From P3P to Data Licenses

Shi-Cho Cha[1] and Yuh-Jzer Joung[2]

[1] csc@mba.ntu.edu.tw
[2] joung@ccms.ntu.edu.tw
National Taiwan University, Taipei, Taiwan

Abstract. P3P provides a standard means for Web sites to disclose their privacy policies when they need users' personal data for processing. A user can then decide whether or not to provide personal data to the sites based on the disclosed policies. The decision process can also be made automatic through an agent or browser via the privacy preferences set by the user. As can be seen, however, this mechanism cannot guarantee that Web sites *do* act according to their policies once they have obtained user's personal data. In light of this, we proposed a new technical and legal approach, called *Online Personal Data Licensing* (OPDL). The idea is that the use of a person's data must be authorized by the person through the issue of data licenses. Licenses can then be checked to prevent personal data from being misused. This paper focuses on the implementation of OPDL. As P3P provides a standard format for expressing privacy practices about personal data, we use it here to implement data licenses.

1 Introduction

Accompanied with the hope of industry and individuals, the first formal specifications of Platform for Privacy Preferences (P3P) was proposed by the World Wide Web Consortium (W3C) in April 2002 [1]. P3P provides a standard means for Web sites to disclose their privacy policies when they need users' personal data for processing. A user can then decide whether or not to provide personal data to the sites based on the disclosed policies. The decision process can also be made automatic through an agent or browser via the privacy preferences set by the user. As can be seen, however, this mechanism cannot guarantee that Web sites *do* act according to their policies once they have obtained user's personal data [2, 3].

This critique applies to other non-P3P compatible Web sites as well. In fact, current Web sites are usually required to disclose their privacy policies online, regardless of whether or not they choose P3P to state their policies. If a user continually uses a site's services or even registers as a member of the site, then this will be considered as an indication that he is aware of the site's policies, and has given his consent to the site for allowing the site to collect and use his personal data. This type of 'passive consent' raises lots of privacy disputes as (1) most people have neglected the disclosed privacy policies, and (2) disclosing privacy policies does not guarantee faithful execution of them.

In light of these problems, we proposed a new technical and legal approach called *Online Personal Data Licensing (OPDL)* in [4]. The idea is that the use of a person's data must be authorized by the person. By allowing people to issue licenses for the use of their data, they can control the way their data are to be used, and be more alert to the exposure of their personal data. Moreover, licenses can be used in the following ways to enforce individuals' consent about privacy practices:

– Licenses can be used to prevent personal data from being used without individuals' consent. To do this, we can request licenses to be shown when the data are to be used. For example, a person's mail server can be modified so that it only accepts an email if a license to use the person's email address is attached in the email. So when a site wants to send advertisements to a person, it must first obtain an agreement from the person. Also, a credit card issuer (such as a bank) can prevent a credit card fraud if an online retailer cannot provide a valid license to use the credit card for a transaction. Note that the license generated by the credit holder contains information to ensure that it can be used only for the transaction, so as to present security attacks such as replay and masquerade.
– Licenses can be used as evidence to see if a site has used a person's data in the same way as the site requested. In contrast, in passive consent that is generally adopted in the cyberspace, it is very hard for users to prove that a site has used their personal data in accordance with their understandings. This is because a site may change its privacy policies, and a user generally does not keep the copy of the policies he saw when he gave the site his consent.
– Licenses can be used in auditing processes to prevent data misuse. For example, an automatic procedure can be developed to check the licenses of personal data stored at a service provider. The procedure can be integrated into an internal audit system to improve their customers' confidence. Like TRUSTe [5, 6], an organization or company can also be established to serve for external auditing to increase service providers' credibility.

In this paper we focus on the implementation issues of licenses. Notice that licenses should be made machine-processable so that they can be issued, processed and checked efficiently and automatically. At this point, P3P provides a standard format for expressing privacy practices about personal data. Its rich vocabulary can be used as a foundation to build OPDL. Before presenting the details and necessary modifications, we first discuss some basic privacy principles in the cyberspace. These principles guide how an online service provider should do when it wishes to collect and use customers' data [7–9]:

– Notice: The service provider must clearly tell its customers its practices about personal data, e.g., what personal data are to be collected? How the data will be used? Whom they can contact with for complaining? And what choices they have to restrict the collection and use of their personal data?

- Choice: Its customers must have options to decide whether or not to allow the data to be collected and used.
- Individual participation: Its customers must also have right to maintain the correctness of their personal data.
- Security: The service provider must adopt reasonable security safeguards to protect collected personal data.

We can see that P3P meets only the notice and choice principles. P3P does not provide a means for individuals to access and update their personal data at the service providers. This is one of the reasons that the European Union has explicitly rejected P3P as part of its privacy protection framework [10]. Besides, when a user agent receives privacy practices in P3P format (or so called a *P3P policy* in P3P specification [1]) from a Web site, the user agent cannot know what security safeguards the Web site take to protect collected personal data. So P3P also fails to meet the security principle. Because of this, in this paper we focus on how to extend P3P to OPDL in these two directions. Other modifications to P3P that are needed to make OPDL a widely acceptable privacy standard in the cyberspace are discussed in [4].

The rest of the paper is organized as follows. Section 2 reviews P3P and OPDL. In OPDL, requests for personal data are represented as *licensing proposals*. They are based on P3P. In Section 3 we discuss how to make a P3P policy into an OPDL licensing proposal by adding security labels to the policy. Section 4 introduces users' preferences and how OPDL deals with licensing proposals based on the preferences. Section 5 discusses the implementation and verification of licenses. Section 6 presents the mechanism for individuals to update their licenses. Section 7 discusses related work. Finally, conclusions and future work are offered in Section 8.

