To appear in IEEE 4th International Workshop on Policies for Distributed Systems and Networks (Policy'03), Lake Como, Italy. June 4–6, 2003.

Translating Privacy Practices into Privacy Promises — How to Promise What You Can Keep

Günter Karjoth, Matthias Schunter, Els Van Herreweghen
IBM Research
Zurich Research Laboratory
{gka,mts,evh}@zurich.ibm.com

Abstract

Enterprises advertise privacy promises using the W3C Platform for Privacy Preferences (P3P). These privacy promises define what recipients can obtain what collected data for what purpose. Internally, enterprises can use fine-grained privacy practices such as defined by the Platform for Enterprise Privacy Practices (E-P3P) to enforce privacy. These internal privacy policies should guarantee and enforce the promises made to the customers. Since privacy practices reflect business internals, they can change frequently. As a consequence, it can be challenging to keep the promises up-to-date with the actual practices. To enable up-to-date privacy promises, we describe a methodology for enterprises to promise what they can keep. This is done by automatically transforming E-P3P privacy practices into corresponding P3P privacy promises that reflect the actual enterprise-internal behavior. These P3P promises can then be published on a regular basis. Whenever the internal policies change, the P3P promises can easily be updated as well.

1 Introduction

Enterprises begin to actively manage and promote the level of privacy they offer to their customers.[1] The goals are to obtain better publicity, to limit liabilities, and to comply with regulations. Visible signs of enterprises' privacy awareness are privacy statements and privacy seals. Customers can read such privacy promises explaining how collected data will be used. They can also examine the privacy seals, TRUSTe [10] for example, certifying that privacy promises exist and are accessible.

In April 2002, the World-Wide Web Consortium (W3C) standardized the Platform for Privacy Preferences (P3P) specification [4]. P3P enables Web sites to describe their data collection practices in a machine-readable XML format, which can then be read and displayed by P3P-enabled browser software or other user agents. A goal of P3P is to enable Web users to understand what data is collected by sites they visit, who can use it for what purposes, and how long it is retained. For a more detailed description of P3P we refer to [5, 7].

Whether or not the data inside the enterprise is used as promised by a P3P statement depends on the enterprise's actual privacy practices as defined by the enterprise's chief privacy officer. E-P3P is a language that aims at formalizing enterprise-internal privacy policies [1, 9]. E-P3P formalizes privacy authorization for actual enforcement within an enterprise. Privacy practices reflect the business processes and should correspond to privacy promises. Today, both are synchronized manually. Since there is no sound notion of what this 'correspondence' means, they can easily get out of sync, especially if the privacy practices change frequently.

In this paper, we show how to ensure consistency between practices and promises through an automatic transformation between privacy practices formalized using E-P3P and privacy promises formalized using P3P. This automated translation ensures that privacy promises are kept up-to-date even if privacy practices change frequently. Another benefit is that enterprises can test or detect whether changes in their practices requires changes to the privacy promises made. This is important as the customer consents to a set of promises and if the actually enforced promises differ, the enterprise may be required to obtain updated consent from the customer.

The remainder of this paper is structured as follows. Section 2 outlines our privacy policy management model. Section 3 briefly defines the languages for formalizing privacy policies: E-P3P formalizes internal privacy practices of an enterprise while P3P formalizes advertised privacy promises. Section 4 gives an example for an E-P3P policy and its corresponding P3P privacy promises. Section 5 presents the transformation procedure from E-P3P practices to P3P promises. Section 6 draws some conclusions.

[1] General introductions to privacy can be found in [2, 5, 6].

2 Managing Privacy Policies

2.1 A Typology of Privacy Policies

We distinguish two types of privacy policies: enterprise-internal *privacy practices* and published *privacy promises* (see Figure 1). Enterprise privacy practices define how data is collected, processed, and used (see Figure 1). They are required to comply with legal regulations. In addition, they need to implement the privacy goals and business processes of the enterprise. Enterprise privacy practices can be formalized using E-P3P [1, 9]. They can be very fine-grained and can define access rights down to individual employees. As a consequence, they may change frequently.

Privacy promises communicate certain privacy guarantees to the enterprise's customer. The most common form are textual privacy statements that explain what data is collected, how it is used, and what other enterprises may use it. Compared to enterprise privacy practices, they do not deal with enterprise-internals but offer a coarser-grained view, considering all the enterprise-internal data users and the enterprise's business agents as one data user. Thus, they are quite stable and change only when major revisions are made. Privacy promises can be formalized using P3P [4].

An enterprise's privacy practices should be consistent with its privacy promises, i.e., they should not allow behavior violating a promise. If, for example, an enterprise promises not to disclose customer addresses to direct marketers, the practices should ensure that this will not happen. Enterprises also want privacy promises to properly advertise good privacy practices, i.e., not to describe data usage or data disclosure that will be prevented by the privacy practices. If, for example, an enterprise never discloses data to a direct marketer then it should not ask its customers for permission to do so.

2.2 Flows of the Policy Management Model

The goal of our policy management model is to ensure consistency of published promises with frequently-changing enterprise-internal privacy practices. This is done by an automated translation of the enterprise-internal practices, specified in E-P3P, into privacy promises, described in P3P. The flows for managing policies are depicted in Figure 2, where dotted arrows denote frequent updates and dashed arrows denote infrequent updates. We now outline each depicted step in more detail.

The enterprise defines its internal terminology formalized as "E-P3P Definitions",[2] which fixes the scope of the enterprise privacy practices. To enable an automated translation, this terminology needs to be augmented with P3P-specific details that cannot be derived from the E-P3P policy. This is depicted in the box "P3P Mapping Info".

[2]The enterprise may also use a pre-defined terminology or a terminology that has been standardized in a certain sector.

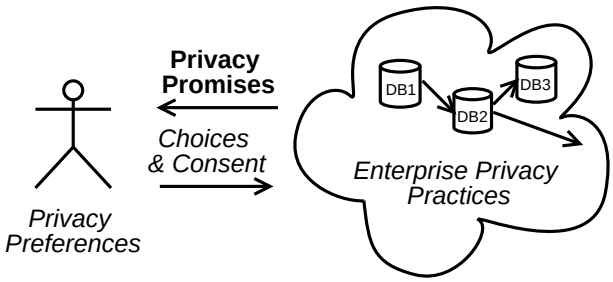

Figure 1. Privacy policy types and negotiation between individuals and the collecting enterprise.

