Capturing P3P Semantics Using an Enforceable Lattice-based Structure

Kambiz Ghazinour
Department of Computer Science
University of Calgary
Calgary, Canada
kghazino@ucalgary.ca

Ken Barker
Department of Computer Science
University of Calgary
Calgary, Canada
kbarker@ucalgary.ca

ABSTRACT
With the increasing amount of data collected by service providers, privacy concerns increase for data owners who must provide private data to receive services. Legislative acts require service providers to protect the privacy of customers. Privacy policy frameworks, such as P3P, assist the service providers by describing their privacy policies to customers (e.g. publishing privacy policy on websites). Unfortunately, providing the policies alone does not guarantee that they are actually enforced. Furthermore, a privacy-preserving model should consider the privacy preferences of both the data provider and collector. This paper discusses the challenges in development of capturing privacy predicates in a lattice structures. A use case study is presented to show the applicability of the lattice approach to a specific domain. We also present a comprehensive study on applying a lattice-based approach to P3P. We show capturing privacy elements of P3P in a lattice format facilitates managing and enforcing policies presented in P3P and accommodates the customization of privacy practices and preferences of data and service providers. We also propose that the outcome of this approach can be used on lattice-based privacy aware access control models [8].

Categories and Subject Descriptors
E.1 [DATA STRUCTURES]: Graphs
K.4.1 [Computers and Society]: Public Policy Issues - *Privacy*

General Terms
Algorithms, Design, Experimentation, Security, Verification.

Keywords
Privacy policy, P3P, lattice, privacy protection, privacy model.

1. INTRODUCTION
Privacy is a leading concern for individuals who utilize computing resources for a variety of reasons (e.g. visiting a health care provider). As more data from customers are being collected, how their data is treated after being collected becomes more crucial to the data providers. Legislative acts implicitly require service providers address privacy concerns by informing customers how their data is being treated after the collection. Much research has proposed privacy policy frameworks and policy languages and models such as P3P [6] and EPAL to assist data owners and collectors in communicating their privacy concerns.

PAIS 2011, March 25, 2011, Uppsala, Sweden.

Although these privacy enhanced technologies inform the data provider of the privacy policies practiced in the organization, description alone is not sufficient, and there are two issues that must be addressed: First, these frameworks do not provide any mechanism to enforce privacy policies. Thus, data providers receive no guarantee that the policies and promises are actually enforced by the access authorization and data processing phases. In fact, data providers have no choice but to rely on data collectors and hope their data is being treated as promised. Second, since both the data provider and collector want to maximize their control over the data, there must be a mechanism to handle the privacy preferences of each party.

1.1. Motivations and contributions
Since P3P is a well-known privacy policy framework which is widely used by a variety of enterprises and researchers, having a privacy-preserving model that enforces the policies mentioned in P3P is a necessity. This paper will have a contribution toward:
a) Identifying the challenges one may face in capturing privacy predicates in a lattice structure.
b) A pattern to capture P3P semantics and policies in a privacy-preserving model using an enforceable lattice-based structure.
The rest of this paper is organized as follows: Section 2 discusses the development of lattices that capture privacy predicates. We describe the challenges one may face in developing lattices for privacy predicates (Section 3) and provide a use case to clarify the challenges (Section 4). In the next step, we review the privacy elements described in P3P and apply a lattice-based approach to them (Section 5). Furthermore, due to the differences in the nature of P3P policies and the lattice-based privacy aware model we discuss the challenges and propose our approach in detail specification (Section 6). Finally we propose a model of transition from P3P policy into a lattice-based privacy aware model (Section 7).

2. PRIVACY REQUIREMENTS
Terminology found in the literature is often ambiguous; hence, we start by providing some definitions and their relationship to this work. The *data owner* is an individual or organization that has ownership of the data. The *data provider* is an individual or organization providing data to be stored or used for particular purposes. Data providers share personal information if they trust the organization and perceive benefits in return. The data provider may or may not be the owner of the data; however, we use the term data provider and data owner in this paper interchangeably. The *data collector* is an individual, enterprise or organization that collects data from the data providers to provide service to them. We also use the term *house* as defined by Barker *et al.* [4] because the data repository houses the data supplied by the data provider. The current ethos of most data collectors is that they can view data provided to them by default. However, since storage is now considered a "service", and is often sold as such, it is critical that a distinction between housing

and viewing be made explicit. A *third-party* is any individual, enterprise or organization that acquires the provided data from the house. Finally, a *data user* is an individual or organization that requests a data item from the database and submits the query. A data user may be a member of the house, third party or the data owner.

