

**A PRIVACY FRAMEWORK TO PROVIDE USERS WITH CONTROL,
ACCURACY AND AUDIT**

Maksym Nozin

Thesis

submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of Master of Computer Science

April 3, 2005

Ottawa-Carleton Institute for Computer Science
School of Information Technology and Engineering
University of Ottawa

© Max Nozin, Ottawa, Ontario, Canada, 2005

Abstract

With the introduction of the Canadian PIPEDA act and similar privacy legislation around the world, it is becoming increasingly important for the e-commerce industry to provide a mechanism for ensuring the protection of privacy while still enabling automatic disclosure of personal information as needed to facilitate on-line processing.

In privacy protection the interests of consumers and businesses can be opposite and conflicting. For consumers, it is important to achieve the highest degree of protection while maintaining usage convenience. For businesses it is important to maximize the potential return from usage of personal information and at the same time to avoid possible legal consequences in connection with improper handling of personal information.

This thesis defines an architectural framework which addresses the interests of both groups. It provides users with control over their personal information and the manner in which it can be used by businesses, as well as an ability to verify the accuracy of that information and to audit the manner in which it is used by businesses. However, it does so in a framework that enables efficient mechanisms for providing consent to businesses to enable fast and flexible access to information that businesses are allowed to use.

The premise of the framework is that businesses form a circle of trust which is a business network of trusted entities cooperating in a B2B environment. A framework of laws, government regulation, self-policing activities of the members, and a technology infrastructure ensure that individuals will rely on the circle of trust to protect their personal information. The framework consists of a number of distributed components such as a Discovery Service, a Policy Decision Point, an Information Transfer Registry, an Attribute Provider and a Customer Gateway. A Customer Gateway is a key contribution of the thesis. Along with aforementioned components

and a number of interaction protocols it helps to improve the user's experience, capture explicit consent, eliminate unnecessary collection of the personal information and provides a means for the user to dynamically update personal information and means to browse audit information.

TABLE OF CONTENTS:

| | | |
|-------|---|-----|
| 1 | INTRODUCTION | 1 |
| 1.1 | Issues addressed | 2 |
| 1.2 | Contributions..... | 5 |
| 2 | BACKGROUND | 7 |
| 2.1 | Forces influencing privacy enhancing technology development..... | 7 |
| 2.2 | Privacy law..... | 9 |
| 2.2.1 | International privacy law overview | 9 |
| 2.2.2 | PIPEDA as a model privacy protection law..... | 12 |
| 2.3 | Relevant technology | 14 |
| 2.4 | Privacy frameworks..... | 19 |
| 2.4.1 | P3P | 20 |
| 2.4.2 | IBM's EPA..... | 23 |
| 2.4.3 | Hewlett Packard Lab framework..... | 26 |
| 2.4.4 | Liberty Alliance Project..... | 28 |
| 3 | A FRAMEWORK FOR PRIVACY ENHANCEMENT | 34 |
| 3.1 | Framework Components and Architecture..... | 35 |
| 3.2 | Security Considerations..... | 43 |
| 3.3 | Some Key Architectural Patterns..... | 45 |
| 3.3.1 | Cross Domain Cookie..... | 46 |
| 3.3.2 | Message Callback/Correlation Strategy..... | 49 |
| 3.3.3 | Service Coordinator | 50 |
| 3.4 | Performance and Scalability | 51 |
| 4 | CASE STUDY..... | 53 |
| 4.1 | Framework implementation..... | 53 |
| 4.2 | Drugstore scenario..... | 54 |
| 4.3 | Use cases for the drugstore scenario | 57 |
| 4.3.1 | Scenario I (main): | 57 |
| 4.3.2 | Scenario II (missing consent):..... | 58 |
| 4.3.3 | Scenario III (individual controls his personal information usage):..... | 59 |
| 4.4 | Joe Self usage scenario from the Project Liberty specification | 59 |
| 5 | ANALYSIS OF RESULTS | 62 |
| 5.1 | Scenario I: test run result analysis..... | 65 |
| 5.2 | Scenario II (missing consent) test run result analysis. | 73 |
| 5.3 | Scenario III (viewing audit information)..... | 76 |
| 5.4 | Joe Self revised scenario..... | 79 |
| 5.5 | Analysis summary | 80 |
| 6 | CONCLUSIONS AND FUTURE WORK..... | 83 |
| 6.1 | Conclusions..... | 83 |
| 6.2 | Future work..... | 86 |
| 7 | REFERENCES | 88 |
| | Appendix 1. Demo system class diagrams..... | 91 |
| | Appendix 2. P3P policy file used for Drugstore demo. | 94 |
| | Appendix 3. APPEL properties file. | 96 |
| | Appendix 4. Individual's profile used in demo system. | 105 |
| | Appendix 5. EPAL policy and vocabulary used in demo system. | 106 |

FIGURES:

| | |
|--|----|
| Figure 2.1 Customer –Vendor interactions through Infomediary | 19 |
| Figure 2.2 IBM Tivoli privacy management solution components..... | 23 |
| Figure 2.3 IBM Enterprise Privacy Architecture [Ashley2002/2] | 24 |
| Figure 2.4 E-commerce scenario | 26 |
| Figure 2.5 HP Labs architecture for privacy protection | 27 |
| Figure 2.6 Liberty Modules..... | 30 |
| Figure 2.7 Liberty user experience [Liberty2002]..... | 32 |
| Figure 3.1 Core framework architecture..... | 36 |
| Figure 3.2 Distributed components-based privacy enhancing system (DCBPES) architecture.... | 38 |
| Figure 3.3 Framework architecture..... | 39 |
| Figure 3.4 Individual gets audit information..... | 40 |
| Figure 3.5 Individual creates/updates PI usage policy..... | 42 |
| Figure 3.6 Business gets Individual’s PI..... | 42 |
| Figure 3.7 Cross-domain cookie pattern..... | 47 |
| Figure 3.8 Customer Gateway single sign-on | 48 |
| Figure 3.9 Message correlation pattern..... | 49 |
| Figure 3.10 Service coordinator | 50 |
| Figure 4.1 Drugstore scenario [Peyton2004] | 55 |
| Figure 5.1 P3P enabled web-site screenshot. | 66 |
| Figure 5.2 Privacy Bird preferences screen..... | 66 |
| Figure 5.3 Rugstore.com welcome page..... | 68 |
| Figure 5.4 Scenario I sequence diagram. | 69 |
| Figure 5.5 Log4J logging messages for the Scenario I..... | 70 |
| Figure 5.6 Individual’s personal attribute screen. | 70 |
| Figure 5.7 Information to be collected notification. | 71 |
| Figure 5.8 Order confirmation screenshot..... | 72 |
| Figure 5.9 Scenario II sequence diagram. | 73 |
| Figure 5.10 Customer Gateway policy update screenshot..... | 74 |
| Figure 5.11 Scenario II Log4J logs..... | 75 |
| Figure 5.12 Scenario III sequence diagram..... | 77 |
| Figure 5.13 View audit information screenshot. | 77 |
| Figure 5.14 Scenario III Log4J log records..... | 78 |
| Figure 5.15 Customer Gateway welcome screen. | 78 |
| Figure 5.16 Joe Self revised scenario | 79 |
| Figure Appendix1.1. Demo system main package diagram..... | 91 |
| Figure Appendix1.2. Demo system Attribute Provider Service class diagram..... | 91 |
| Figure Appendix1.3. Demo system ITR Service class diagram..... | 92 |
| Figure Appendix1.4. Demo system PDP Service class diagram..... | 92 |
| Figure Appendix1.5. Demo system Discovery Service class diagram. | 93 |

TABLES:

| | |
|---|----|
| Table 2-1 Comparison of international privacy legislation..... | 12 |
| Table 2-2 Example EPAL rule | 25 |
| Table 4-1 PIA Drugstore Data Flow table | 56 |
| Table 4-2 PIA Drugstore Summary Table | 57 |
| Table 4-3 Control, Accuracy and Audit in the existing technologies..... | 61 |
| Table 5-1 Supported functionality comparison summary..... | 81 |

Glossary of acronyms:

API – Application Program Interface

APPEL – A P3P Preference Exchange Language

B2B – business to business

COPPA - Children's Online Privacy Protection Act

DCBPS - Distributed Components-Based Privacy Enhancing System

DRM - Digital Rights Management

EIS - Enterprise Information System

EPA – Enterprise Privacy Architecture

EPAL - Enterprise Privacy Authorization Language

EU - European Union

FTC - US Federal Trade Commission

GLB - Gramm-Leach-Bliley Act

HIPAA - Health Insurance Portability and Accountability Act

ID-FF - Identity Federation Framework

ID-SIS - Identity Service Interface Specifications

ID-WSF - Liberty Identity Web Services Framework

IT – Information Technology

ITR – Information Transfer Registry

Log4J – Logging for Java

LTPA - Lightweight Third-Party Authentication

OECD - Organization for Economic Co-operation and Development

P3P - Platform for Privacy Preferences

PDA – Personal Digital Assistant

PDP – Policy Decision Point

PI – Personal Information

PIA – Privacy Impact Assessment

PII - Personal Identifiable Information

PIPEDA - Personal Information Protection and Electronic Documents Act

PKI – Public Key Infrastructure

RBAC - Role-Based Access Control

SOAP – Simple Object Access Protocol

SP – Service Provider

SSO - Single Sign-On

TAA - Tracing and Auditing Authority

TCPA - Trusted Computing Platform Alliance

URL – Universal Resource Locator

WAP – Wireless Access Protocol

XACML - eXtensible Access Control Markup Language

1 INTRODUCTION

The definition of privacy first appeared in the Harvard Law Review in 1890, as a response to the proliferation of photography, which, many believed, was an intrusion of personal privacy. That is why privacy was defined as “the right to be left alone”. According to [Fischer2001] the most common definition of privacy today is by Alan Westin:

“Privacy is the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others.”

There are a few concepts that stem from the base definition of privacy such as; “territorial privacy, privacy of the person and information privacy” [Fischer2001]. In this thesis, the focus is on the concept of informational privacy and how it may be supported by controlling the circumstances under which personal data can be gathered, stored, processed or selectively disseminated.

Some individuals tend to underestimate the importance of privacy protection until they fall victim of privacy infringement. Many researchers agree that in the information technology era, when an enormous amount of data is flowing through global networks, and when conventional communication and information storage methods are going digital, privacy without due protection is at great risk. Such information as a network audit in the wrong hands can pose a risk to the career of the individual when it is analyzed from the perspective of how active was an employee working from his workstation. One of the major threats to privacy is the aggregation of data from different sources. It can be difficult to predict and prevent data from being combined into all possible combinations of data. A control mechanism which protects against privacy infringement must be developed. [Olivier2002].

This thesis will be built in the following way. First, in the Background section, an introduction will be made to existing international privacy protection laws that explain why Canadian PIPEDA should be considered as the model law for deriving requirements for a privacy protection framework. Also in this section, a review of existing and emerging privacy protection technology will be made.

In subsequent sections, an outline of the basic requirements for any privacy enhancement system will be discussed. These requirements will enable compliance with the corresponding legislation and meet the needs of the businesses and individuals that the system protects. These requirements will be the basis for designing a privacy-enhancing framework.

The web-services-based Liberty Alliance Project architecture will be considered as a basis for this framework. After requirements analysis, some drawbacks of the Project Liberty will be shown and it will be enhanced by adding a number of architectural components and interaction protocols, or rather by combining different elements of that work into one.

Finally, a case study will be described in which a prototype of the framework was developed and compared to existing privacy frameworks.

1.1 Issues addressed

Since the legislative infrastructure for privacy protection was introduced, technology experts have tried to propose a solution which allows individuals the nearly full control of their information at the compliance level. This can be challenging especially in a B2B environment where personal information is distributed across the B2B network. In the B2B environment it is extremely difficult for the individual to *administer* and *audit* his or her personal information.

The system proposed in this thesis focuses on a number of issues connected with the usage of personal information. These issues stem from the PIPEDA principles and can be summarized in the following way:

- 1) CONTROL. *An individual must have full control over his personal information;*
- 2) ACCURACY. *Personal information should be accurate and up-to-date;*
- 3) AUDIT. *In order to control compliance and make conscious decisions on how to manage personal information, an individual has to have access to audit information, which contains records about collection, disclosure and keeping of personal information;*

As will be demonstrated later, these issues are not completely addressed by existing solutions.

With the introduction of PIPEDA, the business community in Canada has faced new information technology challenges. What was previously just the concern of a group of privacy-aware individuals now has become a key business issue and has led to the drafting of PIPEDA-based requirements. The biggest challenge here is that PIPEDA doesn't define any particular technology to achieve compliance which is typical for a law but makes the job of achieving compliance more difficult.

Users and businesses have different objectives in sharing and protecting personal information. Past approaches to privacy have either focused on enabling business interests, or on enforcing business policies on privacy. Users need a convenient mechanism for furthering their interests. In particular, they need a framework that is convenient for them to use, that gives them control over their personal information, the ability to ensure its accuracy, and the ability to audit how businesses are using the information. The creation of such a user-centric framework is the primary goal of this thesis.

There are different flavors of privacy-enhancing technology depending on the domain in which it is used. Below there are a few privacy-enhancing techniques, according to Olivier [Olivier2003/2], which are not going to be considered in this thesis. However it must be said that this technology can augment the degree of privacy protection:

- 1) Personal privacy enhancing technologies including infomediary-encryption-based schemas, depersonalization, and anonymization, etc. Although such technology can be helpful in increasing privacy protection, it is a separate area of research. For instance, “there is always a risk of re-identification depending on the entropy of the depersonalized dataset and additional data about the data subject. Developing reliable criteria to estimate this risk is a non-trivial task” [Fischer2001].
- 2) Network-based technologies and organizational safeguards for privacy enforcement. Numerous technologies exist that ensure a piece of personal data can be read by only an entity authorized to do so. These include encryption algorithms, PKI, The Trusted Computing Platform Alliance [TCPA], different Digital Rights Management (DRM) systems, etc. In my view these are complementary technologies when it comes to ensuring personal data protection in compliance with the legislation.

The focus of out thesis is on the features required to enhance control, accuracy and audit on behalf of the user. Security and enforcement is also an important aspect of any such framework, but not one that was a focus for this thesis. For the most part, it has been assumed that participants are well intentioned and each server and access point of the framework is secure. From a practical point of view, it is assumed that the circle of trust will need to provide suitable security and enforcement to be viable as a commercial enterprise. Here are some of the issues

that would need to be touched on that which weren't covered in this thesis (except for a brief discussion in section 3.2 Security considerations):

- Enforcement mechanisms. The proposed framework depends on how all members of the framework perform the actions required. For example, all personal information transfers must be registered with the information transfer registry. A second aspect of enforcement is how to ensure that businesses fulfill their privacy promises they give when collecting user's consent. For instance, there has to be a way to guarantee that personal information is used as declared.
- Security mechanisms to provide safeguards of personal information. As required by PIPEDA all personal information has to be protected in all stages of its usage. This is usually done by applying different cryptographic techniques. For more information on the topics above please refer to the section 3.2 Security Considerations.

1.2 Contributions

The contributions of this thesis are organized around a framework for privacy enhancement focused on PIPEDA compliance in the context of B2B networks and includes:

- A customer service gateway as a single service window (administration, audit, etc.) to give a user full support for control, accuracy and audit.
- A greater role for a discovery mechanism to increase flexibility.
- A mechanism allowing the existence of multiple replicated policies while presenting them to the user as if it is a single one. This makes it possible for the user to easily manage multiple policies.

- A mechanism for capturing individual's explicit consent dynamically without the need of explicit callbacks.
- Propose key extensions to an existing industry framework (Project Liberty) specification to accommodate requirements of both business and individual users. Currently there is no working implementation of the second phase of Project Liberty which covers personal information management; this adds to the motivation for this thesis. The Circle of Trust proposed in the initial stage of the Project Liberty is used as a base for the proposed framework. The Project Liberty [Liberty2003] will be covered in more details in the section 2.4.4 Liberty Alliance Project.
- Technical implementation of several PIPEDA principles not currently supported by the existing frameworks such as "Ensuring Accuracy" and "Providing Individual Access" principles. By providing users with convenient way of access to personal information (individual access) and audit records it is possible to ensure that personal information is up-to-date (accurate) and used as intended.

2 BACKGROUND

2.1 Forces influencing privacy enhancing technology development

The history of privacy law can be used to support the fact that public concerns about privacy have increased due to the threat new technology poses to their privacy.

Frichman et al, provide a definition of Information Rich Commerce [Frichman2003], and highlight several key factors influencing further developments in the e-commerce industry. “Information Rich Commerce” is a process where detailed consumer data, such as preferences, historical records, and different personal information, are used to customize the content offered to the customer including commercials, marketing offers, and new products etc.. This is done in order to add extra value to consumers and service providers.

Merchants benefit from Information Rich Commerce by the increased efficiency of the existing marketing channels, flexible pricing structures and increased customer trust and loyalty.

The authors also describe the complex response to Information Rich Commerce in all its sophistication. Despite the fact that some of DoubleClick’s and Amazon’s initiatives met hostile response from consumers along with the introduction of “chilling” legislation, the prevailing belief continued to be that “while the idea of unified platforms for Information Rich Commerce may seem alarming to some, they are essential for the vision of pervasive, economically efficient, and user-controllable Information Rich Commerce to become a reality” [Frichman2003]. The authors also insist that the benefits of Information Rich Commerce will significantly outweigh the potential risk and associated cost despite the controversial interests of those involved.

A very interesting example of how society can benefit from highly available personal data is given by Rindfleisch [Rindfleisch97]. This author discusses ePhysician, a company specializing in PDA solutions that allow physicians to issue prescriptions electronically.

There is obvious benefit for the physician to have wireless access to patient records, prescription data and insurance data; these benefits include increased convenience and reduced errors. One of the important issues in health care is adverse drug interaction, which increases as the complexity of medical treatment grows. Keeping health records electronically opens an opportunity to dynamically update patients' reactions to drugs, thus improving patient care. Another invaluable asset of ePhysician is the ability to electronically manage complex health insurance rules, allowing physicians to concentrate on their primary function.

This same source shows that all the benefits of using Information Rich Commerce are shown from the point of view of different players:

- Insurers can benefit from the ability to aggregate information which can be used in negotiating better pricing for medicine and treatment. Also, it can reduce the cost of expensive additional treatment by avoiding complications caused by drug allergy;
- Pharmacies can obviously benefit from error reduction. Also, links to the system can bring the drugstore additional marketing opportunity through insurer and doctor observed competition.
- Patients. Not only do patients benefit from better service, but some health care providers allow patients to access their patient records online. In Canada this has caused some controversy, but many agree that patients need some degree of control over their health records.

Many researchers in the medical domain agree that there is a paradox when limiting access to medical records. "While our medical records contain information about us that is of the utmost sensitivity, yet this information is only useful to us when it is shared with the medical providers and systems under which we get our care. Indeed, our physicians need and expect access to our complete medical records in order to help diagnose diseases correctly, to avoid duplicative risky

or expensive tests, and to design effective treatment plans that take into account many complicating factors” [Rindfleisch97].

Indeed, medical records are the most sensitive type of individual data. In [Rindfleisch97], it is pointed out that the potential for the “abuse” of personal genetic information is very sobering. This source refers to one recent study which showed a risk posed by the abuse of such information. This study reported 206 cases of direct discrimination and also employment and insurability problems when genetic information was used improperly. No patients actually exhibited any signs of disease.

All these concerns are especially acute when many health providers have deployed electronic patient record systems.

Ross J. Anderson [Ross2000] identifies medical information that passes outside professional control, such as the payment data collected by both insurers and employers, as the most at risk for abuse. The conclusion was that this problem is impossible to solve without corresponding law and regulations in place. The author concludes that “Now, when we have all that, the technology gets the major role in solving this problem.”

2.2 Privacy law

2.2.1 International privacy law overview

According to S. Fischer-Hübner [Fischer2001], the Parliament of the West German state Hesse was the first state to adopt a modern Data Protection Act in 1970. Other German states, as well as governments outside Germany, used this act as a template for their own similar acts. Sweden followed suit with Sweden’s Data Act in 1973. This act is regarded as the first national data protection act in the world. In 1974, the Congress in the USA adopted the Privacy Act, in acknowledgement of the threat posed to personal privacy from the development of complex information systems.

In 1981, the Council of Europe adopted the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data. This act, along with OECD Guidelines [www.oecd.org], had a major impact on the development of privacy legislation around the world. During the course of the next 20 years, privacy legislation received wide adoption especially between EU and G7 countries. These countries created several laws which used the Convention and Guidelines as their base.

The European Parliament passed the directive 2002/58/EC on July 12, 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [EU2002]. As of June 2003, twelve Member States have enacted laws or regulations implementing the Directive in full. The Directive applies to the collection, transmission, and processing of "personal data" within the EU. Processing of personal data is permitted if the data subject has unambiguously given his or her consent and in some other cases outlined in the Directive.

This Directive is the result of a several decades long legislative process and is considered to be one of the most comprehensive and widely adopted.

The Directive prohibits flow of information from a member country to a country without similar legislation, unless there is proof that due to certain conditions this country constitutes a so called "safe harbor" for personal information. This part of the Directive makes it one of the most powerful driving forces behind adoption of such legislation outside the EU. In fact, the US still falls under the definition of "safe harbor" because it does not have a comprehensive law in place.

Since the adoption of the Privacy Act, the government of the US has passed several laws:

- COPPA (Children's Online Privacy Protection Act of 1998) [COPPA98] to protect children when they are surfing the Internet from unnecessary collection of their personal data without parental consent;

- The Gramm-Leach-Bliley Act Privacy of Consumer Financial Information (GLB) November 12, 1999 [GLB99];
- Health Insurance Portability and Accountability Act (HIPAA) (August 1996) [HIPAA9]. HIPAA allows for health information to be released and used for research based on a patient authorization, an approved waiver of patient, the de-identification of a person's health information as defined by HIPAA, and the de-identification through a limited data set.

Although these acts contain regulations very similar to the EU Directive, they are limited to specific industry domains. This is a major concern for big US corporations which may have trouble exchanging data between branches located in the US and Europe.

Ackerman et al [Ackerman2003] point out the difference in how geographical location privacy is treated in the US and EU. While the EU Directive has a clear definition of geographical location privacy, in the US this issue is obscure and the definition is spread between different acts of different legislatures. However it is pointed out that explicit consent is now the focus of any legislative act.

Considerable efforts toward privacy protection were made in Japan with the adoption of five bills passed into law in the House of Councilors on May 23, 2003. These laws stipulated how private companies should handle personal information and public administrators to protect the rights of individuals. Other Asian countries, such as Singapore, are now making similar efforts.

In Table 2-1 there is a comparison of three privacy models [Hochheiser2002]. The author compares all three models. He notes that they are similar in containing notions of disclosure, prior consent of the individual for collection and use of data, requirements to provide access to data and audit trail on use and collection of personal information. What makes them different is that only OECD and Canadian principles clearly specify requirements for limiting the amount of

data collected, also those two principles require that the purposes for which data is to be used are specified prior to collection of such data. Any other use of the data is prohibited except when required by law. Another similar part of OECD and Canadian principles is the definition of the “Openness” principle when policies and practice as for the use of personal information should be “readily available”. Meanwhile FTC (The US Federal Trade Commission) regulations are more ambiguous. Under FTC regulations sometimes specifying only who collects data can be considered as giving an adequate notice. Also there is no notion of strong purpose binding, which means that data might be used for other than declared reasons.

| OECD | US. FTC | Canadian |
|--------------------------|----------------------|---------------------------------------|
| Openness | Notice | Openness |
| Purpose Specification | | Identifying Purpose |
| Collection Limitation | Choice/Consent | Limited Collection |
| | | Consent |
| Individual Participation | Access/Participation | Individual Access |
| Data Quality | Integrity/Security | Accuracy |
| Security Safeguards | | Safeguards |
| Accountability | Enforcement/Redress | Accountability |
| | | Challenging Compliance |
| Use Limitation | | Limited Use, Disclosure and Retention |

Table 2-1 Comparison of international privacy legislation

2.2.2 PIPEDA as a model privacy protection law

On January 1, 2004 Canada’s government fully enacted Canada’s Personal Information Protection and Electronic Documents Act [PIPEDA2000]. PPIPEDA is quite similar to EU, Japanese and, to some degree, US privacy laws; but what makes it unique is that one PIPEDA bill covers virtually every aspect of data collection and utilization including government organizations, private businesses, health care and the Internet. It regulates private information

protection regardless of the provider, the collector, and the way it was obtained. PIPEDA also regulates the usage of electronic signatures, which is an important issue in e-commerce in general and personal data protection in particular. Later in this thesis, the PIPEDA will be used as a model law when defining the requirements for the data protection framework.

There are ten basic principles outlined in PIPEDA:

- 1) **Accountability.** Organizations are held responsible for the personal information they control;
- 2) **Identifying Purposes.** Prior to the collection of personal information the purpose of such collection should be identified. This principle is also known as “purpose binding” [Fisher2001];
- 3) **Consent.** Wherever appropriate, the consent of an individual for collection of personal information should be obtained;
- 4) **Limiting Collection.** What is necessary for the declared purposes should limit the collection of personal information;
- 5) **Limiting Use, Disclosure, and Retention.** Similarly to the previous principle, declared purposes should limit these actions with collected personal information;
- 6) **Accuracy.** Personal information shall be accurate, complete, and up-to-date;
- 7) **Safeguards.** Personal information should be protected;
- 8) **Openness.** Personal information management practices of the organization should be publicly available;
- 9) **Individual Access.** Individuals should have access to the information about existing personal information, its use and its disclosure;

10) **Challenging Compliance.** Each organization should have a designated person to whom an individual can address concerns about compliance with above-listed principles.

2.3 Relevant technology

Although the PIPEDA principles are seemingly easy to implement in practice, due to a number of factors, it is not a trivial task. One reason is that these requirements appeared only recently with the introduction of the Internet and privacy laws themselves.

Simon Fischer-Hübner [Fischer2001] gives detailed analysis of the existing privacy protecting models. For this purpose he analyzed existing security models under several privacy criteria:

- Protection of confidentiality of personal data;
- Protection of integrity of personal data;
- Purpose binding of access to personal data (the purpose of the user's current task must be contained in the set of purposes for which the personal data was obtained or there has to be a consent by the data subjects);
- Necessity of personal data processing (a user may access personal data only if the access is necessary to perform his current task);
- Right of "implementational" self-determination (individual's right to determine the disclosure and use of his personal data).

However, he separates the aspect of anonymous or pseudonymous system use from the aspect of personal data protection as more relevant to identity protection.

It has been shown that all existing security models are inappropriate for privacy protection assuming additional requirements of recently introduced privacy legislation [Fischer2001]. The least supported functionality and at the same time the most important for privacy protection was "purpose binding". Among those security frameworks were the Bell LaPadula Model, the Lattice

Model of Information Flow, the Biba Model, the Clark Wilson Model, the Chinese Wall Model, the Role-Based Access Control (RBAC) Model (probably most common nowadays), the Task-Based Authorization Models and the Object-Oriented Security Models.