The enterprise develops "E-P3P Rules" that formalize the legal regulations and the business practices of the enterprise. The "E-P3P Practices" result from joining definitions and rules. These formalized practices are then used as the default policy for using data and enforcing privacy throughout the enterprise. This can be done using traditional access control, E-P3P-aware business processes, or privacy-enabled access control systems such as [8].

To derive the corresponding privacy promises, the enterprise uses the mapping process defined in Section 5 to translate "E-P3P Practices" and "P3P Mapping Info" into "P3P Promises" that can be advertised to the customers. Whenever the rules change, this translation can be re-done to either verify that the changed rules had no impact on the promises or else to advertise the updated privacy promises.

3 Formalizing Privacy Policies

3.1 Common Concepts

In general, a "privacy policy" defines what data is collected, for what purpose the data will be used, whether the enterprise provides access to the data, who are the data recipients (beyond the enterprise), how long the data will be retained, and who will be informed in what cases. Privacy promises as well as privacy practices are specified by a policy language, capable to express some or all of the below elements.

Categories identify the types of data that need privacy-aware treatment. Typically, the data types used in privacy policies are high-level descriptions of data, such as customer contact information. *Data users* are parties accessing the data. The person whose data has been collected is a distinguished data user called "data subject". Granting rights to the data subject defines whether the data subject can access and/or update its personal data stored at the enterprise.

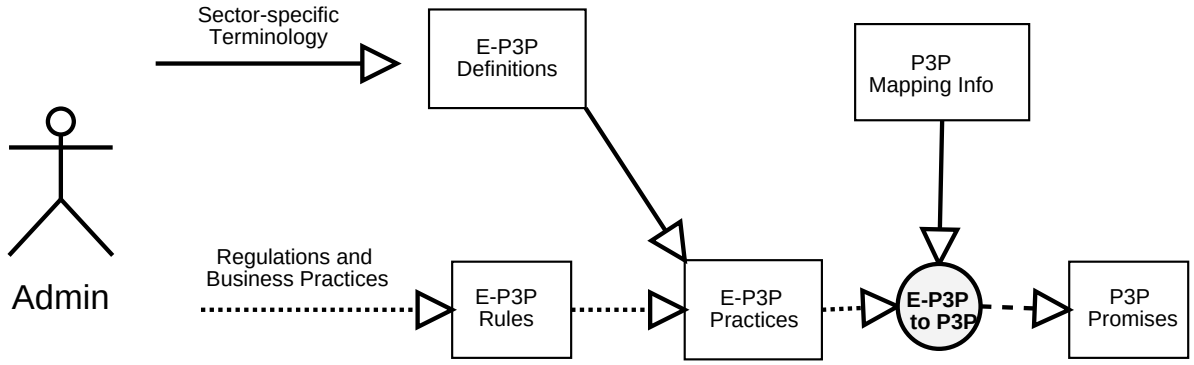

Figure 2. Flows of Enterprise Privacy Policy Management.

Actions model the actual privacy-relevant operations on the data.

Purposes explain for what reason or business purpose the collected data will be used. Most national privacy laws, codes of conduct, codes of ethics of different computer societies, as well as international privacy guidelines or directives, require the principle of purpose binding [6]:

> The purpose for which personal data is collected and processed should be specified and legitimate. The subsequent use of personal data is limited to those specified purposes, unless there is an informed consent by the data subject.

According to this principle, a subject may only access an object for a purpose for which the data has been collected. Thus, unlike in access control, a subject has to specify a *purpose* for accessing an object. A DOCTOR might be able to read a certain object for purpose MEDICATION but not for purpose BILLING.

Access rules can be qualified based upon additional *conditions* referring to specific attributes of the data user as well as of the object. For example, COPPA [3] imposes requirements on data received from persons less than 13 years of age. Another common condition in privacy policies is that the data subject must have consented before personal data can be used for a particular purpose. *Obligations* are duties imposed on the enterprise by the privacy policy, such as timely deletion of data or that all accesses against a certain type of data for a given purpose must be logged.

3.2 The Platform for Enterprise Privacy Practices

An E-P3P policy contains terminology definitions and a list of rules, sorted by priority. E-P3P definitions specify data-categories DC, purposes P, data users DU, privacy actions A, conditions $Cond$, and obligations Obl. Data-categories, data-users, and purposes are ordered in hierarchies. These elements are then used as the terminology to express privacy rules expressing what requested data accesses are allowed or denied, and under what conditions:

`ALLOW/DENY [Data User] to perform [Operation] on [Data Type] for [Purpose] provided [Condition] yielding [Obligation].`

Formally, an E-P3P rule is a tuple[3] $(dc,\ p,\ du,\ \pm a,\ o^*,\ c^*)$ with $dc \in DC$, $p \in P$, $du \in DU$, $a \in A$, $o \in Obl$, and $c \in Cond$, where x^* denotes a set of zero or more elements x. A rule defines that the data-user (and its descendants) can/cannot perform the action on the category (and its descendants) for the given purpose (and its descendants) under the conditions resulting in the obligations. If one rule allows an operation while another denies it, then denial takes precedence. For more details on E-P3P, we refer to [1].

3.3 The Platform for Privacy Preferences

P3P is an XML-based language in which privacy promises of an enterprise can be expressed. It formalizes the data collected and its use by the enterprise. Besides some general policy information,[4] a P3P policy consists of a data schema and privacy statements. The data schema defines abstract data types, called data elements, in the domain DE_{p3p} that can be organized hierarchically. Data elements are used to identify data that is collected from data subjects.

P3P defines a base data schema, re-usable structures and a set of pre-defined data types. A policy is free to define its own data schema (possibly re-using structures defined in the base data schema) or to use only elements of the base data schema. P3P also defines a set of data categories DC_{p3p}={`physical`, `demographic`, `socioeconomic`, ...}. Data elements can then be labeled with one or more categories.

Privacy statements define the permissions granted by a P3P policy. Each statement contains a group of data elements $de_{p3p} \in DE_{p3p}$ (see Appendix B for details), a

[3] For brevity, we omit the precedences in E-P3P rules, as they can be removed by pre-processing.

[4] For example, there is information about the policy's issuer, possible dispute resolution mechanisms, and whether the enterprise grants the data subject access to its data.