We review the key predicates of data privacy. This helps to identify what predicates are involved in a privacy policy that a privacy-preserving model should support. Barker et al. [4] introduce a data privacy taxonomy[1] that captures the purpose, visibility, granularity, and retention as privacy predicates.

Purpose defines the intention of the data owner for how data can be used once it has been collected (e.g., patients provide their medical data to a physician for the purpose "treatment").

Visibility defines who is allowed to see the provided data (e.g., a patient might provide medical information and allow only a family physician to access it to ensure that it should not be made visible to a third party such as a potential future insurance company). Visibility of data is an important key in controlling its use and ensuring appropriate system utility. This is especially important today when networks make data potentially available to more people than the data provider would imagine.

The granularity of data defines how much precision is provided in response to a query. A data provider has the right to specify how detailed the information provided is (e.g., the data provider could define whether an exact age is shown or a range such as child, teenager, adult).

In terms of retention, data providers should have a clear idea of how long their data is kept by the data collector. For example, in a very recent case[2], the Privacy Commissioner of Canada had concerns about the social networking site, Facebook. Their privacy policies did not clearly specify a retention condition, and they had to fix it to continue their activities in Canada.

3. LATTICE STRUCTURE OF PRIVACY

Each privacy predicate of purpose, visibility, granularity and retention can be represented in a lattice format [7].

Although using lattice facilitates managing the predicates, capturing the privacy predicates in a lattice format is a challenging task for the following reasons:

a) Due to the nature of the purpose predicate, defining a hierarchical or partial order is a cumbersome task. In other words, purpose or intention is a concept that is really hard to be materialized and related to each other. In much research [1, 2], purposes are treated as strings. In other works, purpose is captured in a tree format [3, 5].

b) Unlike the purpose predicate, visibility is easier structured in lattice. Visibility is closely related to the organizational structure of the enterprise and the third-parties that collaborate with the enterprise. Relationships between different departments of an organization may assist the Chief Privacy Officer (CPO) to design the visibility lattice. The challenge is some research work do not distinguish between roles in a Role Based Access Control (RBAC) and the visibility levels that exist in an organization and try to extend the RBAC model to capture visibility level. It should be mentioned that roles are orthogonal to the visibility levels and the data should not be visible to an individual solely for their role. Hence, the CPO should be careful not to develop the visibility lattice based on the roles but based on different levels that the collected data items must be visible to.

c) Development of such lattices is application dependant and domain specific. For instance, the visibility lattice in a health care domain contains nodes such as the medical departments, financial section, insurance companies, *etc*. whereas the visibility levels in a social networking domain may include friends, friends of friends, all the network, *etc*. Hence, there cannot be a unique set of lattices that apply to all the domains.

d) Regarding the privacy laws and legislation practiced in different geographical locations, the lattices designed for a specific domain may not necessarily be valid for the same domain in a different geographical location.

e) Storing privacy lattices in a relational database is also important for two reasons. First, the privacy predicates should be stored in system tables that helps the access control system in authorizing access to the data items according to the corresponding privacy policies. Second, lattices must be stored in a way that optimizes finding subsuming and traversing nodes.

4. USE CASE: BANKING SYSTEM

In this section we describe an example use case in a banking system. In particular, we study the privacy policies of the Royal Bank of Canada (RBC)[3] posted on their website. We do not attempt to fully capture all of the privacy predicates appeared there in a lattice format but rather draw from it some characteristics common to other banking systems.