For this reason, a formal Task-based Privacy Model was proposed which served as a base for several frameworks now under development.

Another classification attempt is made by M. S. Olivier [Olivier2003/2]. The author refined and amended the OECD [www.oecd.org] classification of privacy-enhancing technologies to include:

- Personal privacy-enhancing technologies: Cookie managers or blockers (private communications), Ad blockers (personal control), Encryption software (private communications);
- Web-based technologies: Anonymizers (identity management), Platform for Privacy Preferences Project (personal control), Privacy networks (identity management and personal control);
- Information brokers: Infomediaries (identity management and personal control);
- Network-based technologies: Proxies (identity management) and firewalls, Privacy networks (identity management and personal control).

Personal control refers to the use of technology to ensure that an individual's personal information is only used in a manner aligned with the individual's privacy policy.

Also, a definition of organizational safeguards was added to the classification. This is when technology is used to ensure that the organization complies with its own privacy policy as well as the preferences of the individual. For example, this kind of technology can be used to keep track of the individual's opt-in or opt-out choices when receiving unsolicited e-mail.

Finally, four layers of privacy enhancing architecture are defined according to the role each layer plays in the classification. IBM Enterprise Privacy Architecture (see below) is one of the closest implementations of the proposed formal model.

Two main approaches to the implementation of privacy enhancing technologies are outlined by Arnesen. One is to minimize the amount of personally identifiable data through pseudonymization or anonymization, or by simply not collecting any data at all. The other approach is to assure that any privacy agreement, to which both a data subject and a data collector have consented, is enforced [Arnesen2003]. In this work, authors define a framework for protecting privacy. They stress that existing access control models such as Bell LaPadula are inapplicable for private data protection because they do not support purpose binding. Although the authors designed a very feasible architecture, it was done without consideration of the current supporting technology and stakeholders interests. What can be particularly useful is the notion of a Personal Data Broker – the component which is now a part of virtually any personal data protection system.

There is rising support for technology other than traditional technology for providing users with anonymity. “Anonymity has the drawback of preventing users from learning the usefulness of recommendations from particular people, tracking trends over time and using reputations which are built up over repeated interactions” [Hogg2000]. Also, there is a concern about the feasibility of establishing a centralized authority which manages users’ data on their behalf.

Another system was introduced by Grandon and Sadeh. Their proposed "Semantic e-wallet" [Grandon2003] can be used in future systems to automate the discovery of data, based on the description of the data, using a semantic web approach.

An attempt to facilitate “privacy-sensitive ubiquitous computing” was made by Hong and Landay. This work also supports the idea of the necessity to share information and proposes a

way to control unwanted links of the information by “combining data into a 3-tuples: metadata, data itself, and policies on how to use” [Hong2004]. This approach is very similar to the one proposed by HP Labs which will be discussed later in more detail.

A policy-based electronic prescription transmission system is presented by Chadwick et al [Chadwick2003]. The authors claim that none of the currently functioning systems in the UK has an ability to control the access to individual’s prescription information based on the number of roles and dynamic conditions. Unfortunately the authors developed their own language for expressing the access control policy instead of trying to find an adequate existing policy language or at least comparing their language with existing ones.

One interesting approach to protect information flow via decentralized labels (policies defined by the individual) is shown by Myers and Liskov [Myers2000]. Although their system does not comply with the requirements of the privacy protection legislation (since it does not have many essential functions) it represents one of the trends in information security. This trend is to protect information at disclosure time using a variation of the trusted computing platform. Such an approach may be feasible in light of the recent announcement by Microsoft to introduce a completely secure platform for digital rights management ensuring that only authorized entities have access to the data.

To review other efforts in developing privacy enhancing systems please refer to the Privacyright and IDCIDE web sites ([PRight2001] and [IDCIDE2001]).

Before starting the detailed review of the most viable technologies, an explanation of the trusted proxy called “Infomediary” or “data broker” must be included. Having a central entity communicating personal data to others on behalf of the individual would first be beneficial to the individual. This paradigm perfectly fits into the well-established schema of bank-customer relations. Establishing a trusted relationship with the Infomediary is similar to the opening of an

account in the bank. By signing under an agreement, an individual can define a policy governing his personal information usage in almost the same way as when he or she signs an investment agreement with the bank. Instead of money, personal information is deposited. The bank's security measures, regular audits, government regulation and the positive history of the bank contribute to its credibility. Similar indicators can be used to determine trustworthiness of an infomediary. Certain minimum requirements for trustworthiness may be set.

This easy-to-understand paradigm may increase support for the technology among users. However, the drawback is that the paradigm requires a central infomediary which may never be accepted by the industries' key players. A reasonable compromise between the central and distributed approach has to be found. Infomediaries can be effectively used to increase privacy protection, but it is important to find efficient implementations of this concept [Dix2000].

The need for a wallet, which will hold personal information, is supported by Pfitzmann and Waidner [Pfitzmann2002]. They also give a comprehensive review of the existing technology as well as a preview of the technology under development. The paper mainly focuses on the aspect of browser-based attribute exchange – the way a web site can automatically obtain necessary information about the visitor. While the technical possibility of automated exchange is discussed, privacy concerns are left untouched.

A classic example of using Infomediary schema based on PKI cryptography is given by Gritzalis et al [Gritzalis2001]. The approach is illustrated in Figure 2.1. This approach is limited because although it provides protection of private communication, it leaves out personal control, which is crucial for achieving legal compliance.

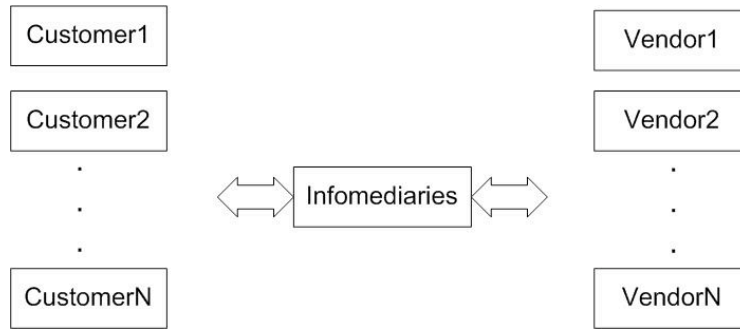


Figure 2.1 Customer –Vendor interactions through Infomediary

All the above-mentioned frameworks, because they originate mainly from academia and were developed without considering legislative requirements, lack industry support. Also, these frameworks lack working implementations to demonstrate the feasibility of the approach, and are limited to one particular field, for instance they are exclusive to wireless applications.

2.4 Privacy frameworks

It would be fair to say that within the context of a single enterprise, the problem of PIPEDA compliance can be effectively solved. One example of such a solution is IBM’s Tivoli Privacy Manager, which is based on Enterprise Privacy Architecture (EPA). This approach enables an employee of the enterprise to fully control and audit the flow of personal information (PI) within one organization. The situation is completely different should PI leave the boundaries of the enterprise [Peyton2004].

The most noticeable attempts to solve the problem of PI uncontrolled disclosure were made by Hewlett-Packard (HP) [Mont2004] and by Liberty Alliance’s Project Liberty Phase 2. Both of these are similar in the way that they enable automatic sharing of PI attributes between businesses in B2B communications: They both allow businesses to better serve customers by reducing the number of times the customer has to disclose his PI. Both frameworks contain a concept similar to the Information Transfer Registry (ITR) [Peyton2004] but there is a significant difference in the way this entity is implemented. Project Liberty’s naturally

decentralized architecture allows for multiple instances of any component of the architecture whereas HP's approach defines its ITR-like audit and tracing authority as a single entity. The latter makes the odds of this architecture receiving wide business community adoption pretty slim. The strength of the Project Liberty lies in its decentralized architecture; however, this has the potential to make the experience of an individual user daunting as it is a well known fact that individual users are more comfortable with a single-entry or customer service counter paradigm. None of the frameworks therefore completely addresses the three main issues outlined before.

2.4.1 P3P

The Platform for Privacy Preferences (P3P) is the most widely endorsed approach to enhance privacy protection. While many are pitching it as a complete solution for privacy protection, some are heavily criticizing it. P3P is often regarded as a "self-regulation tool" assuming the voluntary adoption of the standard [Grimm2000] [Cranor98].

P3P history is surrounded by controversy. Some privacy advocates view it more as a way for corporations to avoid litigation than really protecting privacy. Advocates of P3P claim that inclusion of P3P into Microsoft's IE 6.0 was a culmination of many years of effort aimed towards the development of a technological means to protect privacy. The truth, as usual, lies somewhere in the middle. While P3P is certainly helping users to make conscious decisions on their privacy, P3P doesn't serve the purpose of being the complete solution for all privacy issues [Hochheiser2002].

P3P is a protocol for automating the negotiation of privacy policies between Web sites and user agents that are often Web browsers. It also covers the comparison of those policies with statements of user preferences. Under a typical P3P usage scenario, a web site that collects personal information should have a published P3P policy. In essence, P3P is an XML-style language with a very narrow set of predefined data types and purposes. In plain words, the

privacy policy should be translated into P3P to enable its automatic analysis. The user should have his preferences defined in A P3P Preferences Exchange Language (APPEL) that is a complementary to the P3P standard. When the user visits a web site using an agent, usually the browser, a P3P policy is matched against preferences defined in APPEL and a suggestion is given to the user on the level of compliance of the web site to the user's preferences. It is up to the user to make a final decision on whether to proceed or navigate away from the site [P3P2002].

In the core of P3P is the comparison of P3P policy against user preferences expressed in APPEL. This is usually carried out by the browser. Creators of P3P acknowledge that there is "a wide latitude" on the browser side when such a comparison is done. Often it is difficult to determine with high certainty, which operation can and which operation cannot be done with the data based on the P3P policy - APPEL rules. Because P3P policies themselves are not intended to be human-readable, the analysis of P3P policies has to be performed by an automatic agent. Among the concerns expressed by the creators of P3P, is whether such agents can render correct suggestions and, more importantly the "consequences of the wrong judgment" [Cranor2002].

"Unfortunately, the APPEL constructs interact with the P3P policy language in unintended ways, making it non-trivial to get even simple preferences right. It is easy to write a preference that appears correct and find that it does not accomplish the intended goal" [Agrawal2003]. An alternative language for preferences was proposed but even if it becomes a part of the specification soon it is not going to remove all P3P shortcomings as a privacy tool. The ambiguity of APPEL and lack of a clearly defined algorithm to compare policies and preferences are the main reason why it is still impossible to consider P3P as a viable solution for personal information protection.

Although P3P's guiding principles discuss other aspects of privacy, including use limitations, fairness, and integrity [P3P2003], P3P "does not address any of these principles directly. Instead, P3P's developers see it as a force that might indirectly lead to implementation of privacy principles such as use limitation. In this view, the presentation of P3P policies might motivate changes in practice, as companies work to be more consumer-friendly. Alternatively, a proliferation of P3P policies that do not meet customer needs might be used as evidence to support arguments for stronger privacy legislation" [Hochheiser2002].

Karjoth et al believe that "P3P can be a strong tool to advertise privacy promises to consumers" [Karjoth2003/2]. At the same time they outline a set of amendments to the standards aimed to improve usability of P3P.

The creators of P3P didn't take into account the existing conflict between government regulation and industry self-regulation. While in Germany, P3P serves as a complement to the privacy law, in the USA it reinforces the self-regulatory approach. The major weakness of P3P is that it cannot enforce policies or at least "improve data drift" [Grimm2000].

Finally, Zuigwey et al [Zuidweg2003] highlight another drawback of the P3P language. When an attempt was made to apply the P3P-APPEL combination in a context-aware application, the problem of evaluating dynamic conditions surfaces. To bypass this, the authors proposed to enable P3P to do such evaluations without knowing that such a language already exists [OASIS2002] and is being used in another privacy-enhancing language [Schunter2003/2].

To summarize, P3P being used as a self-regulatory tool, favors corporations rather than ordinary customers transferring the burden of privacy compliance onto the users shoulders. In this thesis an attempt will be made to find some kind of regulatory approach, which is more customer-centric.

2.4.2 IBM's EPA

The IBM Corporation has made the most significant single organization's effort in the area of personal information protection.

The prototype of their enterprise-wide privacy protection system was defined by Karjoth et al [Karjoth2002/1]. The driving force behind such a system was the need to ensure that actual privacy practice within the enterprise corresponds to advertised privacy promises. The number of lawsuits against international retailers such as Toys'R Us and Toysmart contributed to the faster development of privacy management tools. Real-world implementation is described in the IBM Redbook [Bucker2003].

At first, privacy-related code was incorporated into a separate component called Privacy Server, which used to be hard-coded into the system, as shown in Figure 2.2. Through the component called Privacy Monitor, any application can be privacy-enabled through a unified interface.

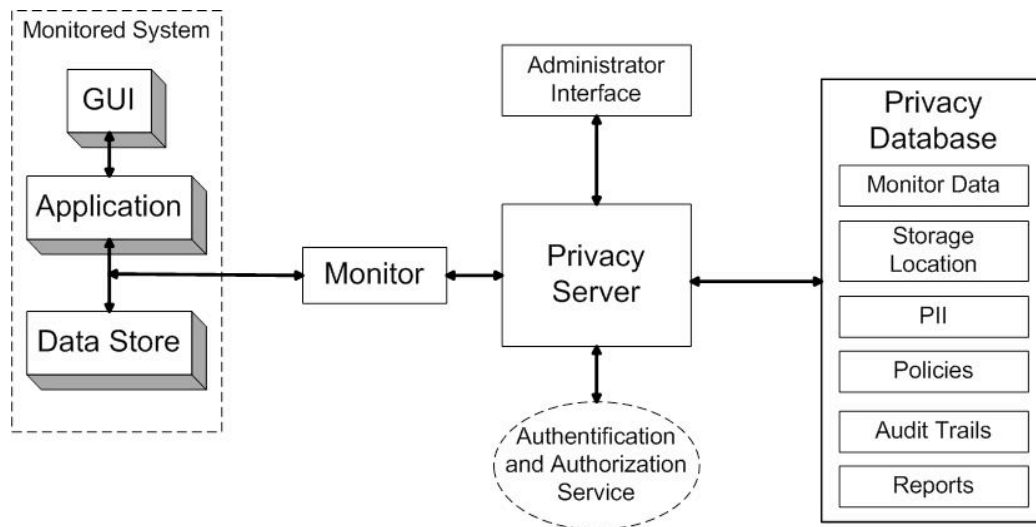


Figure 2.2 IBM Tivoli privacy management solution components

Later, this solution evolved into the Enterprise Privacy Architecture (EPA) shown in Figure 2.3.

In this architecture, the roles played by the Privacy Server were split between two components: Privacy Data Handling Node and Privacy Service Node.

To express privacy policy, a proprietary XML-based language was used up until E-P3P [Ashley2002/2] was introduced. The mechanism to translate privacy promises into privacy policies was designed as well [Karjoth2003/1]. Eventually E-P3P evolved into Enterprise Privacy Authorization Language [Schunter2003/2].

Although Enterprise Privacy Architecture is limited to the scope of a single enterprise it is of a particular interest to this theses by having powerful and comprehensive policy description language EPAL. EPAL is a tool which allows the declaration of fine-grained privacy policies using XML syntax. It also enables the usage of flexible declarations, amendments and transformation of privacy policies [Backes2003].

Enterprise Privacy Architecture is also interesting as a first attempt to develop a system which will cover all aspects of privacy, from collecting user consent and data to enforcing privacy policy.

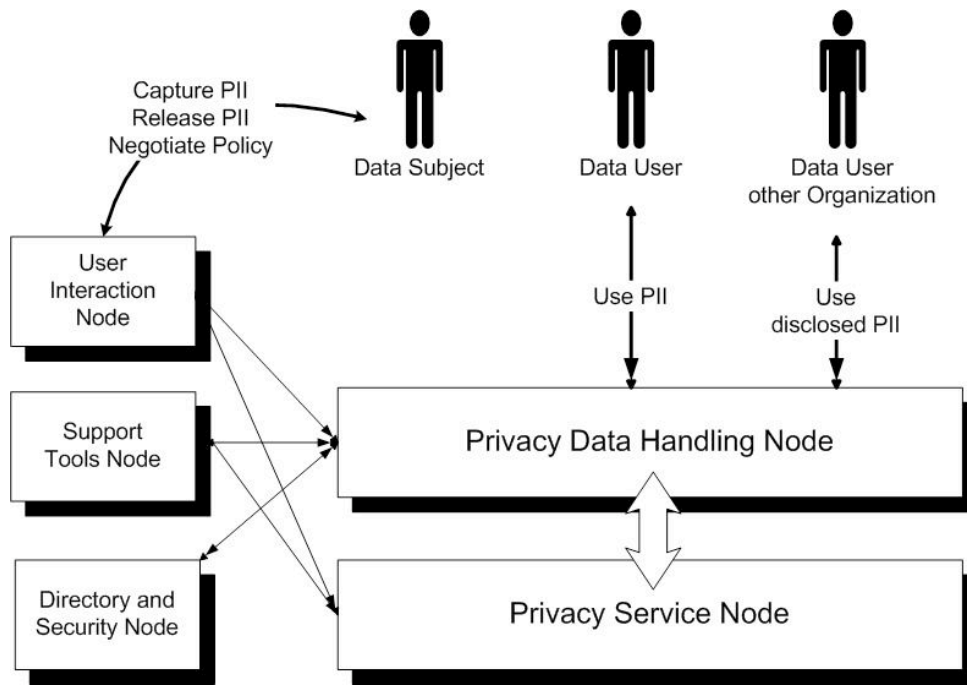


Figure 2.3 IBM Enterprise Privacy Architecture [Ashley2002/2]

PDP stands for Policy Decision Point (PDP) and it is responsible for decision-making. Using policy written in EPAL a decision is made on any action with personal information stored in Privacy Data Handling Node. When policy can be declared rather than hard-coded, the process is called by IBM Declarative Privacy Management [Backes2003]. An example of how plain-text policy can be represented using EPAL rules is shown in the Table 2-2.

| Privacy Policy (informal) | |
|---|--|
| <i>Allow a sales agent or a sales supervisor to collect a customer's data for order entry if the customer is older than 13 years of age and the customer has been notified of the privacy policy. Delete the data 3 years from now.</i> | |
| EPAL Privacy Rule: | |
| ruling | allow |
| user category | sales department |
| action | store |
| purpose | order-processing |
| condition | the customer is older than 13 years of age |
| obligation | delete the data 3 years from now |

Table 2-2 Example EPAL rule

Whereas P3P has as its main purpose to publish privacy promises, the main purpose of EPAL is to express the rules governing the usage of personal information within the enterprise and to allow fine-grained control over personal data and privacy promise enforcement. There is an example EPAL rule in Table 2-2. It provides the necessary purpose binding. EPAL policy consists of a set of rules. Also, the policy is linked to EPAL vocabulary – an XML file containing definitions of possible user categories, actions, data categories, and purposes. Another important feature of EPAL is an ability to evaluate dynamic conditions written in XACML [OASIS2002]. As will be demonstrated later, the XACML conditions enable very fine-grained control over usage of the data, which in turn makes EPAL an indispensable tool especially in contrast with static P3P. And finally, the obligations section enables usage of “sticky policies” developed by HP lab, the formal model for obligation monitoring framework proposed by C.Bettini et al [Bettini2002].

2.4.3 Hewlett Packard Lab framework

Mont gives an example of a typical e-commerce scenario involving user's data provision and disclosure to third party shown in Figure 2.4 [Mont2003]. Obviously, without any additional control mechanism the user has no control over disclosure of his personal data.

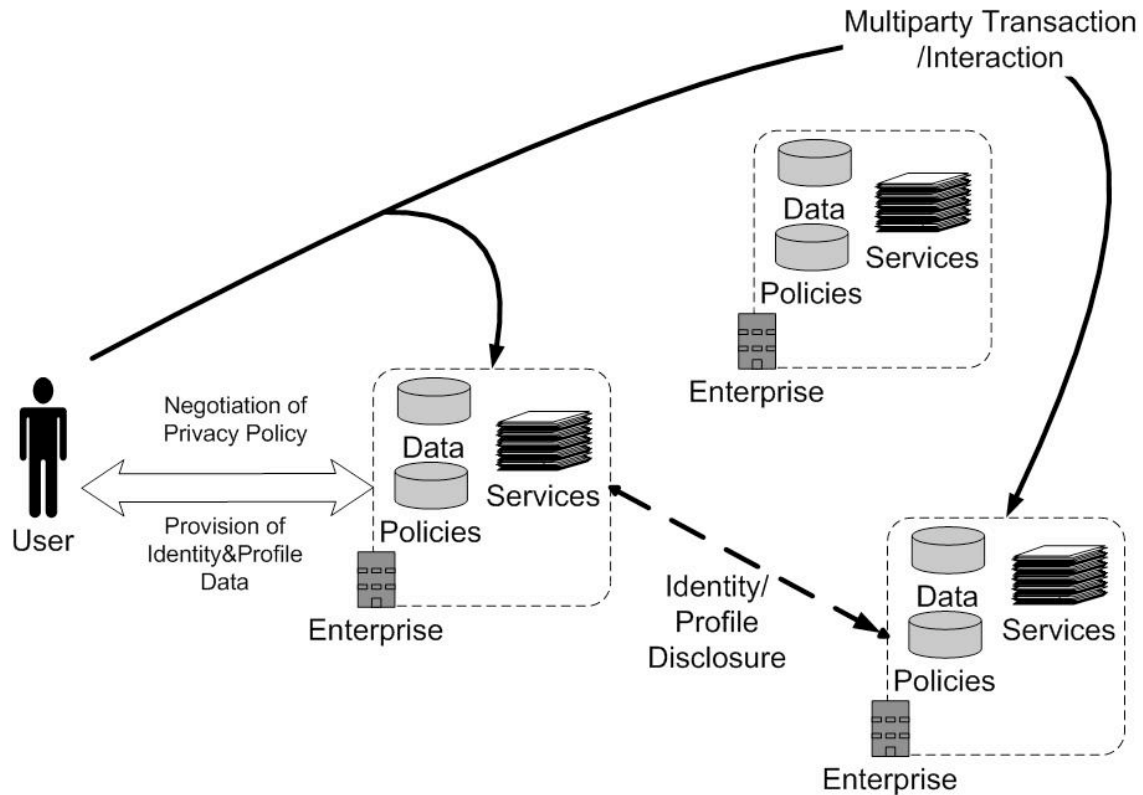


Figure 2.4 E-commerce scenario

The solution for this problem lies in using cryptographic technology to obfuscate and prevent unauthorized leakage of data combined with “sticky policies” and Tracing and Auditing Authority (TAA) as can be seen in Figure 2.5.

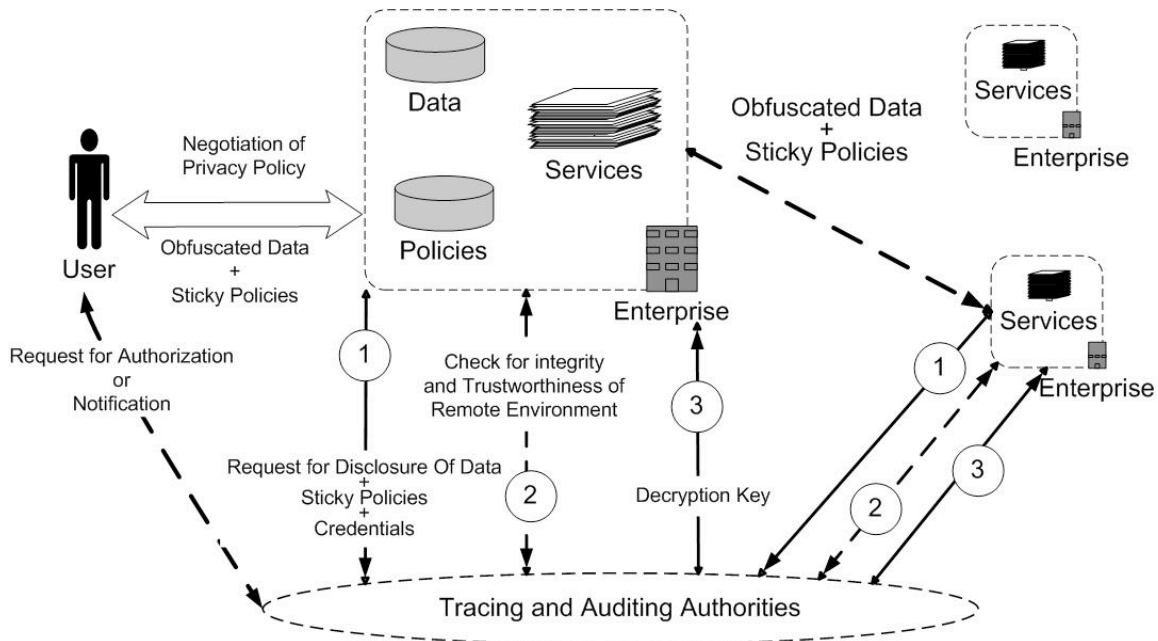


Figure 2.5 HP Labs architecture for privacy protection

The concept of “sticky policies” was introduced in G.Karjoth et al [Karjoth2002/2] and later developed by M.C.Mont [Mont2002]. The main idea is to “encapsulate data into a container and ‘stick’ a policy to it. Later when data travels across enterprises this policy is merged with internal enterprise policy. This way user preference on how to use his personal data can be extended beyond the enterprise which initially acquired the data” [Karjoth2002/2].

This approach heavily relies on Trusted Operating Systems as an enforcement end-point. Although such OS are currently under development [MS2004/1] it is unclear now whether they will be widely adopted.

What is also important is that the architecture permits usage of multiple TAAs even for single transactions. It permits data subjects to maintain such TAAs on their own. A similar approach will be used later in this thesis.

2.4.4 Liberty Alliance Project

Started as a response to Microsoft's Passport initiative, a joint effort of the leading IT and financial companies, The Liberty Alliance Project [Liberty2002], addresses such problems as single sign-on (federated identity) and privacy protection in a service-oriented environment.

Project Liberty is being developed in two major phases. In Phase I, an Identity Federation Framework (ID-FF) was developed allowing different enterprises to operate within the "circle of trust", such that a user can sign-on and continue surfing across multiple service Providers (SP) having his identity accompanying him everywhere.

In Phase II, a second set of specifications was created (ID-WSF) enabling a network of web services and also covering all aspects of automatic personal profile data sharing between multiple Service Providers (SP). Privacy issues are the primary concern of the Liberty Project team.