Data Categories	Data Users	Purposes
/all/customer/financial	/all/internal/accounting	/all/law-enforcement
/all/customer/purchase	/all/internal/sales	/all/admin-r-and-d
/all/customer/browsing	/all/internal/r-and-d	/all/service/transaction/order
/all/customer/contact/postal	/all/external/marketer	/all/service/transaction/delivery
/all/customer/contact/homephone	/all/external/deliverer	/all/service/transaction/payment
/all/business-partners/financial	/all/external/telemarketer	/all/service/crm
/all/business-partners/other	/all/external/law-enforcer	/all/service/marketing/tele
/all/anonprofiles		/all/service/marketing/non-tele

Figure 3. E-P3P data category, purpose, and data user hierarchies.

(**Category**[a]	**Purpose**	**Data User**	**Action**, **Oblig.**	**Condition**)
(/all	//order	//internal/sales	+read, -	-)
(/all	//crm	//internal/sales	+read, -	-)
(//financial	//order	//internal/sales	-read , -	-)
(//financial	//crm	//internal/sales	-read , -	-)
(//financial	//payment	//internal/accounting	+read, delete(30d)	-)
(//purchase	//admin-r-and-d	//internal/r-and-d	+read, -	-)
(//browsing	//admin-r-and-d	//internal/r-and-d	+read, -	-)
(//postal	//delivery	//external/deliverer	+read, -	-)
(//contact	//marketing	//external/marketer	+read, -	opt-in)

[a] The elements are identified using XPath [11]; "//[name]" denotes the unique node in our hierarchies with "name".

Figure 4. E-P3P Rules Reflecting the Merchant's Business Practices.

set of purposes $p_{p3p} \in P_{p3p}$ for which data is collected, a set of data users (called recipients) $du_{p3p} \in DU_{p3p}$ with whom the data will be shared, and a retention policy indicator $ret_{p3p} \in RET_{p3p}$ indicating how long the data will be stored. Note that P3P allows the same data element to occur in many statements. Optionally, a data group can be declared non-identifiable, signaling that the data will be anonymised before being disclosed to this data user. Permissions are inherited down, i.e., if a purpose by a data user is allowed on a data element, it is also allowed on possible sub-elements. A data element can be declared optional, in which case a customer can choose whether or not to provide it. A purpose can be declared as optional ("opt-in", "opt-out") or mandatory ("always"). Also recipients can be declared optional.

For the complete list of data categories, purposes, and data users, we refer to [4].

4 Example: Privacy practices and corresponding promises

In this section, we illustrate our approach by giving a E-P3P policy, which defines a merchant's privacy practices, and a P3P policy, which defines the corresponding privacy statements that can be promised to the customers.

4.1 Merchant's E-P3P privacy practices

The definitions of the E-P3P policy reflecting a merchant's business and privacy practices are depicted in Figure 3. The merchant collects data about customers and business partners. Data about customers is classified as either financial, purchase, browsing, or contact-related. Some of the customer data is used to produce anonymous profiles. Data about business partners could be financial or other.

The merchant has three internal departments, which use customer data: accounting, sales, and R-and-D. Its marketing is done by an external marketer agent. Delivery of the goods sold can be through an external delivery service. The merchant has contacts with an external telemarketer. It may send customer data to a law enforcement entity on request. The merchant has two main classes of purposes: one being related to service to individual customers, the other one related to admin, research and development. Purpose service to customers has sub-purposes marketing (tele- and non-tele-marketing), customer relationship management (CRM), and services related to the transaction (order, delivery and payment). From browsing and purchasing information, the enterprise also derives anonymous behavior profiles. These definitions usually do not change often over time.

The list of rules shown in Figure 4 reflects a simple set of permissions. The sales department can read all the cus-

Data Element	Category	Base Schema Structure
customer.financial	financial	
customer.purchase	purchase	
customer.browsing	navigation	
customer.home-info.postal	demographic, physical	postal
customer.home-info.postal.name	demographic, physical	personname
customer.home-info.postal....	...	...
customer.home-info.telecom	physical	telecom
customer.home-info.telecom.telephone	physical	telephonenum
customer.home-info.telecom....	...	...

Figure 5. Enterprise-specific P3P data schema

(Data Element	Purpose[a]	Recipient	Retention)
(customer.home-info	current, ind-a, ind-d	ours	business-practices)
(customer.purchase	current, ind-a, ind-d	ours	business-practices)
(customer.browsing	current, ind-a, ind-d	ours	business-practices)
(customer.financial	current	ours	stated-purpose)
(customer.browsing	admin, develop, pseudo-a	ours	business-practices)
(customer.purchase	admin, develop, pseudo-a	ours	business-practices)
(customer.home-info.postal	current	same, delivery	business-practices)
(customer.home-info.postal	contact(opt-in)	ours	business-practices)

[a] ind-a, ind-d pseudo-a stand for individual-analysis, individual-decision and pseudo-analysis.

Figure 6. P3P Statements Corresponding to the E-P3P Policy.

tomer data (positive rules) except for financial data (negative rule) for the purposes of CRM and order. The accounting department can read customers' financial data for the purpose of payment but has to delete the data after thirty days, as indicated by a delete obligation. The R&D department can read purchase and browsing data for admin and R&D purposes. The external delivery service can read customer postal contact data for delivery purposes. The external marketing company can read customer postal contact data for non-tele-marketing purposes if the user opted in for that purpose. The enterprise does not share any data with the telemarketing company as there is no rule allowing this.

4.2 Merchant's P3P privacy promises

In the merchant's P3P promises, we make the assumption that all customer data used by the enterprise is collected at some point. The enterprise's data schema is depicted in Figure 5; it only needs to reflect identified customer data (not the anonymous profiles or the business-partner information). The customer data set (an enterprise-defined extension of the P3P base data schema's user data set) re-uses some of the data structures ("postal", "personname") from the base data schema and inherits their subelements (and categories).

Assuming that the customer.financial data element corresponds to the E-P3P /all/customer/financial data category, customer.purchase to /all/customer/purchase etc., the statements in this P3P policy could be the ones shown in Figure 6. E-P3P purposes, data users and opt-in conditions are mapped to sets of pre-defined purposes and recipients, and opt-in declarations. All the internal departments as well as marketer are indicated with ours (ourselves and our agents). The delete obligation is translated in a retention for stated-purpose, whereas the other retention declarations (not explicitly declared in E-P3P) are assumed business-practices.