2 Overview

2.1 P3P

The first public draft of P3P was proposed by W3C [11] in mid 1998. The original concept can be described as follows: P3P defines a vocabulary and specification for a Web site to make its privacy statements. The privacy statements are represented as machine readable "proposals" to describe what personal data will be collected by the site, for what purposes, other recipients of the data, and the destruction timetable. When a user requests a Web page (to which the user has not yet achieved a privacy agreement) from the site, a set of proposals is sent to the user. The user's agent can then choose one proposal that matches the user's preferences, and sends an agreement ID back to the site to express acceptance of the proposal. After receiving the agreement, the site will transfer the requested page to the user. If none of the proposals is accepted by the user, the site can send another set of proposals for further negotiation.

In the above process, the Web site may also request the user's data. This feature was originated from Open Profiling Standard (OPS) [12]. OPS was intended to provide privacy protection for personal profile information exchange over the Web, and was folded into the early P3P. If the user accepts a proposal, the requested data along with the agreement ID are transmitted to the site (in HTTP request header [13]). The automatic transfer of personal data raises some controversies, however. So the P3P Specification Working Group later decided to remove this function [14]. The negotiation module was also simplified due to the complexity of the original process. These two modifications have then established the prototype of current P3P.

2.2 OPDL

Generally speaking, OPDL acts as a delegate for data subjects (the owners of data) to generate licenses to use their personal data. As depicted in Fig. 1, when a requester requests a person's data, the request is sent through the *Data Licensing Interface*. Requests are more or less like a proposal, and so we refer to them as *licensing proposals*. The main components of a licensing proposal include the data that are requested, the purposes, and the retention timetable. Note that personal data may refer to dynamic click-streams a person generates while he is surfing a Web site.

The kernel of OPDL is *Personal Data Licenser*. Its implementation is not compelled to any form: e.g., it can be implemented as a software agent in a person's computer, or as service offered by a service provider or even on a peer-to-peer network. The main components of Personal Data Licenser are also shown in Fig. 1. The components include three databases: the person's data that the licenser manages, the preferences the person set on licensing, and the logs of licensing proposals (called *Access Log* in the figure). There are also six functional components. *Agreement Negotiator* deals with requests based on a person's licensing preferences. The *Notification Mechanism* deals with the following two tasks:

- To some sensitive data, data subjects can specify more strict preference policies so that the Notification Mechanism will inform them to let them make the final decision.
- When people wish to modify the content of issued licenses (including the data content), the related Internet services will be notified through their *Update Notification Interface*. The details are discussed in Section 6.

After verifying that a licensing proposal has met a data subject's preferences, the Agreement Negotiator generates a license for requested personal data. To prevent the content of the license from being altered and to ensure that the license is indeed issued by the genuine data subject, the license will be signed with the data subject's digital signature. The license (that also contains the requested personal data) is then sent back to the requester. Licensing proposals and the issued licenses are recorded in Access Log to allow data subjects to trace them through *Administration Interface*. Administration Interface provides functions for data subjects to set or modify their personal data and preferences. The actual processing of the preferences is done by the *Preferences Processor*.

Security mechanisms must be employed to protect the data. The Authentication mechanism verifies users' identities. Access Control checks if requesters have the authorization to access the requested objects. Secure channels can be established to protect communications. For example, a requester can build SSL communications with the Personal Data Licenser, and X.509 certificates can be exchanged for both parties to authenticate each other.

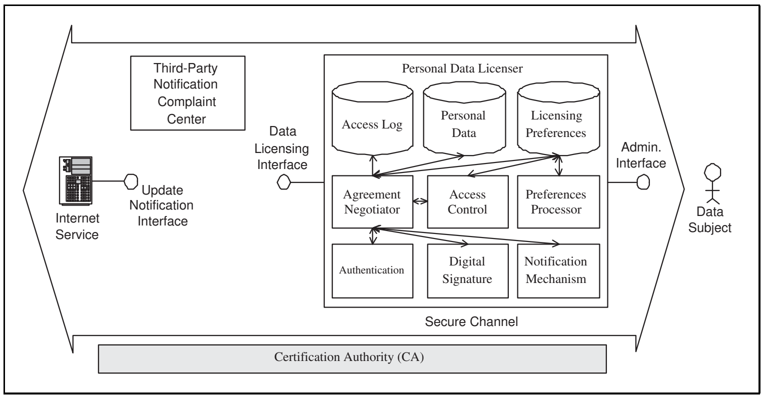

Fig. 1. Architecture of Online Personal Data Licensing.

3 Security Principle in Licensing Proposals

Recall that the security principle requires service providers to adopt reasonable security safeguards to protect collected personal data. P3P, in general, does not support this principle, and so we need to make some modification in order to adopt it into our licensing proposals. For this, a security evaluation framework is needed for users to evaluate the degree of data security in a service provider. In tradition, several such frameworks have been proposed for this purpose. For example, the Trusted Computer System Evaluation Criteria (TCSEC) [15]. TCSEC is proposed by the U.S. Department of Defense (DoD) to provide a metric to evaluate the degree of trust it can place in computer systems for the secure processing of sensitive information. TCSEC, together with other similar standards like the German Green Book and the British CLEF, was later folded into a single world-wide standard called the Common Criteria in 1994 [16].