To better understand the functionality of the Liberty framework it is important to know the definitions of basic roles each Liberty-Enabled Provider can play in Liberty framework. These definitions are taken from the Liberty Project Specification [Liberty2002]:

The Project Liberty specification defines several important components/entities:

- **Individual** – an individual as defined in PIPEDA;
- **Service Provider** – an entity providing service to Individuals. They are also the main consumers of personal information;
- **Attribute Provider** – An entity that stores the Individual's personal attributes along with his usage policies and releases those attributes upon request to third parties including Service Providers;

- **Discovery Service** – “A Discovery Service is an entity that has the ability to direct attribute requesters to the relevant Attribute Provider who provides the requested classes of attributes for the specified Individual. The Discovery Service should register only those Attribute Providers in accordance with the consent or usage directives of the Individual. The Discovery Service should permit the Individual to see which Attribute Providers have been registered on the Individual’s behalf. An attribute requester can locate the Attribute Providers for a given Individual, even though the attribute requester and Attribute Providers do not have a common name for the Individual” [Liberty2002].

Please note that in the context of this thesis the above-mentioned terms should be understood according to the definitions given.

However, it is important to understand that Liberty specification just enables a person’s privacy but doesn’t ensure it. To increase the level of privacy protection, a set of recommendations that all Liberty-enabled providers should follow was proposed. The following shows that Liberty’s main principles map well with the PIPEDA’s principles:

- **Notice.** Same as Canadian PIPEDA “openness” principle;
- **Choice** = “Consent”;
- **Individual Access to Personally Identifiable Information;**
- **Correctness** = “Accuracy”;
- **Relevance** = “Identifying Purpose”;
- **Timeliness** = “Limited Collection”;
- **Complaint Resolution** = “Accountability” and “Challenging compliance”;
- **Security** = “Safeguards”.

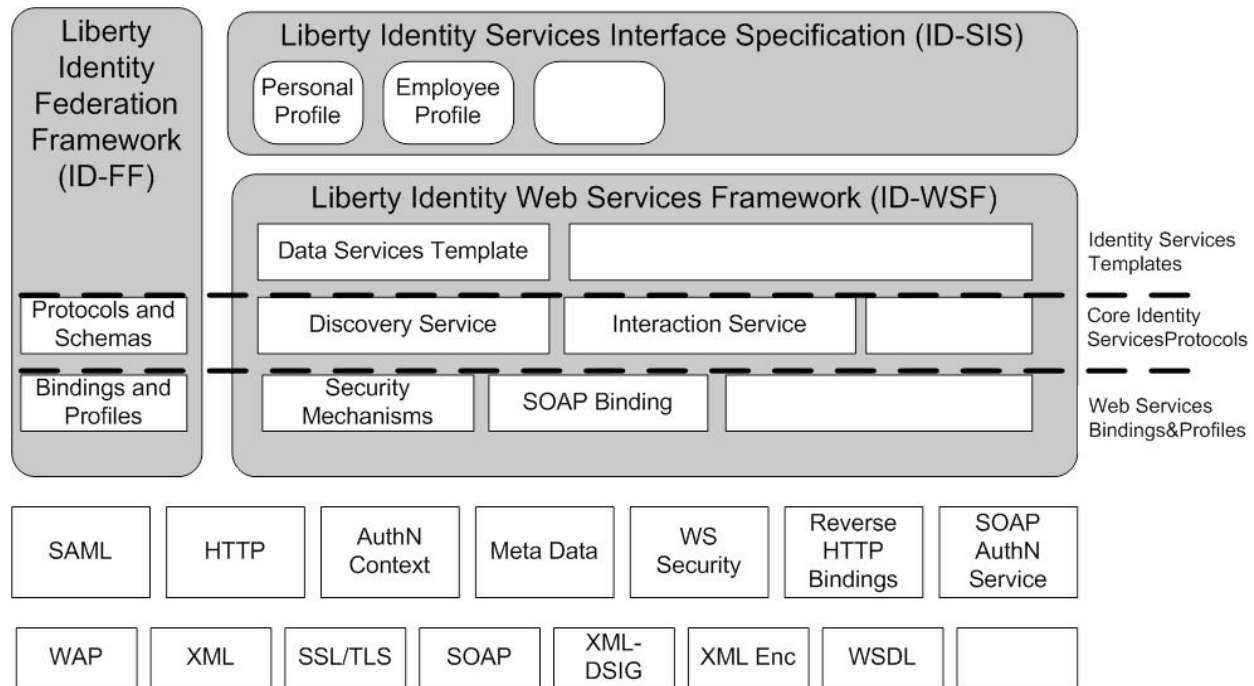


Figure 2.6 Liberty Modules

As shown in Figure 2.6, the Liberty Identity Federation Framework (ID-FF) specifies core protocols, schemata and concrete profiles that allow implementers to create a standardized, multi-vendor, identity federation network.

The Liberty Identity Web Services Framework (ID-WSF) consists of a set of schemata, protocols and profiles for providing a basic framework of identity services, such as identity service discovery and invocation.

Liberty Identity Service Interface Specifications (ID-SIS) utilize the ID-WSF and ID-FF to provide networked identity services, such as contacts, presence detection or wallet services that depend on networked identity.

The Liberty ID-WSF architecture (Figure 2.6) has the notion of a usage directive facility. It is supposed to play a very important role in personal information attribute exchange allowing the requestor and data provider to negotiate intended use and purpose. The specification remains

open ended on the protocol for negotiation and policy format. This also can be carried out in each particular case between two interested parties.

Much like a P3P-APPEL case, a typical policy negotiation scenario in Project Liberty unfolds in the following way. Intended usage is sent along with the request for attributes. Then this usage declaration is evaluated against usage policy maintained by attribute providers. This leaves space for the same ambiguity which plagues the P3P standard.

To interact or “callback” the personal data owner, the Interaction Service was included into the specification. If there is a need for some additional consent from the user, the Interaction Service is supposed to contact him by different means including email and WAP.

To understand the significance of Project Liberty it is important to know the project’s basic engineering requirements. Below is a brief summary of those requirements relevant to this thesis:

- Service Discovery Mechanisms;
- Support for gathering consent from the Individual;
- Support for Usage Directives.

The Individual’s personal privacy is the focus of Project Liberty. Because privacy, security and consumer consideration are the most powerful driving forces in e-commerce the following decisions are fundamental to the specification:

- To use a de-centralized architecture;
- To use a federated architecture, where parties are free to link networks as business judgment dictates;
- To support and promote “permissions-based attribute sharing” [Liberty2002].

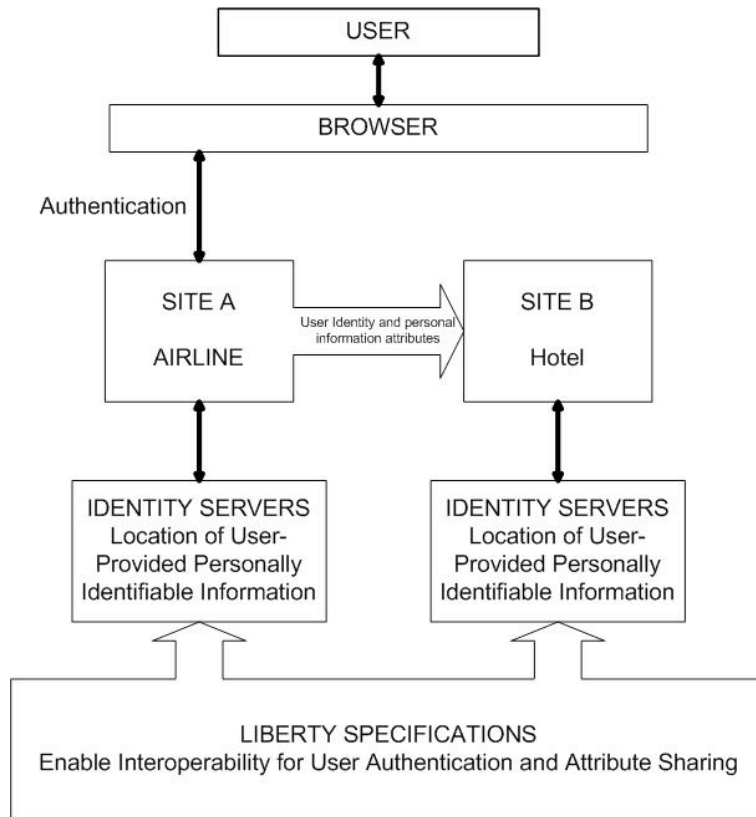


Figure 2.7 Liberty user experience [Liberty2002]

The Figure 2.7 shows a simple example of user experience using Liberty Alliance framework.

Airline (site A) acts not only as an Identity Provider (authenticating user’s identity for Hotel), but also as an Attribute Provider (providing name and address information to Hotel) and as a Discovery Service (letting Hotel know that user’s credit card information may be obtained from Bank, for example). Bank can also be an Attribute Provider because it provides credit card information to Hotel.

User is buying an air ticket from the Airline. She authenticates herself to the Airline. Eventually she decides to book a hotel. For this she goes to the Hotel web site. Because Hotel is Liberty-enabled it is able to identify the user via a unique handle and receive all necessary personal information from different Attribute Providers assuming that user’s usage directives (policies) allow for that.

The Pros and Cons of centralized versus distributed approaches are discussed by Hogg et al [Hogg2000]. Having one trusted third party approach is unacceptable for two main reasons. First, it is unlikely that all interested parties can agree on one single centralized entity. Second, accumulating all information in one place increases the risk of a system-wide failure. A good example of two approaches is Microsoft Password and Liberty Alliance frameworks for single sign-on. It is too early to say anything about Project Liberty – the first working prototype appeared only recently. However, it is a matter of fact that despite the leadership of Microsoft in the operating systems market and all its efforts in promoting Passport, the desired goal – wide adoption of Passport – hasn't been reached.

3 A FRAMEWORK FOR PRIVACY ENHANCEMENT

Because none of the existing technology for privacy protection on the Internet provides automatic disclosure of personal information attributes while providing a high level of compliance with existing privacy legislation and convenience for users, there is a need for a new solution.

An approach to these situations is to create a framework that extends the ITR [Peyton2004] approach by incorporating personal profiles used in Project Liberty to store personal information attributes. Moreover, an ITR will be placed in the context of the type of decentralized architecture consistent with the Liberty project. In order to eliminate the drawbacks of using the P3P/APPEL (Liberty PPEL) language to express policies and preferences, it is proposed to use EPAL as a policy description language in the PDP (policy decision point) component.

The framework consists of a number of components. Each component represents certain roles (e.g. Attribute Provider, PDP, ITR) that each business participant can assume. Because each business can assume more than one role, multiple businesses can play the same role, and businesses themselves are distributed over the network, the components of the framework will be distributed as well. Participants can join and leave the framework if they follow certain rules. The framework will have no centralized entity, hence it will be decentralized and organically growing. Components of the framework will communicate with each other using the Discovery Service.

Each player in the framework has its own unique identity. This is provided by a federated identity framework, often called a single sign-on framework. One example of such a framework is Liberty Alliance ID-FF [Liberty2003]. This framework is also decentralized and organically growing which allows for seamless integration with the proposed framework.

3.1 Framework Components and Architecture

The system essentially consists of a large number of collaborating distributed components which are usually implemented as web services or in some cases as web sites. All those components are a part of the particular enterprise information system with the difference that those components primarily focus on handling personal information. Under a normal deployment scenario, those components would be plugged into the existing system to serve as an enterprise front end for all personal information related operations.

Independently from the physical configuration, the system has a core distributed components part with access gateway called Customer Gateway used by individual users (individuals) as a one-stop shop for all privacy needs and single entry point in the framework (Figure 3.1).

Businesses communicate with the component of the framework directly using the API each component provides. Components expose their APIs as a web services. Each business information system has a support for all necessary protocols, components and APIs to handle such interaction.

For an individual, there is no need to install any extra components as everything is done via the web interface offered by a Customer Gateway.

Using the Customer Gateway interface, an individual can manage and audit his personal information usage in one place independently of the number of businesses in which this information is kept. The Customer Gateway collects all necessary information distributed around and creates a composite view of it to present it to the user. This means that if enterprise A keeps one part of the audit information, enterprise B keeps another part and C keeps the rest, a Customer Gateway will locate them all by using the discovery service, obtain all necessary information and compile one full audit report for the individual.

It is important to consider that the information collected by the Customer Gateway will be personal information in its own right, so the enterprise which maintains this particular gateway may not have a right to collect such information.

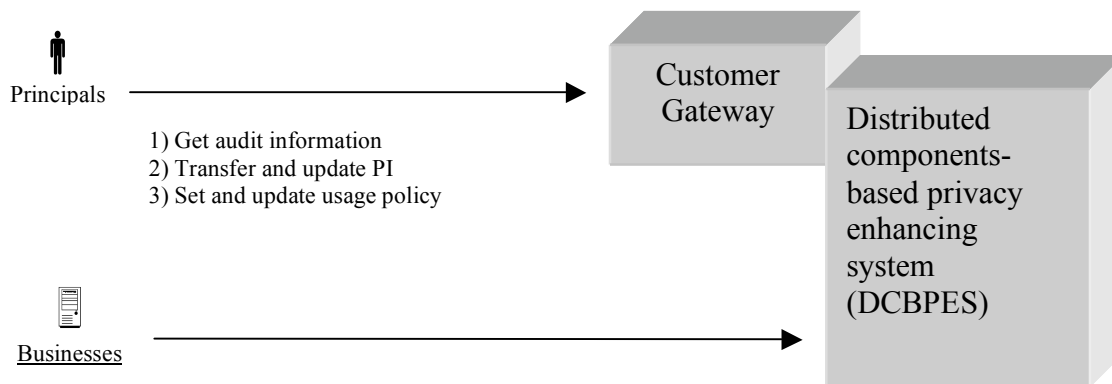


Figure 3.1 Core framework architecture

Individuals can obtain information on how their Personal Information (PI) is used. This includes who keeps the PI, which PI is kept, and when and who requested PI and PI transfers.

To allow PI sharing, Individuals should provide the system with their PI. To ensure that the PI is up to date, Individuals have an ability to modify the PI.

Using the Customer Gateway, Individuals can create and edit a PI usage policy. Although the Individual's PI and the policy may be replicated across the system, with the help of the Customer Gateway he or she sees them as if they are kept in one place.

If when a business joins the system it is already in possession of PI, it is required that the business registers as an Attribute Provider that allows the Individual to have access to his or her PI kept by the business.

Apart from the role of Attribute Provider, there are several other functions a business can perform. The collection of roles each business can play is shown in Figure 3.2.

These roles include:

- A Customer Gateway.
- An Attribute Provider. This is one of the main functional roles of the system. This component stores personal information attributes and releases them on demand from another service provider based on the decision of a Policy Decision Point (PDP).
- A PDP. This has a usage policy for each particular individual. This component makes a decision about whether to allow or deny each particular operation involving personal information based on a number of conditions such as the requestor's identity, the information requested, the purpose of the request, and other dynamic factors such as the current time for instance.
- The Discovery service. This is a key for the whole framework's functionality. Thanks to it, it is possible now to locate each and every component of the framework which is relevant to the individual. One of the possible usages is discovery of all attribute providers who have personal information attributes of one particular individual. There may be an ability to perform advanced search for, say, some particular attribute of a particular Individual.
- An Information Transfer Registry (ITR). This is used to register all actions involving PI, in order to be able to create detailed reports for the Individual.

Shown on the Figure 3.2 is how DCBPES might look in the ultimate case in which there is only one business as a member of the framework. In this case this business has to play all the roles at once, to make the framework functional.

A shadow around components is used to underline (wherever it is important) the fact that they normally exist in multiple instances.

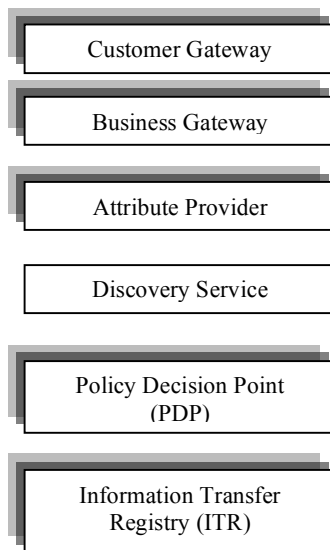
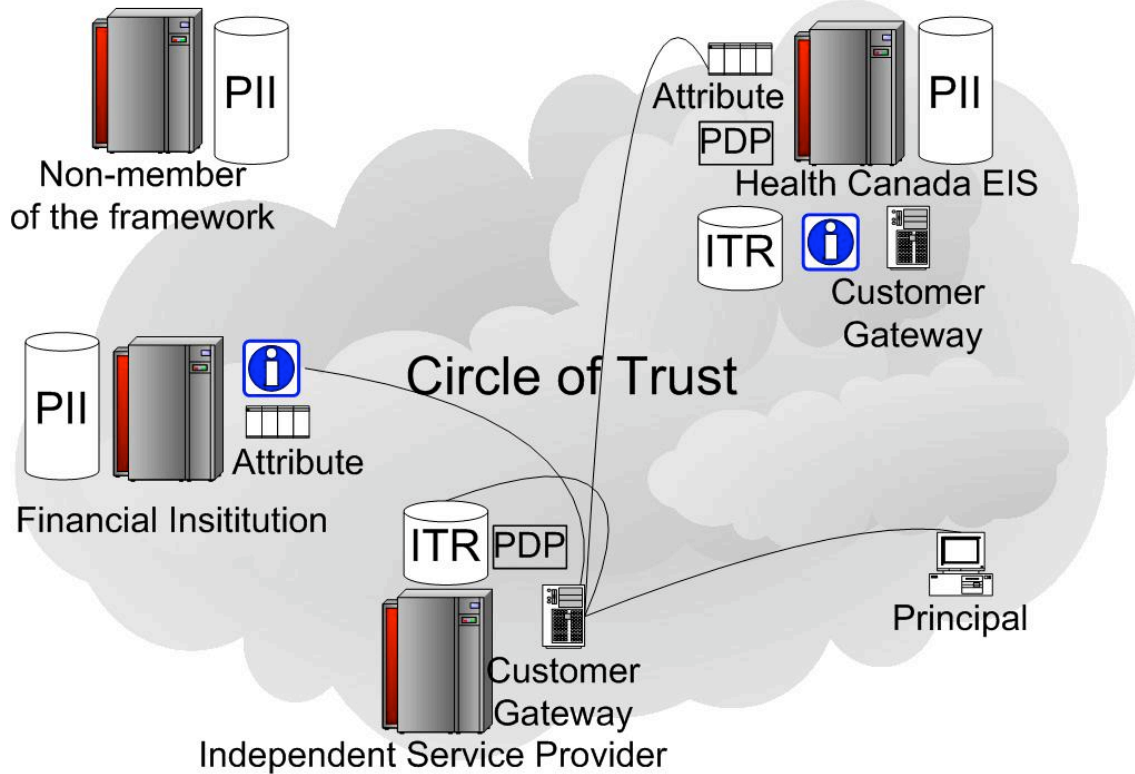


Figure 3.2 Distributed components-based privacy enhancing system (DCBPES) architecture

Before joining the framework, participants are in the form shown as “Non-member of the framework” – just like any standalone Enterprise Information System (EIS) with a databank of Personal Identifiable Information (PII). In order to join the framework, a candidate will first join the Circle of Trust by obtaining a secure identity and implement all necessary security protocols (for instance single sign-on).

Then a candidate has to assume one or more of the framework’s roles: Attribute Provider, PDP, ITR, Discovery Service or Customer Gateway. For instance, Health Canada may assume all possible roles, as the Government is the initiator and biggest promoter of safe privacy practice and privacy legislation. Some entities such as “Financial Institution” and “Independent Service Provider” may assume fewer roles. However there are certain mandatory roles to assume. For instance “Financial Institution” holds PI in the database; hence, it has to play the role of an Attribute provider while “Independent Service Provider” doesn’t have to. It plays the roles of Customer Gateway, ITR and PDP probably hoping to make a profit from an advertisement on the Customer Gateway or by cross-selling some security or privacy protection products.








-  - Attribute Provider
-  - Policy Decision Point (PDP)
-  - Information Transfer Registry (ITR)
-  - Discovery Service
-  - Customer Gateway

Figure 3.3 Framework architecture.

The Individual's possible interaction links are also shown in Figure 3.3. Knowing the Customer Gateway URL of the Independent Service Provider, the Individual uses it as a one-stop service

window for all privacy needs. Using a number of communication and information update/retrieval protocols, the Customer Gateway application (acting on behalf of a Individual) can use service components located all over the circle of trust. All interactions are hidden from the Individual. From the Individual's perspective, it looks as if the data were located in one place.

Figure 3.4 shows simplified interactions in the case where the Individual wants to see his or her audit information. Typically, a individual will use the Customer Gateway's audit facility. On his behalf, the Customer Gateway queries a Discovery Service to find all ITRs which hold the Individual's PI. After getting all records form all ITRs the Customer Gateway will present the Individual with the audit report.

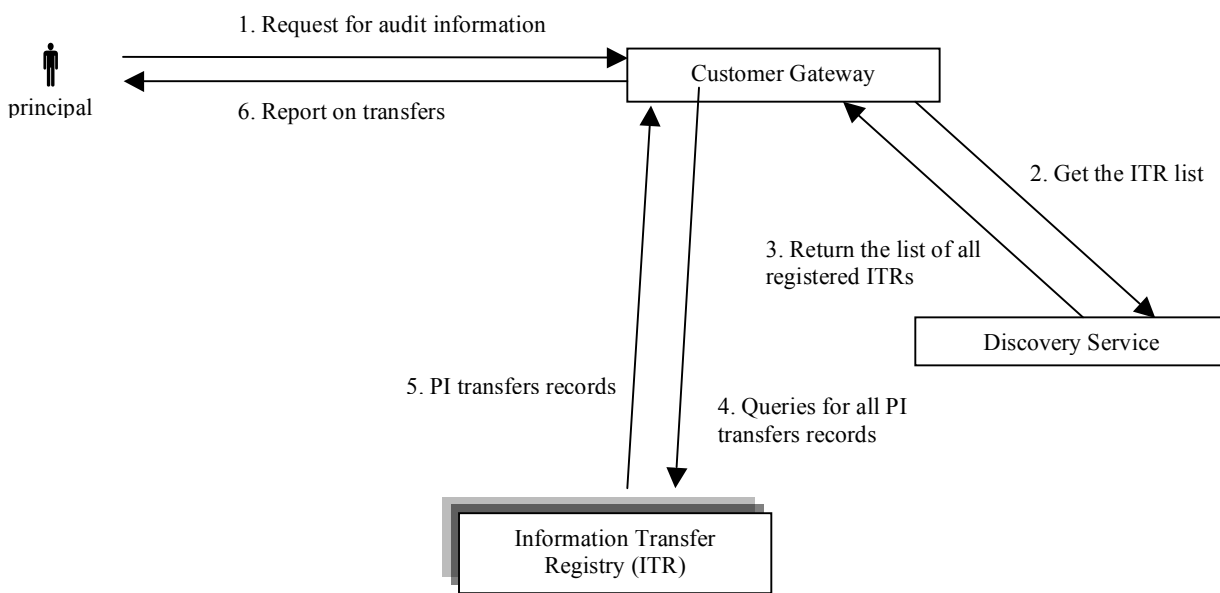


Figure 3.4 Individual gets audit information

Very often, the Individual may have to update or edit or her/his usage policy as shown in Figure 3.5. The Individual may initiate this operation on his or her own or may be asked to do so by one of the service providers when it is necessary for some parties participating in business transaction to obtain certain attributes. Every time any business transaction is about to begin, the business

initiating the transaction checks with the PDP by sending “trial” requests for the data. The main purpose of “trial” requests is to make sure all members of the business chain will get the attributes later during the course of the transaction. For instance, each business transaction is a predetermined chain of the transactions involving several businesses. The business which initiates transaction is aware of all potential consumers of personal information. Instead of collecting personal information it doesn't need in order just to transfer it further down the chain, initiating business sends trial requests to make sure that all information is accessible for the businesses down the chain. This concept helps to solve an excessive collection problem as well as alleviates the need to perform a "call-back" to the Individual in order to receive his consent if later during the transaction one of participants discovers that either information or consent is missing.

If it is determined that some attributes are impossible to get due to the Individual's policy restrictions, or just because there are simply no rules covering these attributes, or because the attributes themselves are missing, then the Individual is forwarded to the Customer Gateway of his or her preference to modify the policy or enter missing attributes. Information about all interested parties is forwarded as well. This is done in order to enable the determination of whose Attribute Provider service should be used to store the Individual's attributes. This is made in the form of a suggestion to the Individual who makes the final decision.

