Trust and Privacy in Digital Business - TrustBus 03. In 14th Int'l Workshop on Database and Expert Systems Applications (DEXA), Prague, 2003.

Amending P3P for Clearer Privacy Promises

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

Abstract

The Platform for Privacy Preferences (P3P) can be a viable tool for organizations to clarify their privacy promises. In this paper, we summarize our experiences and describe some of the problems we have encountered when using P3P. Our main criticisms on P3P are its complicated structure, ambiguities in the specification, and missing guidelines for user agents. We suggest several improvements such as an extended but simplified syntax and a revised consent model that groups opt-in/opt-out choices into one 'consent block', which can be associated with multiple statements.

1 Introduction

In April 2002, the World-Wide Web Consortium (W3C) standardized the Platform for Privacy Preferences (P3P) specification [2]. 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 browsers. Further, users can configure their browsers to accept or reject certain types of policies. The P3P specification includes a base data schema and a standard set of uses, recipients, data categories, and other privacy disclosures. 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. Thus, like a textual privacy statement, a P3P policy should provide a precise and well-defined way for enterprises to advertise their privacy promises to their customers.

P3P has been an early sign of broadening the scope of privacy technologies. Traditional privacy-enhancing technologies enable individuals to enhance their privacy; ideally without help of enterprises or other third parties. Enterprise privacy technologies such as P3P enable enterprises to manage and protect the level of privacy that they decide to offer their customers [1, 3]. The main difference is that privacy-enhancing technologies try to maximize privacy for a single individual while enterprise privacy technology tries to maximize the benefits for an individual enterprise. This means that enterprises use privacy technologies to retain customers, to improve the quality of collected data, to maximize consented usability of collected data, and to adhere to legal regulations.

Having been controversial since its announcement in 1997, critics have decried P3P as an industry attempt to avoid meaningful privacy legislation while developers have portrayed the proposal as a tool for helping users make informed decisions [5]. The level of privacy that enterprises offer largely depends on business decisions, such as the particular market. Since P3P describes this wide range of different promises, it does not necessarily enhances or decreases the privacy of the consumer. P3P merely reflects and advertises the promises of a particular enterprise.

A major contribution of P3P is *enhanced transparency*. P3P forces enterprises to describe precisely their privacy promises and to summarize all usages of the data. Even though P3P still allows for some ambiguities, they are much less than ambiguities that can be built into a written text. For example, the written statement "we comply with a certain law" hides a variety of usages that are allowed by this law and that should be made explicit when formalizing this textual fragment into P3P.

By increasing transparency, P3P is a tool that raises privacy-awareness. As a consequence, rising customer demand for better privacy protection can lead to better promises in the long run. However, the enforcement of the promised privacy practices is out of P3P's scope. As a consequence, P3P can give customers a false sense of privacy that may lead to more data being disclosed and collected and to less privacy.

P3P aims at an organization-independent way to formalize privacy statements. It constitutes a formal language that shall enable organizations to advertise *privacy promises* in a well-defined way. To identify potential improvements, we first clarify our understanding of the scope of P3P. From our point of view, there are at least three different scenarios for privacy policies:

Data subject preferences describe the preferences of a

particular person who's data may be collected by an organization. Preferences can be formalized with APPEL [10].

Privacy promises are the privacy statements advertised by organizations. They enable a user to determine whether his data subject preferences or policy matches and whether data shall be released or submitted. Privacy promises can be formalized with P3P. However, since P3P promises can be interpreted as a legally binding statement, such statements are usually reviewed or created by the legal department and describe only basic guarantees that the organization is willing to give.

Privacy practices are the access- or privacy-control policy that governs the actual usage of data by one or more organization [8]. Such policies can be used to formalize the privacy policies associated with data that is handled inside an enterprise or exchanged between two organizations in one or two enterprises. Privacy practices are more detailed and restrictive than promises. They are usually defined and implemented by security and privacy administrators.

We claim that P3P 1.0 only partially satisfies its goal to advertise clear privacy promises. Ambiguities in the specification of the language limits P3P's usability. Section 2 summarizes criticism that we identified while trying to translate enterprise-internal privacy policies into P3P privacy promises to be advertised [7]. Section 3 describes improvements that we propose to make P3P a language for clearer privacy promises. Note that we only address observations and improvements that strengthen P3P's ability to formalize privacy promises.

2 Inadequacies of P3P

This section surveys limitations of P3P 1.0, divided into syntactical and semantical issues of the language and issues related to P3P user agents.