4.3 Some Observations

A typical P3P policy is more coarse-grained than the merchant's P3P policy defined above, as usually each data element (such as customer.home-info.postal) only appears in one statement. Also, it would group customer.home-info, customer.purchase and customer.browsing in one statement as their P3P statements are identical. This is the result of the fact that ours does not distinguish between different departments or enterprises' agents.

Even an a-typical P3P policy of the granularity level above, with data types closely mapping E-P3P categories, cannot be as fine-grained as its E-P3P equivalent. Whereas the E-P3P policy defines exact data users and

data flows within the enterprise, the P3P policy classifies data recipients according to notions of their privacy policy (`same`, `unrelated`), business relationship with the enterprise (`ours`), or service (`delivery`). Whereas the E-P3P policy can define an exact retention time by mandating a deletion at a certain point in time, P3P policies have generic retention classes (`stated-purpose`, `business-practices`). This requires a mapping or transformation from E-P3P to P3P to transform fine-grained to coarse-grained, and concrete and absolute to abstract and relative.

5 Translating E-P3P into P3P

Whereas E-P3P focuses on enterprise-specific enforcement, P3P focuses on enterprise-independent information. Therefore, the P3P policy stated by an enterprise should never publish better (stricter) privacy practices than actually enforced through the E-P3P policy. However, the P3P policy should always adequately reflect the E-P3P practices.

Our transformation procedure transform an E-P3P policy into a 'best-approximation' P3P policy, using a chosen (base or enterprise-specific) data schema and P3P-specific mapping information. The core of the transformation translates each E-P3P rule into an P3P statement. This transformation assumes that the E-P3P policy is 'fine-grained': it contains only positive authorizations ('allow') for all element-vectors where access is allowed and defaults to the ruling 'deny' if no rule is applicable. A fine-grained E-P3P policy can be derived from a generic E-P3P policy (with positive and negative authorizations and precedences) by pre-processing.

The P3P policy that is output by the transformation is fine-grained, too. This means that multiple statements may govern the use of the same P3P data element. A fine-grained P3P policy can be aggregated to a coarser-grained P3P policy where each data element is only governed by one statement. If applicable, we give hints for this aggregation.

5.1 Data Categories and Elements

The P3P base data schema defines four data type hierarchies (`user`, `dynamic`, `business`, `third-party`), which can be augmented by additional data schema (see Appendix B for details). The P3P categories (such as `physical`, `demographic`, `financial` ...) are flat and used as labels into these data type hierarchies. P3P policy statements about the usage of data can be applied on fixed-category data elements or variable-category data elements. A statement about a fixed-category data element `user.home-info.postal` gives information about the data type (postal contact information) and implicitly (through the base data schema) about its categories (`physical` and `demographic`). A statement about a variable-category data element such as `dynamic.miscdata` needs to be accompanied by the categories associated with it in this statement; it only communicates that this is miscellaneous data with these categories attached. In addition, the fact that multiple data elements are grouped into one policy statement specifies common collection and usage practices for these data.

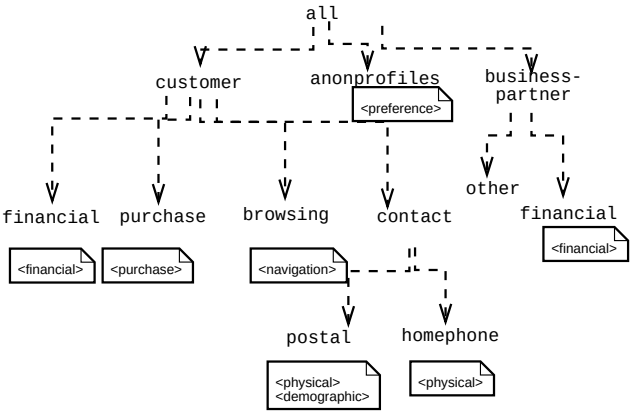

Figure 7. Mapping of E-P3P data categories to P3P data categories only

For a mapping from E-P3P to P3P, we need to express E-P3P categories in terms of P3P categories and/or data elements. A detailed projection of E-P3P rules, including obligations and conditions, to P3P entails:

1. a mapping between E-P3P categories and P3P data schema elements;
2. for E-P3P categories which do not map to P3P base data schema elements, the definition of an enterprise-specific P3P data schema is needed;
3. an assignment of P3P categories to variable-category data elements in the base data schema as well as any enterprise-specific data schema.

Some assumptions and decisions may simplify the mapping. For example, one could omit most of 1 and 2 by only mapping E-P3P categories to (one or more) P3P categories, as represented in Figure 7, where each leaf in the E-P3P data hierarchy is labeled with one or more P3P categories; categories accumulate upward in nodes: each node collects all the categories of its children (not shown in Figure 7). Using such a mapping, the E-P3P rule

`(//customer-contact-postal, //non-tele-marketing, //marketer, +read, -, -)`

allowing marketer (one of our agents) to read customer-contact-postal data for non-tele-marketing would be translated into a P3P statement

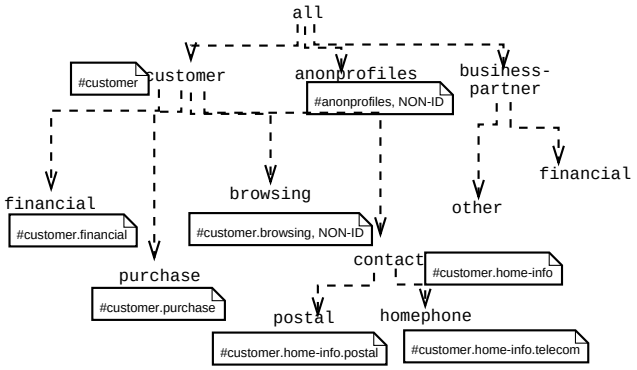

Figure 8. Mapping of data categories: Each E-P3P category is labeled with the collected P3P data elements and an optional NonID **tag.**

(dynamic.miscdata(physical, demographic), contact, ours)

about a miscellaneous data element with categories demographic and physical, purpose contact, and recipient ours. The advantage of this mapping is that it bypasses any data modeling in P3P, and the resulting P3P policy can be interpreted well by P3P user agents specialized in interpreting category information. However, it does not allow user agents to make interpretations and decisions based on data types.