The above frameworks are designed to evaluate computer systems. A standard is also required to evaluate and verify the data security of an organization. Some recent emerging standards such as BS 7799 and ISO 17799 enable a person to judge whether or not an organization has defined and put in place effective security processes [17]. (BS 7799 was first announced in 1995 under the auspices of British Standard Institution. It was later replaced by and internationalized as ISO 17799.) A BS 7799/ISO 17799 compliant service provider should have a well-defined policy about data security. Main security threats (risks) to the data should also be identified, and appropriate controls should be implemented to manage the identified risk.

Of course, a certification organization is needed to let individuals trust the security policy claimed by a service provider. In P3P, Web sites can specify which seal programs they have passed by giving URLs of the pages that contain the validation information. This can also be used for a service provider to specify the certification organization that has verified its security. However, P3P specification does not specify any automatic verification procedure for a user agent to decide the truth of a seal. If a user himself needs to read the validation page for verification, then this would conflict with a design philosophy of P3P—to allow its user to automate decision-making [1]. Therefore, we recommend that there should be a verification interface between certification organizations and user agents. The name of a service provider that needs to be verified can be sent to a certification organization through HTTP GET method. Then, the certification organization can transmit responses (in Web pages) with well-defined structure to the requesting user agent directly.

To illustrate how the above issues are taken into account in licensing proposals, Fig. 2 shows an example of a P3P-based licensing proposal. In this proposal, an online shopping site (exampleshop) wishes to collect some personal data from its members and to use these data for some purposes. It has another version of human readable proposal disclosed in "http://exampleshop/human-readableproposal.html". The proposal is signed by the site to avoid masquerade. In this example, the URLs of the site's security policy, risk assessments, and controls about the risks are offered in the SECURITY-POLICY element so that users can evaluate the data security of the site. In the DISPUTES element the site states that the proposal is verified by "certification.example.org".

The above security information disclosed by a proposal is generally very difficult for a user to understand. So some characterization about the degree of security safeguards a site uses should be attached in its licensing proposals. The characterization should also be machine-readable so that user agents can process it automatically. For the characterization, we borrow the concepts of divisions in TCSEC to provide individuals a simple yardstick for judging the degree of security safeguards.

In TCSEC, there are four divisions of security, A, B, C, D, among which division A has the highest level of security. Division D is reserved for information systems that cannot meet the requirements of the higher level divisions. The remaining three introduce three concepts about security policies: Division C shows the concept of 'discretionary protection' where an information system must ensure that the access of collected data are accountable, and data and users (the persons that are to use the data, not the owner of the data) are identifiable. Furthermore, audit information about what action is initiated and by whom is kept so that actions affecting security can be traced to the responsible party. Then, division B brings in the concept of 'mandatory protection'. That is, a division B-compliant information system must enforce a set of mandatory access control rules. Finally, division A is characterized by the use of 'formal security verification' methods to assure that the security controls employed by the information systems can effectively protect personal data used by the information systems.

The above features are used by OPDL and represented as tag elements of DISCRETIONARY, MANDATORY, FORMAL-VERIFIED, respectively. The tags are attached to the SECURITY-POLICY element based on what division level the site has met. For example, in Fig. 2 the site claims to have achieved the discretionary protection.

The other information in the example expresses the site's practices about personal data. This is put in the STATEMENT elements. In the example, the site needs id/password to authenticate its members. The information is retained until its members decide to opt-out (stop using the service). The site will trace its users' click-streams to secure and improve the site. The click-streams will also be used to determine their habits, interests, or other characteristics for the purpose of research, analysis, reporting, generating recommendations, and tailoring the site. These data will be destructed after two years. Similarly, some demographic data, such as birthday and gender, will also be used to provide tailored services. Finally, the site will use its members' address, telephone and e-mail address to contact with them and to provide some recommendations and advertisements.

4 Automatic Proposal Processing

To automate proposal processing in OPDL, users can set their privacy preferences so that software agents can decide whether or not to accept a proposal based on the privacy preferences a user set. The privacy preferences are presented in rules based on APPEL [18]. This section describes preference rules, and discusses how the rules can be used by a Personal Data Licenser to process a proposal. For example, Fig. 3 shows a person's privacy preferences setting. The example contains four preference rules: the first rule allows requesters to obtain the person's birth date, gender, and job title; the second rule says that a proposal will be rejected directly if the proposal is not verified by *trustcertorg* or the requested data are not to be protected under a discretionary security policy; the third rule allows 'TrustOrganization' to collect the person's click-stream for administration purpose and the data will not be retained after the stated purpose is vanished; the fourth rule says that all other information requests should inform the person to allow him to make the final decision.