In general, the CPO of an organization (in this case RBC) reviews the privacy practices of the bank, the business activities, organizational structure, *etc*. and derives the privacy predicates. Here is a sample of the privacy policies found on the RBC website which is illustrated in Figure 1 in a lattice format:

"We use your personal and financial information for the purposes communicated to you in your agreement(s) with us, for example to: Verify your identity; Provide you with the financial products and services requested; Communicate to you any benefit, feature and other information about products and services you have with us; Respond to any special needs or inquiries you may have; Better understand your financial situation and determine your eligibility for products and services we offer; Manage our risks and operations; Meet regulatory and legal requirements; If we have your social insurance number or social security number, we may use it for tax related purposes if you hold a product generating income and share it with the appropriate government agencies. We may also share it with credit reporting agencies as an aid to identify you."

In terms of the visibility level, the RBC specifies several parties where the customer's personal information may be shared. According to these policies, some of these visibility levels are: RBC employees, government agencies, domestic and international authorities, credit reporting agencies, outside service suppliers, and other third parties. Figure 2 illustrates a list of extracted visibility levels captured in a lattice format.

Unfortunately, most privacy policies published by the data collectors (including RBC) do not encompass elements of retention and granularity predicates. Most organizations keep the data collected from their customers for an indefinite period of time. Furthermore, the microdata provided by the customer is generally accessed by the house even if an aggregated data is sufficient for their purpose. For instance, a phone service provider might only need to know the city the customers live in and if they are over eighteen years old to determine their eligibility for a particular service. Hence, they do not need to collect the customers' exact address or exact age for that specific purpose.

While reviewing the use case example, we faced challenges that are noteworthy: First, it is more challenging to derive the purpose lattices from the privacy policies (see Section 3).The position of the nodes in the lattice also depends on one's interpretation from the policies. For example, should the purpose of *mail distribution* be under the *communicate to you any benefit*

[1] Throughout this paper we refer to it as *the taxonomy*.

[2] http://www.priv.gc.ca/media/nr-c/2009/nr-c_090827_e.cfm, access date: June 21st, 2010

[3] http://www.rbc.com/privacysecurity/ca/our-privacy-principles.html, Access date: 22nd June 2010.

or feature node or the purpose of *promote services* or both? Second, deriving lattices should be supervised by an expert in that specific domain. For instance, in Figure 1, the dashed line between the *verify your identity* purpose and the *marketing* purpose shows the unclearness about this relation.
In the RBC privacy policies this relation is not explicitly mentioned. However, an expert in this field may admit that a part of the marketing activities is a targeted marketing where the organization identifies its target audience before starting the marketing process (e.g. telemarketing).

5. P3P PRIVACY POLICIES

P3P [6] is a standard framework recommended by the World Wide Web Consortium. It was initially proposed to help data providers to understand and evaluate the privacy policies of the service providers on the Internet. An agent on the data provider's browser checks the privacy policies of the data collector in P3P and compares that with the privacy preferences of the data provider. If the policies are in compliance with the preferences, the agent allows the browser to open the website.
P3P uses an XML-based format. A P3P policy file consists of a sequence of *statements*, each containing *purpose*, *recipient*, *retention*, *data group*, and *consequence*.
A *consequence* gives a brief description of that privacy statement in a human readable form (i.e. natural language).
Purpose shows the intention for which the data will be collected and used. P3P has 12 pre-defined values available for this element. Each type of purpose (with the exception of current) can have an optional attribute called *required*. This attribute may have one of the following values:

- *Always*: The purpose is always required and the data provider cannot opt-in or opt-out of this use of their data. In the lattice-based model this is equal to a mandatory policy where the MinAL and MaxAL are defined on the same node and the data provider must agree to that.
- *Opt-in* : Data may be used for this purpose only when the data provider requests it (e.g. when a customer asks Amazon to keep his credit card information for ease of future purchases).
- *Opt-out*: Data may be used for this purpose unless the data provider requests that it not be used in this way.

