Privacy-enabled Services for Enterprises

Günter Karjoth, Matthias Schunter, and Michael Waidner
IBM Research
Zurich Research Laboratory
http://www.research.ibm.com/privacy

Abstract

The IBM Enterprise Privacy Architecture (EPA) is a methodology for enterprises to provide an enhanced and well-defined level of privacy to their customers. EPA is structured in four building blocks. The privacy regulation analysis identifies and structures the applicable regulations. The management reference model enables an enterprise to define and enforce an enterprise privacy strategy and the resulting privacy practices. The privacy agreement framework is a methodology for privacy-enabling business process re-engineering. It outputs a detailed model of the privacy-relevant players and activities as well as the privacy policies that govern these activities. The technical reference architecture defines the technology needed for implementing the identified practices.

1 Introduction

Consumer privacy is a growing concern in the market place. While privacy concerns are most prominent for e-commerce (see for example [1, 7]), the concerns for traditional transactions are increasing as well. Some enterprises are aware of these problems and of the market share they might loose if they do not implement proper privacy practices. As a consequence enterprises publish privacy statements that promise fair information practices. Written in natural language or formalized using the World Wide Web Consortium's "Platform for Privacy Preferences Project (P3P)" [8], they only constitute privacy promises and are not necessarily backed-up by technological means. In addition, laws increasingly impose baseline privacy regulations.

Enterprises willing to implement fair privacy practices usually face the following problems:

- Business processes are designed without considering privacy requirements. Thus, enterprises are forced to create stockpiles of personally identifiable information (PII or short *personal data*) instead of collecting personal data when needed for the business at hand.
- Existing services often identify users even though their identity is not needed for the business at hand. Privacy-enhancing security technology that provides security with less data is rarely used.
- Enterprises store a variety of personal data. Larger enterprises may not know what types of PII are collected and where it is stored.
- Enterprises may neither know the consent a customer has given nor the legal regulations that apply to a specific customer record.

Enterprises that want to respect the privacy of consumers need three main technologies.

Privacy-enabling design includes techniques that make services more privacy friendly. A core building block is data minimization. The goal of data minimization is to minimize the amount of personal data than needs to be collected to achieve the objectives of an enterprise. This includes privacy-enabling applications that are designed to provide services while minimizing the data needed. Tools for such applications are pseudonymity systems and anonymous authentication schemes [3].

Even with privacy-enabled design, an enterprise still stores a certain amount of personal data. The customers are required to trust the enterprise to use this data as promised in a privacy policy. *Privacy management services* help enterprises to enforce the promised practices in an auditable way.

Without computer security, a company cannot guarantee privacy. *Privacy-enabled security services* are needed to secure the infrastructure running the enterprise privacy management services. Nevertheless, existing security technology often provides security without privacy. Examples are non-anonymous identifica-

tion and authentication schemes, data collected by intrusion detection systems, and coarse access control. In order to enable privacy, these technologies need to be transformed into privacy-enabling security services. Analogously to privacy-enabled applications, the goal is to provide security without collecting personal data about honest users.

The goal of the IBM Enterprise Privacy Architecture (EPA) is to solve these problems while concentrating on the enterprise-related aspects. Essentially, EPA is a methodology for enterprises to provide a well-defined and enhanced level of privacy to its customers. It provides the foundation for the privacy part of IBM's Security and Privacy Services.

2 The IBM Enterprise Privacy Architecture

The IBM Enterprise Privacy Architecture is a methodology that allows enterprises to maximize the business use of personal information while respecting privacy concerns and regulations. It provides a sustainable privacy management system, which can be customized to the total set of privacy regulations and privacy choices facing an enterprise.