In general, the components of a preference rule can be categorized into two types: (1) the actions taken when a proposal matches the rule, and (2) one or more expressions. The actions of a rule are (a) reject or accept a proposal directly, and (b) inform the user to make the final decision. An expression can be viewed as a 'predicate' asserting some subjects in an incoming licensing proposal. The possible subjects include:

```
⟨POLICY name="example proposal"
   discuri="http://exampleshop/humanreadableproposal.html" sigalgorithm="DSA"
   signature=" sbxJTQ/YWtQGf75ay/2E6ybTo51gYeInMC0CFQCHapEh+cL14zg5fY
      eyl580uj6=" date="2003/3/6 12:34:23 GMT" ID="f1099c8f2f"⟩
⟨SECURITY-POLICY discuri="http://exampleshop/securitypolicy.html"
   risks="http://exampleshop/mainrisks.html"
   controls="http://exampleshop/riskcontrol.html"⟩
   ⟨POLICY-TAG⟩⟨DISCRETIONARY /⟩ ⟨/POLICY-TAG⟩⟨/SECURITY-POLICY⟩
⟨DISPUTES-GROUP⟩
   ⟨DISPUTES resolution-type="independent"
      service="http://www.certification.example.org"
      verification="http://www.certification.example.org/verify?entity=exampleshop&
         proposalid=f31e0e5fea" short-description="certification.example.org"⟩
      ⟨REMEDIES⟩⟨correct/⟩⟨/REMEDIES⟩ ⟨/DISPUTES⟩
⟨/DISPUTES-GROUP⟩
⟨STATEMENT⟩ ⟨CONSEQUENCE⟩ The id/password used to login and access our
   Web site.⟨/CONSEQUENCE⟩
   ⟨PURPOSE⟩⟨individual-decision/⟩⟨/PURPOSE⟩
   ⟨RETENTION⟩⟨stated-purpose/⟩⟨/RETENTION⟩
   ⟨DATA-GROUP⟩⟨DATA ref="#user.login.id"/⟩
      ⟨DATA ref="#user.login.password"/⟩ ⟨/DATA-GROUP⟩
⟨/STATEMENT⟩
⟨STATEMENT⟩⟨CONSEQUENCE⟩ To secure and improve our Web site. To determine
      your habits, interests, or other characteristics for the purpose of research, analysis,
      reporting, generating recommendations, and tailoring our Web site.
   ⟨/CONSEQUENCE⟩
   ⟨RETENTION duration="2y"⟩⟨stated-purpose/⟩⟨/RETENTION⟩
   ⟨PURPOSE⟩⟨admin/⟩⟨develop/⟩⟨tailoring/⟩⟨individual-analysis/⟩⟨individual-decision⟩
      ⟨/PURPOSE⟩
   ⟨DATA-GROUP⟩⟨DATA ref="#dynamic.clickstream"/⟩ ⟨/DATA-GROUP⟩
⟨/STATEMENT⟩
⟨STATEMENT⟩⟨CONSEQUENCE⟩ To determine your habits, interests, or other
      characteristics for the purpose of research, analysis,reporting, generating
      recommendations, and tailoring our Web site.⟨/CONSEQUENCE⟩
   ⟨PURPOSE⟩⟨stated-purpose/⟩⟨individual-analysis/⟩⟨individual-decision⟩⟨/PURPOSE⟩
   ⟨RETENTION⟩⟨indefinitely/⟩⟨/RETENTION⟩
   ⟨DATA-GROUP⟩⟨DATA ref="#user.bdate"/⟩⟨DATA ref="#user.gender"/⟩
   ⟨/DATA-GROUP⟩
⟨/STATEMENT⟩
⟨STATEMENT⟩ ⟨CONSEQUENCE⟩ We will send you some advertisements you might
   be interested in.⟨/CONSEQUENCE⟩
   ⟨PURPOSE⟩⟨contact/⟩⟨telemarketing/⟩⟨/PURPOSE⟩
   ⟨RETENTION⟩⟨stated-purpose/⟩⟨/RETENTION⟩
   ⟨DATA-GROUP⟩ ⟨DATA ref="#user.name"/⟩ ⟨DATA ref="#user.home-info.postal"/⟩
      ⟨DATA ref="#user.home-info.telecom.telephone"/⟩
      ⟨DATA ref="#user.home-info.online.email"/⟩⟨/DATA-GROUP⟩
⟨/STATEMENT⟩
⟨/POLICY⟩
```

Fig. 2. An example of licensing proposal based on P3P.

```
⟨RULESET⟩
⟨RULE behavior = “request” prompt =“no”⟩
    ⟨POLICY⟩⟨STATEMENT⟩
    ⟨DATA-GROUP connective= “or”⟩
        ⟨DATA ref = “#user.bdate”/⟩
        ⟨DATA ref = “#user.gender”/⟩
        ⟨DATA ref = “#user.jobtitle”/⟩
    ⟨/DATA-GROUP⟩
    ⟨/STATEMENT⟩⟨/POLICY⟩
⟨/RULE⟩
⟨RULE behavior = “block” prompt =“no”⟩
    ⟨POLICY connective = “non-or”⟩
        ⟨POLICY-TAG⟩⟨DISCRETIONARY /⟩⟨/POLICY-TAG⟩
        ⟨DISPUTES-GROUP⟩
                ⟨DISPUTES service=“http://trustcertorg”/⟩⟨/DISPUTES-GROUP⟩
    ⟨/POLICY⟩
⟨/RULE⟩
⟨RULE behavior = “request” prompt =“no”⟩
    ⟨POLICY⟩
        ⟨ENTITY⟩⟨DATA ref=“#business.name”⟩TrustOrganization⟨/DATA⟩
        ⟨/ENTITY⟩
        ⟨STATEMENT⟩
            ⟨PURPOSE connective=“or-exact”⟩⟨admin/⟩⟨/PURPOSE⟩
            ⟨RETENTION connective=“or-exact”⟩⟨stated-purpose/⟩⟨/RETENTION⟩
        ⟨/STATEMENT⟩
    ⟨/POLICY⟩
⟨/RULE⟩
⟨RULE behavior = “limited” prompt = “yes”⟩
    ⟨OTHERWISE/⟩
⟨/RULE⟩
⟨/RULESET⟩
```

Fig. 3. A user's privacy preferences setting.

– Requested data, such as a user's gender (with name #user.gender), birthday (with name #user.bdate), or job title (with name #user.jobtitle).
– The name of the requester.
– The certification organizations of the requester.
– The security requirement adopted by the requester.
– The purposes of the request.
– The retention policies of the requested data.