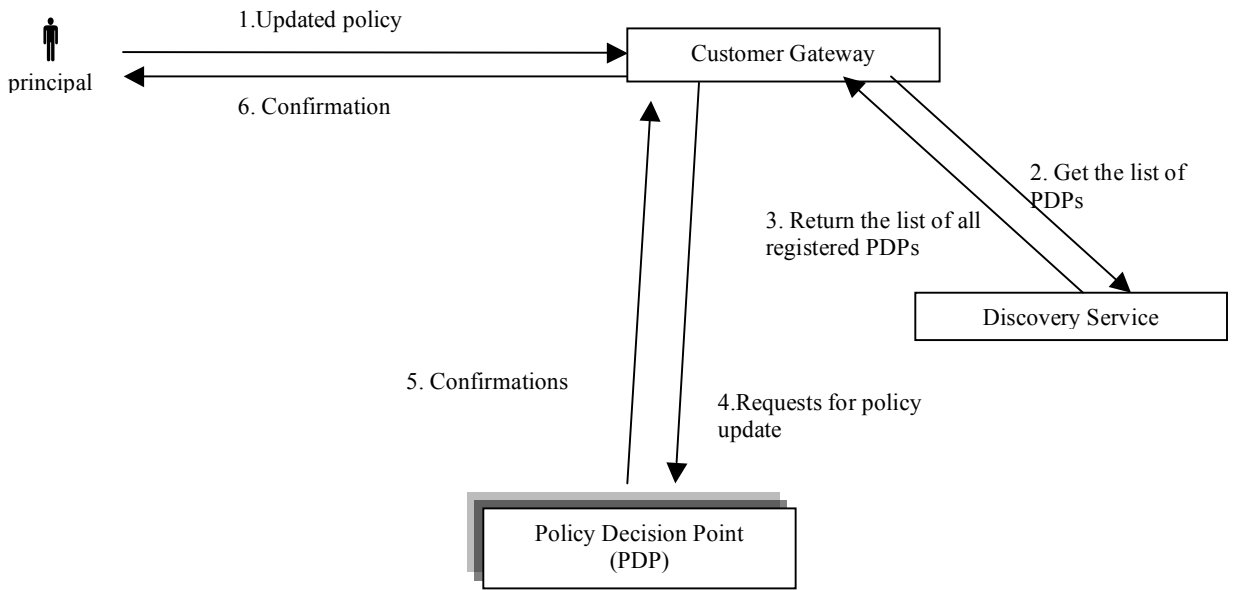


Figure 3.5 Individual creates/updates PI usage policy

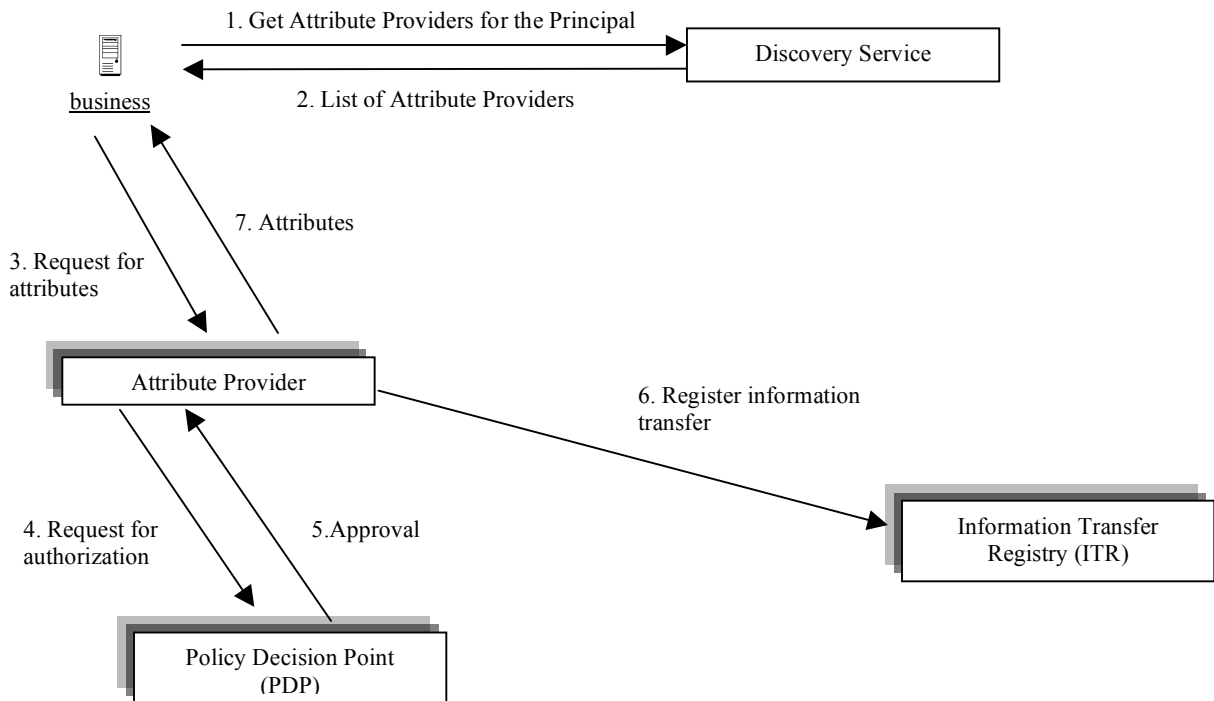


Figure 3.6 Business gets Individual's PI

The most important interaction scenario (Figure 3.6) occurs when it is necessary for the business to obtain the Individual's PI attributes. Through the Discovery Service, all Attribute Providers holding necessary attributes are found and a request is sent to the necessary ones for attributes. Each Attribute Provider sends a request for resolution to the PDP which makes a decision to allow or deny the request. Finally, attributes are disclosed and a ticket is sent to the ITR in order to register the transfer for any future audit.

The information audit can be used to verify that all members of the framework are performing steps 4 and 6 during each transaction. However a security mechanism should be applied to ensure that all framework members follow these steps. As well, all those actions should be made a part of a single transaction to provide for integrity of data across the framework.

Up until now, no clear binding has been made to any particular technology although it could be inferred that HTTP protocol is at the core of all interactions. While there is no question about Individual-system interaction which is done via HTTP and a web interface, a question still exists about inter-component interaction.

3.2 Security Considerations

The focus of this thesis is on a privacy framework to provide users with control, accuracy and audit. In any real world business situation, security and trust have to be carefully considered in order to make the framework viable. A detailed analysis of security and trust issues is beyond the scope of this thesis. However, it is important to understand that an underlying assumption of our approach is that an appropriate social and legal framework has been put in place to regulate and control security and trust. In particular, the "Circle of Trust" is created by appropriate legal contracts and regulations. The companies responsible for Attribute Provider services would be legally liable if they did not do a request for authorization (step 4 Figure 3.6) or register

information transfers (step 6, Figure 3.6). A mechanism for a legal or business audit to take place and verify compliance would need to be in place as well. The “Circle of Trust” is a business framework as well as a technical framework, in which membership is a privilege and there are appropriate legal and social penalties in place to ensure compliance. Market forces would be expected to dictate that such a business framework relies on a continuing good reputation and relationship with customers in order to survive. The onus would be on the members of the Circle of Trust to establish appropriate security and trust measures and ensure they are followed.

Here are some of the assumptions about security and trust that are made in our framework, that a Circle of Trust would have to ensure were met through appropriate safeguards to protect integrity and security of the personal information.

1. The framework is built on top of the existing infrastructure which in turn uses the Internet for communication. Therefore all of the components of the framework are inherently susceptible to the usual kind of attacks faced by any internet application.
2. All components of the framework are trusted. That is they will behave as described in the framework. This means not only will they behave cooperatively, but they will have mechanisms in place to protect them from outside attempts to compromise them. At this point there are still many issues with providing reliable means of protection from all kinds of security threats. These issues are the focus of much current research and are not covered in this thesis.
3. It is possible to compromise personal information due to human error or intentionally breaking the rules of the framework. That is why the framework has a built-in audit mechanism to provide a record that can be used to validate and certify the behavior in the framework and assist in the prosecution of those who have behaved illegally.

4. A mechanism exists to provide each individual and corporate player with uncompromised secure authentication and signing keys. So when it is said that consent is given it is assumed that there is a secure ticket signed by the party whose consent is collected and when it is said that the user is authenticated there is a reliable way to provide this authentication. Moreover, the technology used for digital signature has been endorsed by all legal and social regulatory bodies. Secure digital identity and signature is a very comprehensive topic which lies outside of the scope of this thesis.
5. The identity of any business member of the Circle of Trust is known and recognized by other business members of the Circle. The identity of all business members is open and verifiable by consumers. Consumers have a mechanism available for taking legal action against members of the Circle of Trust.

Over time, it is possible for the Circle of Trust to mature into a global mechanism for regulating on-line commercial environments incorporating the majority of organizations. However that may never happen due to the number of unresolved cornerstone social and technical issues. Being a member of the Circle is a privilege because it gives considerable competitive advantage but it also comes with a price - every member has to comply and follow strict rules of the framework. There is transparent and reliable mechanism to evaluate whether a member is in compliance with the rules or not. Violators can have their membership suspended or permanently revoked as well as facing legal and criminal sanctions.

3.3 Some Key Architectural Patterns

This section is a review of a few important patterns. The Cross Domain Cookie pattern demonstrates how single sign-on can be achieved across different domains. The concept of single sign-on is fundamental for building a circle of trust. Using cookies for achieving single sign-on

may impose certain constraints on the implementation of the framework as well as it may introduce additional security threats.

In the proposed framework, some interactions are transactional in nature which may pose a problem in case all components of the framework are implemented as a web service. The second pattern, called Message Correlation/Callback Strategy demonstrates that there is a possible solution for this issue.

And finally third pattern (Service Coordinator) was the key pattern used in the framework to implement a customer service gateway.

3.3.1 Cross Domain Cookie

The concept of Single Sign-on (SSO) is the basis of “circle of trust” creation. A user session within the framework can be maintained by the means of a security ticket which is in essence an HTTP cookie. HTTP cookies have a number of limitations, one of which is the so-called “double dot” rule allowing one to set cookies at the highest granularity of a second domain level (e.g. *.mydomain.com). The server residing in *.onedomain.com is not able to read the cookies from *.anotherdomain.com. And it is impossible to set cookie for just the *.com top level domain. This limitation becomes a serious obstacle in SSO if the framework spans multiple domains, which is always a case in our framework. This way when user's session is authorized within one domain it is difficult to pass the security ticket onto another domain due to second domain level limitation.

An approach to circumvent the single domain limitation is called the “Cross Domain Cookie”. Very little has been done to formalize this pattern, however it is being used in some systems including a similar approach used in Project Liberty. One description has been given by Matt Pouttu-Clarke [Pouttu-Clarke2005].

This pattern has the following components:

- *Client*: A client communicating using the HTTP protocol, usually web-browser but also can be any client using SOAP over HTTP;
- *Master Domain*: The original cookie domain;
- *Vanity Domains*: Domains within the same hierarchy as the customer's domain;
- *Cross Domain Cookie*;
- *Vanity Servers*: Servers from the Vanity Domain;
- *Master Server*: A single server from Master Domain.

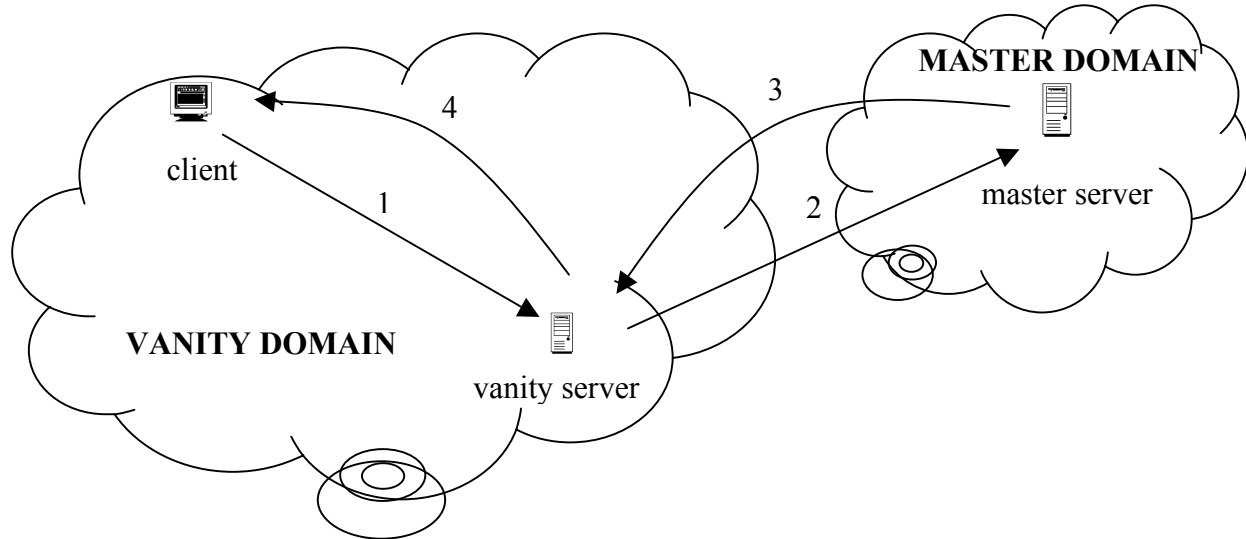


Figure 3.7 Cross-domain cookie pattern

Figure 3.7 shows basic client-server interaction for this pattern.

1. The client's request to a vanity server;
2. An HTTP redirect containing the originally requested URL as a parameter;
3. The request is returned with a cookie attached as a parameter;
4. The response is returned with a clone of the cookie for the vanity domain.

This pattern allows for transparent cookie propagation between different domain members and proxies.

Another important side effect of this pattern is its use by the Customer Gateway component. The fact is that the Customer Gateway being a trusted entity in many cases acts on behalf of a Individual yet must not compromise the security of the Individual's identity. As shown in Figure 3.8, the Customer Gateway can "spoof" an Individual session upon his or her agreement by redirecting him or her to the identity provider service and obtaining from the identity provider a secure ticket copy (as an attachment) in order to present it to other members of the framework. This ticket is invalidated (it is only valid for one session) after the Individual logs off from the framework. Ticket may have properties similar to the LTPA (Lightweight Third-Party Authentication) token.

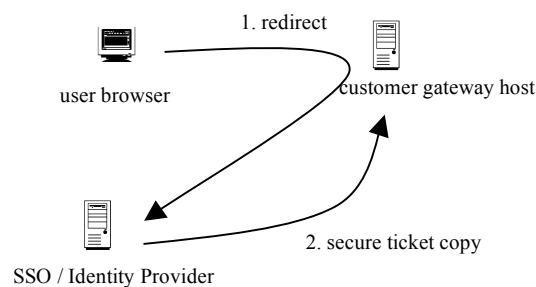


Figure 3.8 Customer Gateway single sign-on

3.3.2 Message Callback/Correlation Strategy

All distributed components of the framework are envisioned as web services. Some of the components of the framework will call methods on remote services which require extensive processing. For instance, policy update across multiple PDPs can take a long time making synchronous call infeasible. This results in a problem caused by asynchronous callback calls when it is difficult to match initial request with subsequent response(s) which happened at different moments in time. The pattern called Message Correlation Strategy [Sun2004] helps to solve this problem.

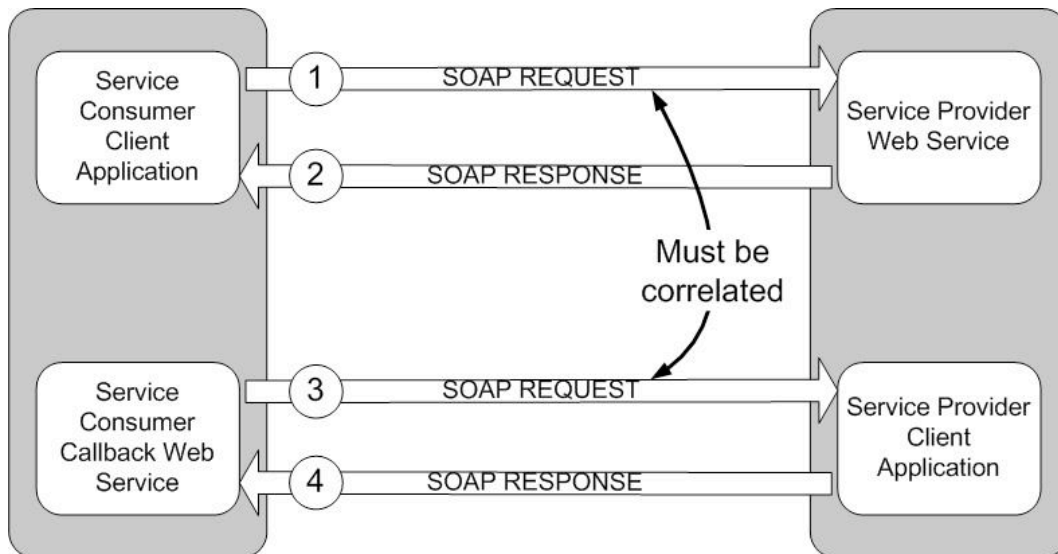


Figure 3.9 Message correlation pattern

Figure 3.9 shows an example. After sending the initial request, the client gets confirmation (2) instead of the business response which is sent later (4). To maintain the correlation between (2) and (3) a unique identifier is attached to (1) and (3) usually as a part of the message.

This simple-to-implement pattern is indispensable in case of asynchronous SOAP calls, however some additions have to be made when implementing it. One of the main drawbacks is that in this

form the pattern doesn't guarantee message delivery and it has to be done using additional mechanisms.

It also may have a security implications should the unique session ID be made available to the third party. This pattern is used in combination with other authentication mechanisms, therefore security risk is minimized.

3.3.3 Service Coordinator

Proposed by M.Bigatti [Bigatti2004] this pattern may be used to orchestrate communication among different components, as shown in Figure 3.10. It may be useful to encapsulate all logic to communicate among components into one component, the "Service Coordinator". When all communication aspects are isolated in the Service Coordinator, other components can concentrate on their primary concerns. In the proposed framework, the Customer Gateway is an example of a Service Coordinator pattern.

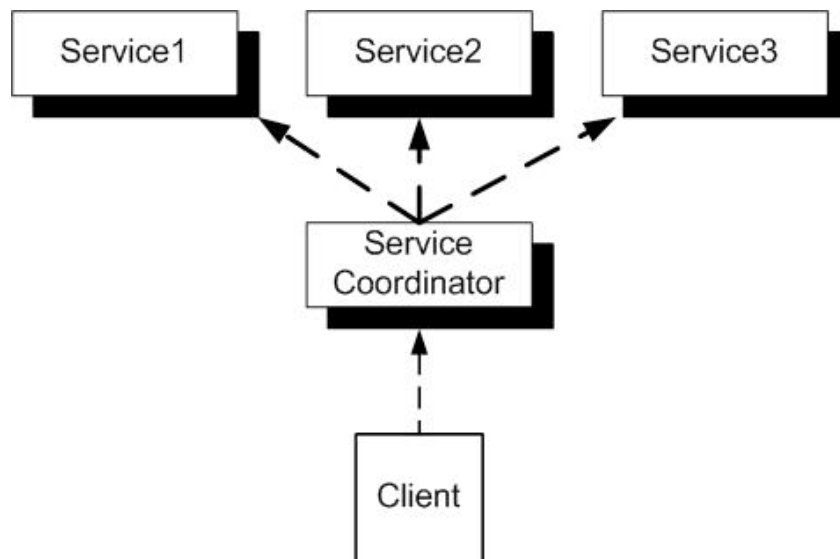


Figure 3.10 Service coordinator

3.4 Performance and Scalability

The distributed architecture of the framework makes performance evaluation a non-trivial task.

The overall performance of the framework will be much depending on the performance of each individual component and users' behavior.

There are several major factors which can impact performance:

- total number of users and businesses in the framework.
- amount and granularity of personal information kept by each business participant.
- complexity of the policies defined by each user.
- number of participants in a typical business transaction. The bigger the chain the more “maintenance” requests will be sent to the PDP to determine if all participants down the chain are eligible for receiving personal information.
- amount of personal information collected during business transaction
- preferences of users on whether to define complete policy at the time when they join the framework or define it as they go.
- performance of the network.
- organically growing nature. Being one of the strongest points of the framework it poses a threat when the framework grows out of proportion. Too many policies, audit information scattered around will cause an overhead when updating and retrieving it. Limiting this number too much will also hurt performance and availability bottom lines. A protocol has to be developed to keep this numbers within optimal limits.

To collect performance metrics a full-scale environment has to be build with emulation of simultaneous users' activity which in correspondence with real-world pattern. Due to the limited scope of this thesis it is presumed to be a part of the future work.

At the same time the distributed architecture of the framework eases off the task of making it scalable. There is no obvious bottleneck in the framework as the framework is de-centralized. Large number of participants may however make some supplementary tasks such as for example monitoring trustworthiness or auditing participants resource intensive. But this task is secondary to the primary framework tasks – automated disclosure of personal information when preserving privacy. After all, the framework can be divided on the sub-domains of the fixed size which should decrease overhead associated with auditing and monitoring tasks. All aforementioned, however, need to be thoroughly evaluated and rested on the real-world prototype of the framework.

4 CASE STUDY

This Chapter is organized in the following way. It begins with the description of the implementation and testing environment we created for validating our framework in direct comparison with P3P. Then we describe a drugstore scenario that we used to drive the three main use cases used in our validation. These use cases (sections 4.3.1, 4.3.2, 4.3.3) are used to compare the proposed framework with a P3P framework. A Privacy impact assessment was also conducted on the drugstore scenario to highlight potential privacy threats associated with it.

In addition to the drugstore scenario, we also describe the Joe Self scenario from the literature of the Liberty Project [Liberty2003]. No working implementation of the Liberty Alliance framework is available, but we compare our analysis of the Joe Self scenario based on our framework and principles with the analysis that has been done in the context of the Liberty Alliance Framework.

4.1 Framework implementation

In order to validate the approach proposed in this thesis, a prototype has been developed along with a testing environment. Three scenarios derived from the drugstore scenario are run in this environment. The same scenarios are also executed in a P3P framework. Then those results are compared.

This environment consists of two separate players from diagram Figure 4.1 (not all participants are used for test scenarios):

- Individual;
- Drugstore.

These players will perform the following functions:

- Individual: Will initiate all testing scenarios. He or she will interact with the Drugstore's web site according to testing scenarios.
- Drugstore: Is represented by the drugstore web site which allows an Individual to fulfill the prescription on-line.

For simplicity, the assumption was made that all participants are part of a single sign-on framework hence all authentication issues can be left outside the scope of the testing framework. Also, instead of using fully functional web services in testing the framework, plain objects implementing the same API are used instead.

Validation of the system is done by analyzing logs from the logging facility and by analyzing the content of the databases where the Individual's information is stored along with the ITR database reflecting transfers of PI.

4.2 Drugstore scenario

The scenario (Figure 4.1) was used to derive use cases which were later executed using implementation of the framework and using P3P technology. In this scenario, the Individual visits a drugstore in order to get prescription medicine. The drugstore needs certain personal information to fulfill the prescription. The information required is the name of the Individual, the address, the employer name, the Insurance Company name and policy number, and (optionally) the Individual's medical records.

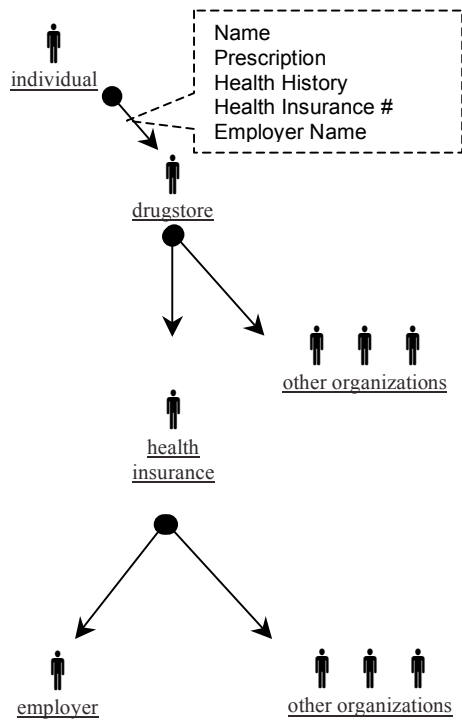


Figure 4.1 Drugstore scenario [Peyton2004]

The Individual fills out a form with all necessary information. The Drugstore may even ask for the permission to transfer the information to the Insurance Company. What is not clear is that the Insurance Company may later transfer, for example, prescription information to the Employer.

As a privacy protection measure each organization is recommended to conduct a Privacy Impact Assessment (PIA). The PIA isn't mandatory but may be so for large organizations or governmental departments. It is especially recommended if the organization is going to offer a new service or change the way of delivering it's service, for example from paper-based to electronic.

The scenario (Figure 4.1) was assessed in accordance with the Treasure Board of Canada Privacy Impact Assessment Guidelines. The ultimate objective of the PIA is a PIA Report which, among other things, contains a Summary table and a Data Flow table. The PIA Report is produced in order to give management a systematic view of the operations with personal information within

the organization and possible threats associated with it. The Data Flow (Table 4-1) contains a list of all possible ways personal information is collected, disclosed and stored. It is used later to produce the Summary table (Table 4-2).

| Description of personal information cluster | Collected by | Type of format (e.g. paper, electronic) | Used by | Purpose of collection | Disclosed to | Storage or retention site |
|--|---------------------|--|---------------------------------|--|-------------------------------------|----------------------------------|
| Name and Address | pharmacist | electronic | Pharmacist, Delivery, Insurance | Validate prescription, Delivery, Insurance | Delivery Company, Insurance Company | Own database |
| Prescription | pharmacist | electronic | pharmacist | Validate prescription | Insurance Company | Own database |
| Health History | pharmacist | electronic | pharmacist | Validate prescription | nobody | Own database |
| Health Insurance | pharmacist | electronic | accounts Receivable | Claim Insurance Money | Insurance Company | Own database |
| Employer Name | pharmacist | electronic | accounts Receivable | Claim Insurance Money | Insurance Company | Own database |

Table 4-1 PIA Drugstore Data Flow table

| Element | Nature of risks | Level of risks | | | Comments | Mitigating Mechanisms |
|---|---|----------------|--------|------|--|---|
| | | Low | Medium | High | | |
| Collection and disclosure Address and Prescription and Employer information | Combination of three different pieces of data by Employer | * | | | Can cause considerable legal and financial consequences if proved that such combination cost somebody a job. | Avoid collection of such information or obtain consent for such disclosure. |
| Storing all pieces of information in the database | Improper use of the information by staff members | | | * | Can be abused by staff members and lead to a leak of sensitive information. | Build proper safeguards and avoid storing of personal information |
| Disclosure of prescription to health insurance | Collection of prescription patterns by a health provider | * | | | Some doctors and patients may consider prescription their personal sensitive information. | Avoid direct disclosure of such information or ask explicit consent to do so. |

Table 4-2 PIA Drugstore Summary Table

4.3 Use cases for the drugstore scenario

This section contains a description of the use cases used for comparison of the proposed approach with the P3P framework.

4.3.1 Scenario I (main):

Preconditions:

The Individual's PI is stored by the Attribute Provider. The Individual has predefined a policy with the PDP allowing the Health Insurance Company to use his or her PI including prescription

information and the physician's name. This policy also allows the Pharmacy to receive prescription information and the physician's name as well.

Scenario:

The Individual logs on to the Drugstore web site in order to get a prescription. After authenticating the Individual, the drugstore requests all necessary attributes from the Attribute Provider and successfully receives them.

The Drugstore calls a business method on the Health Insurance Company service in order to claim insurance money. Before that, the Drugstore checks with the PDP to make sure that the Health Insurance company is allowed to get all the PI it needs.

The Health Insurance Company requests all necessary PI from the Attribute Provider and successfully receives it.

4.3.2 Scenario II (missing consent):

Preconditions:

The Individual's PI is stored by the Attribute Provider. The Individual has a predefined policy with the PDP which *doesn't* allow the Health Insurance Company to use his or her PI including prescription information and the physician's name. This policy, though, allows the Drugstore to receive prescription information and the physician name as well.

Scenario:

The Individual logs on to the Drugstore web site in order to get a prescription. After authenticating the Individual, the Drugstore requests all necessary attributes from the Attribute Provider and successfully receives them.

The Drugstore checks with the PDP whether the Health Insurance Company is allowed to receive all necessary information. After getting a negative answer the Individual is redirected to the Customer Gateway's web site where he or she is prompted for the consent. If consent is obtained the Individual's PI usage policy is modified accordingly.

The Drugstore calls a business method of the Health Insurance Company's service in order to claim insurance money.

The Health Insurance Company requests all necessary PI from the Attribute Provider and successfully receives it.

4.3.3 Scenario III (individual controls his personal information usage):

Preconditions:

Scenarios II and I were executed successfully.

Scenario:

The Individual logs on to the Customer Gateway web site and requests a report on his or her personal information usage. The Customer Gateway finds all ITRs holding information on transfers of Individual's PI using the Discovery Service. The Customer Gateway queries all ITRs for information about the Individual's PI transfers. The report is compiled and presented to the Individual.

4.4 Joe Self usage scenario from the Project Liberty specification

Presently there is no existing working implementation of the Project Liberty which makes it impossible to run validation scenarios. Nevertheless this separate scenario is given to demonstrate the advantages of the proposed framework.