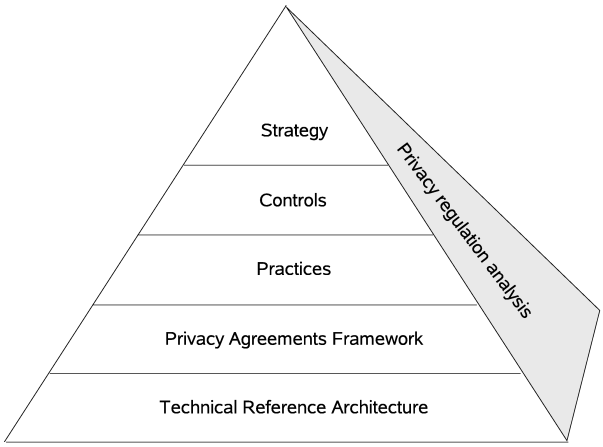

Figure 1. Building Blocks of the IBM Enterprise Privacy Architecture

EPA introduces privacy-awareness and privacy services into enterprises in a systematic and complete way. Figure 1 illustrates its components, outlined in form of a pyramid. As a prerequisite, the EPA *privacy regulation analysis* identifies and structures the applicable regulations. The *Management Reference Model* (top 3 layers in Figure 1) constitutes the tip of the EPA pyramid, defining the privacy strategy and practices of the enterprise. The *Privacy Agreements Framework* provides a privacy-enabled model for privacy-enhanced business process re-engineering. The lowest layer is the *Technical Reference Architecture* that defines the technology for implementing the required privacy services.

2.1 Privacy Regulation Analysis

Regulatory compliance is a primary driver of privacy-related activity in the marketplace. Thus, it is clear that a useful picture of the regulatory landscape is a pre-requisite to developing any kind of privacy architecture. The challenge is that regulations are typically written in dense legal style with formats and terminology that tend to differ depending on their origin and purpose.

EPA addresses this challenge by *regulatory summary tables* and *regulation rules tables*. Regulatory summary tables summarize the applicable regulations using a unified terminology. The regulation rules tables identify data that is in the enterprise as well as the legal restrictions on using such data. The regulation rules tables are enterprise-specific and more formal than the regulation summary table. An entry describes which party can perform which action on which type of data, the resulting privacy obligations, and a reference to the legal regulation. In addition, the four business-use phases Collection, Retention, Processing and Use ("CRPU") are used to categorize the scope of privacy regulations.

2.2 Management Reference Model

The EPA Management Reference Model addresses the enterprise-wide processes necessary for a comprehensive privacy management program. These processes are structured and linked to drive the program starting from a strategic view down through the implementation of privacy practices (see Figure 1).

Strategy defines the privacy philosophy, the high-level policies and identifies the applicable regulations.

This represents the highest level of an enterprise's privacy program and embodies its philosophy, its policies and the regulations it will adhere to. The outputs are a privacy strategy as well as a security strategy. Both define what an enterprise will do for protecting privacy and security.

Control defines the general controls necessary to enforce policy.

Its components are a Privacy Requirements Process, the Information Asset Classification and Control, a Compliance Enforcement Process, a definition of the Organizational Roles and Responsibilities as well as an Employee Education Program.

Practices defines the incorporation of policy into business processes.

This represents the level of an enterprise's privacy program that translates privacy policy obligations into the general processes, programs and activities that will implement them. Its components are a Privacy Statement declaring the enterprise policy, a Customer Preference Program for defining opt-in and opt-out choices, an Individual Participation Process that enables customers to access their data, a Dispute Process, an External Communication Program that advertises the privacy efforts of an enterprise, and Information Access Controls that protect the enterprises' data and resources.

2.3 Privacy Agreements Framework

The Privacy Agreements Framework models the transaction level management of privacy at the points where enterprises use personal information within business processes. This includes processes that connect the individual to the enterprise, processes linking people and departments within the enterprise, and processes linking the enterprise with third parties. This model can then be used to identify privacy agreements that are required between the players involved. The main parts of the model are players, data, and rules:

Players The players are the entities that interact while processing collected data. Basic players are data subjects (persons about whom data is collected) and different data users (enterprises or employees using the data). The player model uses an object-oriented modeling technique to identify the players, their operations on the data, as well as the interactions among the players. The result is documented using UML [2] class and collaboration diagrams.

Data The data model identifies the data needed for the processes. Besides identifying the fields collected in forms, it classifies data into at least three categories:

- Personally identifiable information is the most sensitive kind of information that can be linked to a real-world identity. Examples include a tuple name/surname or a U.S. social security number.
- Depersonalized Information is PII where the identifying information has been replaced by a pseudonym. Even though this data is less sensitive, some parties are able to re-personalize it by replacing the pseudonym with the identifying information. Examples include the age with a customer number.
- Anonymized Information contains no identifying information or pseudonyms. It is the least sensitive kind of information that can be obtained by removing all personal data from a set of data. For anonymized data, it is required that identifying the data subject given the data is virtually impossible. Examples include the town of residence or an age in years on its own (i.e., without any other information that may enable identification of the data subject).

Rules The rules model identifies the rules that govern the usage of data by players and their operations. It defines what player may perform which operation for what purpose. In addition, rules may impose conditions and may define obligations that result from performing an operation.