An expression is evaluated to TRUE/FALSE depending on if the 'predicate' asserted by the expression is true or not. For example, an expression ⟨DATA ref = "#user.gender"/⟩ is TRUE if a licensing proposal does request the user's gender. If a rule contains more than one expression, the evaluated values are logically combined based on the *connective* attribute of the rule. A proposal matches a rule if the result of the rule is TRUE.

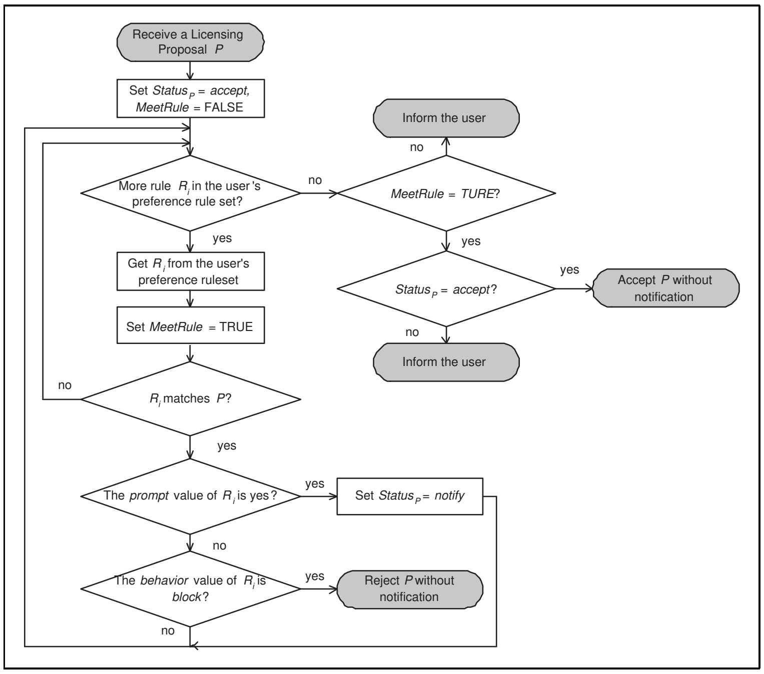

Fig. 4. Proposals processing.

Finally, Fig. 4 shows how a Personal Data Licenser processes a proposal based on a user's preference rules. Basically, if any one of the rules blocks the proposal, then the proposal is rejected. A notification form will be generated and sent to the user if the action of a matching rule is to notify the user (in the *prompt* attribute of the rule). The details about a notification form are given in [4]. Moreover, a notification form will also be generated if no preference rule has been set by the user.

5 Licenses

This section discusses the design and implementation issues of licenses. Fig. 5 shows the logical view of a license. Suppose that a license is issued from a person X (with private key SK_X) to a service provider Y. The main components of the license include a header H and a number of clauses $C_1, \ldots, C_n$. The header contains the general information about the license, e.g., the licenser, the licensee, the time the license is issued, and the security level claimed by the licensee. The clauses describe the licenser's allowed practices about personal data. Clauses are generally independent, meaning that one clause can be extracted from the license without effecting the integrity of another. The extracted clause can also be treated as a license (by appending an appropriate header). This decomposable feature has the following benefits:

- An audit or verification mechanism may only need part of the license, and so there is no need for a service provider to present the whole license. For example, when using email address of a user, the user's mail server may only need to verify if the service provider has the license to use his email address. Providing only the necessary part of a license speeds up the verification process, and may also protect the privacy of the licenser.
- If a person wishes to update some part of his issued license, the person can update related clauses (see Section 6) instead of reissuing the whole license.

For each clause C_i, it contains the allowed privacy practices about a set D_i of data items; that is, the purposes P_i for using the data, with whom (S_i) the information may be shared, and the destruction time T_i of the data (or validation period of the clause). In addition, the person may also assign a set of values V_i to D_i. (Some data item may not have an initial value to be provided by the licenser, e.g., behavior logs and click streams).

Finally, to prove that the license is indeed issued by the person and that clauses are not altered, each C_i together with the license header H is signed with the licenser's private key SK_X. Therefore, the clauses can be verified separately as discussed above.

Licenses can be implemented based on the vocabulary of P3P. An example of a license based on the proposal in Fig. 2 is given in Fig. 6. As shown in this figure, a license includes the following components:

First of all, the whole license is stored in an XML document. Each license has a unique ID as its name. To make sure that the name is unique, it is composed of the licenser's unique identity and the timestamp when the license is issued. The name will then be used for update notification.

In the header of the license, ISSUER and HOLDER show the licenser and the licensee. ISSUE_DATE records the time the license is issued. It can then be

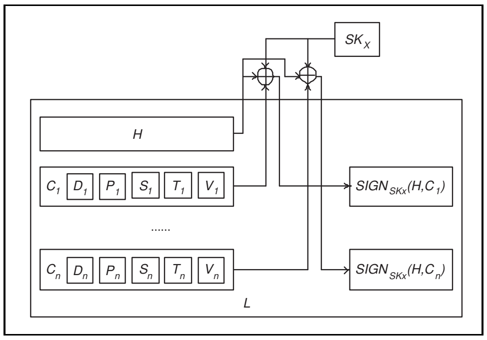

Fig. 5. Logical view of a license.

used to calculate the validation period of the license. Furthermore, SECURITY-LEVEL describes the security level that the licensee adopts to protect the target data in this license.