2.1 Syntax of P3P Privacy Statements

The core of a P3P policy is a data schema and (privacy) statements. The data schema defines data elements that can be used in privacy statements. Each privacy statement defines what data elements can be used by what recipients for what purpose as well as the retention period for the data elements. The meaning of a statement with multiple elements of each type is that *all* listed data elements can be used by *all* listed recipients for *all* listed purposes (i.e., corresponding to a cross-product). If multiple statements contain the same data element, both can be applied (i.e., implementing a union).[1]

[1] We were unable to determine whether P3P allows multiple statements for the same data element or not.

In general, P3P allows statements to be ambiguous and redundant without associating any meaning to such statements. In particular, statement elements recipient, retention, and purpose are not clearly separated.

Consent and Choice. Single elements (data-users, purposes, data-elements) inside a P3P statement can be declared opt-in or opt-out. Opting in for, e.g., a purpose allows this purpose for all other data-users and data-elements. This makes statements with multiple elements of each type and some of them being optional very hard to understand. In addition, if all recipients and purposes are optional but its collection is required, it is unclear whether the data may be collected (but not used) or whether it should not be collected in the first place.

Retention. P3P uses pre-defined labels to advertise how long data can be retained. An actual maximum time-span is only required in the human-readable text. Label `stated-purpose` signals that the data will be retained for the stated purposes. However, with multiple purposes in a statement, this means that the data can be retained as long as *any* of the stated purposes is still active. The consequence is a variety of retention times for the same data element.

Recipients and Purposes. The pre-defined values for recipients and purposes are fuzzy and mix notions of business relationship and policy. The recipient `delivery`, for example, mixes purpose and recipient. This can easily mislead user agents. The names are sometimes misleading. The purpose `contact`, e.g., conceals that the data will be used for marketing. To promote trust by the consumers, the goal of the P3P terms should be to clearly and unambiguously communicate the recipient and purposes, i.e., if a user does not want to release data for marketing, P3P must clearly express this privacy practice.

The following list describe ambiguities of pre-defined tags in the recipient element.

`delivery`: This tag mixes recipient with purpose because the recipient may use data for different practices. Delivery services "may use data for purposes other than completion of the stated purpose" or "for delivery services with unknown practices." Thus, if `purpose=current` and `recipient=delivery` then something related to the current transaction is done by a service who may use your data for any other purposes.

`same`: Giving data to an entity who uses it only once seems not to be considered as a disclosure.

`other-recipient`: The recipient can use the data in a way not specified in the service providers practices but it's in the service provider's interest that the data is not

used in a way considered abusive to the users' and their own interests.

P3P's pre-defined purposes are mainly relevant for collecting data on the web. Some unclear purposes are:

`pseudo-analysis, pseudo-decision`: The data collected is not pseudonymous but the record created from it is pseudonymous. P3P's definition of pseudonymous is "without tying identified data (such as name, address, phone number, or email address) to the record". It will not be used to attempt to specify specific individuals.

`contact`: This purpose allows the data collector to contact visitors, i.e., data subjects, "for marketing of services or products." P3P purpose 'telemarketing' is the same except that 'contact' is via telephone. The chosen name of this purpose could easily cause the misconception that allowing contact is desirable in most cases (e.g., like answering a user's request for access). Why not call it 'marketing'?

Categories. Categories are elements inside data elements that provide hints to users and agents as to the intended **uses** of the data. The multitude of categories give an impression of a lot of choice/granularity, but definitions are un-intuitive and data cannot but overlap between categories, or is at least linkable. The privacy-relevant question "what can be derived from all that collected data" is even not addressed. For example, collecting click-stream data is usually harmless unless it is used to link other data.

2.2 Unclear Meaning of a P3P Policy

A P3P policy should make clear what recipient is allowed to perform what purpose on which data element. In addition, it should define what data can be collected, whether it needs to be anonymized at collection, and how long can it be retained. Unfortunately, the P3P specification only describes the meaning of a policy that restricts itself to the most primitive case. Complicated cases, like conflicts, are not sufficiently addressed. As a consequence, each reader that implements a user agent needs to perform his own educated guess how to resolve such issues. Writing policies without knowing what they mean and building user agents that interpret meaningless policies is complex and error-prone. We identify two semantical ambiguities that we had difficulties resolving.

Overlapping Statements. The specification of P3P does not define whether the same data element can be included in multiple statements. If each data element can only be used once, the resulting policy would be very coarse grained. The following statement could no longer be expressed: " 'ours' can use the data for 'stated-purpose' and 'contact' while 'delivery' can use the data for 'stated-purpose' only.".

