phi\}\rangle$
Address	$\langle\{G\}, \{T\}, \{\Phi\}\rangle$	$\langle\{G\}, \{\Phi\}, \{T\}\rangle$	$\langle\{G\}, \{\Phi\}, \{\Phi\}\rangle$
Phone	$\langle\{G\}, \{T\}, \{\Phi\}\rangle$	$\langle\{G\}, \{\Phi\}, \{T\}\rangle$	$\langle\{G\}, \{\Phi\}, \{\Phi\}\rangle$
Age	$\langle\{G\}, \{T\}, \{\Phi\}\rangle$	$\langle\{G\}, \{\Phi\}, \{T\}\rangle$	$\langle\{G\}, \{\Phi\}, \{\Phi\}\rangle$
Income	$\langle\{G\}, \{T\}, \{\Phi\}\rangle$	$\langle\{G\}, \{\Phi\}, \{T\}\rangle$	$\langle\{G\}, \{\Phi\}, \{\Phi\}\rangle$

data access also serves a certain purpose. So it is a natural expectation that a privacy policy should concern which data object is used for which purposes. Many authors indicated that purpose is a central part in many privacy preserving access control model.

3.1 Definition of Purpose

For preserving the privacy of customers, each and every data access must obey with the privacy policies on which customers have conditionally or unconditionally agreed. A representative privacy policy for a data element includes purposes, retention, condition and obligation. This means that the particular data element can be conditionally or unconditionally accessed only for the specific purposes with certain conditions. The retention indicates how long the data element can be reserved, and the obligation designates the actions that must be followed after an access to the data element is approved. So purpose is the most interesting thing to researchers as it directly shows how access to data elements has to be controlled. P3P defines purpose as "the reason(s) for data collection and use" and specifies a set of purposes (World Wide Web Consortium). In commercial surroundings purposes normally have a hierarchical associations among them; i.e., generalization and specialization relationships. For instance, a group of purposes such as direct-marketing and third party marketing can be represented by a more general purpose, marketing. We borrow the purpose definition from Byun et al. (2008).

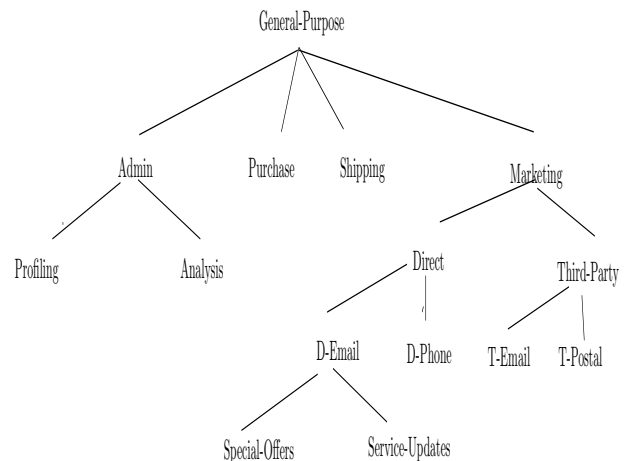


Figure 1: Purpose Tree

Definition 1: (Purpose and Purpose Tree): A purpose describes the intentions for data collection and data access. A set of purposes, denoted as ω , is organized in a tree structure, referred to as Purpose Tree and denoted as Ω , where each node represents a purpose in ω and each edge represents a hierarchical relation between two purposes. Let r_i, r_j , be two purposes in Ω . We say that r_i is an ancestor of r_j (or r_j is a descendent of r_i) if there exists a downward path from r_i to r_j in Ω . Figure 1 is an example of

purpose tree, where each node represents a purpose in ω and each edge represents a hierarchical relation between two purposes.

Purposes, depending on their association with objects and subjects, may be called intended purposes or access purposes respectively.

Definition 2 (Access Purpose): An access purpose is intensions for accessing data objects, and it must be determined by system when data access is requested. So access purpose specifies the purpose for which a given data element is accessed.

Definition 3 (Intended Purpose): An intended purpose is the specified usages for which data objects are collected. That is, purpose associated with data and thus regulating data accesses as intended purpose. According to our approach an intended purpose consists of the following three components.