In the Project Liberty ID-WSF Web Services Framework Overview [Liberty2003], an example usage scenario is given to illustrate the importance of user control over accuracy. Joe Self

decides to change his flight, but discovers that the airline he usually uses is booked and he tries to switch to another airline. The new airline is able to discover his personal profile but the attributes there are outdated. Joe has to update his attributes before booking a flight. Eventually he decides to create another profile on his mobile device in case he needs it again.

The Joe Self and drugstore (called Rugstore.com to avoid trademark law infringement) scenarios have many common points as well as differences.

In the drugstore case, the primary concern is to prevent unwanted disclosure of personal information, whereas in the case of Joe Self it is important to provide disclosure while ensuring that disclosed information is correct.

Another difference is the environment of the two cases. The drugstore case is taken from the healthcare domain and the Joe Self case is taken from the travel industry.

A common point is that in both cases it is important to provide the user with audit and control over personal information.

In both cases, information being disclosed is extremely sensitive: prescription information in case of the drugstore example, and financial information the Joe Self case.

Despite existing differences, the framework handles both scenarios in the same way.

During the execution of the scenarios the most attention will be paid to the features partially or not fully supported in existing frameworks for the consequent comparison to the support by the proposed framework. These features are summarized in Table 4-3.

| | Control | Accuracy | Audit |
|------------------|---------|----------|-------|
| P3P | no | no | no |
| HP Framework | no | no | yes |
| Liberty Alliance | yes | no | yes |
| IBM EPA* | no | yes | yes |

* designed to function only within one enterprise

Table 4-3 Control, Accuracy and Audit in the existing technologies

The comparison of the results will demonstrate how this functionality (which is lacking in the frameworks from Table 4-3 can be better covered in the proposed framework.

5 ANALYSIS OF RESULTS

Before starting the analysis of the results, a few words should be given regarding the approach taken to analyze the results. We will also discuss the particular implementation of the demo system used both to demonstrate the main functionality of the framework and also to compare existing privacy enhancing technology, such as P3P, with the proposed framework.

In order to validate the proposed solution, a simplified version of the framework was built along with necessary test infrastructure to make it possible to run the main scenarios. At the same time, a parallel test infrastructure was developed which is based on the P3P specification. All scenarios, where possible, were run on both systems. The outcome of each test was documented and later used to compare the two approaches.

P3P technology was chosen for comparison because it is the most widely accepted and perhaps the only functioning privacy enhancing technology for the Internet now available.

The P3P-based system consists of a fictitious Drugstore (called Rugstore.com to avoid trademark law infringement) web site with a P3P policy associated with it. The full policy can be seen in *Appendix 2. P3P policy file used for Drugstore demo*. In order to analyze the P3P policy, the user must use an agent. Due to limited capabilities of the browser, most of which are focused only on analysis of short policies which accompany browser cookies, a more advanced user agent had to be found. Preference was given to the Privacy Bird [Privacy Bird] developed by the creators of P3P in AT&T Labs as a tool demonstrating the P3P specification in action and promoting privacy awareness among Internet users.

Privacy Bird is essentially a browser plug-in. It has three predefined APPEL rulesets, which range from low to high privacy restriction level. The APPEL preference file used for testing and corresponding to a high restriction level setting is shown in *Appendix 3. APPEL properties file*.

There is no tool for editing the APPEL preferences provided although a third-party utility might be used for that purpose.

When the user navigates to web site, Privacy Bird retrieves the P3P policy associated with this site and compares it to the set of user preferences defined in APPEL. In the top right corner of the browser window a little bird head icon appears in different colors. A green color indicates that the P3P policy of the web site corresponds to the user's preferences.

As for the proposed framework, a simplified version of it was implemented for testing purposes along with a web-based application: the fictitious Rugstore.com – an online store selling medicinal drugs.

Components of the system implementing core interfaces are included in *Appendix 1. Demo system class diagrams*. The package diagram for the whole testing framework shown in Figure Appendix1.1.

The Javadoc for the demo system can be found at Demo Javadoc (supplied with the source code).

Some assumptions and simplifications were made to allow rapid development of the testing framework, namely:

- In the real framework, all components such as PDP, Attribute Provider, ITR, Customer Gateway and Discovery Service can exist in multiple instances, usually web services, for the same user distributed over the Internet. Users or their agents have to know a binding URL of a single Discovery Service to use it as a starting point of entry to the framework. In the demo system, all those components are represented by a single instance which is a Java class implementing one of the core interfaces. Logically, there is no difference between a Java class or web service implementing same interface. In the real-world, a

complicated search and combining algorithm should be used in order to first find all services pertaining to the current user and then combine the same or different personal attributes avoiding possible conflicts if the content of the same attribute isn't the same across different providers.

- Identity management issues were left outside of the scope of the thesis and demo application as a completely separate research area. Real-world applications will rely on distributed identity frameworks such as Project Liberty and Microsoft Passport. The demo application supports only single-user operating mode. All scenarios are shown to work for one fictitious user.
- EPAL was the language of choice to implement the user-defined policy as the closest solution to satisfy the needs for the framework's policy language. The IBM implementation engine was used to evaluate conditions under this policy. This implementation doesn't support the XACML condition definition and evaluation but this doesn't impair illustrating the ability of the demo system. However it is quite possible that EPAL will need some modifications in order to completely fit the proposed framework because it still remains a submission to a standards body and hasn't been standardized or widely recognized.
- Although a work have to done in order to standardize the way the Individual information is stored, for simplicity the demo system uses an arbitrary XML container to store personal attributes along with an arbitrary taxonomy for the meta-information about the attributes. The same pertains to the ITR database which is just a serialized form of a Java collection stored on the hard drive. The demo system wasn't designed as an application capable of handling large amounts of data or high request volume.

For the complete user's personal information profile XML file please refer to *Appendix 4. Individual's profile used in demo system*. For the EPAL policy and vocabulary used for testing please see *Appendix 5. EPAL policy and vocabulary used in demo system*.

5.1 Scenario I: test run result analysis

First let's consider this scenario for the P3P enabled system.

The user visits Rugstore.com to fulfill prescription by pointing his browser to the store URL http://localhost:8080/drugstore/enter_info.html.

As you can see on the screenshot in

Figure 5.1, the Privacy Bird user agent retrieved the P3P policy associated with this part of the Rugstore.com web portal. As indicated by the arrow, the bird sings a tune, which is an indication that the policy matched the user's preferences. Figure 5.2 shows the Privacy Bird preferences screen. The Privacy Level is set to "high" which means that the Bird should warn the user if a web site collects and/or transfers personally identified medical information to a third party.

At the same time the P3P policy contained a declaration that health prescription information is being collected by this site for the purpose of claiming insurance money as you can see on the P3P policy fragment below. The full policy can be found in *Appendix 2. P3P policy file used for Drugstore demo.*:

```
.....
<xmlns="http://www.software.ibm.com/P3P/editor/extension-1.0.html" name="Health Drug
Prescription Information"/>
<CONSEQUENCE>We need this information in order to claim insurance money from your
insurer.</CONSEQUENCE>
<PURPOSE><historical required="opt-in"/><other-purpose required="opt-in">Claim Insurance
Money</other-purpose></PURPOSE>
.....
```

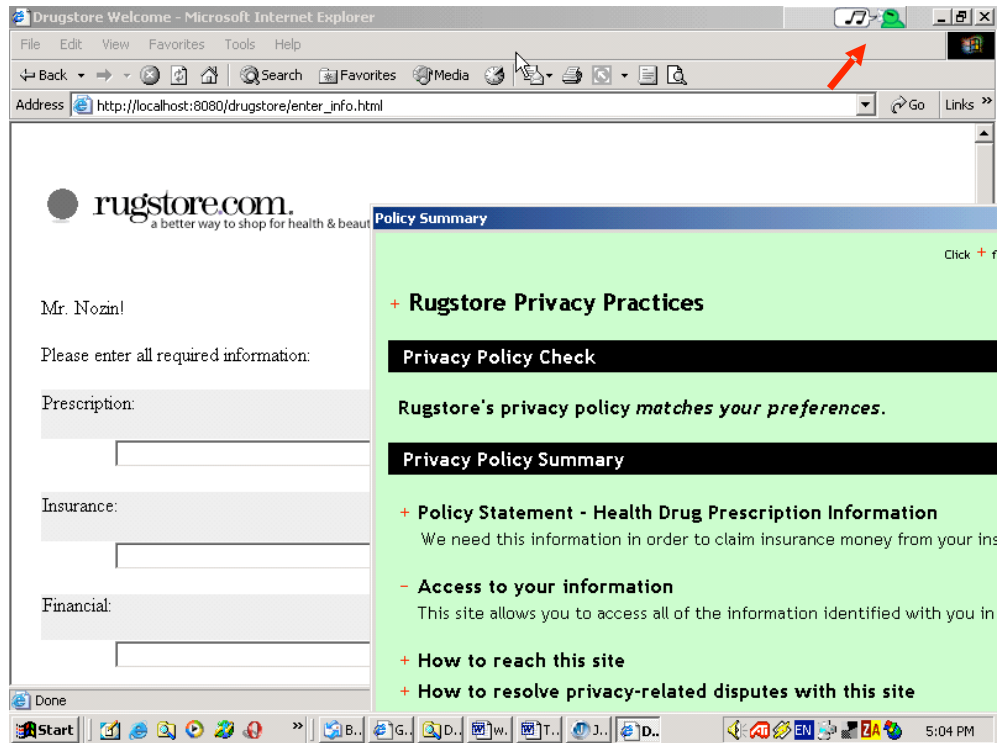


Figure 5.1 P3P enabled web-site screenshot.

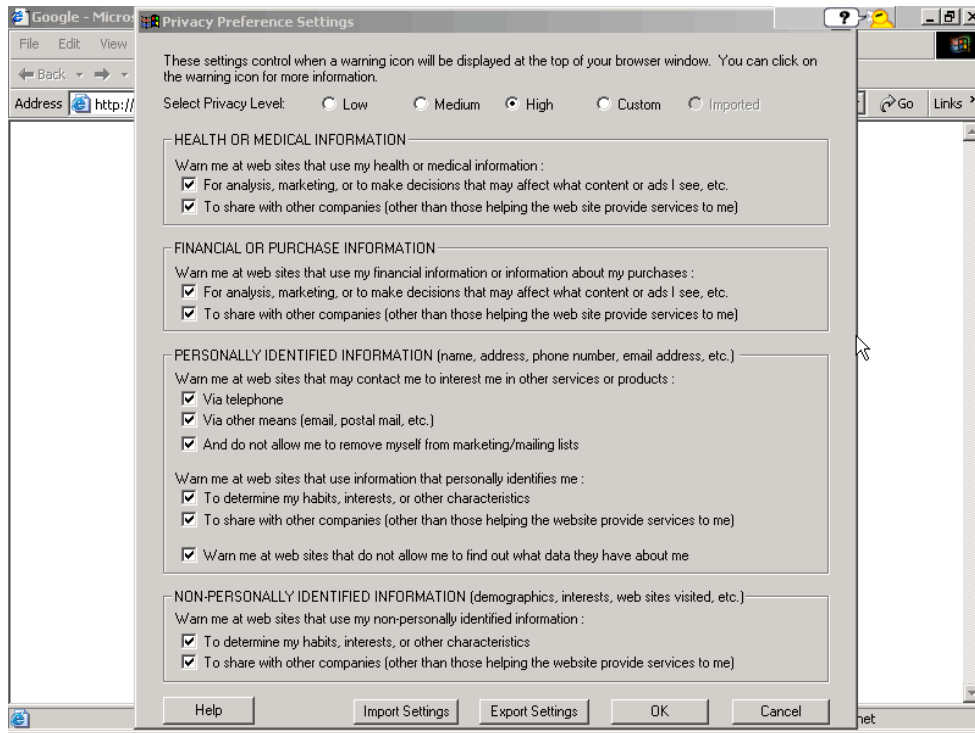


Figure 5.2 Privacy Bird preferences screen.

Due to well-known limitations of P3P, the Privacy Bird failed to warn the user about sensitive information collected. That is because the P3P policy defined a custom data category through an extension mechanism. To achieve better compatibility it is recommended to use a very limited set of predefined categories, which severely limits expressive ability of the policy. It is most likely that any user agent will be biased towards interpreting policies in a more restricted manner fearing that a loose interpretation may cause unwanted disclosure of personal information.

It should also be pointed out that input fields on a screen form have to be manually filled out as the P3P specification doesn't provide any means of automatically exchanging the Individual's personal attributes. It will be demonstrated that the proposed framework better addresses this issue as well as a number of others.

The above-mentioned example almost fully covers all the functionality of P3P defined in the first version of the specification, which defines the main purpose of P3P as a way to give the user notice of the privacy practices followed by a web site.

Now let's see what happens if the same scenario is run in the proposed framework. The user starts by typing in the drugstore web site URL - <http://localhost:8080/pharmacy/index.html>.

After the user clicks the <Continue> button in the client application, which is Rugstore.com, it performs a request to the Discovery Service in order to obtain the Attribute Provider list of all attribute providers which serve personal attributes for the user as shown on Figure 5.4.

After the list is obtained, Rugstore.com sends requests to the Attribute Provider of choice; in our case we have only one Attribute Provider from which to get personal information attributes. The Attribute Provider sends the request for a ruling to the PDP (also obtained through the Discovery Service and not shown on the diagram for simplicity). When a positive ruling is obtained, events are logged with the ITR and attributes are released to the requestor – Rugstore.com. This

scenario sequence can also be seen on Log4J log messages as shown on the Figure 5.5. Please note that for evaluation purpose the most important was to see the sequence of actions. That is why Log4J logs show only actions in order of appearance. Log4J logs are implementation specific and are used to ensure that actual flow of actions in the demo implementation of the framework matches those of testing scenarios. At the same time logs captured by ITR are framework specific and used by individuals to control personal information usage.

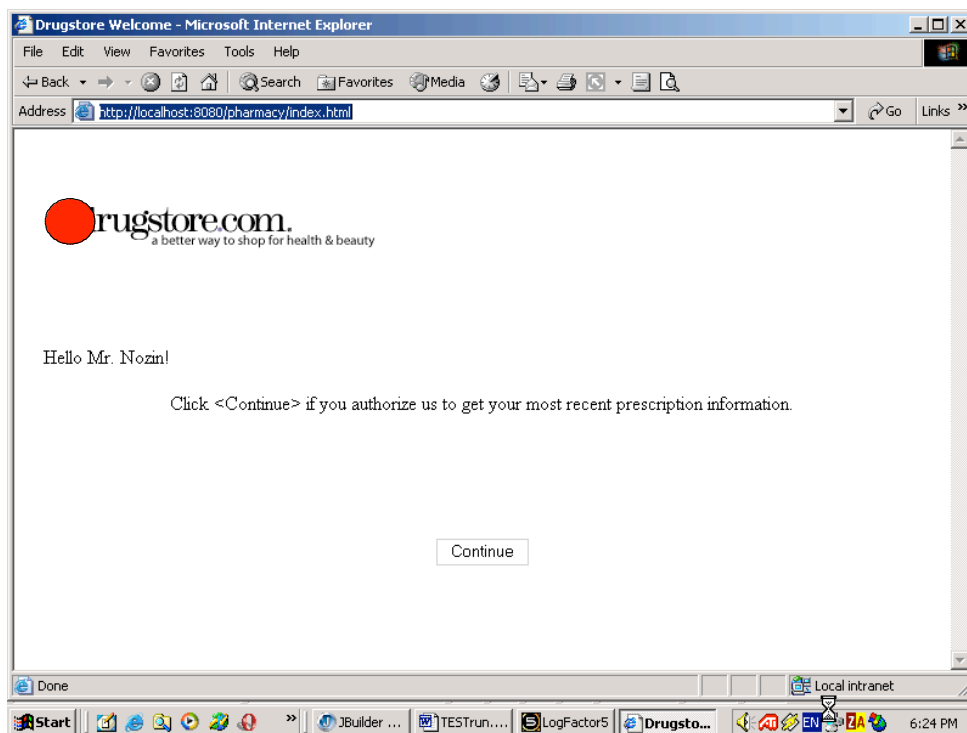


Figure 5.3 Rugstore.com welcome page.

First the Attribute Provider list was requested from the Discovery Service. In our case there is just one Attribute Provider – log record (9).

Then the first personal attribute was requested from the Attribute Provider - (10). The Attribute Provider requested the EPAL policy ruling from the PDP (11). Both the PDP and the Attribute Provider logged this event with the ITR - (12,13).

These steps are repeated for each personal attribute. Finally, the user is presented with the screen showing his personal attributes that the Drugstore needs to fulfill his prescription:

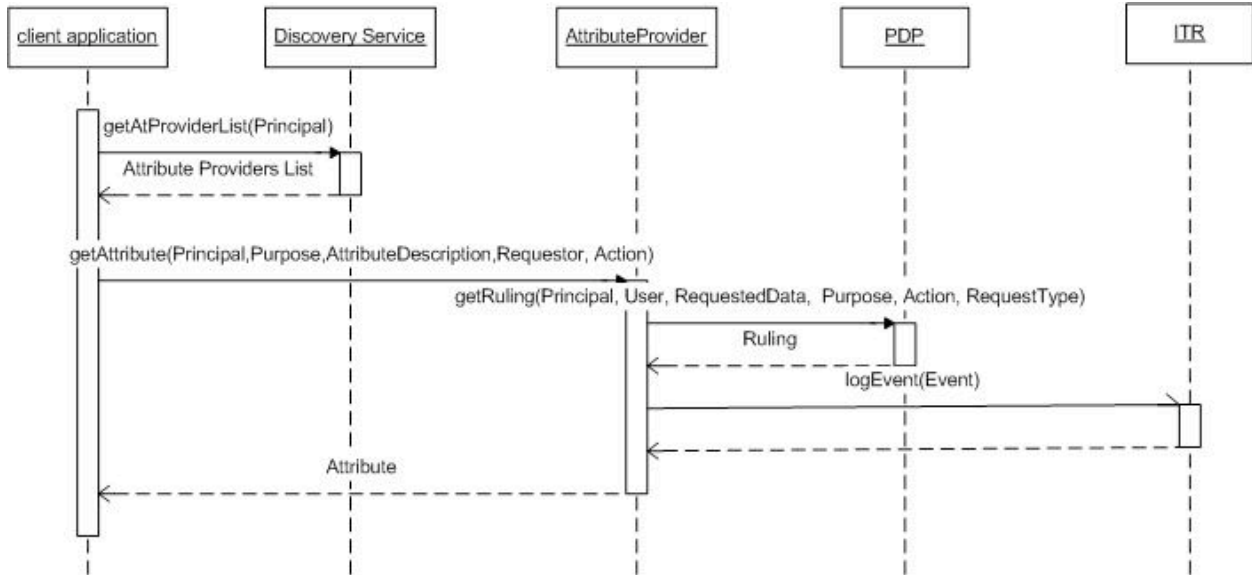


Figure 5.4 Scenario I sequence diagram.

Eventually the user can see all his personal attributes retrieved from Attribute Providers located elsewhere on the Internet. That might be the Individual’s financial institution, the hospital, a government information system etc., as demonstrated in Figure 5.6.

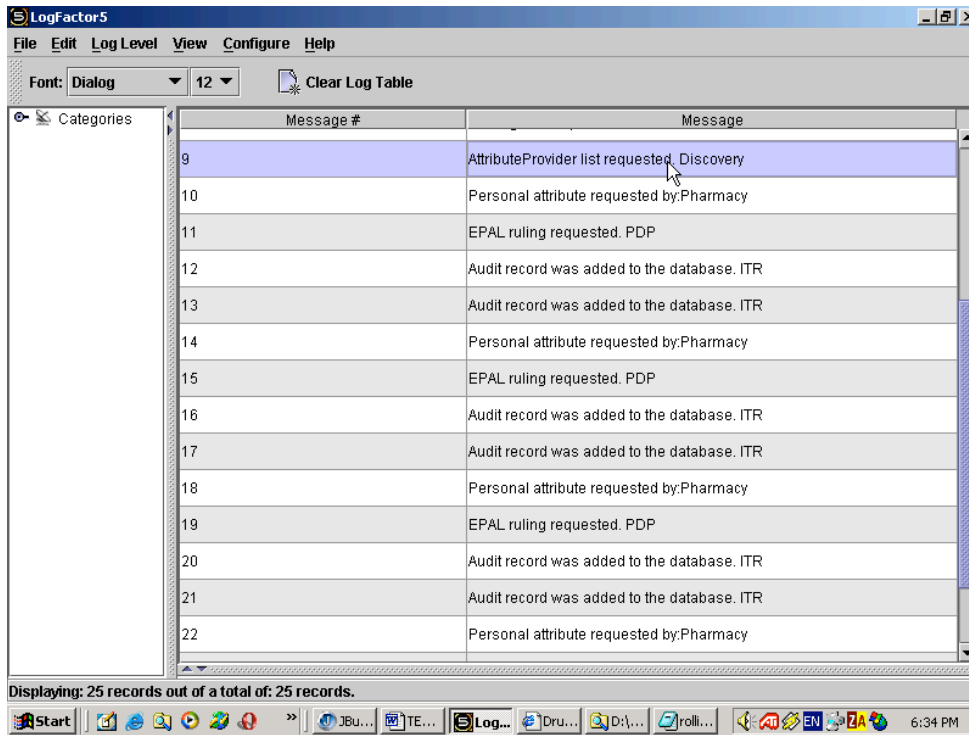


Figure 5.5 Log4J logging messages for the Scenario I.

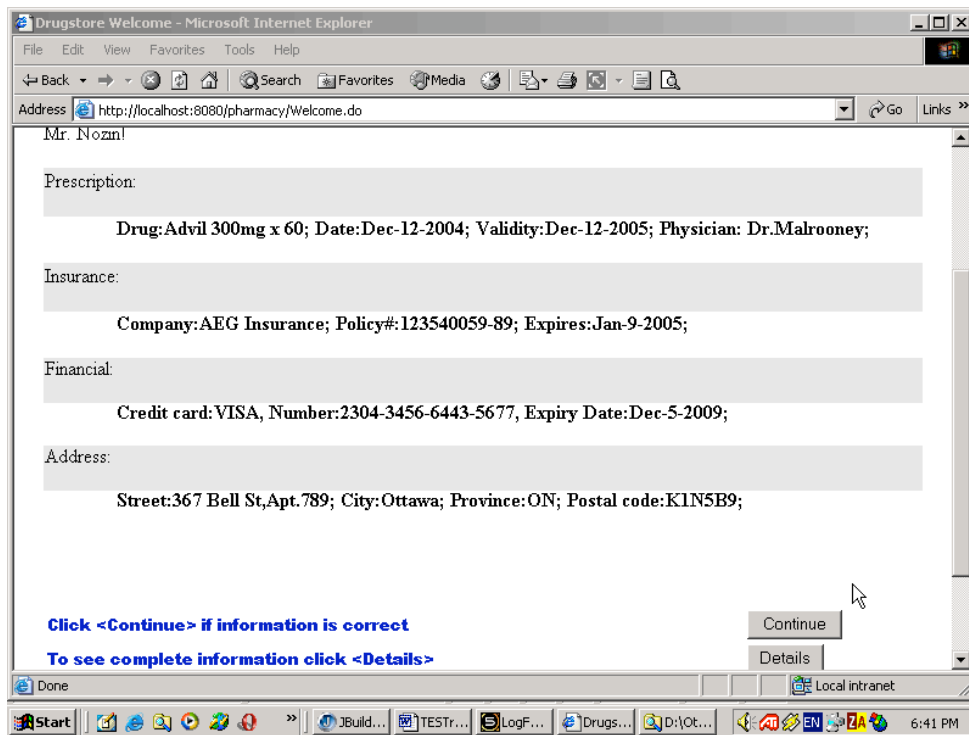


Figure 5.6 Individual's personal attribute screen.

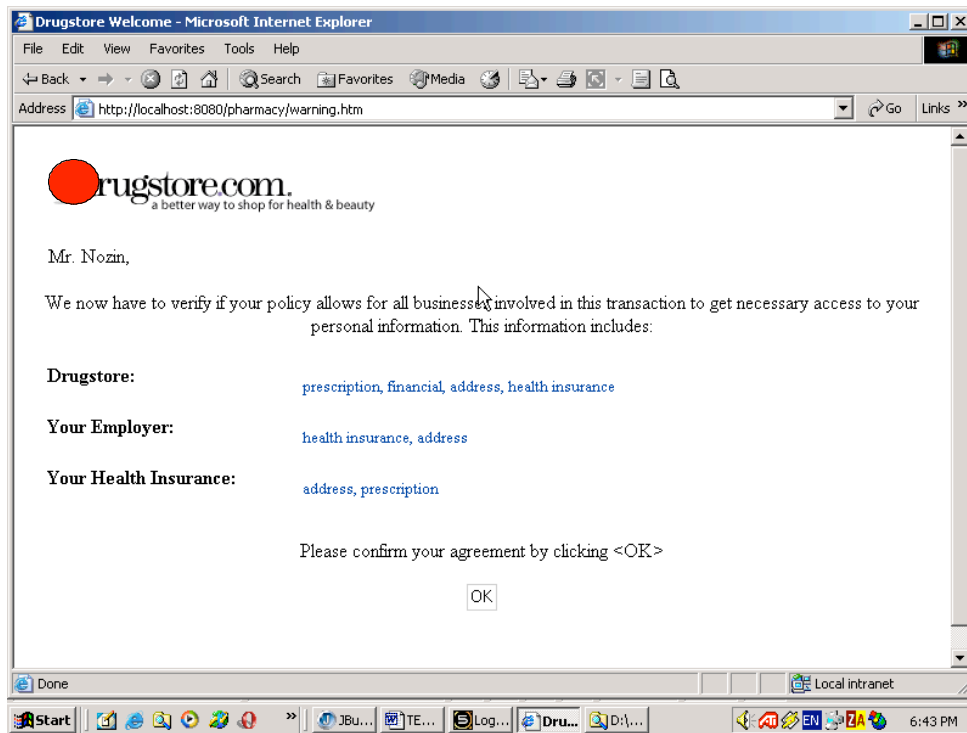


Figure 5.7 Information to be collected notification.