If we assume that the same data element can be used in different statements, P3P would need a notion of conflict resolution, which defines what statements are considered as conflicting and how conflicts are resolved. This holds, in particular, if the same data element is collected as identifiable as well as non-identifiable, or as always as well as opt-in and/or opt-out.

Pseudonymous and Anonymous Use of Data. The issue of data-anonymization is largely unclear. And what is P3P's definition of "anonymizing data"? It is unclear what pseudonymous use mean in combination with other purposes. Another open issue is what a `non-identifiable` label at a data element means if the same data element is also collected without this label.

2.3 Guidelines for User Agents

Interpretation by user agents is completely undefined. As a consequence, the meaning and the expressiveness of a policy mainly depends on the user agent interpreting it. There is no guidance or recommendations for writing 'agent-friendly' policies and how agents must interpret them. Two problems for user agents are described below.

- The P3P schema does not define whether it is desirable to re-use elements from the base data schema. If my organization's data does not fit the base data elements, I can either use `dynamic.misc` from the base data schema or I could write my own data schema that may not be interpretable by user agents.

- P3P focuses on categories instead of the exact data that is collected. A category summarizes many different pieces of information. If an organization claims to collect socioeconomic data, a user agent may assume that all socioeconomic data (including salaries) is collected even though the site collects only the gender under this category.

3 Enhancing P3P

We believe that P3P can be a strong tool to advertise privacy promises to consumers. As a consequence, all considerations for the augmented P3P should take the user-agents and their user interfaces into account. A larger fraction of our criticism can be addressed by clarifying the P3P specification. We now sketch some more general improvements that go beyond clarifications of the specification. Our proposed changes can be applied to the improvements and extensions that have been described in [4, 6, 9].

3.1 New Features

Improved Consent Model Currently, data subjects opt-in or opt-out to elements within a statement. For example, they can opt-out of a certain recipient for a given set of statements and retention policies. This implies that they automatically opt-in or opt-out to the resulting cross product with this recipient and all purposes and retentions. This is usually not what a user wants. In practice, a customer usually opts in for a abstract textual description that reflects many uses.

Since opt-in and opt-out usually correspond to certain business processes in an organization that require multiple data elements for multiple purposes, it is advisable to introduce 'consent blocks' that enable to opt-in or opt-out to a set of statements. This can be formalized by named consent descriptors that can be opt-in or opt-out and describe (in text) what the consent means. Each statement can then specify a consent descriptor. If this particular consent has been given, the statement is applicable. Otherwise, it is not applicable.

Another advantage of consent blocks is that they can reflect the actual business process. Consent could be collected one block at a time if and only if the process is actually started. Such real-time consent enables the organization to better illustrate why and how the data will be used. A user agent could then display an opt-in block only if the organization requires opt-in. For example, a policy may contain a consent-block 'newsletter', which is connected to all statements that specify usage related to distributing the newsletter of the company.

Other options are that an opt-in/opt-out choice is attached to a statement (or set of statements) or is related to one purpose or one recipient but then have the semantics of covering all statements and possibly all data. For example, "I never want any of my data to go to a tele-marketer."

Augmented elements that can record consent. Much of the important contents of a policy is placed in the human-readable policy only, real retention policies and their destruction time tables for example. This kind of information should be expressible inside optional elements of the policy.

All elements in privacy statements should be augmentable by additional qualifying information. Recipients should be augmentable by a contact address to enable the customer to find out to what organizations his data might be forwarded. The same holds for the purposes that should be augmentable with a concrete purpose and retention that should be augmentable by a concrete time-span.

An important advantage of such augmented P3P policies is that they can be fixed and context-independent. They can, for example, contain an exact business purpose and exact recipients. This is a first step toward using P3P to store privacy promises given to a particular data subject.

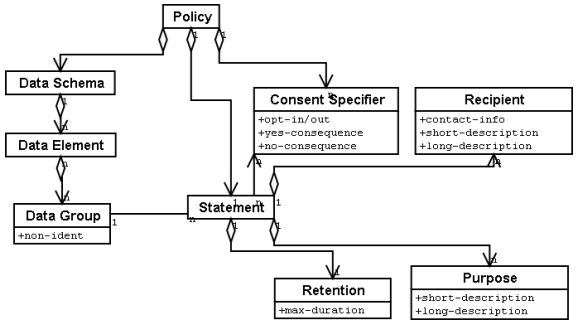

Figure 1. Simplified P3P Syntax in UML

Two additional changes improve P3P's applicability as a promise storage format: (1) consent blocks that are augmentable by the actual chosen value of a particular data subject and (2) an element with an opaque data subject identifier that can be used by organizations for promises management.