Allowable Intended Purpose (AIP): This means that data providers explicitly allow accessing the data for a particular purpose. For example data providers may consider that his/her information can be used for marketing purpose without any further restrictions.

Conditional Intended Purpose (CIP): This means that data providers allow accessing the data for a particular purpose with some conditions. For example data providers may consider that his/her income information can be used for marketing purpose by hiding his/her personal identification information (e.g. id or name etc.) or his/her income data can be reveal through generalization. or only the first letter of name can be used for marketing purpose.

Prohibited Intended Purpose (PIP): This means that data providers strictly disallow accessing the data for a particular purpose. For example data providers may consider that his/her income information cannot be used for marketing purpose. In that case data provider's income attribute is strictly prohibited to use for marketing purpose. An example of how AIP, CIP and PIP works is illustrated through a hypothetical database in Table 3.

So an intended purpose IP is a tuple $\langle AIP, CIP, PIP \rangle$, where $AIP \subseteq \omega$, $CIP \subseteq \omega$ and $PIP \subseteq \omega$ are three sets of purposes. The set of purposes implied by IP, denoted by IP^* , is defined to be $AIP^\downarrow \cup CIP^\downarrow - PIP^\uparrow$, where

R^\downarrow , is the set of all nodes that are descendants of nodes in R, including nodes in R themselves,

R^\uparrow , is the set of all nodes that are ancestors of nodes in R, including nodes in R themselves, and

R^\downarrow , is the set of all nodes that are either ancestors or descendants of nodes in R, that is, $R^\downarrow = R^\uparrow \cup R^\downarrow$.

The following example explains the definition of AIP, CIP and PIP.

Example 1: Suppose $IP = \langle \{\text{Admin, Direct}\}, \{\text{Third-party}\}, \{\text{D-mail}\} \rangle$, then $IP^* = \{\text{Admin, Profiling, Analysis, D-Phone, Third-party, T-E-mail}_c, \text{T-Postal}_c\}$. where subscripts c indicates that customers information can be used for the purpose with some conditions.

Definition 4 (Access Purpose Compliance): Let Ω be a purpose tree. Let $IP = \langle AIP, CIP, PIP \rangle$ and AP be an intended purpose and an access

purpose defined over Ω , respectively. AP is said to be compliant with IP according to Ω , denoted as $AP \leftarrow_{\Omega} IP$, if and only if $AP \in IP^*$.

Example 2: Suppose a company established the following privacy policies:

- We use your information for purchasing purposes, to provide services to you, and to inform you of services that may better meet your needs.
- We will disclose, conditionally disclose or will not disclose your information to third parties according to our privacy requirements.

In Table 2, Group 3 represents customers who have given consent for third-party marketing, Group 1 represents customers who have given consent for third-party marketing by removing personal identification information or via generalization/suppression (conditionally given consent) and Group 2 represents customers who have not given consent for third-party marketing.

4 Conditional purpose based access control

In our model data providers are asked three options for their data usage, permissible, prohibited and conditional permissible usages of each data item. For example, a data provider may select his/her name is permissible for **Admin** purpose, address is not permissible for **Shipping** purpose but income information is conditionally permissible for **Marketing** purpose. That is, data provider does not have any privacy concern over the name when it is used for the purpose of administration, great concern about privacy of the address information (and so does not want to disclose address) when it is used for the purpose of shipping, but his/her income information can be used for marketing purpose with some conditions. Here the term "conditions" means that data provider ready to release his/her certain information for certain purpose by removing his/her name or id or through generalization. This information is then stored in the database along with the collected data, and access to the data is tightly governed according to the data provider's requirements. For using the term condition data providers feel more comfortable to release their data. So according to our approach enterprise can establish an attractive privacy policy and it is possible to extract more information from data providers. Table 4 illustrates some imaginary records and intended purposes stored in a conceptual data base relation. Notice that each data element is stored in three different purposes each of which corresponds to a particular intended purposes.