License clauses are represented as CLAUSE elements in the LICENSE_BODY. Basically, other than some validation information, the CLAUSE element has the same structure as the STATEMENT element in the licensing proposal. The signature of each clause is encoded in BASE64 format and is attached as attributes in the clause for verification. Note that an XML document usually has multiple representation forms. For example, an empty element (e.g., ⟨CONSEQUENCE /⟩) can also be represented as a start-end tag pair (e.g., ⟨CONSEQUENCE⟩⟨/⟩). Therefore, contents of licenses must be canonicalized [19] while licenses are signed and verified. Finally, the default values of data items can be set as contents of the corresponding DATA elements in the license.

6 Update Notifications

To meet the individual participation principle, individuals can inform licensees of the update to the contents of their licenses. To achieve this, OPDL-compatible services should provide an Update Notification Interface for receiving the notifications of updates so that individuals can send update notifications to them.

The detailed process of an update notification can be depicted in Fig. 7. If a person X wishes to update his license issued to a service provider Y, he can send an update request U_X to the service provider through his Personal Data Licenser. The request may belong to one of the following types:

- addLicenseClause(L_{ID}, C_{new}, $SIGN_{SK_X}(H, C_{new})$): To add a new clause C_{new} into a license with ID = L_{ID}. The content of the new clause C_{new} and its signature $SIGN_{SK_X}(H, C_{new})$ is also transmitted. These data can be appended to the original license because clauses of a license are mutually independent.
- deleteLicenseClause(L_{ID}, C_{ID}): To invalidate a clause C_{ID} in a license.

⟨LICENSE ID="CSC_f1099c5ea6"⟩ ⟨LICENSE_HEADER⟩
⟨ISSUER⟩⟨NAME⟩Shi-Cho Cha⟨/NAME⟩⟨CERT_SERIAL⟩12345678
⟨/CERT_SERIAL⟩ ⟨/ISSUER⟩
⟨ISSUE_DATE⟩Fri Jul 14 16:16:58 GMT+8:00⟨/ISSUE_DATE⟩
⟨HOLDER⟩⟨NAME⟩exampleshop⟨/NAME⟩
⟨CERT_SERIAL ⟩87654321⟨/CERT_SERAL⟩⟨/HOLDER⟩
⟨SECURITY-REQUEST⟩⟨DISCRETIONARY /⟩⟨/SECURITY-REQUEST⟩
⟨/LICENSE_HEADER⟩
⟨LICENSE_BODY⟩
⟨CLAUSE sigalgorithm="DSA" signature="MC0CFQCHapEh+cL14In5fYeyl580uj6
tcwIUP/ZVsOFg64zw/1F7 waTo51gWFzg=" ID="f1099c8f2f"⟩
⟨CONSEQUENCE⟩ The id/password used to login and access our Web site.
⟨/CONSEQUENCE⟩
⟨PURPOSE⟩⟨individual-decision/⟩⟨/PURPOSE⟩
⟨RETENTION⟩⟨stated-purpose/⟩⟨/RETENTION⟩
⟨DATA-GROUP⟩⟨DATA ref="#user.login.id"⟩ shichoc ⟨/DATA⟩
⟨DATA ref="#user.login.password"/⟩⟨/DATA-GROUP⟩
⟨/CLAUSE⟩
⟨CLAUSE sigalgorithm="DSA" signature="MC0CFQCWsmkv6kpEKCeKhVswPDb
R9TN7GgIUP3k8rDZPlahL0QkCv0BizABFeiI=" ID="f10b83dbb9"⟩
⟨CONSEQUENCE⟩ To secure and improve our Web site. To determine your
habits, interests, or other characteristics for the purpose of research, analysis,
reporting, generating recommendations, and tailoring our Web site.
⟨/CONSEQUENCE⟩
⟨RETENTION duration="2y"⟩⟨stated-purpose/⟩⟨/RETENTION⟩
⟨PURPOSE⟩⟨admin/⟩⟨develop/⟩⟨tailoring/⟩⟨individual-analysis/⟩
⟨individual-decision⟩⟨/PURPOSE⟩
⟨DATA-GROUP⟩⟨DATA ref="#dynamic.clickstream"/⟩
⟨/DATA-GROUP⟩⟨/CLAUSE⟩
⟨CLAUSE sigalgorithm="DSA" signature="MCwCFFKcscvKr5PJApq+zcuSGuqcc6
Z4AhQWxV4hOxAMPKtfZAKhZPTd6qZbYQ==" ID="f11e0955e5"⟩
⟨CONSEQUENCE⟩ To determine your habits, interests, or other characteristics
for the purpose of research, analysis, reporting, generating recommendations,
and tailoring our Web site.⟨/CONSEQUENCE⟩
⟨PURPOSE⟩⟨stated-purpose/⟩⟨individual-analysis/⟩⟨individual-decision⟩
⟨/PURPOSE⟩
⟨RETENTION⟩⟨indefinitely/⟩⟨/RETENTION⟩
⟨DATA-GROUP⟩⟨DATA ref="#user.bdate"/⟩⟨DATA ref="#user.gender"⟩
Male ⟨/DATA⟩⟨/DATA-GROUP⟩
⟨/CLAUSE⟩
⟨CLAUSE sigalgorithm="DSA" signature="MCwCFHzXv5zI9wTqd5Dbg+wkr1tG7A
opAhQVaaK88Wq8JJYGTAi6472MFc8SCw==" ID="f19068a0e6"⟩
⟨CONSEQUENCE⟩ We will send you some advertisements you might be
interested in.⟨/CONSEQUENCE⟩
⟨PURPOSE⟩⟨contact/⟩⟨telemarketing/⟩⟨/PURPOSE⟩
⟨RETENTION⟩⟨stated-purpose/⟩⟨/RETENTION⟩
⟨DATA-GROUP⟩ ⟨DATA ref="#user.name"⟩ Shi-Cho Cha ⟨/DATA⟩
⟨DATA ref="#user.home-info.postal"/⟩
⟨DATA ref="#user.home-info.telecom.telephone"/⟩
⟨DATA ref="#user.home-info.online.email"⟩csc@eland.com.tw⟨/DATA⟩
⟨/DATA-GROUP⟩
⟨/CLAUSE⟩⟨/LICENSE_BODY⟩ ⟨/LICENSE⟩

Fig. 6. A license based on the licensing proposal in Fig. 2.