To exploit the full potential of P3P, a general mapping should enable the use of data types as well as of categories. It should allow re-use of pre-defined categories as well as the other category. It should use the P3P base data schema and its category assignments but also allow for the definition of a new P3P data schema (with appropriate P3P category associations).

The most general data mapping labels each E-P3P leaf category representing *P3P-relevant* data (identifiable customer data) with zero or more P3P data types (data elements). These data elements can be taken from the base data schema or from an enterprise-specific data schema, where data elements are appropriately labeled with P3P categories. This way, we associate with each E-P3P data category the corresponding P3P data elements as well as P3P categories, giving user agents the choice whether to use only data type information, only category information, or both.

The P3P enterprise-specific data schema is depicted in Figure 5. The actual mapping information set is depicted in Figure 8 and formalized as a mapping and a subset of E-P3P categories that identify "non-identifiable" categories of E-P3P:

$$\begin{aligned} DataMap &= \{CategoryMap, NonIdentMap\} \\ &\text{with } CategoryMap \subseteq DC \times DE_{p3p} \\ &\text{and } NonIdentMap \subseteq DC \end{aligned}$$

In the example, this mapping maps rules about /all/customer/contact/postal to a P3P statement about customer.home-info.postal.

Data element labels in $CategoryMap$ accumulate upward as each node category collects its children's' data element labels (not shown in Figure 8). A non-identifiable label in $NonIdentMap$ does not propagate upward to a parent node unless all the children of the parent node are non-identifiable. The set of E-P3P categories in DE_{p3p} which are part of $CategoryMap$ contains at least the leaf elements of DE_{p3p} which correspond to P3P-relevant P3P data elements.

The policy administrator creating $CategoryMap$ may decide to also include non-leaf elements of DC in $CategoryMap$. For example, in Figure 8, /all/customer and /all/customer/contact are also labeled with P3P data elements. This later facilitates automatic aggregation of the resulting fine-grained P3P policy: Assume that the data schema also contains customer.home-info.online, but the enterprise currently does not collect this information. If the E-P3P rules about //postal and //homephone were identical, the translation would lead to identical P3P statements about customer.home-info.telecom and customer.home-info.postal. These statements could not, however, be automatically aggregated into a statement about customer.home-info as this could lead a user agent to interpret that the enterprise also collects e-mail). The node labeling indicates that such an aggregation is allowed (either because the administrator knows that this situation cannot occur, or because he wants to allow it).

Note that the mapping can be many-to-many: we cannot exclude that the enterprise's data storage system stores the same P3P data element as part of multiple E-P3P data categories. Specifically, a data element could be stored in a non-identifiable way as a part of a non-identifiable E-P3P data category, and in an identifiable way as part of an identifiable E-P3P data category. This results in multiple P3P statements about the same data. When aggregating such seemingly conflicting statements, one needs to make a worst-case approximation by retaining the stronger statements granting the maximum permission to the enterprise while discarding weaker statements. The same approach will be applied for the mapping of E-P3P data users and purposes to their P3P equivalents.

5.2 Data Purposes and Data Users

Data purpose and data user mappings are similar. Each mapping labels P3P-relevant leaf elements of the E-P3P hierarchies (data users or purposes acting on P3P-relevant data) to one or more elements from the corresponding P3P set (purposes P_{p3p} or recipients DU_{p3p}):

$$UserMap \subseteq DU \times DU_{p3p}$$

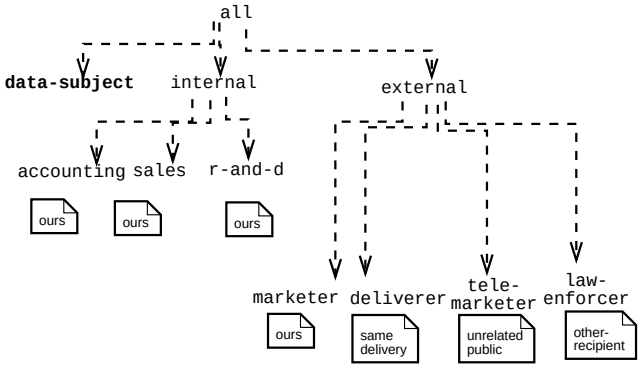

Figure 9. Data User Mapping

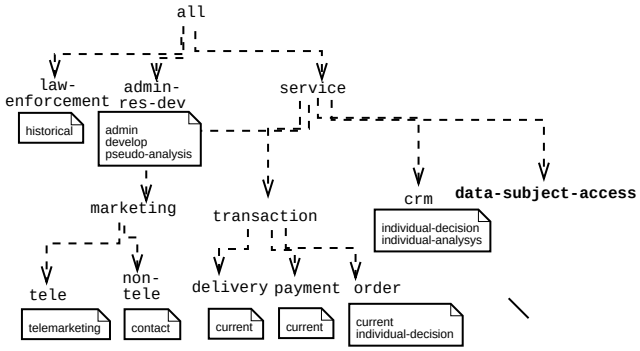

Figure 10. Purpose Mapping

$$PurposeMap \quad \subseteq \quad P \times P_{p3p}$$

As P3P only mandates purpose and recipient elements for statements about identifiable data elements, the labeling is optional for E-P3P purposes or data users acting only on `non-identifiable` data. Also here, labels accumulate upward into parent nodes.

In Figure 9, we have labeled all the internal departments `ours` as well as the marketing service, which acts as the merchant's agent. The external data users have similar, unknown, or other privacy practices. We added a distinguished purpose `/all/service/data-subject-access` and data-user `/all/data-subject` to the E-P3P hierarchies, which allows us to formulate E-P3P rules expressing data subjects' access rights (P3P ACCESS element) as discussed in Section 5.5.

Two other special purposes deserve special attention. While `non-identifiable` seems to be a feature of the data during or after collection, the P3P purposes `pseudo-decision` and `pseudo-analysis` may act on identifiable data but with the purpose of making pseudonymous decisions or building pseudonymous profiles.

In Figure 10, E-P3P purpose `/all/admin-res-dev` is labeled with `admin`, `develop` and `pseudo-decision` to indicate that all processing for this E-P3P purpose is pseudonymous. The labeling of the same E-P3P purpose with both `pseudo-decision` and `individual-decision` or `individual-analysis`, however, is defined to be semantically invalid.