Recipients are those who the data will be disclosed and made visible to them. There are 6 pre-defined values for this element. Consistent with purpose, the recipient element has an optional attribute that can have values of *always*, *opt-in* and *opt-out*.
Retention shows the expiry condition of the data after which the data will not be accessible.
Data group defines the set of data items to which the privacy statement apply. P3P also has predefined types of data items. It is also possible to assign *categories* to data items. Using categories facilitates aggregating the data items and helps data collectors to specify a category of data rather than specifying every single data item.
Although P3P is a standard that facilitates communicating privacy policies and preferences in a structured format which is readable by both machine and human, there are some issues regarding its usability: First, P3P is a specification and does not have an enforcement mechanism. Second, there is no hierarchical purpose in P3P which makes it difficult for the applications that generalize purposes. Third, some values in the privacy elements are not clear and are subject to different interpretations by data collectors and providers. For instance, the purpose of *current* essentially describes whatever the organization is doing at that moment. Pre-defined purposes limit the use of this predicate. Furthermore, since these purposes are generalized, they may cause compromising the individuals' privacy. As another example in the recipient element, the term *ours* does not clearly define who has access to the data (i.e. all the departments in an organization are treated as a whole unit). Fourth, the choices that are provided by the data collector for the privacy elements of purpose and recipient are limiting and may not truly reflect the desire of the data provider. For instance, the opt-in or opt-out condition forces the data provider to either accept or refuse the offer. Whereas, the MinAL, and MaxAL [8] attributes provide a negotiation opportunity between the data provider and collector over a range of choices.

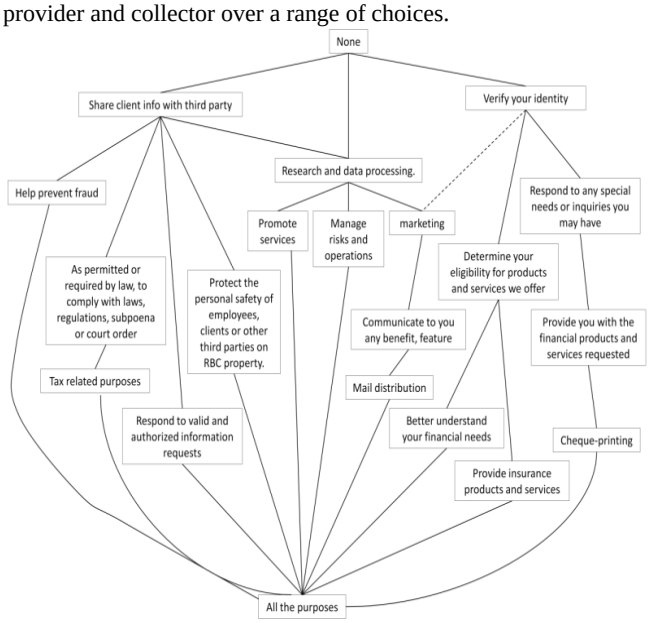

Figure 1 - A sample purpose lattice from RBC privacy policies

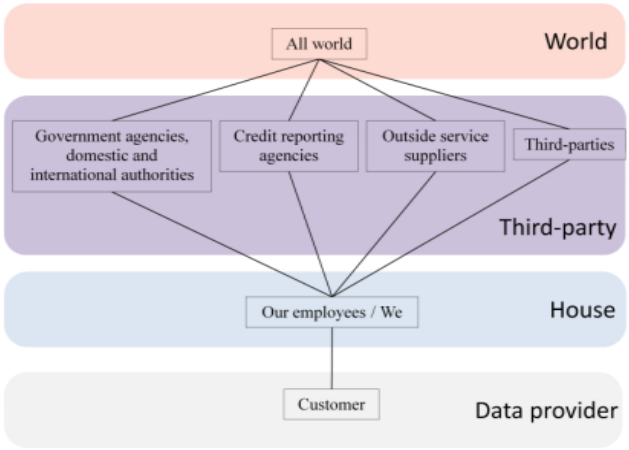

Figure 2- A sample visibility lattice from RBC privacy policies

The P3P predicates and the corresponding ones in the taxonomy are shown in Table 1.