- updateClause(L_{ID}, C_{ID}, C_{new}, $SIGN_{SK_X}(H, C_{new})$): To update the content of a clause in a license.
- withdrawLicense(L_{ID}): To withdraw a license.

The P3P vocabulary can be used in the same way as in licenses to express an update request, and so a detail demonstration is skipped here.

After receiving an update request, the service provider signs its response R_{U_X} and the update request with its private key SK_Y. The response along with the signature $SIGN_{SK_Y}(R_{U_X}, U_X)$ are transmitted back to the person. As in data licenses, the signature can be used to solve disputes between the person and the service provider about the genuineness of the update request.

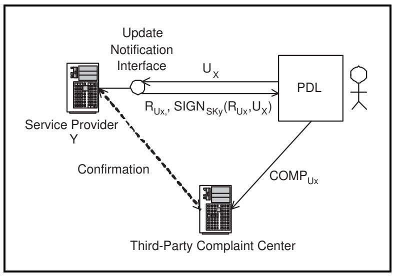

Fig. 7. Update notifications.

In some situations, updates to licenses must be guaranteed, e.g., the revocation of a credit card. A malicious service provider may deny an update request by ignoring response to the request. A Third-Party Notification Complaint Center is therefore required here to deal with the complaints from individuals and to distinguish the situation in which the service provider's Update Notification Interfaces is transiently unavailable, from the situation in which the service provider has ignored the update request on purpose. To do so, when a person does not receive the response to its update request from the service provider after a number of attempts, the person sends his complaint $COMP_{U_X}$ about the update request to the Notification Complaint Center. The Notification Complaint Center then takes the responsibility of confirming the truth of the complaint. Some remedies may be taken if the complaint has been confirmed.

One may wonder who will serve for the complaint service. We first observe that current seal programs also have similar functions. For example, TRUSTe's WatchDog Dispute Resolution Service [5] enables a person to file a complaint about a TRUSTe-approved Web site. If the complaint has been confirmed, TRUSTe may require the site to change its stated privacy policy, to remove the site from the TRUSTe program, or even to sue the site for breach of its contract with TRUSTe. For OPDL, as described before, a third-party certification organization plays an important role in verifying the faithful execution of a site's proposal and its security policy. The organization can also serve for the complaint service.

Note that depending on whether or not licenses can be transferred (directly or indirectly), license revocation may become a complicated process. For example, suppose a service provider Y has a license L_y that is dependent on a license L_x hold by X. Then, a revocation of L_x at X may incur the revocation of L_y at Y. That is, a revocation may trigger a sequence of revocations at different sites. This type of chained revocations will be considered in the future.

A more fundamental question is whether or not we should allow a license to be modified or even withdrawn after it is issued. In fact, the answer may depend on the laws of a country, as well on the content of the license. Here we discuss only the general concept of data licenses. Their applications to specific circumstances are left to the future work.

7 Related Work

Digital Right Management (DRM) technologies [20] (e.g., Windows Media Right Manager (WMRM) [21] and IBM's Electronic Media Management System (EMMS) [22]) have recently been developed as a protection against the illegal distribution of copyrighted online publications such as music, e-books and videos. The technologies embed copyright information into online publications so that users must obtain licenses in order to read the publications. The publications may also be encrypted so that they can only be accessed based on the keys in the licenses. Obviously, licenses play an important role to allow a system to automatically check whether or not end users have permissions to access the content of a publication. Several standards of licenses have also been designed for this purpose, e.g., the eXtensive Media Commerce Language(XMCL) [23] (used by IBM's EMMS), and the eXtensive rights Markup Language (XrML) [24] (used by Microsoft's WMRM).

Both DRM and OPDL aim to regulate the use of data. However, because their targets are different (DRM on digital media, while OPDL on personal data), there are differences between the two. In particular, the types of elements in personal data are far more complex than that of digital media. In fact, personal data may contain digital media (e.g., a person's portrait or its voice self-introduction), but not vice versa. Because of this, licenses in OPDL need to deal with more issues than licenses in DRM, especially on privacy. For example, a license in OPDL may involve several elements (e.g., email address, credit card number, pictures), and different applications may need only to verify the license of some of the elements. Therefore, the decomposition feature of licenses proposed in Section 5 is important to protect personal data from being unnecessarily exposed to different verification mechanisms. In contrast, a license in DRM is usually for a whole, non-decomposable digital publication.

Moreover, because of the complexity of personal data, several issues (e.g., the purposes of using the data, security levels of the requesters, and update notifications) need to be taken into account when issuing a license. So a licensing proposal needs to be presented and verified before issuing a license.