5.3 Optional Data, Opt-In and Opt-Out

In both E-P3P and P3P, notions of choice and options are mixed with notion of user consent for specific usages of data. The approach taken here is that a user consents to a policy including (or modified with) specific opt-in or opt-out choices. A condition such as "if consented to by the user" then means "if the user consented to the policy and made this specific (opt-in or opt-out) choice."

In P3P, opt-in and opt-out choices are attached to purposes and/or to data users within the same statement. Each data element within a statement can also be optional or mandatory for the set of purposes and recipients in that statement.

As data use (or, in P3P, 'sharing') is always associated with a recipient and a purpose, the difference in semantics between an optional purpose and recipient disappears when considering tuples with atomic elements (one data element, one purpose and one recipient). In addition, the collection of data, the use of which is optional and not consented to by the user, should always be optional, regardless of whether it is declared as such: whether or not data collection is optional should be consistent with (choices about) its use.

In E-P3P, data subject consent (or, more specifically, opt-in or no opt-out of certain uses of data) is tied to a specific rule and thus to a combination of data category, data user, data purpose and action. The need for the presence of an opt-in choice or the absence of an opt-out choice is represented by a condition verified at run-time.

To map E-P3P's opt-in and opt-out conditions to P3P choices, we first define which E-P3P conditions are interpreted as opt-in and opt-out conditions. Let $OptMap$ define the E-P3P conditions testing the presence of opt-in and the absence of opt-out:

$$OptMap = \{opt\_in\_cond, opt\_out\_cond\}$$

When transforming a fine-grained E-P3P policy to a fine-grained P3P policy, an E-P3P rule with an opt_in_cond or opt_out_cond condition is transformed into a P3P statement with `opt-in` or `opt-out` for the stated purpose and an `optional`='yes' for the data: the data collection is optional (for this purpose and recipient). When aggregating statements about the same data into one statement, we can only assign `optional`='yes' if it is 'yes' for all occurrences.

5.4 Data Retention and Deletion

P3P uses a set of abstract values expressing how long data is retained: RET = {`no-retention`, `stated-purpose`,

legal-requirement, indefinitely, business-practices}; several of these may apply to the same data. For all the retention values other than no-retention (which is basically "current session") and indefinite, the site's human-readable policy must give more information.

An E-P3P policy specifying a retention period should enforce that the E-P3P authorization engine mandates deletion of data corresponding to the targeted retention policy. As a consequence, we mandate a 'delete' obligation to any 'store' rule about data that has a finite retention. The deletion may be conditional on consent obtained by the data subject. The transformation then uses these obligations to derive the appropriate P3P retention label to be assigned to each data element that is collected. If data can be used for several purposes, some of which are optional, and these purposes have different retention times or policies, the actual deletion of the data should occur at the maximum retention time for the purposes to which the user consented (or which were required). As consent may not be known at collection and store time, this implies that the 'store' rule execution creates delete obligations for each of the data use purposes, and that each of the delete obligations, at scheduled execution time, only actually deletes data if no other pending delete obligations for the same data for consented use exist.

For example, data /all/customer/contact/postal can be used by deliverer (for current purpose) and by marketer (for non-tele-marketing purpose). Assume retention periods for these purposes are two, respectively twelve months. Storing /all/customer/contact/postal puts two delete obligations on the obligation stack; the first one (executed after two months) will delete the data only if the user has not opted-in for non-tele-marketing, in which case also the second delete obligation is taken from the obligation stack. In P3P, the published retention in the fine-grained P3P policy will be stated-purpose for both statements. In Appendix A, we elaborate on the detailed translation of retention limitations.

5.5 Data Subject Access

The "ACCESS" element in a P3P policy describes data subjects' access (read or update) rights to identified data that has been collected from them. P3P does not specify a mechanism for it, although it seems implied that data subjects access their data by contacting a representative of the enterprise. Indeed, a real enforcement by giving concrete E-P3P access rights to data subjects is not desirable. However, we can model the notion of access in E-P3P by defining a purpose, data-subject-access for example, and a data user or role, data-subject for example, which can be used by the authorized enterprise representative to access data on behalf of data subjects (after appropriate authentication of the data subject).

Possible values of "ACCESS" are nonident (the Web site does not collect identified data), all (access is given to all identified data), contact-and-other (access is given to (some[5]) identified online and physical contact information as well as to certain other identified data), ident-contact (access is given to (some) identified online and physical contact information), and none (no access to identified data is given). To derive a P3P access statement from an E-P3P policy, mapping information has to specify which

- E-P3P data user $AccessSubject$ and purpose $AccessPurpose$ correspond to data subject access. For example, $AccessSubject$=/all/data-subject and $AccessPurpose$=/all/data-subject-access.
- values for the subsets $AccessMapAll$, $AccessMapContact \subseteq AccessMapAll$, $AccessMapContactAndOther \subseteq AccessMapAll$, $AccessMapOtherIdent \subseteq AccessMapAll$, $AccessMapIdentContact \subseteq AccessMapContact$ indicating which sets of data correspond to P3P ACCESS element values. For example,, $AccessMapAll$ = /all/customer and $AccessMapContact$ = /all/customer/contact;
- action(s) $AccessMapActions$ correspond to data-subject access. For example, $AccessMapActions = \{\texttt{read}, \texttt{update}\}$.

If the E-P3P data hierarchy contains no identifiable customer information, the value of P3P element ACCESS is nonident. Else, if $AccessMapAll$ is defined and appropriate authorizations exist for access to $AccessMapAll$ by $AccessSubject$ for executing any action in $AccessMapActions$ for purpose $AccessPurpose$, the P3P value is all. Otherwise, authorizations for the other defined sets are checked until a set is found which is defined and has corresponding authorizations.

5.6 Actions

We define as 'P3P-relevant' actions the ones that can be interpreted as 'usage' or 'sharing' in the P3P sense. This identification defines which of the E-P3P rules need to be transformed. In our example, from the E-P3P actions {read, update, store, delete}, only store and delete are relevant for retention but need not be translated. Thus, the P3P-relevant actions are $ActionMap = \{\texttt{read}, \texttt{update}\}$. In addition, a rule about an action a will only be transformed into a P3P statement if the data user is not the dedicated $AccessSubject$.

[5] Any disclosure (other than all) is not meant to imply that access to all data is possible, but that some of the data may be accessible and that the user should communicate further with the service provider to determine what capabilities they have.