As discussed before, our design of intended purposes supports permissive, conditions and prohibitive privacy policies. This construction allows more squash and flexible policies in our model. Moreover, by using CIP and PIP, we can assure that data access for particular purposes are allowed with some conditions and never allowed. Note that an access decision is made based on the relationship between the access purpose and the intended purpose of the data. Access is allowed only if the access purpose is included in the implementation of the intended purpose; in that case the access purpose is compliant with the intended purpose. The access is accepted with conditions if the implementation of intended purpose includes the access purpose with conditions; in this case we say that access purpose is conditionally complaint with intended purpose. The access is denied if the implementation of the intended purpose

Table 3: Hypothetical data base illustrating AIP, CIP and PIP

name	age	address	income	name _{ip}	age _{ip}	address _{ip}	income _{ip}
Alice	35	21, West St., TBA, QLD 4350	35000	$\langle\{G\}, \{\Phi\}, \{\Phi\}\rangle$	$\langle\{\Phi\}, \{M\}, \{A\}\rangle$	$\langle\{G\}, \{\Phi\}, \{A, S\}\rangle$	$\langle\{G\}, \{A\}, \{M\}\rangle$
Bob	29	45, Fay CT., TBA, QLD 4350	23000	$\langle\{G\}, \{\Phi\}, \{\Phi\}\rangle$	$\langle\{G\}, \{M\}, \{\Phi\}\rangle$	$\langle\{G\}, \{M\}, \{A, S\}\rangle$	$\langle\{G\}, \{M\}, \{A\}\rangle$
Ron	56	20, Anita Dr., TBA, QLD 4350	56000	$\langle\{G\}, \{\Phi\}, \{\Phi\}\rangle$	$\langle\{G\}, \{M\}, \{\Phi\}\rangle$	$\langle\{G\}, \{\Phi\}, \{A, S\}\rangle$	$\langle\{G\}, \{S\}, \{A\}\rangle$
Jak	48	25, Wuth St., TBA, QLD 4350	48000	$\langle\{G\}, \{\Phi\}, \{\Phi\}\rangle$	$\langle\{G\}, \{M\}, \{\Phi\}\rangle$	$\langle\{G\}, \{M\}, \{A\}\rangle$	$\langle\{G\}, \{M\}, \{A\}\rangle$

G={General purpose}, A={Admin purpose}, S={Shipping purpose}, P={Purchase purpose},
M={Marketing purpose}, ip={Intended purpose}=(AIP, CIP, PIP)

Table 4: Fictional records and intended purposes

	name	age	address	income
AIP	Alice	35	21, West St., TBA, QLD 4350	35000
CIP	A	30-40	West St., TBA, QLD 4350	30000-40000
PIP	*	*	*	*
AIP	Bob	29	45, Fay CT., TBA, QLD 4350	23000
CIP	B	20-30	Fay CT., TBA, QLD 4350	20000-30000
PIP	*	*	*	*
AIP	Ron	56	20, Anita Dr., TBA, QLD 4350	56000
CIP	R	50-60	Anita Dr., TBA, QLD 4350	50000-60000
PIP	*	*	*	*
AIP	Jak	48	25 Wuth St., TBA, QLD 4350	48000
CIP	A	50-60	Wuth St., TBA, QLD 4350	40000-50000
PIP	*	*	*	*

* means data providers are reluctant of any usage of their data items

does not include the access purpose, in this case access purpose is not complaint with the intended purpose. Suppose in the online marketing system, an enterprise collects name, age, address and income of customers along with other information and the enterprise uses customer's information for the purpose of admin, shipping, purchase and marketing. Consider the hypothetical database in Table 3.

In Table 3, the value of Alice's attribute income_{ip} is $\langle\{G\}, \{A\}, \{M\}\rangle$ which means that Alice income could be used for General purpose but strictly prohibited to use for Marketing purpose. It also means that Alice income could be used for Admin purpose by hiding her personal identification information or through generalization. Similarly, Bob, Ron and Jak's income information could be used conditionally for Marketing purposes but their income information is strictly prohibited for Admin purpose.