8 Conclusions and Future Work

This paper discusses the implementation issues about OPDL. Generally speaking, OPDL concretizes a user's consent on allowing another people to use his personal data by letting the user to issue his own licenses to use the data. Compared to P3P, OPDL not only lets individuals know the privacy practices of a Web site, but can also legally enforce the practices.

Our future work includes the following directions. First, the success of OPDL relies on legislation: *giving the licenses the evidence power in law and requiring non-owners of personal data to obtain licenses from their owners before collecting, processing, and using the data*. How the laws should be made requires deliberate integration between information technologies and jurisprudence, and remains an important and interesting future work.

Second, we observe that OPDL focuses on the collection and process of personal data in the Internet. However, Internet is not the only source of personal data for enterprises to collect. For example, they may request their customers to fill in paper with their personal data for communication or other purposes. Other forms may also be used in the physical world. This brings up an integration problem about the exchange and use of licenses in different formats, which will be considered in the future.

Moreover, because OPDL adopts P3P for users to express their privacy preferences, it also takes the critique about the specification: too complex for users to manage their privacy preferences. Therefore, a friendly user interface or a set of default preference settings is needed to remove this concern.

The current prototype of OPDL we have designed provides only a simple request and reply architecture. More complex features can be added in the future. For example, a negotiation mechanism can be designed for requesters and Personal Data Licenser to reach an agreement both parties can accept. Moreover, personal data may be treated as properties [25]. So a pricing mechanism may then be investigated for requesters to pay for the license fee (i.e., to lease personal data). For example, people may charge commercial e-mail senders for using their e-mail address to send advertisements. OPDL also opens up another direction for future work: design and develop applications to incorporate data licenses. For example, email servers can be modified to filter out unlicensed use of email addresses.

References

1. Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M., Reagle, J.: Platform for Privacy Preference (P3P). In: W3C Recommendations. (2002) Retrieved from `http://www.w3c.org/TR/P3P/`.

2. EPIC, Junkbuster: Pretty poor privacy: An assessment of p3p and internet privacy. http://www.epic.org/reports/prettypoorprivacy.html (2000)
3. Isenberg, D.: The GigaLaw—Guide to Internet Law. Random House Trade Paperbacks (2002)
4. Cha, S.C., Joung, Y.J.: Online personal data licensing. In: The 3rd International Conference of Law and Technology (LAWTECH2002). (2002) 28–33
5. TRUSTe: Retrieved from `http://www.truste.org`. (2002)
6. Benassi, P.: TRUSTe: an online privacy seal program. Communications of the ACM **42** (1999) 56–59
7. Organization for Economic Cooperation and Development (OECD): Guidelines on the protection of privacy and transborder flows of personal data. Committee for Information, Computer, and Communication Policy (1980)
8. U.S. Federal Trade Commission: Privacy online: a report to congress. Retrieved from: http://www.ftc.gov/reports/privacy3/index.htm (1998)
9. U.S. Department OF Commerce: Safe harbor privacy principles. http://www.export.gov/safeharbor/SHPRINCIPLESFINAL.htm (2000)
10. European Comission: Platform for privacy preferences and the open profiling standard. Draft opinion of the working party on the protection of individuals with regard to the processing of personal data. http://www.epic.org/privacy/internet/ec-p3p.html (1998)
11. World-Wide Web Consortium: W3C publishes first public working draft of P3P 1.0. http://www.w3.org/Press/1998/P3P (1998)
12. Hensley, P., Metral, M., Shardanand, U., Converse, D., Meyers, M.: Proposal for an open profiling standard. In: W3 Consortium (available as `http://www.w3.org/TR/NOTE-OPS-FrameWork.html`). (1997)
13. Kristol, D.M.: HTTP Cookies: Standards, privacy, and politics. ACM Transactions on Internet Technology (TOIT) **1** (2001) 151–198
14. W3C: Removing data transfer from P3P (1999) Retrieved from `http://www.w3c.org/P3P/data-transfer.html`.
15. US Department of Defense: Trusted Computer System Evaluation Criteria. Technical Report 5200.28, US Department of Defense (1985)
16. Kaufman, C., Perlman, R., Speciner, M.: Network Security: Private Communication in a Public World. Prentice Hall (2002) ISBN: 0-13-046019-2.
17. Calder, A., Watkins, S.: IT Governance: Data Security & BS 7799/ISO 17799. Kogan Page Ltd. (2002) ISBN: 0-7494-3845-2.
18. Cranor, L., Langheinrich, M., Zurich, E.: A P3P Preference Exchange Language 1.0 (APPEL1.0). In: W3C Working Draft. (2002) Retrieved 20 Aug. 2002 from `http://www.w3c.org/TR/P3P-preferences.html`.
19. Boyer, J.: Canonical XML. W3C Recommendation Version 1.0, W3C (2001)
20. Sonera Plaza Ltd MediaLab: Digital Rights Management white paper. Technical report, Sonera Plaza Ltd (2002) http://www.medialab.sonera.fi.
21. Microsoft Corporation: Windows Media Rights Manager 9 series - Live DRM. Technical report, Microsoft Corporation White Paper (2002) http://www.microsoft.com/windows/windowsmedia/drm/livedrm.pdf.
22. IBM Corporation: Electronic Media Management System. Technical report, IBM Corporation (2000) http://www-1.ibm.com/industries/media/pdf/emms_brochure_in_english.pdf.
23. Ayars, J.: XMCL - the eXtensible Media Commerce Language. W3c note, W3C (2002)
24. ContentGuard: XrML 2.1 overview. Technical report, ContentGard (2002)
25. Lessig, L.: Code and other Laws of Cyberspace. Basic Books (1999)