After all attributes are successfully retrieved and user has confirmed their correctness, a notification window is shown to inform user that in order to complete this business transaction some third parties will need to access certain pieces of personal information Figure 5.7. To make sure that all permissions are defined in the user policy, trial requests are sent to the PDP by Rugstore.com on behalf of all transaction participants as the code fragment below shows:

pharmacywebactions.Verfy.java

```

.....
boolean perm1 = "ALLOW".equalsIgnoreCase(PDP.getRuling("Mr.Nozin",
    "Insurance",
    "health-prescription",
    "redeem-insurance-money",
    "store",
    "Trial"));

```

Under test Scenario I the user policy has all necessary rules defined to allow access. The order confirmation screen is shown Figure 5.8

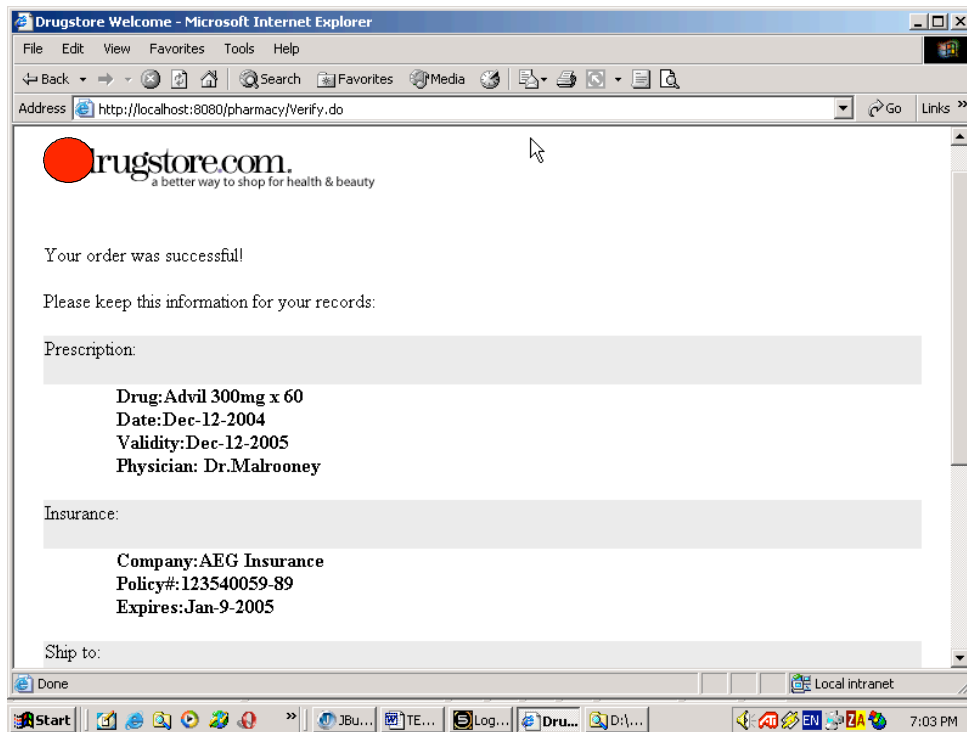


Figure 5.8 Order confirmation screenshot.

Conclusion: Scenario I demonstrated the obvious advantage the proposed framework brings to the table. Explicit consent (more on explicit consent is discussed in section *Scenario II (missing consent) test run result analysis.*) and notice is provided to the user. Users have the convenience of their attributes being automatically retrieved. There is also the means to ensure that all information stored is up-to-date. Scenario III will also demonstrate how the user's attributes can be reviewed and edited using the Customer Gateway. Another important benefit of the proposed framework is that it is not limited to the scope of one enterprise as EPAL is or to one particular protocol as P3P is limited to HTTP only according to the specification. (However there were certain attempts to apply it outside of HTTP.) Finally, no excessive information is collected by

Rugstore.com with the purpose to transfer it further down the transaction business chain. For the full summary of the comparison with P3P please refer to Table 5-1.

5.2 Scenario II (missing consent) test run result analysis.

The initial part of this scenario is similar to Scenario I, up to the point where a check is performed to ensure that all necessary permissions were given as shown in Figure 5.7. During the check it is detected that Individual's policy is missing one (or more) permissions. In this case the user is warned and then redirected to the Customer Gateway application (Figure 5.9). Information about what is missing in the policy is passed along. When in the Customer Gateway, user is asked to review, and confirm/sign changes to his policy.

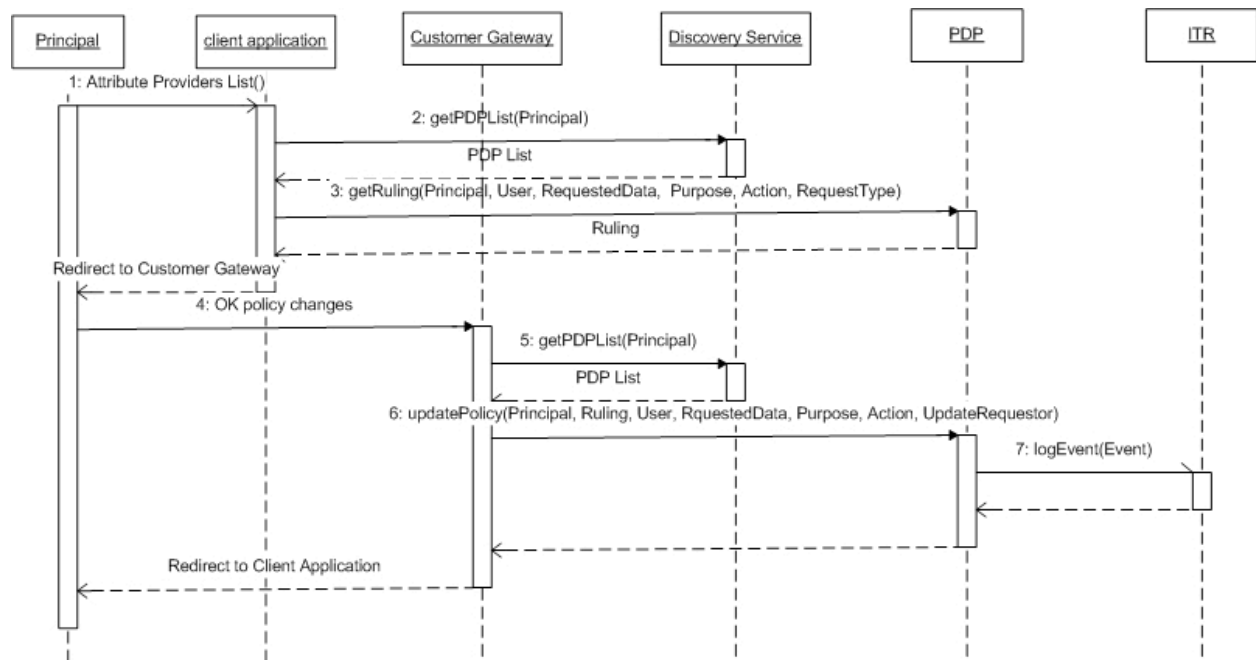


Figure 5.9 Scenario II sequence diagram.

After agreeing to the changes by clicking the <ALLOW> button, as demonstrated in Figure 5.10, the user gives an explicit agreement to his information being obtained by the party specified in the policy. Also, a signing mechanism may be involved allowing signing a sort of receipt and

storing it as an ITR log entry to support non-repudiation. The syntax of the rule below is very simple. However, EPAL has support for the XACML conditions which allows for defining virtually any possible conditions including temporal ones. Usage of complex conditions raises a number of other issues, including finding reliable sources for input data to evaluate such conditions.

The rule shown below is added to the Individual's policy:

```
<rule id="addeerule1" ruling="allow">
  <short-description language="en"/>
  <long-description language="en"/>
  <user-category refid="Insurance"/>
  <data-category refid="health-prescription"/>
  <purpose refid="redeem-insurance-money"/>
  <action refid="store"/>
</rule>
```

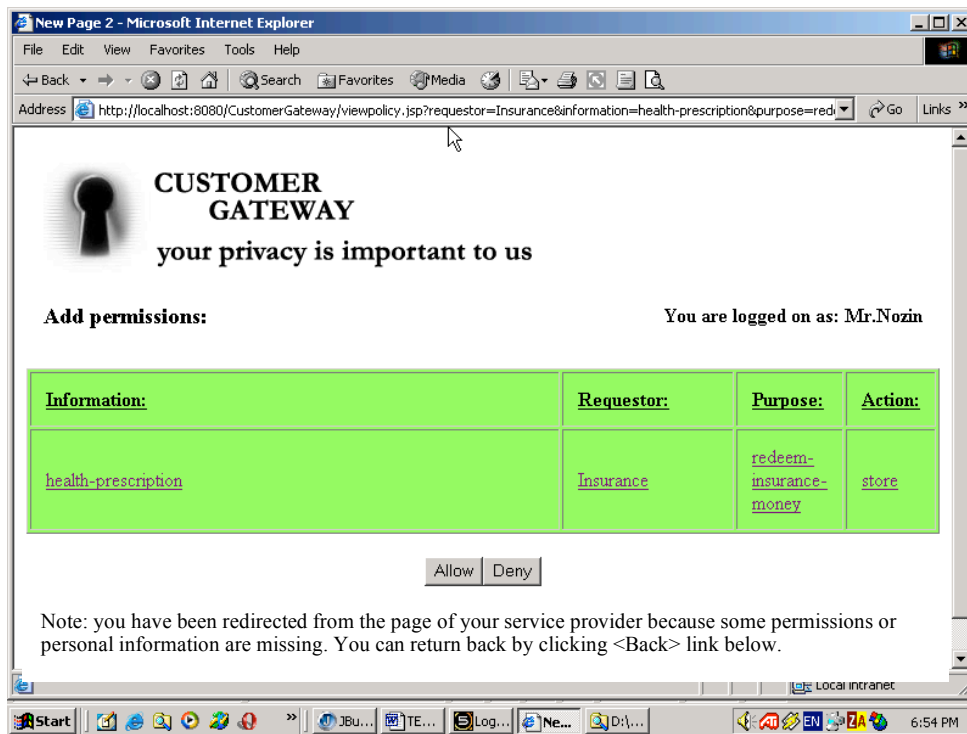


Figure 5.10 Customer Gateway policy update screenshot.

Logging records 32 to 34 show the action sequence when a new rule is added to the policy (see Figure 5.11).

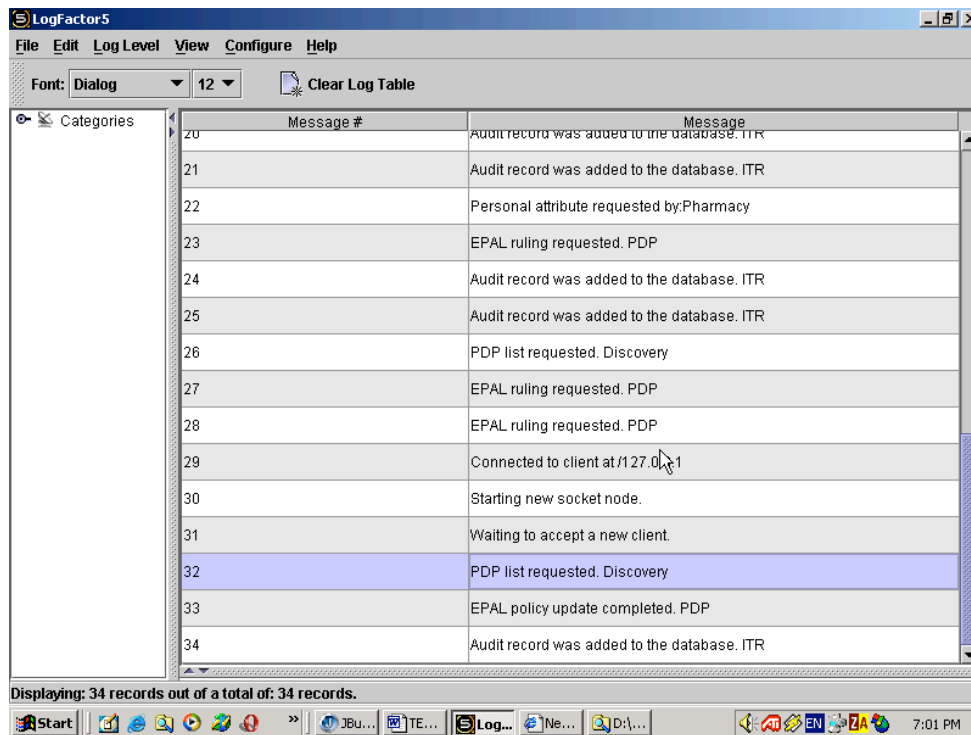


Figure 5.11 Scenario II Log4J logs.

Conclusion: Scenario II demonstrated a number of important points. First, there is a way to collect explicit consent from the user by prompting him to add a rule to the policy. It is also possible to extend this protocol by adding the Individual’s digital signature to the transaction with subsequent logging of this signed transaction with the ITR.

Second, the Individual’s policy is stored by distributed PDPs. This brings an advantage over P3P where APPEL preferences can be stored only in one place, otherwise managing them is going to become a hectic task. In the proposed framework all burden associated with storing and managing policies is hidden from an individual. Using several protocols, the Customer Gateway application is able to retrieve policies from different PDPs, to present a composite view of them to the user, to make changes and to synchronize an update of the policy across multiple PDPs.

5.3 Scenario III (viewing audit information)

In this scenario, another important piece of functionality is demonstrated which is completely unsupported by P3P. That is the ability to capture and view audit information. In this framework, audit consists of a number of records which store an event description, logger identity and a timestamp. What makes it different from the way policy is stored is that audit information is logged with only one ITR out of many and not necessarily all audit records are synchronized among different ITRs. When information is retrieved, all duplicate records are removed and a composite view of audit is presented to the user as if there were the only audit database available.

The action sequence for the demo system is shown in Figure 5.12. Actual execution logs are shown in Figure 5.14 (records 38,39). And the screenshot demonstrating how audit information is shown is in Figure 5.13. Log4J logger was used in the demo implementation to trace an application flow. Privacy events are captured by the ITR. Identity of an entity which sent event to the ITR and audit events are clickable. This means that in the real-world application, the user can get expanded information by clicking on the event. Moreover, events can be used as supplementary information for the user when considering policy modifications. For example, based on certain audit events, the user can make a decision on whether to tighten or loosen his or her policy rules. The Customer Gateway welcome screen for the demo system is shown in Figure 5.15. It gives an idea of the functionality of the Customer Gateway. Apart from that, Individuals should have the ability to enter personal information attributes and store them with an Attribute Provider of choice and all that through the interface provided by the Customer Gateway.

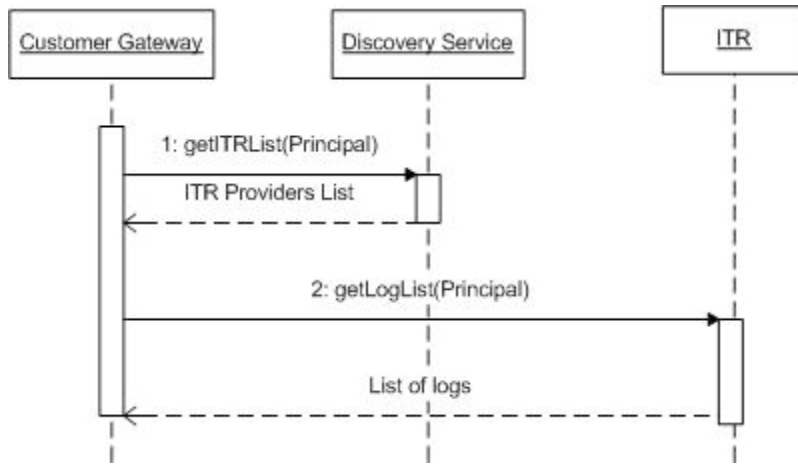


Figure 5.12 Scenario III sequence diagram.

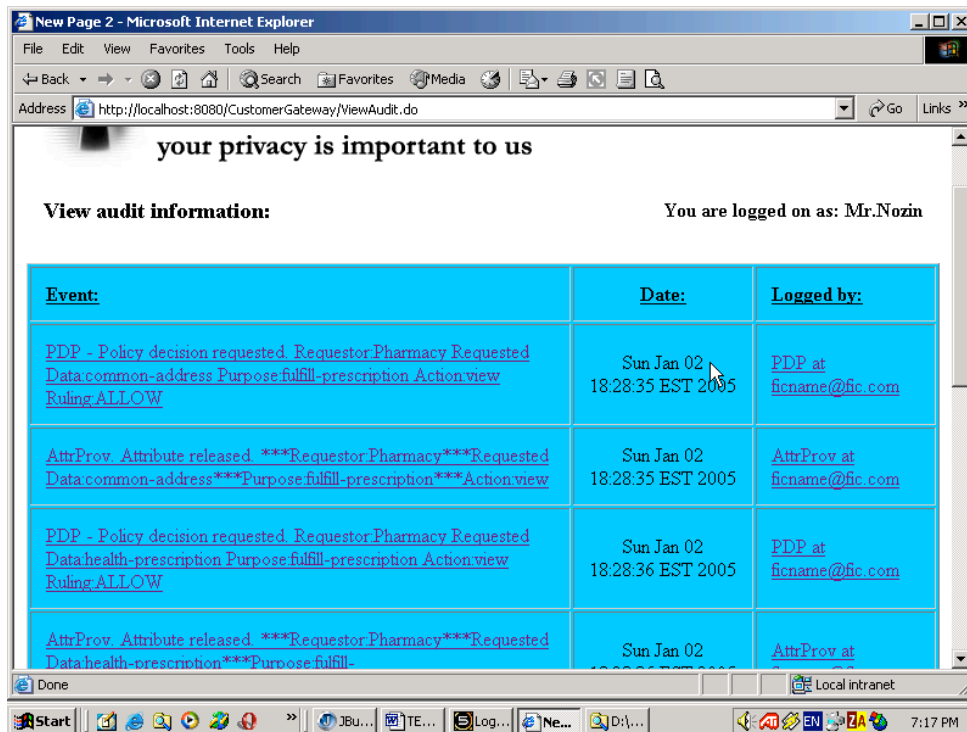


Figure 5.13 View audit information screenshot.

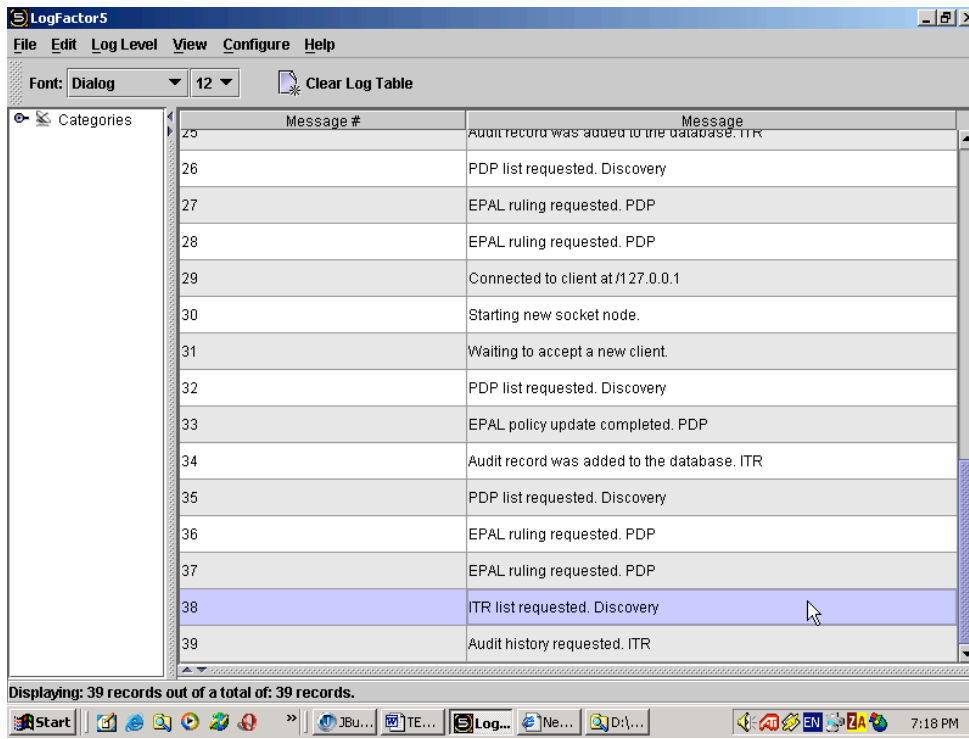


Figure 5.14 Scenario III Log4J log records.

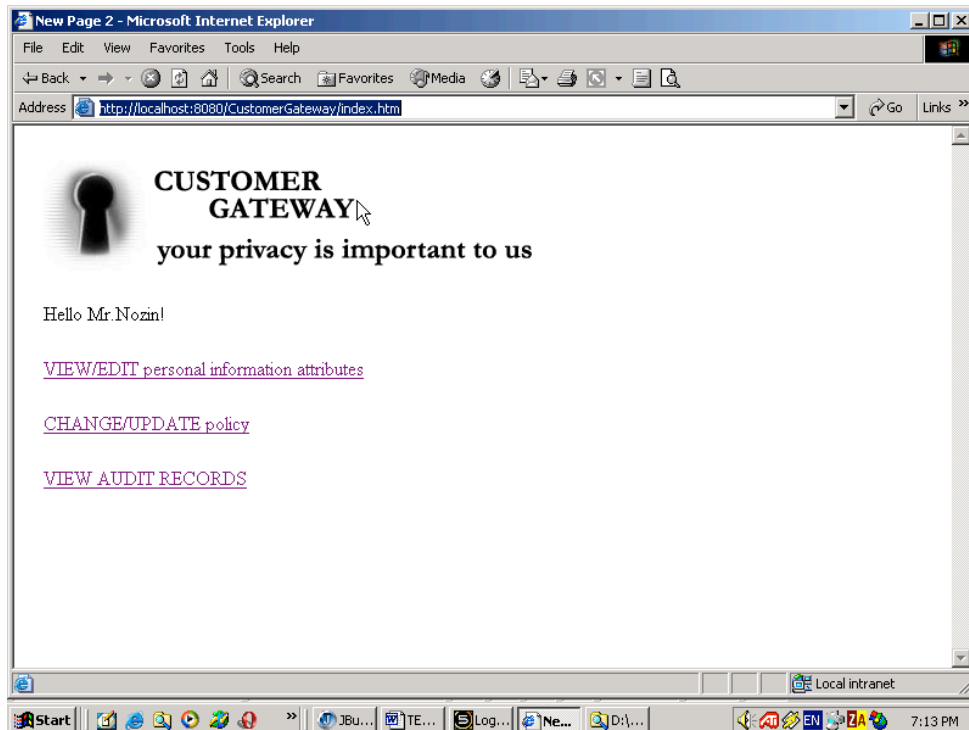


Figure 5.15 Customer Gateway welcome screen.

Conclusion: Scenario III demonstrated a very important feature of the framework – the ability to capture and view audit information which contributes to compliance with the PIPEDA full control over personal information and accountability.

5.4 Joe Self revised scenario

In the revised scenario shown in Figure 5.16, Joe Self is trying to buy a ticket from other than his usual airline (Usual Airline). When some changes occurred to his personal information he updated it using the Customer Gateway simultaneously at various Attribute Providers, including his Usual Airline which has also assumed the role of an Attribute Provider. What is missing is a permission (consent) allowing the New Airline to retrieve personal information attributes.

Under the revised scenario, Joe’s actions would be as follows:

1. Joe logs on to New Airline’s online booking system and picks an airfare he wants to buy.

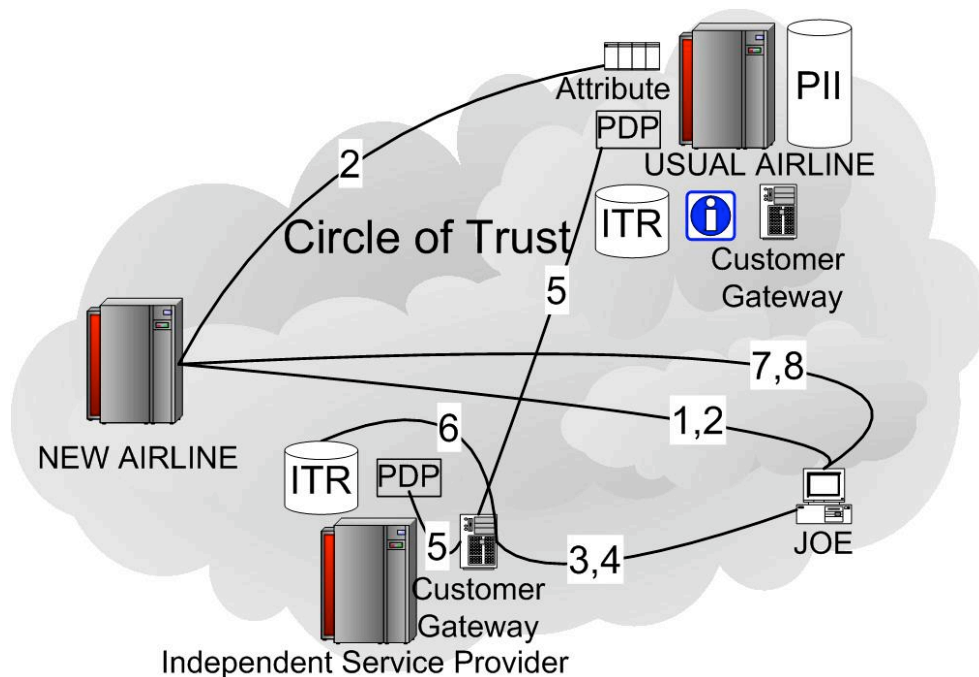


Figure 5.16 Joe Self revised scenario

2. New Airline asks his permission to access his personal information and tries to get attributes from an Attribute Provider which happens to be Usual Airline.

3. Because permissions are missing from the EPAL policy, New Airline redirects Joe to the Customer Gateway he picked from the list New Airline found using the Discovery Service.

4. Along with the redirection, New Airline sends information about permissions which have to be added, including permissions needed for New Airline partners such as a taxi service, to access the necessary information. If the taxi service needs some information it will be able to receive it directly from an Attribute Provider. This way New Airline doesn't have to collect any information it doesn't need.

5. At the Customer Gateway, Joe is presented with a wizard question-answer style interface. Each question answered triggers changes to his EPAL policy. The wizard can even propose to add provisions allowing all airlines to receive the same personal information attributes. Changes to the EPAL policy are synchronized among all PDPs. This means if there were multiple PDPs holding Joe's EPAL policy they all would have to make changes to the policy.

This way Joe will never have to worry about adding permissions for a participating airline again.

6. Joe consents to all changes to the policy. Those changes are logged with ITR (one or several of them).

7. Joe is redirected back to New Airline where he can see his personal attributes which were received after EPAL policy update.

8. After ensuring that all information is correct, Joe confirms booking and receives booking confirmation.

5.5 Analysis summary

In the previous chapters a comparison of the existing privacy enhancing technology with the proposed framework has been made. This included running the same test scenarios on the implementation of the proposed framework and using a P3P framework. Also an analysis of the

framework was performed based on the Joe Self scenario from the Liberty Alliance specification to demonstrate weak spots of the Liberty Alliance framework and advantages of the proposed framework. The results of the comparison are summarized in Table 5-1.

The Liberty Alliance framework delivers a higher degree of privacy support than P3P. However it still needs to be amended in a manner similar to the Customer Gateway concept to provide higher levels of Control and Audit with the need to change the way of storing user profiles to ensure that information stored in them is up-to-date to increase support for Accuracy.

| Functionality sought | support | | | |
|--|--------------------|-----------------|-----------------|------------------|
| | Proposed Framework | Project Liberty | P3P | EPA ² |
| 1) Up-to-date personal information for Accuracy | yes | no | no | yes |
| 2) Dynamic explicit consent | yes | no | no ¹ | no |
| 3) Not collecting excessive information when collecting on behalf of a third party | yes | yes | no | no |
| 4) Easy policy management ³ for Control | yes | no | no | N/A |
| 5) Easily available audit trail for Audit | yes | no | no | yes |

¹ Expected to be supported in a future version of the P3P specification.