5.7 Disputes, Contact and Other Policy-Specific Statements

Except for the "access" element, most of the general policy information (such as dispute and some contact information) can not or only partially be derived from the E-P3P policy and thus has to be added by the mapping information. Therefore, the mapping set $GenMap$ contains appropriate values for general policy information which is not present in E-P3P, such as: the name of the P3P policy, the location for a human-readable version, the URL for opting-in and opting-out, and information about dispute resolution and remedies: $GenMap = \{PolName, PolOptURI, \ldots\}$

5.8 The Transformation Procedure Summarized

The complete procedure for transforming a generic E-P3P policy to a corresponding P3P policy consists of the following two preparation steps that need to be done once:

1. The designer of the transformation defines the P3P data schema to be used. It may be the P3P base data schema or an enterprise-specific data schema. The mapping is easier and yields finer-grained results the more the data sets in the P3P data schema correspond to sub-hierarchies in the E-P3P hierarchy. On the other hand, re-using the base data schema should result in a better interpretation by user agents.

2. The designer of the transformation defines the different mappings. Depending on the E-P3P policy, some of these mappings may be empty: for mapping elements such as $AccessPurpose$, $AccessSubject$, $RetTimeMap$ if may be impossible to define values if the E-P3P policy was not written with retention or access goals in mind. This leads to a none value for access, and to indefinitely values for retention.

Whenever a given E-P3P policy shall be translated into P3P, this information is then used in the actual transformation. The transformation consists of the following steps:

3. The E-P3P policy is translated into a fine-grained E-P3P policy.

4. The fine-grained E-P3P policy is transformed into a fine-grained P3P policy. The general P3P policy information is extracted partially from the E-P3P policy (e.g., contact information), partially from $GenMap$; and the data schema (or a pointer to it) is inserted. Each of the fine-grained E-P3P rules with a P3P-relevant action and with a data-user not being the designated data-subject, is translated into a P3P statement where data group, recipients and purposes correspond to the P3P labels of the corresponding E-P3P elements; and where retention as well as data, purpose and recipient optionality are determined as described in Sections 5.3 and 5.4.

5. The fine-grained P3P can optionally be aggregated into a coarser-grained P3P policy. An automatic (one statement per data-element) or semi-automatic (the administrator identifying data to be grouped in a statement) data aggregation process can aggregate statements about the same or multiple data elements into one statement.

   To avoid ambiguities, the aggregation procedure may group statements about the same data by defining unions of its sub-elements (e.g., the union of two "optional" values is their logical AND, the union of "opt-in" and "opt-out" is "opt-out", the union of "opt-out" and "" is ""), make statements about parent data types resulting from equal statements for children, and group statements about groups of data collected together if so required, by using the same union mechanisms.

6 Conclusions

We presented a transformation from privacy practices stated in E-P3P to privacy promises formulated in P3P. This translation guarantees that changes of the enterprise-internal privacy practices are reflected in the corresponding P3P policy.

Whenever changes in the E-P3P do not violate the current P3P policy, the generated P3P policy should be textual equal or at least equivalent. Thus, the publication of a new privacy promise is only necessary when there is a 'fundamental' change in the privacy practices. Since the process is automated and E-P3P driven, it may not produce the 'desired' P3P statements like 'we grant data subject access to all its data'. As a consequence, it can be useful to adopt the E-P3P policy with the transformation in mind in order to achieve the desired results.

A major obstacle we had to resolve is the unclear semantics of P3P. To describe a sound mapping, we made several assumptions that fill ambiguities in the P3P specification.

We think that this transformation of policies is a first but important step into the direction of Enterprise Privacy Management, which will enable enterprises to manage privacy like they manage systems security today.

Acknowledgements

We thank Birgit Pfitzmann, Michael Waidner, and Calvin Powers for helpful discussions. This work has been partially funded by the IBM Privacy Institute (see www.research.ibm.com/privacy).

References

[1] P. Ashley, S. Hada, G. Karjoth, and M. Schunter. E-P3P privacy policies and privacy authorization. In *Workshop on Privacy in the Electronic Society*, Washington, DC, USA, Nov. 2002. ACM Press. To appear.

[2] A. Cavoukian and T. J. Hamilton. *The Privacy Payoff: How Successful Businesses build Customer Trust*. McGraw-Hill Ryerson Lim., 2002.

[3] COPPA. Children's Online Privacy Protection Act of 1998 (COPPA), October 1998. Available at `www.cdt.org/legislation/105th/privacy/coppa.html`.

[4] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle. The platform for privacy preferences 1.0 (P3P1.0) specification, Apr. 2002. W3C Recommendation, `http://www.w3.org/TR/2002/REC-P3P-20020416/`.

[5] L. F. Cranor. *Web-Privacy with P3P*. O'Reilly & Associates, 2002.

[6] S. Fischer-Hübner. *IT-Security and Privacy : Design and Use of Privacy-Enhancing Security Mechanisms*. Lecture Notes in Computer Science 1958. Springer Verlag, 2001.

[7] H. Hochheiser. The Platform for Privacy Preferences as a social protocol: An examination withoin the U.S. policy context. *ACM Transactions on Internet Technology*, 2(4):276–306, 2002.

[8] IBM Corporation. *Tivoli SecureWay Privacy Manager Version 3.6*.

[9] G. Karjoth, M. Schunter, and M. Waidner. The platform for enterprise privacy practices – privacy-enabled management of customer data. In *2nd Workshop on Privacy Enhancing Technologies (PET 2002)*, Lecture Notes in Computer Science 2482. Springer Verlag, 2003. To appear.

[10] TRUSTe. Privacy Certification. Available at `www.truste.com`.

[11] W3C. XML Path Language (XPath 1.0), 1999. Available at `www.w3.org/TR/xpath`.