4.1 Implementation

In our proposed model, users query the database using standard SQL statements. In this article we assume that each query is connected with a specific purpose. The data is accessible to each query varies depending on the data providers agreement and the purpose of the query. For example, any query against Table 3 with any purpose returns a result that is equivalent to the result of the query. As our proposed model directly reflect the information that is allowed, conditionally allowed or prohibited by each data provider, querying against these model does not violate privacy. This model is quite different from the conventional access control model as different sets of data may be returned for the same query depending on the purpose of the query and the data providers' agreements. Thus from the hypothetical database in Table 3, if we take the query

```
SELECT name, income
FROM Table 3
FOR Marketing Purpose,
```

Table 5: Filtering information

Ron	56000
Bob	20000-30000
Jak	40000-50000

then by using Table 4, we get the information in Table 5.

We can see from Table 5 that it gives name and income of Ron as he allows to disclose his name and income information for Marketing purpose. It also shows other two incomes via generalization as they conditionally allowed to disclose their income. This clearly shows the utility of using our proposed model. It demonstrates that it can extract more information from data providers.

Theorem 1: Let p , q and r denote the probability that a data provider gives consent of a particular attribute for AIP, PIP and CIP respectively. Assuming that these probabilities remain the same from data provider to data provider. Then the conditional based access control model extracts more information than the model proposed by Byun et al. (2008).

Proof: Let n be the total number of data providers. If p and q are the probabilities that a given data provider gives consent of a particular attribute for AIP and PIP. Then the average numbers of data providers who give consent for AIP is np . That means by using the model of Byun et al. (2008), the average number of data providers who give consent for AIP of a particular attribute is np . If we use our model then the average number of data providers who give consent to disclose their data for a certain purpose with some conditions is nr . So by using the conditional based access control model total average number of data providers whose information is accessible is $(np + nr)$. Since n and p both are positive so $(np + nr)$ is always greater than np . This means that it is enable to extract more information from customers by using the conditional

based access control model.

In our model, the collected data is used for different purposes on the basis of the data providers requirements. For using the CIP, both privacy and usability of data can be achieved as it filters out the values by performing a purpose compliance. By using a hypothetical database and the extracted outcome in Table 5, it shows clearly the data utility and data providers information is protected. It showed by Theorem 1 that our proposed model extracts more information with assuring privacy.

5 Access control

Among the various possible techniques to determine access purpose, in this paper we utilize the method where the users are required to explicitly state their access purposes when they try to access data. That is, users provide an access purpose for each query they issue.

5.1 Compliance Check

Consider the purpose tree in Figure 2 and it encoded into a relation *pt-table* as shown in Table 6. The first column *p_id* represents the identification number of each purpose node, the second column *p_name* represents the name of each purpose node, and the third column *parent* is used to capture the hierarchical relationships among purpose nodes. The column code is the binary encoding of each purpose. For example, in Table 6 the purpose B is encoded as '0x100' in hexadecimal representation, while the purpose E is encoded as '0x020' in hexadecimal form. The last three columns *aip_code*, *cip_code* and *pip_code* are precalculated encodings of purpose implications. As we know, when a purpose r_i is used as an AIP, it means that every descendant of r_i , including r_i itself is allowed. For example, the purpose D in Figure 2 used as an AIP implies that access is allowed for the purpose of D as well as G, H, I and J. Thus, the *aip_code* of D contains the implied set of D, which is the sum of the encodings of D, G, H, I and J. Note that *aip_code* and *cip_code* of each purpose is same as in the long run both are allowed. The *pip_code* of a particular purpose r_j is computed similarly by summing the encodings of every descendant and ancestor of r_j with the encoding of r_j itself.

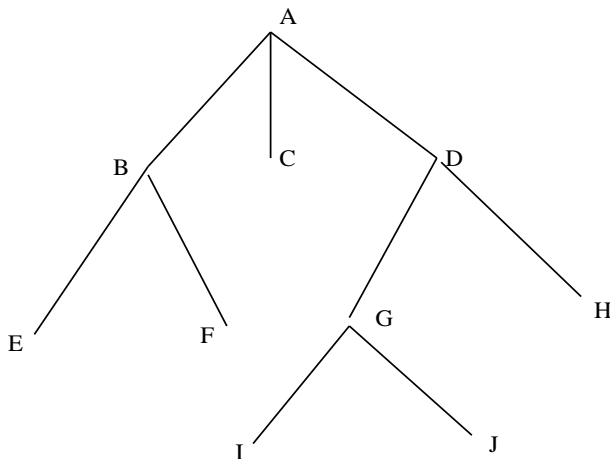


Figure 2: Purpose Tree

An access purpose is compliant with an intended purpose if and only if the access purpose is not prohibited by PIP and it is allowed by both AIP and CIP. Thus, the purpose compliance check can be done with two bitwise AND an operation as follows:

AIP, CIP and PIP, say *ap_code*, *aip_code*, *cip_code* and *pip_code* respectively, the access purpose is compliant with the intended purpose if and only if

$$\begin{aligned}
 & (ap_code \& cip_code)=0 \wedge (ap_code \& aip_code) \neq 0 \vee \\
 & (ap_code \& pip_code)=0 \wedge (ap_code \& aip_code) \neq 0 \vee \\
 & (ap_code \& pip_code)=0 \wedge (ap_code \& cip_code) \neq 0.
 \end{aligned}$$

where, $\&$ is bitwise AND operator, \wedge is logical AND operator and \vee is logical OR operator. Conflicts among the AIP, CIP and the PIP for the same data element are resolved by applying the denial-takes-procedure policy where PIP overrides AIP and CIP, and CIP overrides AIP. The computation for purpose compliance check is illustrated in Table 7.

5.2 Query modification

It is a natural expectation that privacy-preserving access control techniques ensures a query result contains only the data items that are allowed or conditionally allowed or completely prohibited for the access purpose of the query. This expectation is achieved in this paper using query modification Stonebraker & Wong (1974). It is important to notify that query modification provides powerful and flexible controls without requiring any alteration in underlying mechanisms and that it is supported in a major commercial Data Base Management System (Oracle Corporation 2002). Our query modification algorithm is outlined in Table 7.

The complexity of our query modification algorithm is in $O(n)$, where n is the number of attributes accessed by a given query. The method *Modifying_Query* is invoked only if the access purpose of the query is verified to be acceptable by the *validate* function. If the access purpose is unacceptable, then query is rejected without further being processed. The query modification algorithm checks both the attributes referenced in the projection list and the attributes referenced in predicates. As the attributes in the projection list determine what data items will be included in the result relation of a query, it may seem enough to enforce privacy policy based only on the attributes in the projection list. However, the result of a query also depends on the predicates, and not enforcing privacy constraints on the predicates may introduce inference channels. The abounding algorithm filters out a tuple if any of its elements that are accessed is conditionally allowed or prohibited with respect to the given access purpose. For example, consider a query,

```

SELECT name, income, address
FROM Table 3
FOR Marketing Purpose.
  
```

Suppose there is a customer record of which name is allowed for marketing, income is conditionally allowed for Marketing but the address is prohibited for this purpose. Then our algorithm only excludes address of this record from the query result and income information is visible anonymizing the customer's name or income information reveal via generalisation. So according to our proposed model

Table 6: Pt-table

p_id	p_name	parent	code	aip_code	cip_code	pip_code
1	A	-	0×200	0×3FF	0×3FF	0×3FF
2	B	1	0×100	0×130	0×130	0×330
3	C	1	0×080	0×080	0×080	0×280
4	D	1	0×040	0×04F	0×04F0	0×24F
5	E	2	0×020	0×020	0×020	0×320
6	F	2	0×010	0×010	0×010	0×310
7	G	4	0×008	0×00B	0×00B	0×24B
8	H	4	0×004	0×004	0×004	0×244
9	I	7	0×002	0×002	0×002	0×24A
10	J	7	0×001	0×001	0×001	0×249

income information of this customer is still usable for Marketing purpose instead of excluding other records.