² IBM EPA's scope is limited to a single enterprise. It has no concept of a user policy. One single policy exists which reflects the organization's privacy policy. Consent is generally captured in paper-based form.

³ Think of the situation in which a principal uses more than one system.

Table 5-1 Supported functionality comparison summary

The EPA framework deserves to be mentioned separately. EPA has earned respect for its approach in the privacy community for managing personal information within a single enterprise but fails to deliver when the interests of an individual have to be considered. In addition, it can not improve convenience for the user, especially in the context of information sharing across many enterprises in a B2B network. Information may become unaccounted for and may become outdated should it leave the boundaries of the single enterprise. A user would have to deal with multiple enterprises to get a full account of how his information was used. EPA has no concept

of individual user policy and does not provide individual access to audit records and to view the accuracy of personal information. It is possible that some operations in EPA are performed in a paper-based form. In fact, there is no mechanism to capture or request consent from the individual if needed.

From the chart we can see the proposed framework has incorporated features from several existing frameworks into a unified whole. More importantly it has a mechanism for dynamic explicit consent, and easy policy management in the control of the user that no other framework has incorporated.

6 CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

The aim of this thesis has been to create a privacy-enhancing framework that focused on key principles from PIPEDA. In particular, we focused on issues that had not yet been adequately addressed by other frameworks (P3P, EPA, Project Liberty) and proposed some extensions. In particular:

- *The Customer Service Gateway as a single point of entry* (as demonstrated in scenario III). It enables a user to get all information in one spot increasing *convenience* to the user of the framework and giving extra means for the user to *control* privacy information. None of the existing frameworks currently provides such facility. In those frameworks users can potentially have several places where their profiles are kept. However additional research needs to be done to determine how many profiles a typical user will have and if this number makes it difficult for the user to control them.
- *Ability to use Customer Gateway's service to specify the policy and study its impact by looking at the logs* (scenario III). This opens an opportunity to fine-tune the policy to suit the user's needs the best. The Initial policy can be chosen from a set of predefined policies if the user doesn't want to spend too much time designing his own from scratch. All of the existing frameworks, except for P3P, have the notion of a collection of the audit information but there is no means for the user to analyze it and make changes to the policy accordingly. To make it even more useful Customer Gateway should be intelligent enough to analyze audit information, identify potential problems and present the user with suggestions on how to eliminate them. Also an enforcement mechanism should be

developed to enforce each member of the framework to register correct and sufficient audit information.

- *A bigger role for the discovery mechanism as a starting point of every scenario has been implemented in the framework (scenario I), adding to the flexibility of the framework and enabling easy addition of new participants to the framework without the need for user involvement. Should any new entity join the framework, the user instantly becomes aware of this fact, which makes subsequent *audit* easier. This concept was borrowed from Project Liberty where it was solely used for discovery of personal information attributes across multiple providers. In the proposed framework it is also used for the discovery of the audit information, policies, and services of the framework such as ITR, PDP etc. This gives the framework a flexibility allowing for organic growth. At the same time it may pose a potential threat to the overall performance of the framework as it heavily relies on this service. Additional work has to be done to assess the discovery mechanism's impact on scalability and performance.*
- *A mechanism that allows existence of multiple replicated policies while presenting them to the user as if it is a single one. This simplifies policy management. Easy policy management is a fundamental concept of this framework and a crucial point in achieving a high degree of *control* over personal information. Keeping the policy in one place may discourage business users because they don't want to face the condition that the system which has the policy is offline.*

As in the case with multiple user profiles a policy regulating usage of personal information can exist in many instances in the proposed framework. It also applies to

P3P, Project Liberty, and HP Lab's framework. EPA is in exception because it doesn't have a concept of user's policy at all. In EPA there is a single policy for the enterprise.

What makes the proposed framework different is that it provides a way to manage multiple policies as if they were one. As in the case with multiple user profiles detailed comparison should be done of the user's experience managing multiple policies in different frameworks.

- *A mechanism for capturing Individual's explicit consent dynamically without the need for explicit callbacks* (scenario II). Each framework covered in this thesis handles this problem in a different way. Most attention was given to this issue by Liberty Alliance which defines a special protocol for callbacks. However the problem still exists if the user is offline during a callback. That is why an attempt was made in this framework to capture user consent within one session without the need for explicit callbacks. This approach requires detailed understanding of each business process by the initiating party which is responsible for sending "trial" requests to the PDP to ensure that all participants down the business chain have access to the information they will need.

By use-case scenarios it was demonstrated that the system is capable of serving as a technical implementation of the PIPEDA principles not currently supported by existing frameworks:

- Accuracy. This is achieved by giving the user the ability to control each and every bit of his personal information from one location.
- Individual Access. The Customer Gateway makes it easy to expose stored personal information to individuals in a convenient and unified manner.

Comparison of the prototype and the scenario based on P3P has shown that the proposed framework has certain very important advantages over the existing P3P standard:

- It increases the level of compliance with PIPEDA and similar legislation by providing access to the audit information, as well as the means to ensure that personal information is up-to-date;
- It accommodates the natural preferences of industry players because it is distributed and does not have any centralized entities.

6.2 Future work

The following is some of the future work that can be built on the results of this thesis:

- 1) Develop a sophisticated negotiation protocol allowing users to easily define their PI usage policy and monitor the usage of PI according to the policy with the ability to make an adjustments to the policy based on audit information.
- 2) Extend the current taxonomy for privacy information artifacts to allow all participants to operate using a common vocabulary. In fact, the functioning of the whole framework relies on the existence of such taxonomy.
- 3) Develop a set of interaction protocols to accommodate several important scenarios such as:
 - Entering new information into the system;
 - Resolving conflicts when discrepancies exist in attributes stored by different attribute providers;
 - Synchronizing user policies stored by different PDPs;
 - Resolving ownership-of-information issues. While the Individual definitely owns each and every piece of his personal information, it could have been already entered or already owned, before the legislation came into effect, by a third party. There is a

- potential for conflict over who has a right to modify such information. A mechanism to resolve such conflicts should be developed;
- After the system is implemented and the proof of concept is completed, validation of the alignment of technical approaches with legislative requirements should be conducted by an expert in law.
- 4) Perform comprehensive analysis of the threat model and security aspects of the framework.
- 5) Research into some issues outlined in the previous section. In particular:
- Detailed usability assessment of the proposed framework.
 - Evaluation of performance and scalability limitations of the crucial components of the proposed framework such as the discovery service.
 - An introduction of enforcement mechanisms to ensure that all participants are in compliance with the rules of the framework should be studied.

7 REFERENCES

- [Ackerman2003] Ackerman L., Kempf, J., Miki, T., Wireless Location Privacy: Law and Policy in the U.S., EU and Japan, Internet Society, 2003. <http://www.isoc.org/briefings/015/index.shtml>
- [Agrawal2003] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, Yirong Xu, An XPath-based preference language for P3P, Proceedings of the twelfth international conference on World Wide Web, 2003
- [Arnesen2003] Arnesen, R. and Danielsson, J., "A Framework for Enforcement of Privacy Policies", Nordic Security Workshop 2003. http://publications.nr.no/A_Framework_for_Enforcement_of_Privacy_Policies.pdf
- [Ashley2002/1] Paul Ashley, Calvin Powers, Matthias Schunter, From Privacy Promises to Privacy Management, The proceedings of the 2002 workshop on New security paradigms, 2002
- [Ashley2002/2] Paul Ashley, Satoshi Hada, Gunter Karjoth, Matthias Schunter, E-P3P Privacy Policies and Privacy Authorization, Proceeding of the ACM workshop on Privacy in the Electronic Society, 2002
- [Backes2003] Michael Backes, Birgit Pfützmann, and Matthias Schunter, A Toolkit for Managing Enterprise Privacy Policies, Proceedings of the 8th European Symposium on Research in Computer Security (ESORICS 2003) <http://www.zurich.ibm.com/security/publications/2003/BaPfSc2003-PrivacyToolbox-ESORICS.pdf>
- [Bettini2002] C. Bettini, S. Jajodia, X. Wang, D. Wijesekera, Obligation Monitoring in Policy Management, 3rd International Workshop on Policies for Distributed Systems and Networks (POLICY'02), 2002
- [Bigatti2004] Massimiliano Bigatti, Web Services Integration Patterns, O'Reilly, [Webservice.xml.com](http://webservice.xml.com), <http://webservice.xml.com/pub/a/ws/2004/06/16/patterns.html?page=2>, 2004
- [Bonatti2001] P. Bonatti, E. Damiani, S. de Capitani di Vimercati, P. Samarati, A Component-based Architecture for Secure Data Publication, Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC'01), 2001
- [Bonatti2000] Piero Bonatti, Sabrina de Capitani di Vimercati, Pierangela Samarati, A modular approach to composing access control policies, Proceedings of the 7th ACM conference on Computer and communications security, 2000
- [Bucker2003] Axel Bucker, Bill Haase, David Moore, Martin Keller, Dr. Otto Koblinger, Hai-Fun Wu, IBM Tivoli Privacy Manager Solution Design and Best Practices, IBM Redbook, 2003
- [Chadwick2003] Chadwick, D.W., Mundy, D., Policy based electronic transmission of prescriptions, The proceedings. POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks, 4-6 June 2003
- [COPPA98] Children's Online Privacy Protection Act of 1998, Federal Trade Commission, United States. <http://www.ftc.gov/ogc/coppa1.htm>
- [Cranor 98] L. Cranor, J. Reagle, Designing a Social Protocol: Lessons Learned from the Platform for Privacy Preferences, Telecommunications Policy Research Conference, Alexandria, VA, 1998 <http://www.w3.org/People/Reagle/papers/tprc97/tprc-f2m3.html>
- [Cranor2002] L. Cranor and J. Reidenberg, Can user agents accurately represent privacy notices?, Proceedings of the 30th Research Conference on Communication, Information, and Internet Policy, MIT Press, 2002. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=328860
- [Demo Javadoc] Automatically generated source code documentation for the demo application. <http://www.site.uottawa.ca/~mnoz/in/doc/>
- [Dix2000] A. Dix. Infomediaries and negotiated privacy techniques. In 10th Conference on Computers, Freedom and Privacy, page 167, Toronto, Ontario, Canada, April 2000.
- [EU2002] Directive on Privacy and Electronic Communications, European Union, 2002. http://europa.eu.int/eur-lex/pri/en/oj/dat/2002/l_201/l_20120020731en00370047.pdf

- [Fischer2001] S. Fischer-Hübner. IT-Security and Privacy : Design and Use of Privacy-Enhancing Security Mechanisms. Number 1958 in Lecture Notes in Computer Science (LNCS). Springer Verlag, Berlin, 2001. ISBN: 3540421424
- [Frichman2003] Frichman, R.G., Cronin, M.J., Information-Rich Commerce at a Crossroads: Business and Technology Adoption Requirements, Communications of the ACM Sept. 2003, Vol. 46, No. 9
- [GLB99] The Financial Modernization Act, Federal Trade Commission, United States, 1999.
<http://www.ftc.gov/privacy/glbact/>
- [Gowadia2003] Vaibhav Gowadia, Csilla Farkas, XML access control: RDF metadata for XML access control, Proceedings of the 2003 ACM workshop on XML security, 2003
- [Grandon2003] Gandon and N. Sadeh, A Semantic e-Wallet to Reconcile Privacy and Context Awareness, Second International Semantic Web Conference, 2003, USA. http://www-2.cs.cmu.edu/~sadeh/Publications/Small Selection/ ISWC2003_camera ready.pdf
- [Grimm2000] R. Grimm and A. Rossnagel. Can P3P help to protect privacy worldwide? In International Multimedia Conference, ACM Press, 2000, pp. 157–160.
- [Gritzalis2001] Dimitris Gritzalis, Nikolaos Kyrloglou, Consumer Online-Privacy and Anonymity Protection using Infomediary Schemes, Proceedings of the XXI International Conference of the Chilean Computer Science Society (SCCC'01), 2001
- [HIPAA96] Health Insurance Portability and Accountability Act (HIPAA), United States, 1996.
<http://www.hipaa.org/>
- [Hochheiser2002] Harry Hochheiser, The platform for privacy preference as a social protocol: An examination within the U.S. policy context, ACM Transactions on Internet Technology (TOIT), Volume 2 Issue 4, 2002
- [Hogg2000] T. Hogg, B. Huberman, M Franklin, Protecting Privacy While Sharing Information in Electronic Communities, Proceedings of the tenth conference on Computers, freedom and privacy: challenging the assumptions, Toronto, Ontario, Canada, 2000 <http://www.cfp2000.org/papers/hogg.pdf>
- [Hong2004] J. Hong, J. Landay, An Architecture for Privacy-Sensitive Ubiquitous Computing, Berkeley EECS Annual Research Symposium, 2004 www.eecs.berkeley.edu/BEARS/STARS/final/hong.pdf
- [Hull 2003] Hull, R.; Bharat Kumar; Lieuwen, D., Towards federated policy management, Proceedings of the POLICY 2003, IEEE 4th International Workshop on Policies for Distributed Systems and Networks, 2003
- [IDCIDE2001] IDCIDE. IDcide introduces corporate privacy compliance software. Press release. 2001.
http://www.idcide.com/pages/press_releas.htm#6
- [Karjoth2002/1] Günter Karjoth, Matthias Schunter, A Privacy Policy Model for Enterprises, Proceedings of the 15th IEEE Computer Security Foundations Workshop (CSFW'02), 2002
- [Karjoth2002/2] G. Karjoth, M. Schunter, M. Waidner, Platform for Enterprise Privacy Practices: Privacy-enabled Management of Customer Data, 2nd Workshop on Privacy Enhancing Technologies, Lecture Notes in Computer Science, Springer Verlag, 2002
- [Karjoth2003/1] Günter Karjoth, Matthias Schunter, Els Van Herreweghen, Translating Privacy Practices into Privacy Promises -How to Promise What You Can Keep, Proceedings of the 4th International Workshop on Policies for Distributed Systems and Networks, 2003
- [Karojth2003/2] Karjoth, G., Schunter, M., Van Herreweghen, E., Waidner, M., Amending P3P for clearer privacy promises, Proceedings of the 14th International Workshop on Database and Expert Systems Applications, 2003.
- [Kudo2000] M. Kudo and S. Hada, XML Document Security based on Provisional Authorization, 7th ACM Conference on Computer and Communication Security 2000. www.trl.ibm.com/projects/xml/xacl/ccs2k-kudo.pdf
- [Liberty2003] Liberty Alliance Project, www.projectliberty.org, 2003
- [LibertyPPEL2003] Liberty architecture framework for supporting Privacy Preference Expression Languages (PPELs), www.projectliberty.org, 2003
- [Maheu v. IMS] Maheu v. IMS Health Canada, Federal Court of Canada, FCT 647 (T-1967-01) 27th May, 2003

- [Mont2002] Marco Casassa Mont, Identity Management: On the “Identity = Data + Policies” Model, Trusted Systems Laboratory HP Laboratories
- [Mont2003] M. Mont, S. Pearson, P. Bramhall, Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services, 8th European Symposium on Research in Computer Security, Norway, 2003. <http://www.hpl.hp.com/techreports/2003/HPL-2003-49.pdf>
- [MS2004/1] Microsoft Corporation, Next Generation Secure Computing Base (NGSCB), <http://www.microsoft.com/resources/ngscb/>, 2004
- [Myers2000] Andrew C. Myers, Barbara Liskov, Protecting privacy using the decentralized label model, ACM Transactions on Software Engineering and Methodology (TOSEM), Volume 9 Issue 4, 2000
- [OASIS2002] OASIS. eXtensible Access Control Markup Language (XACML), 2002. Available at www.oasis-open.org/committees/xacml
- [Olivier2002] Olivier, M. S., Database privacy, ACM SIGKDD Explorations Newsletter, Volume 4 Issue 2, 2002
- [Olivier2003/1] Martin S. Olivier, Using organisational safeguards to make justifiable privacy decisions when processing personal data, The proceedings of the 2003 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology, 2003
- [Olivier2003/2] Olivier, M. S. A layered architecture for privacy-enhancing technologies. In Proceedings of the Third Annual Information Security South Africa Conference (ISSA2003), J. H. P. Eloff, H. S. Venter, L. Labuschagne, and M. M. Eloff, Eds. Sandton, South Africa, 113–126.
- [P3P2002] The Platform for Privacy Preferences 1.0 Specification, World Wide Web Consortium Recommendation, April 2002. <http://www.w3.org/TR/P3P/>
- [Peyton2004] Liam Peyton, and Max Nozin, Tracking Privacy Compliance in B2B Networks, The proceedings of the Sixth International Conference on Electronic Commerce, 2004, pp.376-382.
- [Pfitzmann2002] Birgit Pfitzmann, Michael Waidner, Privacy in browser-based attribute exchange, The proceeding of the ACM workshop on Privacy in the Electronic Society, 2002
- [PIPEDA2000] The Personal Information Protection and Electronic Documents Act (PIPEDA), Department of Justice, Canada, 2000. http://e-com.ic.gc.ca/epic/internet/inecic-ceac.nsf/vwGeneratedInterE/h_gv00045e.html
- [Pouttu-Clarke2005] Matt Pouttu-Clarke, Corss Domain Cookie Provider, TheServerSide.com online journal, Jan-19-2005, http://www.theserverside.com/user/userthreads.tss?user_id=545755
- [PRight2001] PRIVACYRIGHT. 2001. Control of personal information —the economic benefits of adopting an enterprise-wide permissions management platform. White Paper. <http://www.privacyright.com/info/economic.html>
- [Privacy Bird] AT&T Lab’s Privacy Bird. <http://privacybird.com/>
- [Rindfleisch97] T.C. Rindfleisch. Privacy, Information Technology, and Health Care. Communications of the ACM, 40(8):93-100,1997
- [Ross2000] Ross J. Anderson, Privacy Technology Lessons from Healthcare, 2000 IEEE Symposium on Security and Privacy (S&P 2000), 2000
- [Schunter2003/1] Schunter M., Van Herreweghen E., Waidner M., Translating EPAL to P3P, IBM, March 2003, <http://www.w3.org/2003/p3p-ws/pp/ibm2.html>
- [Schunter2003/2] Schunter M., Powell C., The Enterprise Privacy Authorization Language (EPAL), IBM, June, 2003. <http://www.zurich.ibm.com/security/enterprise-privacy/epal/>
- [Sun2004] Patterns and Strategies for Building Document-Based Web Services, Sun Microsystems Technical Articles, <http://java.sun.com/developer/technicalArticles/xml/jaxrpcpatterns/index5.html#10>, 2004
- [TCPA] The Trusted Computing Platform Alliance (TCPA)
- [Zuidweg2003] M. Zuidweg, J. Filho, M. van Sinderen, Using P3P in a web services-based context aware application platform, Ninth EUNICE Workshop on Next Generation Networks, Hungary, Budapest, September, 2003 www.w3.org/2003/p3p-ws/pp/utwente.pdf

Appendix 1. Demo system class diagrams.

For detailed description of API calls please refer to Javadoc for the Demo [Demo Javadoc].

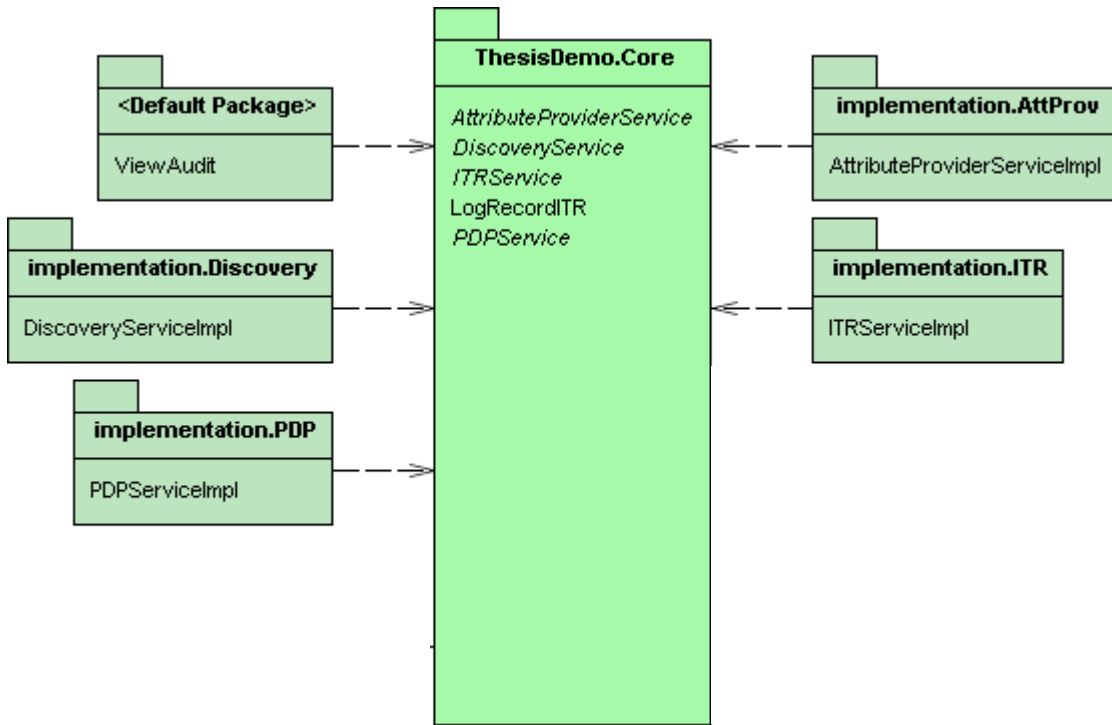


Figure Appendix1.1. Demo system main package diagram.

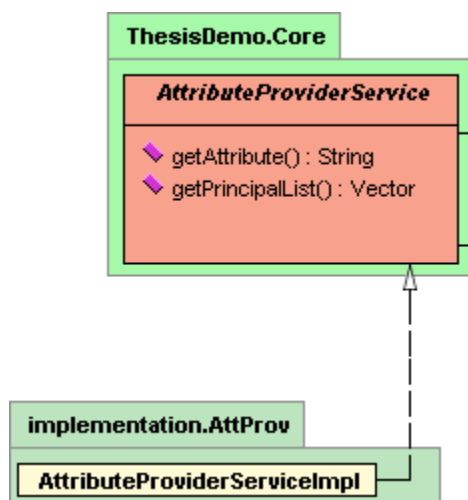


Figure Appendix1.2. Demo system Attribute Provider Service class diagram.

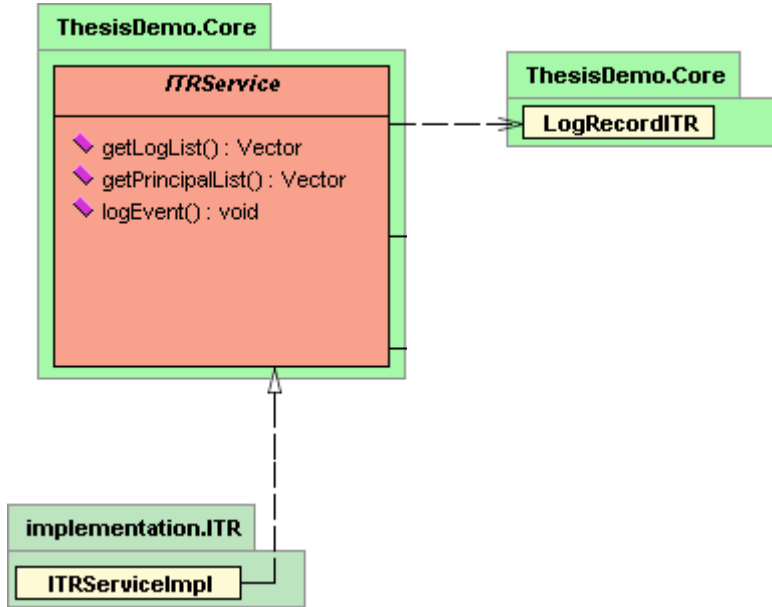


Figure Appendix1.3. Demo system ITR Service class diagram.

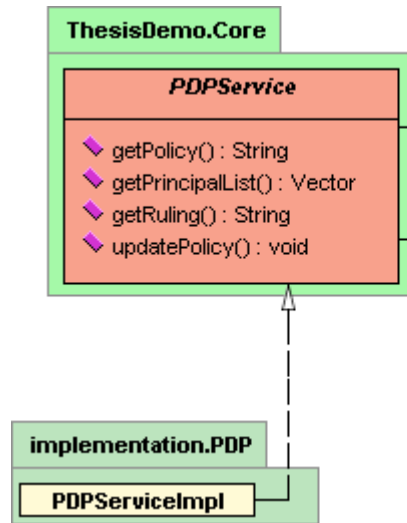


Figure Appendix1.4. Demo system PDP Service class diagram.

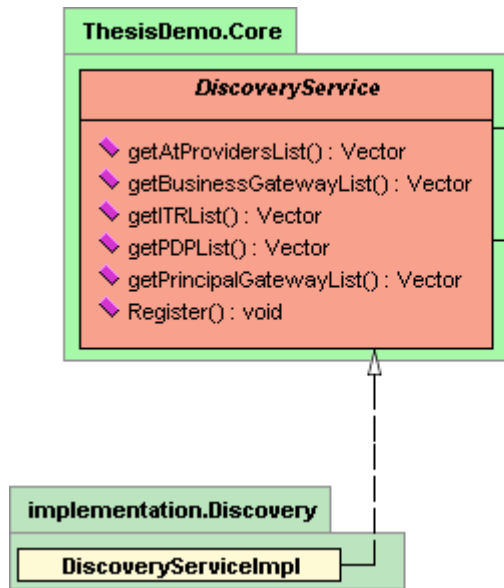


Figure Appendix1.5. Demo system Discovery Service class diagram.