A Transforming Retention Limitations

This section elaborates on the mapping and transformation achieving retention and deletion consistency between the E-P3P and P3P policies. We assume that E-P3P policy writers create the appropriate store rules and obligations.We do not consider a mapping to `business-practices`: a company can either state its business practices in the form of purposes (and thus can claim `stated-purpose` retention) or keeps the data for purposes not consented to by the data subject, in which case we consider the retention to be equivalent to `indefinitely`. The mapping is defined as follows:

- Let $RetLawMap \in P$ indicate the E-P3P purpose that is associated with law enforcement. Thus, retention for law enforcement can be treated like retention for any other purpose.
- For each pair (data category, purpose) occurring in an E-P3P rule with a P3P-relevant action (see Section 5.6), define the retention time and a human-readable explanation of the use:
  $RetTimeMap \subseteq T \times P \times \{String\} \times \{Time\}$

When transforming an E-P3P rule to P3P statement, we now proceed as follows:

- If the data in the rule is not stored by any store rule, retention is `no-retention`.
- Else if, among the possible multiple delete obligations in the data's store rule, there is an obligation to delete the data after the purpose-specified time in $RetTimeMap$, retention is `stated-purpose` (or `law enforcement` if the purpose is $RetLawMap$) with explanation of the use as in $RetTimeMap$.
- Otherwise, retention is `indefinitely`.

A P3P data-aggregation procedure can then derive the retention value for a data element occurring with different retention values as follows.

- If a data element has a retention of `indefinitely` in any of the statements, then the retention value of the grouped statement is `indefinitely`.
- Else, if the data element has a retention of `law-enforcement` or `stated-purpose` in any of the statements, these are copied into the retention for the aggregated statement.
- Otherwise, retention in the aggregated statement is `none`.

B P3P data elements and categories

Data elements in P3P can be unstructured or structured; the data schema definition facilitates building hierarchically structured elements through an associated hierarchical naming scheme (`personname.given`, `personname.suffix`, ...). Most data elements (whether at the top level of an element hierarchy or not) have categories assigned to them when defined in a data schema. Implementers can extend the data schema and its categories mapping but all P3P implementations are required to understand the P3P base data schema. There are rules about how category definitions propagate up in an element hierarchy or how a category assignment on a structure overrides categories of its sub-elements.

Figure 11 gives an overview of the `user` data-element hierarchy and category assignments of its sub-elements.[6]

[6] This is for illustration purposes only; the `String` type was added to the base elements to be UML compliant; not all of the sub-elements are depicted (for example, `user` has more sub-elements; `date.ymd` has sub-elements not depicted

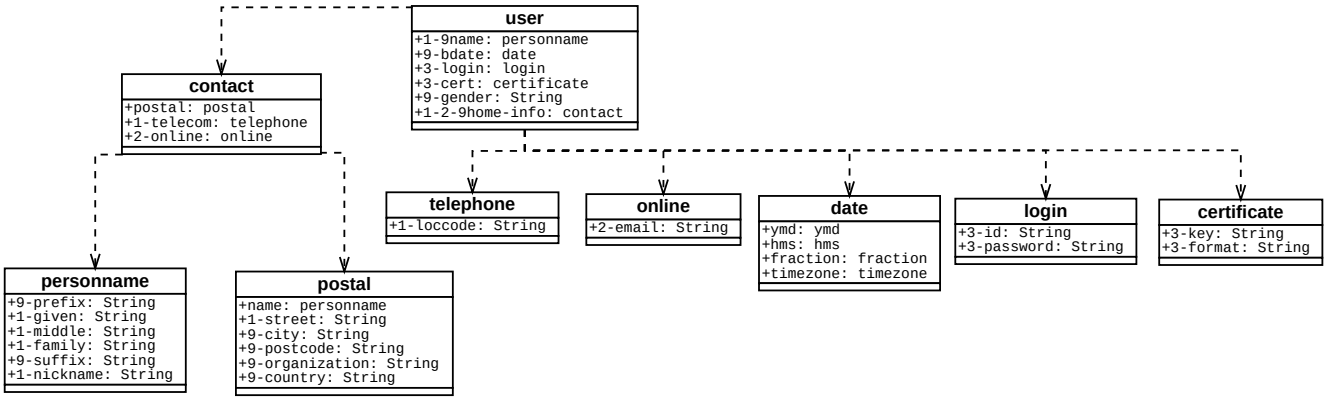

Figure 11. The P3P `user` data element

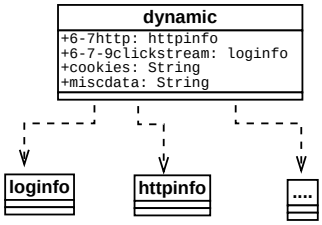

Figure 12. The P3P `dynamic` data element

The categories are represented as numerical prefixes in front of the sub-elements (1=physical, 2=online, 3=uniqueid, 4=purchase, 5=financial, 6=computer, 7=navigation, 8=interactive, 9=demographic, 10=content, 11=state, 12=political, 13=health, 14=preference, 15=location, 16=government, 17=other-category).

The P3P base data schema defines four (hierarchical) data sets. The data sets `user`, `thirdparty` and `business` include all elements that users and businesses may provide values for while data set `dynamic` includes elements that are dynamically generated during a browsing session. Elements in these data sets use the structures defined in the data schema. For example, `user.name` and `thirdparty.name` use `personname`.

The `dynamic.miscdata` element, depicted in Figure 12, references information collected by the service that the service does not reference using a specific data element. Categories have to be used to describe the data.

P3P practice statements are about data-groups, which contain data elements. Data groups can be more or less fine-grained. For example, a statement can be about `user.name.prefix` (category `demographic`, or about `user` (categories `demographic`, `physical`, `online`, `uniqueid`). P3P assumes that enterprises define their own data groups based on commonalities how data in that group is used or collected.

Actually, P3P groups or categorizes data in three different ways, and P3P user agents can derive information from all three ways of categorizing or grouping information:

- Within a data schema, P3P groups sub-elements into elements in the data element hierarchy (abstract data type hierarchy). Data elements referred to in data groups as part of policy statements refer to elements in this schema.
- Within a data schema (or also within a policy, by using variable-category elements), P3P can quite arbitrarily assign (usage) categories to elements or sub-elements (obeying certain propagation rules).
- Within a policy, P3P can arbitrarily group data elements into statements specifying collection and usage practice.

User agent interpretation of P3P policies may depend on several factors:

- How much use one makes of the P3P base data schema and how well user agents can interpret additional data schemas;
- How the elements in possible new data schemas are labeled with categories;
- How variable-category elements from the base or other data schemas (such as `dynamic.miscdata`) are labeled with categories when used in a policy statement.

Section 5.6 of the P3P specification [4] describes the pros and cons of using the P3P base data schema as opposed to newly defined data schema.