The following example illustrates how our algorithm modifies queries. This example is a revised version of Byun et al. (2008) where purpose encoding of Marketing is assumed to be '0×200'. For the query

```
SELECT name, income
FROM table 3
FOR Marketing Purpose,
```

modified query becomes

```
SELECT name, income
FROM Table 3
WHERE Comp_Check('0×200',
name_aip, name_cip, name_pip)
AND Comp_Check('0×200',
income_aip, income_cip, income_pip).
```

Table 7: Query Modification Algorithm

```
Comp_Check (ap, aip, cip, pip)
/* This function is required for query modification */
Returns Boolean
if (ap & cip)=0 and (ap & aip)≠ 0
return True;
else if (ap & pip)=0 and (ap & aip)≠ 0
return True;
else if (ap & pip)=0 and (ap & cip)≠ 0
return True;
else False;

Modifying_Query (Query Q)
Returns a modified privacy-preserving query Q
Let R1, ..., Rn be the relations referenced by Q
Let P be the predicates in WHERE clause of Q
Let a1, ..., am be the attributes referenced in both
the projection list and P
Let AP be the access purpose encoding of Q
for each Ri where i=1,..,n do
if (Comp_Check (AP, Ri.aip, Ri.cip, Ri.pip)=False) then
return ILLEGAL-QUERY;
end if;
end for;
return Q without modified P;
```

6 Comparison

There are some related works on privacy preservation. The closest works related to this article are Hippocratic databases (Agrawal et al. 2002) and purpose based access control model (Byun et al. 2008). In this section we will compare our proposed model with these two models.

Agrawal et al. (2002) proposed Hippocratic databases that incorporate privacy protection within relational database system. The proposed technique uses privacy metadata, which consist of privacy policies and privacy authorizations stored in two tables. The authors proposed a strawman design for

Hippocratic databases. This design identified the technical challenges and problems in designing such databases. But the authors did not consider the concepts of purpose. By contrast, in our proposed model we investigated more sophisticated concepts of purpose. We used conditional purpose and the association of different purposes with a data element which are not considered in their work.

Byun et al. (2008) provided a comprehensive framework for purpose and data management. They argued that in order to protect data privacy, the notion of purpose must play a major rule in access control model. The authors proposed approach is based on intended purposes, which specify the intended usage of data, and access purposes, which specify the purposes for which a given data element is accessed. They also argued that traditional access control models focus on which user is performing which action on which data objects but privacy policies are concerned with which data object is used for which purposes. The authors proposed purpose based access control model (PBAC) allows multiple purposes to be associated with each data element and also supports explicit prohibitions. Although their proposed model designed on the basis of customers requirements and so does not violate privacy, the main drawback of this model is the information loss. In that model customers are given only two options whether their private data can be used or not for certain purposes instead of giving more possible options. But we strongly believe that by giving more options to customers data extractions can be easily achieved. By contrast, the proposed model in this paper provides three more options that help enterprises to extract more information from customers, assuring privacy. This criteria is achieved theoretically by Theorem 1 in Subsection 4.1. This clearly shows the utility and usability of our proposed model in a effective and trustworthy way.

7 Conclusion

Although privacy preserving desires a secure infrastructure and relies on access control technology, it is not a security problem but it is related to a data management problem. Purposes play a significant role in the field of database management system privacy preserving techniques. In this paper we introduced conditional based access control model for privacy protection in database system that enables enterprise to operate as a reliable keeper of their customers data. The basic concepts of the proposed conditional based access control model are discussed and it has shown the possibility to extract more information from customers by providing a secure privacy policy. The study reveals that this model achieves a better progress than the other access

control models in the area of privacy preserving in data mining environment. The utility of the proposed conditional based access control model is illustrated through a practical example. We also discussed the algorithm to achieve the compliance check between access purpose and intended purposes. The effect of the proposed access control can be extremely useful for internal access control within an organization as well as well as information sharing between organizations. The enterprise can use this technology to enforce the privacy promises they make and to enable their customers to maintain control over their data. It would also help researchers, users and the associated people in the area of data mining.

Our proposed approach provides a complete structure for privacy preserving access control model. On the basis of this approach, a significant further work still needs to be done. Our future work includes extending this model in the Role-based Access Control (RBAC), Dynamic Purpose-based Access Control (DPBAC) and in the other access control systems.