Table 1 – P3P vs. Privacy Taxonomy predicates

P3P	Privacy Taxonomy
Purpose	Purpose
Recipient	Visibility
Retention	Retention
Data group	Granularity

6. MAPPING P3P PRIVACY ELEMENTS TO LATTICE-BASED PRIVACY PREDICATES

6.1. Purpose

As described in Section 5 there is no hierarchical relation between the pre-defined values in P3P and none of the values have priority over the others or provide a more specific purpose to use a data item. Since there are 12 purposes available and any combination of them might be used in a privacy statement, the

approach is to capture all the 2^{12} in a lattice (see Figure 3). However, there is a value called *other-purpose* in the list which gives a choice to define a purpose that is not captured by the other 11 values. Therefore, if there are *n* purposes available (including those ones defined by the data user) and a new purpose is added, then 2^n nodes will be added to the lattice. Hence, capturing the P3P purpose in a lattice format seems a proper choice for environments that have a fixed number of purposes in a static domain.

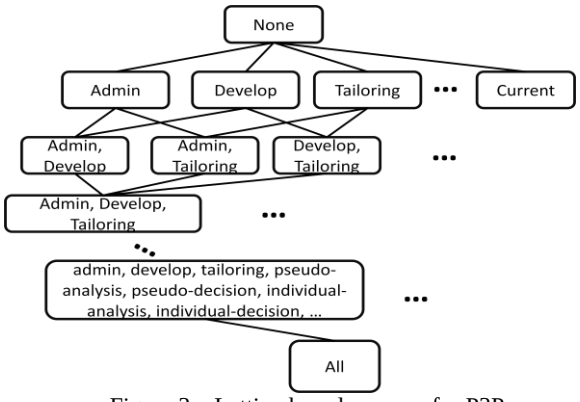

Figure 3 – Lattice-based purpose for P3P

6.2. Visibility

As shown in Table 2, P3P makes no distinction between the owner and the house. In other words, P3P assumes when the data items are provided to the house they must be visible to the house as well. However, the taxonomy emphasizes that the data providers (owners) may have the option to use services while service providers do not have access to some of their data. An example that illustrates the importance of making such a distinction is a service called iStorage that permits online storage of subscriber data on their servers where they keep the right to see the files the subscribers upload.

In terms of defining the third-party, P3P has several values that are equal to the definition of third-parties in the taxonomy. P3P distinguishes between the third-parties (i.e. some of the third-parties follow the practices of the house, some have their own practices known by the house, and some have practices that are unknown by the house). The *public* value in P3P is equal to the *All/world* level in the taxonomy which means data is visible to the public. The proposed recipient/visibility lattice for P3P is illustrated in Figure 4. Notice that if something is visible to the third-party it is not necessarily visible to the *public* but it is definitely visible to the house (*ours*) since the house provided the third-party with the data in the first place.

Table 2- Visibility vs. Recipient

Taxonomy	Owner	House	Third-party	All/world
P3P	-	<ours>	<delivery>, <same>, <other recipient>, <unrelated>	<public>

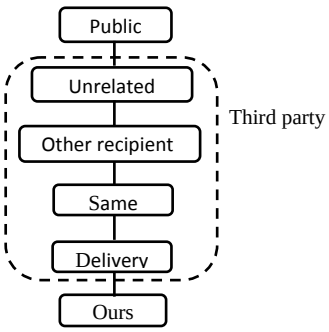

Figure 4 – Lattice-based recipient for P3P

6.3. Retention

As shown in Table 3, P3P recognizes different levels of retention for its privacy policies. Considering the definition of each retention attribute in P3P, the corresponding lattice that captures this predicate in P3P is illustrated in Figure 5. In this lattice the subsumption feature helps manage and define the retention periods. For instance, the *stated-purpose* retention is implicitly part of the *business-practices*. Obviously, the *legal-requirement* retention overrides all the retention periods, which means regardless of the type of the retention the data should be kept for the legal requirements.

Table 3 – Retention predicate in the Taxonomy and P3P

Taxonomy	now	until an expiry condition
P3P	<no-retention/>	<stated-purpose/>, <legal-requirement/>, <business-practices/>, <indefinitely/>

Notice that this is an abstract model of the retention lattice and each of the nodes are aggregated for simplicity (i.e. there might be several levels of retention condition in the *business-practices* node).