Appendix 2. P3P policy file used for Drugstore demo.

```
<?xml version="1.0"?>
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1">
  <!-- Generated by IBM P3P Policy Editor version Beta 1.12 built 2/27/04
1:19 PM -->

  <!-- Expiry information for this policy -->
  <EXPIRY max-age="86400"/>

<POLICY
  name="our_main_policy"
  discuri="http://localhost:8080/drugstore/drugstore.html"
  xml:lang="en">
  <!-- Description of the entity making this policy statement. -->
  <ENTITY>
    <DATA-GROUP>
<DATA ref="#business.name">Rugstore</DATA>
    </DATA-GROUP>
  </ENTITY>

  <!-- Disclosure -->
  <ACCESS><all/></ACCESS>

  <!-- Disputes -->
  <DISPUTES-GROUP>
    <DISPUTES resolution-type="service"
service="http://localhost:8080/drugstore/privacy_dispute.html" short-
description="Privacy_Dispute">
      <LONG-DESCRIPTION></LONG-DESCRIPTION>
      <REMEDIES><law/></REMEDIES>
    </DISPUTES>
  </DISPUTES-GROUP>

  <!-- Statement for group "HealthDrug Prescription Information" -->
  <STATEMENT>
    <EXTENSION optional="yes">
      <GROUP-INFO
xmlns="http://www.software.ibm.com/P3P/editor/extension-1.0.html"
name="Health Drug Prescription Information"/>
      </EXTENSION>

  <!-- Consequence -->
  <CONSEQUENCE>
We need this information in order to claim insurance money from your
insurer.</CONSEQUENCE>

  <!-- Use (purpose) -->
  <PURPOSE><historical required="opt-in"/><other-purpose required="opt-
in">Claim Insurance Money</other-purpose></PURPOSE>
```

```
<!-- Recipients -->
<RECIPIENT><same required="opt-in"/></RECIPIENT>

<!-- Retention -->
<RETENTION><business-practices/></RETENTION>

<!-- Base dataschema elements. -->
<DATA-GROUP>
  <DATA ref="#dynamic.miscdata"><CATEGORIES><health/></CATEGORIES></DATA>
</DATA-GROUP>
</STATEMENT>

<!-- End of policy -->
</POLICY>
</POLICIES>
```

Appendix 3. APPEL properties file.

```
<?xml version="1.0"?>
  <appel:RULESET xmlns:appel="http://www.w3.org/2001/02/APPELv1"
xmlns:p3p="http://www.w3.org/2000/12/P3Pv1" crtddb="WorldNet Privacy Tool"
crtndon="Thu Oct 11 15:58:40 2001
" >
  <appel:RULE behavior="limited" description="Site may use health or
medical information for analysis or to make decisions that may affect what
content or ads you see, etc." >
    <p3p:POLICY >
      <p3p:STATEMENT >
        <p3p:PURPOSE appel:connective="or" >
          <p3p:pseudo-analysis required="always" />
          <p3p:pseudo-decision required="always" />
          <p3p:individual-analysis required="always" />
          <p3p:individual-decision required="always" />
        </p3p:PURPOSE>
        <p3p:DATA-GROUP >
          <p3p:DATA >
            <p3p:CATEGORIES >
              <p3p:health />
            </p3p:CATEGORIES>
          </p3p:DATA>
        </p3p:DATA-GROUP>
      </p3p:STATEMENT>
    </p3p:POLICY>
  </appel:RULE>
  <appel:RULE behavior="limited" description="Site may use health or
medical information for marketing" >
    <p3p:POLICY >
      <p3p:STATEMENT >
        <p3p:PURPOSE appel:connective="or" >
          <p3p:contact required="always" />
          <p3p:telemarketing required="always" />
        </p3p:PURPOSE>
        <p3p:DATA-GROUP >
          <p3p:DATA >
            <p3p:CATEGORIES >
              <p3p:health />
            </p3p:CATEGORIES>
          </p3p:DATA>
        </p3p:DATA-GROUP>
      </p3p:STATEMENT>
    </p3p:POLICY>
  </appel:RULE>
  <appel:RULE behavior="limited" description="Unless you opt-out, site may
use health or medical information for analysis or to make decisions that may
affect what content or ads you see, etc." >
    <p3p:POLICY >
      <p3p:STATEMENT >
        <p3p:PURPOSE appel:connective="or" >
```



```

    <p3p:pseudo-analysis required="opt-out" />
    <p3p:pseudo-decision required="opt-out" />
    <p3p:individual-analysis required="opt-out" />
    <p3p:individual-decision required="opt-out" />
  </p3p:PURPOSE>
  <p3p:DATA-GROUP >
    <p3p:DATA >
      <p3p:CATEGORIES >
        <p3p:health />
      </p3p:CATEGORIES>
    </p3p:DATA>
  </p3p:DATA-GROUP>
</p3p:STATEMENT>
</p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Unless you opt-out, site may
use health or medical information for marketing" >
  <p3p:POLICY >
    <p3p:STATEMENT >
      <p3p:PURPOSE appel:connective="or" >
        <p3p:contact required="opt-out" />
        <p3p:telemarketing required="opt-out" />
      </p3p:PURPOSE>
      <p3p:DATA-GROUP >
        <p3p:DATA >
          <p3p:CATEGORIES >
            <p3p:health />
          </p3p:CATEGORIES>
        </p3p:DATA>
      </p3p:DATA-GROUP>
    </p3p:STATEMENT>
  </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Site may share health or
medical information with other companies (other than those helping the site
provide services to you)" >
  <p3p:POLICY >
    <p3p:STATEMENT >
      <p3p:RECIPIENT appel:connective="or" >
        <p3p:same required="always" />
        <p3p:other-recipient required="always" />
        <p3p:unrelated required="always" />
        <p3p:public required="always" />
      </p3p:RECIPIENT>
      <p3p:DATA-GROUP >
        <p3p:DATA >
          <p3p:CATEGORIES >
            <p3p:health />
          </p3p:CATEGORIES>
        </p3p:DATA>
      </p3p:DATA-GROUP>
    </p3p:STATEMENT>
  </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Unless you opt-out, site may
share health or medical information with other companies (other than those
helping the site provide services to you)" >

```

```

<p3p:POLICY >
  <p3p:STATEMENT >
    <p3p:RECIPIENT appel:connective="or" >
      <p3p:same required="opt-out" />
      <p3p:other-recipient required="opt-out" />
      <p3p:unrelated required="opt-out" />
      <p3p:public required="opt-out" />
    </p3p:RECIPIENT>
    <p3p:DATA-GROUP >
      <p3p:DATA >
        <p3p:CATEGORIES >
          <p3p:health />
        </p3p:CATEGORIES>
      </p3p:DATA>
    </p3p:DATA-GROUP>
  </p3p:STATEMENT>
</p3p:POLICY>
</appel:RULE>
  <appel:RULE behavior="limited" description="Site may use financial
information or information about your purchases for analysis or to make
decisions that may affect what content or ads you see, etc." >
  <p3p:POLICY >
    <p3p:STATEMENT >
      <p3p:PURPOSE appel:connective="or" >
        <p3p:pseudo-analysis required="always" />
        <p3p:pseudo-decision required="always" />
        <p3p:individual-analysis required="always" />
        <p3p:individual-decision required="always" />
      </p3p:PURPOSE>
      <p3p:DATA-GROUP >
        <p3p:DATA >
          <p3p:CATEGORIES appel:connective="or" >
            <p3p:purchase />
            <p3p:financial />
          </p3p:CATEGORIES>
        </p3p:DATA>
      </p3p:DATA-GROUP>
    </p3p:STATEMENT>
  </p3p:POLICY>
</appel:RULE>
  <appel:RULE behavior="limited" description="Site may use financial
information or information about your purchases for marketing" >
  <p3p:POLICY >
    <p3p:STATEMENT >
      <p3p:PURPOSE appel:connective="or" >
        <p3p:contact required="always" />
        <p3p:telemarketing required="always" />
      </p3p:PURPOSE>
      <p3p:DATA-GROUP >
        <p3p:DATA >
          <p3p:CATEGORIES appel:connective="or" >
            <p3p:purchase />
            <p3p:financial />
          </p3p:CATEGORIES>
        </p3p:DATA>
      </p3p:DATA-GROUP>
    </p3p:STATEMENT>
  </p3p:POLICY>
</appel:RULE>

```

```

    </p3p:POLICY>
  </appel:RULE>
  <appel:RULE behavior="limited" description="Unless you opt-out, site may
  use financial information or information about your purchases for analysis or
  to make decisions that may affect what content or ads you see, etc." >
    <p3p:POLICY >
      <p3p:STATEMENT >
        <p3p:PURPOSE appel:connective="or" >
          <p3p:pseudo-analysis required="opt-out" />
          <p3p:pseudo-decision required="opt-out" />
          <p3p:individual-analysis required="opt-out" />
          <p3p:individual-decision required="opt-out" />
        </p3p:PURPOSE>
        <p3p:DATA-GROUP >
          <p3p:DATA >
            <p3p:CATEGORIES appel:connective="or" >
              <p3p:purchase />
              <p3p:financial />
            </p3p:CATEGORIES>
          </p3p:DATA>
        </p3p:DATA-GROUP>
      </p3p:STATEMENT>
    </p3p:POLICY>
  </appel:RULE>
  <appel:RULE behavior="limited" description="Unless you opt-out, site may
  use financial information or information about your purchases for marketing"
  >
    <p3p:POLICY >
      <p3p:STATEMENT >
        <p3p:PURPOSE appel:connective="or" >
          <p3p:contact required="opt-out" />
          <p3p:telemarketing required="opt-out" />
        </p3p:PURPOSE>
        <p3p:DATA-GROUP >
          <p3p:DATA >
            <p3p:CATEGORIES appel:connective="or" >
              <p3p:purchase />
              <p3p:financial />
            </p3p:CATEGORIES>
          </p3p:DATA>
        </p3p:DATA-GROUP>
      </p3p:STATEMENT>
    </p3p:POLICY>
  </appel:RULE>
  <appel:RULE behavior="limited" description="Site may share financial
  information or information about your purchases with other companies (other
  than those helping the site provide services to you)" >
    <p3p:POLICY >
      <p3p:STATEMENT >
        <p3p:RECIPIENT appel:connective="or" >
          <p3p:same required="always" />
          <p3p:other-recipient required="always" />
          <p3p:unrelated required="always" />
          <p3p:public required="always" />
        </p3p:RECIPIENT>
        <p3p:DATA-GROUP >
          <p3p:DATA >

```

```

        <p3p:CATEGORIES appel:connective="or" >
            <p3p:purchase />
            <p3p:financial />
        </p3p:CATEGORIES>
    </p3p:DATA>
</p3p:DATA-GROUP>
</p3p:STATEMENT>
</p3p:POLICY>
</appel:RULE>
    <appel:RULE behavior="limited" description="Unless you opt-out, site may
share financial information or information about your purchases with other
companies (other than those helping the site provide services to you)" >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:RECIPIENT appel:connective="or" >
                <p3p:same required="opt-out" />
                <p3p:other-recipient required="opt-out" />
                <p3p:unrelated required="opt-out" />
                <p3p:public required="opt-out" />
            </p3p:RECIPIENT>
            <p3p:DATA-GROUP >
                <p3p:DATA >
                    <p3p:CATEGORIES appel:connective="or" >
                        <p3p:purchase />
                        <p3p:financial />
                    </p3p:CATEGORIES>
                </p3p:DATA>
            </p3p:DATA-GROUP>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>
    <appel:RULE behavior="limited" description="Site may contact you via
telephone to interest you in other services or products" >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:PURPOSE >
                <p3p:telemarketing required="always" />
            </p3p:PURPOSE>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>
    <appel:RULE behavior="limited" description="Unless you opt-out, site may
contact you via telephone to interest you in other services or products" >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:PURPOSE >
                <p3p:telemarketing required="opt-out" />
            </p3p:PURPOSE>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>
    <appel:RULE behavior="limited" description="Site may contact you through
means other than telephone (email, postal mail, etc.) to try to interest you
in other services or products" >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:PURPOSE >

```

```

        <p3p:contact required="always" />
    </p3p:PURPOSE>
</p3p:STATEMENT>
</p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Unless you opt-out, site may
contact you through means other than telephone (email, postal mail, etc.) to
interest you in other services or products" >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:PURPOSE >
                <p3p:contact required="opt-out" />
            </p3p:PURPOSE>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Site may contact you to
interest you in other services or products and does not allow you to remove
yourself from marketing/mailing list" >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:PURPOSE appel:connective="or" >
                <p3p:contact required="always" />
                <p3p:telemarketing required="always" />
            </p3p:PURPOSE>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Site may keep information
that personally identifies you to determine your habits, interests, or other
characteristics for research and analysis purposes" >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:PURPOSE >
                <p3p:individual-analysis required="always" />
            </p3p:PURPOSE>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Site may keep information
that personally identifies you to determine your habits, interests, or other
characteristics for making decisions about what content or ads you see at the
site, etc." >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:PURPOSE >
                <p3p:individual-decision required="always" />
            </p3p:PURPOSE>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Unless you opt-out, site may
keep information that personally identifies you to determine your habits,
interests, or other characteristics for research and analysis purposes" >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:PURPOSE >

```

```

        <p3p:individual-analysis required="opt-out" />
    </p3p:PURPOSE>
</p3p:STATEMENT>
</p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Unless you opt-out, site may
keep information that personally identifies you to determine your habits,
interests, or other characteristics for making decisions about what content
or ads you see at the site, etc." >
    <p3p:POLICY >
        <p3p:STATEMENT >
            <p3p:PURPOSE >
                <p3p:individual-decision required="opt-out" />
            </p3p:PURPOSE>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Site may share information
that personally identifies you with other companies (other than those helping
the site provide services to you)" >
    <p3p:POLICY >
        <p3p:STATEMENT appel:connective="and" >
            <p3p:RECIPIENT appel:connective="or" >
                <p3p:same required="always" />
                <p3p:other-recipient required="always" />
                <p3p:unrelated required="always" />
                <p3p:public required="always" />
            </p3p:RECIPIENT>
            <p3p:DATA-GROUP >
                <p3p:DATA >
                    <p3p:CATEGORIES appel:connective="or" >
                        <p3p:physical />
                        <p3p:online />
                        <p3p:government />
                    </p3p:CATEGORIES>
                </p3p:DATA>
            </p3p:DATA-GROUP>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Unless you opt-out, site may
share information that personally identifies you with other companies (other
than those helping the site provide services to you)" >
    <p3p:POLICY >
        <p3p:STATEMENT appel:connective="and" >
            <p3p:RECIPIENT appel:connective="or" >
                <p3p:same required="opt-out" />
                <p3p:other-recipient required="opt-out" />
                <p3p:unrelated required="opt-out" />
                <p3p:public required="opt-out" />
            </p3p:RECIPIENT>
            <p3p:DATA-GROUP >
                <p3p:DATA >
                    <p3p:CATEGORIES appel:connective="or" >
                        <p3p:physical />
                        <p3p:online />
                        <p3p:government />
                    </p3p:CATEGORIES>
                </p3p:DATA>
            </p3p:DATA-GROUP>
        </p3p:STATEMENT>
    </p3p:POLICY>
</appel:RULE>

```

```

        </p3p:CATEGORIES>
        </p3p:DATA>
        </p3p:DATA-GROUP>
        </p3p:STATEMENT>
        </p3p:POLICY>
    </appel:RULE>
    <appel:RULE behavior="limited" description="Site does not allow you to
    find out what data they have about you" >
        <p3p:POLICY >
            <p3p:ACCESS >
                <p3p:none />
            </p3p:ACCESS>
        </p3p:POLICY>
    </appel:RULE>
    <appel:RULE behavior="limited" description="Site may keep information
    that does not personally identify you to determine your habits, interests, or
    other characteristics for research and analysis purposes" >
        <p3p:POLICY >
            <p3p:STATEMENT >
                <p3p:PURPOSE >
                    <p3p:pseudo-analysis required="always" />
                </p3p:PURPOSE>
            </p3p:STATEMENT>
        </p3p:POLICY>
    </appel:RULE>
    <appel:RULE behavior="limited" description="Site may keep information
    that does not personally identify you to determine your habits, interests, or
    other characteristics for making decisions about what content or ads you see
    at the site, etc." >
        <p3p:POLICY >
            <p3p:STATEMENT >
                <p3p:PURPOSE >
                    <p3p:pseudo-decision required="always" />
                </p3p:PURPOSE>
            </p3p:STATEMENT>
        </p3p:POLICY>
    </appel:RULE>
    <appel:RULE behavior="limited" description="Unless you opt-out, site may
    keep information that does not personally identify you to determine your
    habits, interests, or other characteristics for research and analysis
    purposes" >
        <p3p:POLICY >
            <p3p:STATEMENT >
                <p3p:PURPOSE >
                    <p3p:pseudo-analysis required="opt-out" />
                </p3p:PURPOSE>
            </p3p:STATEMENT>
        </p3p:POLICY>
    </appel:RULE>
    <appel:RULE behavior="limited" description="Unless you opt-out, site may
    keep information that does not personally identify you to determine your
    habits, interests, or other characteristics for making decisions about what
    content or ads you see at the site, etc." >
        <p3p:POLICY >
            <p3p:STATEMENT >
                <p3p:PURPOSE >
                    <p3p:pseudo-decision required="opt-out" />
                </p3p:PURPOSE>
            </p3p:STATEMENT>
        </p3p:POLICY>
    </appel:RULE>

```

```

    </p3p:PURPOSE>
  </p3p:STATEMENT>
</p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Site may share information
that does not personally identify you with other companies (other than those
helping the site provide services to you)" >
  <p3p:POLICY >
    <p3p:STATEMENT >
      <p3p:RECIPIENT appel:connective="or" >
        <p3p:same required="always" />
        <p3p:other-recipient required="always" />
        <p3p:unrelated required="always" />
        <p3p:public required="always" />
      </p3p:RECIPIENT>
    </p3p:STATEMENT>
  </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Unless you opt-out, site may
share information that does not personally identify you with other companies
(other than those helping the site provide services to you)" >
  <p3p:POLICY >
    <p3p:STATEMENT >
      <p3p:RECIPIENT appel:connective="or" >
        <p3p:same required="opt-out" />
        <p3p:other-recipient required="opt-out" />
        <p3p:unrelated required="opt-out" />
        <p3p:public required="opt-out" />
      </p3p:RECIPIENT>
    </p3p:STATEMENT>
  </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="This site collects data for
an unknown purpose" >
  <p3p:POLICY >
    <p3p:STATEMENT >
      <p3p:PURPOSE >
        <p3p:other-purpose required="always" />
      </p3p:PURPOSE>
    </p3p:STATEMENT>
  </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="limited" description="Unless you opt-out, this site
collects data for an unknown purpose" >
  <p3p:POLICY >
    <p3p:STATEMENT >
      <p3p:PURPOSE >
        <p3p:other-purpose required="opt-out" />
      </p3p:PURPOSE>
    </p3p:STATEMENT>
  </p3p:POLICY>
</appel:RULE>
<appel:RULE behavior="request" >
  <appel:OTHERWISE />
</appel:RULE>
</appel:RULESET>

```


Appendix 4. Individual's profile used in demo system.

```
<?xml version="1.0"?>
<PersonalProfile>
<DataCategory id="health-prescription">
  <DataItem id="Drug"> Advil 300mg x 60;</DataItem>
  <DataItem id="Date"> Dec-12-2004;</DataItem>
  <DataItem id="Validity"> Dec-12-2005;</DataItem>
  <DataItem id="Physician"> Dr.Malrooney;</DataItem>
</DataCategory>
<DataCategory id="financial-creditcard">
  <DataItem id="Credit card"> VISA,</DataItem>
  <DataItem id="Number"> 2304-3456-6443-5677,</DataItem>
  <DataItem id="Expiry Date"> Dec-5-2009;</DataItem>
</DataCategory>
<DataCategory id="common-address">
  <DataItem id="Street"> 367 Bell St,Apt.789;</DataItem>
  <DataItem id="City"> Ottawa;</DataItem>
  <DataItem id="Province"> ON;</DataItem>
  <DataItem id="Postal code"> K1N5B9;</DataItem>
</DataCategory>
<DataCategory id="health-insurance">
  <DataItem id="Company"> AEG Insurance;</DataItem>
  <DataItem id="Policy#"> 123540059-89;</DataItem>
  <DataItem id="Expires"> Jan-9-2005;</DataItem>
</DataCategory>
</PersonalProfile>
```

Appendix 5. EPAL policy and vocabulary used in demo system.

EPAL policy:

```
<?xml version="1.0" encoding="UTF-8"?>

<epal-policy default-ruling="deny" global-condition="none" version="1.0"
xmlns="http://www.research.ibm.com/privacy/epal"
xmlns:xacml="urn:oasis:names:tc:xacml:1.0:policy"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.research.ibm.com/privacy/epal
D:\OttawaU\Peyton\EPAL\epal.xsd">
  <policy-information id="test-policy">
    <short-description language="en">simple test policy</short-
description>
    <long-description language="en"/>
    <issuer>
      <name>Max </name>
      <organization>OttawaU</organization>
      <e-mail/>
      <address>777 King Edvard</address>
      <country>Canada</country>
    </issuer>

    <version-info end-date="2004-07-26T00:00:00" last-modified="2004-
07-26T00:00:00" revision-number="1" start-date="2004-07-26T00:00:00"
superseded-by-id="test-policy" superseded-by-revision="1" test="false"/>
  </policy-information>
  <epal-vocabulary-ref id="test-vocab" location="D:\OttawaU\Thesis
DEMO\WORK\MainThesisDemo\test-vocab_1.xml"/>

  <rule id="pharmacy1" ruling="allow">
    <short-description language="en"/>
    <long-description language="en"/>
    <user-category refid="Pharmacy"/>
    <data-category refid="common-address"/>
    <purpose refid="fulfill-prescription"/>
    <action refid="transfer"/>
  </rule>
  <rule id="pharmacy2" ruling="allow">
    <short-description language="en"/>
    <long-description language="en"/>
    <user-category refid="Pharmacy"/>
    <data-category refid="common-address"/>
    <purpose refid="fulfill-prescription"/>
    <action refid="view"/>
  </rule>
  <rule id="pharmacy3" ruling="allow">
    <short-description language="en"/>
    <long-description language="en"/>
    <user-category refid="Pharmacy"/>
    <data-category refid="health-prescription"/>
```

```

    <purpose refid="fulfill-prescription"/>
    <action refid="view"/>
</rule>
<rule id="pharmacy4" ruling="allow">
  <short-description language="en"/>
  <long-description language="en"/>
  <user-category refid="Pharmacy"/>
  <data-category refid="health-prescription"/>
  <purpose refid="fulfill-prescription"/>
  <action refid="store"/>
</rule>
<rule id="pharmacy5" ruling="allow">
  <short-description language="en"/>
  <long-description language="en"/>
  <user-category refid="Pharmacy"/>
  <data-category refid="health-insurance"/>
  <purpose refid="fulfill-prescription"/>
  <action refid="transfer"/>
</rule>
<rule id="pharmacy6" ruling="allow">
  <short-description language="en"/>
  <long-description language="en"/>
  <user-category refid="Pharmacy"/>
  <data-category refid="health-insurance"/>
  <purpose refid="fulfill-prescription"/>
  <action refid="view"/>
</rule>
<rule id="pharmacy7" ruling="allow">
  <short-description language="en"/>
  <long-description language="en"/>
  <user-category refid="Pharmacy"/>
  <data-category refid="financial-creditcard"/>
  <purpose refid="fulfill-prescription"/>
  <action refid="view"/>
</rule>

<rule id="principal1" ruling="allow">
  <short-description language="en"/>
  <long-description language="en"/>
  <user-category refid="Principal"/>
  <data-category refid="common-address"/>
  <purpose refid="accurate-information"/>
  <action refid="edit"/>
</rule>
<rule id="principal2" ruling="allow">
  <short-description language="en"/>
  <long-description language="en"/>
  <user-category refid="Principal"/>
  <data-category refid="health-prescription"/>
  <purpose refid="accurate-information"/>
  <action refid="edit"/>
</rule>
<rule id="principal3" ruling="allow">
  <short-description language="en"/>
  <long-description language="en"/>
  <user-category refid="Principal"/>
  <data-category refid="health-insurance"/>

```

```

        <purpose refid="accurate-information"/>
        <action refid="edit"/>
    </rule>
<!-- This rule was added to allow the user to retrieve and edit his own
information -->
    <rule id="principal4" ruling="allow">
        <short-description language="en"/>
        <long-description language="en"/>
        <user-category refid="Principal"/>
        <data-category refid="financial-creditcard"/>
        <purpose refid="accurate-information"/>
        <action refid="edit"/>
    </rule>
<!--these rules were defined according to the Drugstore Scenario [Peyton2004]
-->
    <rule id="employer1" ruling="allow">
        <short-description language="en"/>
        <long-description language="en"/>
        <user-category refid="Employer"/>
        <data-category refid="health-insurance"/>
        <purpose refid="redeem-insurance-money"/>
        <action refid="view"/>
    </rule>
</epal-policy>

```

EPAL vocabulary:

```

<?xml version="1.0" encoding="UTF-8"?>

<epal-vocabulary xmlns="http://www.research.ibm.com/privacy/epal"
xmlns:xacml="urn:oasis:names:tc:xacml:1.0:policy"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.research.ibm.com/privacy/epal
D:\OttawaU\Peyton\EPAL\epal.xsd">
    <vocabulary-information id="test-vocab">
        <short-description language="en">short-description</short-
description>
        <long-description language="en">long-description</long-
description>
        <issuer>
            <name>Max Nozin</name>
            <organization>OttawaU</organization>
            <e-mail/>
            <address>address</address>
            <country>country</country>
        </issuer>
        <location/>
        <version-info end-date="2001-12-31T12:00:00" last-modified="2001-
12-31T12:00:00" revision-number="1" start-date="2001-12-31T12:00:00"
superseded-by-id="test-vocab" superseded-by-revision="" test="false"/>
    </vocabulary-information>
    <user-category id="Employer"/>
    <user-category id="Insurance"/>
    <user-category id="all-other"/>
    <user-category id="Pharmacy"/>

```

```

<user-category id="Principal"/>

<data-category id="common-contact">
  <short-description language="en"/>
</data-category>
<data-category id="health-prescription">
  <short-description language="en"/>
</data-category>
<data-category id="health-insurance">
  <short-description language="en"/>
</data-category>
<data-category id="financial-creditcard">
  <short-description language="en"/>
</data-category>
<data-category id="common-address">
  <short-description language="en"/>
</data-category>

<purpose id="redeem-insurance-money">
  <short-description language="en"/>
  <long-description language="en"/>
</purpose>
<purpose id="fulfill-prescription">
  <short-description language="en"/>
  <long-description language="en"/>
</purpose>
<purpose id="accurate-information">
  <short-description language="en"/>
  <long-description language="en"/>
</purpose>

<action id="store">
  <short-description language="en"/>
  <long-description language="en"/>
</action>
<action id="transfer">
  <short-description language="en"/>
  <long-description language="en"/>
</action>
<action id="view">
  <short-description language="en"/>
  <long-description language="en"/>
</action>
<action id="edit">
  <short-description language="en"/>
  <long-description language="en"/>
</action>

<container id="DynData">
  <attribute auditable="false" id="employerID" maxOccurs="1"
minOccurs="1" simpleType="http://www.w3.org/2001/XMLSchema#string">
    <short-description/>
    <long-description/>
  </attribute>

```

```
        <attribute auditable="false" id="requesterID" maxOccurs="1"
minOccurs="1" simpleType="http://www.w3.org/2001/XMLSchema#string">
            <short-description/>
            <long-description/>
        </attribute>
    </container>
</epal-vocabulary>
```