3.2 Simplified Syntax

We propose a clarified and more structured P3P syntax, which is depicted in Figure 1 and described below in more detail. The basic elements of our revised P3P syntax are a data schema that defines data elements, data groups that group data and define whether they are anonymized during collection, named consent specifiers that allow users to opt-in and opt-out to multiple statements at once, and privacy statements that define retention, recipients and purposes while identifying the required consent choices.

Simplified Statements. The syntax should either prevent or else help to identify conflicting or ambiguous statements. A simplified statement could identify a data group, a list of purposes for each recipient of the data group and an optional name of the required consent. For expressiveness, one should allow that the same elements can contain in multiple statements.

Data Schema. While categories may be useful to give hints about the type of data collected especially in case a user agent cannot interpret a data schema element, the hint may be too coarse-grained to be practical. We think that categories are not very useful without more concrete definitions of their possible contents.

Data Groups. Data elements should be grouped into named data groups that identify a set of data elements, its retention policies for each recipient, and whether they are identifiable or non-identifiable. If there is only a non-identifiable group for a data element, this data element must be anonymized at collection time. If both groups exist, data may be anonymized before being disclosed.

3.3 Specification with a well-defined Semantics

P3P should be augmented by a concise and well-documented semantics. Given a P3P policy, it needs to define answers to the following questions: "Can a given recipient use a given data element for a given purpose?" and "What is the retention policy for some data?". The semantics must answer these questions for all P3P policies that are considered to be syntactically correct. If some P3P policies are excluded (i.e., are meaningless by definition), this should be made explicit and such policies should detected by a P3P-complaint tool. One particular problem that must be addressed is how to resolve conflicts if, for example, statements overlap. The benefit of a clear semantics is that it can guarantee that all implementations of P3P (e.g., [6]) interpret P3P in the same way.

3.4 User Agents

P3P is mainly built for user agents. Today, user agents define the meaning of a P3P policy. A user agent (with slide-bar type privacy settings) is going to interpret a subset of the P3P syntax and while ignoring parts of it. If such an agent becomes a de-facto standard, this agent (and not the specification) will define what P3P policies are privacy protecting and which are not.

We would prefer if the P3P specification makes clearer statements and guidelines on what or what should not be assumed by user agents. The main purpose of user agents is to unambiguously communicate the meaning of the P3P policy. This is even more difficult without a semantics and guidelines on how to communicate a P3P policy. To enable a wide range of agent implementations, such guidelines should only describe requirements without defining the actual user-interface.

4 Conclusions

Even though P3P 1.0 is an important step toward clear and machine-readable privacy statements, it contains some ambiguities. As a consequence, it is hard to understand for user-agents or humans (such as us). We have outlined some directions that can help to simplify and improve P3P.

We identified problems (like a well-documented meaning and clearer user agent guidelines) that could be resolved by editorial clarifications of the P3P specification. Of course, these updated descriptions might still suffer from ambiguity. Thus, we currently work on a formal definition of the presented amended P3P. Other improvements (like a simplified syntax and a blocked consent model) will require changes to the P3P syntax.

In the future, we hope that P3P will be backed by comprehensive enterprise privacy policies as well as proper enforcement. This technology can then be used as a sound foundation of meaningful privacy seals. A meaningful privacy seal can actually rate the privacy promises as well as the strength and security of the enforcement system that enforces the promises throughout the enterprise.

References

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

[2] 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/`.

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

[4] R. Grimm and A. Rossnagel. Can P3P help to protect privacy worldwide? In *Int'l Multimedia Conference*, pg. 157–160. ACM Press, 2000.

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

[6] G. Hogben, T. Jackson, and M. Wilikens. A fully compliant research implementation of the P3P standard for privacy protection: Experiences and recommendations. In *Computer Security – ESORICS 2002*, Lecture Notes in Computer Science 2502, pg. 104–125. Springer, 2002.

[7] G. Karjoth, M. Schunter, and E. Van Herreweghen. Translating privacy practices into privacy promises — how to promise what you can keep. In *4th Int'l Workshop on Policies for Distributed Systems and Networks (Policy 2003)*. IEEE Computer Society Press, 2003.

[8] 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, pg. 69–84. Springer, 2002.

[9] G. Myles, A. Friday, and N. Davies. Preserving Privacy in Environments with Location-Based Applications, *IEEE Pervasive Computing*, 2(1):56–64, 2003

[10] W3C. A P3P preference exchange language 1.0 (APPEL1.0), 2002. Working Draft. Available at `http://www.w3.org/TR/P3P-preferences/`.