6.4. Granularity and Data group

The granularity dimension is addressed by P3P as data group. As shown in Figure 6, a group of data items can be categorized as a group in P3P, the data groups themselves do not have any particular relation to each other. For instance, *physical contact information* and *navigation information* do not necessarily relate to each other. Hence, since there is no hierarchical relationship between the data groups, using a lattice format will only facilitate handling a combination of 2^n nodes, where *n* is the number of data groups available in the system. Notice that P3P allows users to define arbitrary categories that are not captured in the pre-defined values. Therefore, like the purpose element in P3P, using a lattice format would be mainly advantageous in a static environment with a fixed number of data groups.

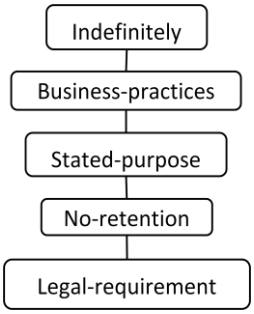

Figure 5 – Lattice-based retention in P3P

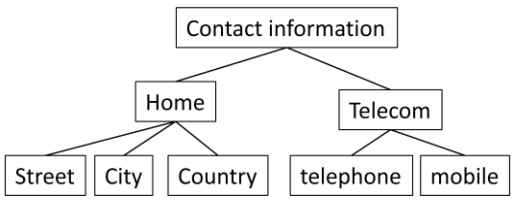

Figure 6 – A sample hierarchy of data group in P3P

7. MAPPING P3P TO LATTICE-BASED MODEL

The process of translating P3P policies to a lattice-based model such as LPAAC involves the following steps:

As illustrated in Figure 7, the first step is to parse the P3P document that reflects the privacy practices of the data collector.

Since P3P is in XML format it is easy to locate the privacy predicates by searching the corresponding tags.
The next step is building the privacy lattices and storing them in the privacy catalogues of a lattice-based model such as LPAAC. As discussed in Section 3, in these privacy catalogues privacy practices of the data collector are stored. Upon data collection, these practices are presented to the data provider and depending on their comfort level, they select their privacy preferences within the specified range (using MinAL and MaxAL) [8]. Once the data providers specify their preferences, the corresponding privacy catalogues store these preferences with the data. When a data user submits an access request of an individual, the privacy preferences of that individual on the requested data is checked and the appropriate response is given to the data user.

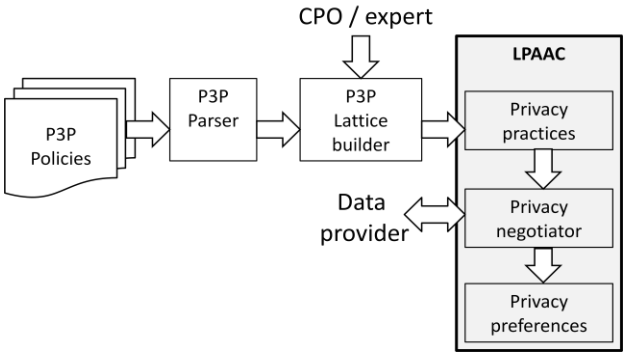

Figure 7 -Translating P3P policies to the LPAAC model

Figure 8, shows a pseudo-code that builds the privacy lattices. In the algorithm, *s, p, r, t, d,* and *i* represent the statement, purpose, recipient, retention, data group, and data item appear in a P3P privacy policy respectively.
As mentioned earlier this phase is supervised by an expert or the CPO of the organization to assist the system in building an appropriate range for the data providers to select their preferences. Due to the characteristics of P3P, as demonstrated in this algorithm, defining ranges are only possible for the purpose and the recipient predicates. If the predicate is always required it means that the data collector does not allow the providers select their own preferences and in the lattice MinAL and MaxAL are defined on the same node.
After all the values of the predicates are extracted, they are inserted in the house privacy practices catalogue (in Figure 8 we call it *SysHousePolicies*). Note that since retention period is not negotiable in the P3P framework, therefore, both MinAL and MaxAL have the same value. Finally, since granularity level is not addressed by P3P, the zero value is inserted into the privacy catalogue which means the microdata is revealed to the user.

```
∀s∀p∀r∀t∀d∀i:
        IF p.required == always THEN
            pmin = p
            pmax = p
        ELSE
            pmin,pmax = Get_From_CPO(Min,Max)
        END IF
        IF r.required == always THEN
            rmin= r
            rmax= r
        ELSE
            rmin,rmax = Get_From_CPO(Min,Max)
        END IF
        INSERT INTO SysHousePolicies(data, MinPur,
                MaxPur,MinVis,MaxVis, MinRet,MaxRet,
                MinGrn,MaxGrn)
        VALUES(i, pmin,pmax,rmin,rmax,t,t,0,0)
```