References

- Agrawal, R., Kiernan, J., Srikant, R. & Xu, Y. (2002), Hippocratic databases, *in* '28th International Conference on Very Large Databases (VLDB)'.
- Agrawal, R., Bird, P., Grandison, T., Kiernan, J., Logan, S. & Xu, Y. (2005), Extending relational database systems to automatically enforce privacy policies, *in* 'ICDE', pp.1013-1022.
- Al-Fedaghi (2007), Beyond Purpose-based privacy access control, *in* '18th Australian Database Conference (ADC)'.
- Barker, S. & Stuckey, P.N. (2003), Flexible access control policy specification with constraint logic programming, *in* 'ACM Transaction on Information and System Security', Vol. 6(4), November.
- Bertino, E., Jajodia, S. & Samarati, P. (1996), Database security: Research and practice, *in* 'Information systems'.
- Bertino, E., Byun, J.W., & Li, N. (2005), Privacy-Preserving database system, *in* 'FOSAD', pp. 178-206.
- Byun, J., Bertino, E. & Li, N. (2005), Purpose based access control of complex data for privacy protection, *in* 'Symposium on Access Control Model And Technologies (SACMAT)'.
- Byun, J., Bertino, E. & Li, N. (2008), 'Purpose based access control for privacy protection in relational database systems', *VLDB J* 17(4), 603-619.
- Denning, D., Lunt, T., Schell, R., Shockley, W. & Heckman, M. (1988), The seaview security model, *in* 'The IEEE Symposium on Research in Security and Privacy'.
- Federal Trade Commission(2000), Privacy online: Fair information practices in the electronic marketplace: A report to congress, May. Available at www.ftc.gov/reports/privacy2000/privacy2000.pdf.
- Forrester Research (2001), Privacy concerns cost e-commerce \$15 billion. Technical report, September.
- IBM, The Enterprise Privacy Authorization Language (EPAL). Available at www.zurich.ibm.com/security/enterprise-privacy/epal.
- LeFevre, K., Agrawal, R., Ercegovac, V., Ramakrishnan, R., Xu, Y. & DeWitt, D. (2004), Disclosure in Hippocratic databases, *in* 'The 30th International Conference on Very Large Databases (VLDB)', August.
- Marchiori, M. (2002). The platform for privacy preferences 1.0 (P3P1.0) specification. Technical report, W3C, April.
- Massacci, F., Mylopoulos, J. & Zannone, N. (2005), Minimal Disclosure in Hierarchical Hippocratic Databases with Delegation, *in* 'The 10th Europran Symposium on Research in Computer Security', September 12-14.
- OASIS, Core and hierarchical role based access control (rbac) profile of xacml v2.0. Available at <http://www.oasis-open.org/>.
- OASIS, Extensible access control markup language (xacml) 2.0. Available at <http://www.oasis-open.org/>.
- OASIS, Privacy policy profile of xacml v2.0. Available at <http://www.oasis-open.org/>.
- Oracle Corporation (2002), The Virtual Private Database in Oracle9iR2: An Oracle Technical White Paper, January, Available at www.oracle.com.
- Rizvi, S., Mendelzon, A. O., Sudarshan, S. & Roy, P. (2004), Extending query rewriting techniques for fine-grained access control, *in* 'SIGMOD Conference', pp.551-562.
- Powers, C.S., Ashley, P. & Schunter, M. (2002), Privacy promises, access control, and privacy management, *in* 'The 3rd International Symposium on Electronic Commerce'.
- Sandhu, R. & Jajodia, S. (1991), Toward a multi-level secure relational data model, *in* 'ACM Transactional Conference on Management of Data (SIGMOD)'.
- Sandhu, R. & Chen, F. (1998), The multilevel relational data model, *in* 'ACM Transaction on Information and System Security'.
- Stonebraker, M. & Wong, E. (1974), Access control in a relational database management system by query modification, *in* 'ACM CSC-ER Proceedings of the 1974 Annual Conference', January.
- World Wide Web Consortium (W3C), Platform for Privacy Preferences (P3P), Available at www.w3.org/P3P.
- Yang, N., Barringer, H. & Zhang, N. (2007), A Purpose-Based Access Control Model, *in* 'Third International Symposium on Information Assurance and Security', pp. 143-148.