Figure 8 -An algorithm for transforming P3P policies into the privacy catalogues

As illustrated in Figure 8, the algorithm contains 6 nested loops which in the worst case scenario has the complexity of $O(n^6)$. However, in the implementation phase run time can be reduced noticeably by changing the order of the loops. For instance, if the policy is related to a particular data item then the loop that applies the policies on the data item is moved to the outer loop.

8. RELATED WORK

Barker *et al.* [4] introduce a data privacy taxonomy that describes key elements of data privacy and categorizes current research around a conceptual framework for database privacy.
After the introduction of P3P [6], much research has been done to enforce the privacy policies described with this framework. Ashley *et al.* [3] introduce the Platform for Enterprise Privacy Practices (E-P3P) which enforces enterprise-internal privacy policies. It has a formal semantic and captures hierarchies of data categories, purposes, and data users. This framework cannot capture and enforce the preferences of the data providers.
While P3P provides a statement about the promises that the data collector is making about the way it will treat the data after collection, E-P3P focuses on the enterprise side and the data providers cannot negotiate and express their preferences.
Agrawal *et al.* [1] propose a mechanism that conventional database management systems can be extended to privacy-preserving ones. Their model studies policies in a machine readable format and build restrictions on the data items.

9. FUTURE WORK

Although the proposed model is in preliminary stage, it opens future research directions such as developing proper interface for interacting with lattice builder, developing a conflict resolution algorithm, building a proper user interface for the policy negotiator and more interesting projects.

10. REFERENCES

[1] R. Agrawal, P. Bird, T. Grandison, J. Kiernan, S. Logan, and W. Rjaibi. "Extending relational database systems to automatically enforce privacy policies." In ICDE '05: Proceedings of the 21st International Conference on Data Engineering, pp. 1013–1022, Washington, DC, USA, 2005.

[2] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. "Hippocratic databases." In VLDB '02: Proceedings of the 28th International Conference on Very Large Databases, volume 28, pp. 143–154, Hong Kong, China, 2002.

[3] P. Ashley , S. Hada , G. Karjoth & M. Schunter, "E-P3P privacy policies and privacy authorization." Proceedings of the 2002 ACM workshop on Privacy in the Electronic Society, pp.103-109, November 21-21, 2002, Washington.

[4] K. Barker, M. Askari, M. Banerjee, K. Ghazinour, B. Mackas, M. Majedi, S. Pun, and A. Williams. "A data privacy taxonomy." In BNCOD: Proceedings of the 26th British National Conference on Databases, pp. 42–54, Berlin, Heidelberg, July 2009. Springer-Verlag.

[5] J.-W. Byun and N. Li. "Purpose based access control for privacy protection in relational database systems." The VLDB Journal, the International Journal on Very Large Data Bases, pp. 603-619, September 2006.

[6] L. Cranor. "The Platform for Privacy Preferences 1.1 (P3P1.1) Specification". 2006. Available at: http://www.w3.org/TR/P3P11/#RECPNT.

[7] B. A. Davey, and H. A. Priestley, "Introduction to lattices and order." Cambridge University press, pp. 33-35, 2002.

[8] K. Ghazinour, M. Majedi, K. Barker. "A Lattice-based Privacy Aware Access Control Model", *in Proceeding of the IEEE International Conference on Privacy, Security, Risk and Trust*, Canada, 2009. pp. 135 -141.