2.4 Technical Reference Architecture

To guarantee that an enterprise provides sufficient privacy to its customers, privacy-enforcement needs to be deployed on an enterprise-wide scale. All applications that handle personal data need to make sure that the handling adheres to the promised policies. An enterprise-wide privacy-management system uses at least three types of systems (see Figure 2):

The Policy Management System enables the administrators of the system to define, change and update privacy policies. It distributes the privacy policies to the privacy enforcement systems.

The Privacy Enforcement System enforces the privacy protection for each individual resource that stores privacy-relevant data. It obtains policies from the policy management system and offers auditing data to the audit console. The privacy enforcement system is usually split into two parts: A resource-specific resource monitor shields the resource and a resource-independent authorization director evaluates the policies and decides whether requests are granted or not.

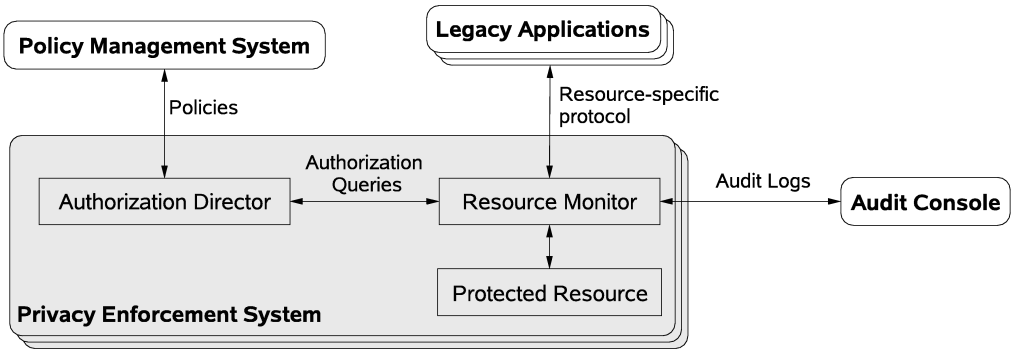

Figure 2. Components of a Enterprise Privacy Enforcement System

The authorization director authorizes operations on the collected data. After evaluating the policy, the authorization director returns whether the request is authorized or not and whether an authorized request implies any privacy obligations.

Each kind of protected resource (database, CRM system, ...) uses a corresponding resource monitor. This monitor shields the resource from direct access. Each incoming request is translated into a call to the authorization director. Only if the authorization director authorizes the request, the request is forwarded to the resource. The resource monitor records audit data and tracks pending privacy obligations.

Audit Console This system enables the Privacy Officer to review the audit information stored in the enforcement nodes and the policies distributed by the policy management systems.

The Platform for Enterprise Privacy Practices (E-P3P) is a refinement of the EPA Technical Reference Architecture [5, 6]. It enables enterprises to formalize and enforce privacy practices and to manage the consent of their customers.

3 Benefits for the Enterprise

EPA helps enterprises leverage personal data while protecting privacy. It recognizes the issues and investment relating to existing personal data in legacy systems as well as those generated by e-business initiatives. As such, EPA's value revolves around the following:

Enhance and preserve the value of data assets. The data model of the Technical Reference Model provides for the identification and categorization of personal data within the organization and thereby allows the establishment of appropriate protection measures.

Operate consistently with multiple privacy regulations and standards. The privacy regulation analysis helps to identify compliance obligations across different jurisdictions and express these in common terms. The applicable regulations are formalized by an enterprise privacy policy that is associated with any collected data. This so-called "sticky policy paradigm" supports identifying the applicable regulations and privacy promises for all personal data in an enterprise.

Build and promote trust in the marketplace. EPA enables customers to retain control over their data. The Management Reference Model enables and promotes responsiveness and privacy awareness of the enterprise. Together with external auditing, these measures promote trust of the customers.

Realize substantial privacy management choices. The regulatory analysis reveals compliance choices. This analysis highlights choices for uses of less sensitive data types and shows high risk and redundant privacy relationships.

Operate a sound platform for persistent privacy management. The Requirements Process within the Management Reference Model ensures ongoing environmental input on privacy.

4 Conclusions

The IBM Enterprise Privacy Architecture enables enterprises to provide an increased and well-defined level of privacy to their customers. It enables enterprises to act as the custodians of their customer's personal data that protect against privacy violations by themselves or others. Note that the Enterprise Privacy

Architecture assumes that the customers trust the privacy administrators and privacy enforcement systems of the enterprise to some extent[1]. The systems then protect the customer's data against privacy violations by regular employees or systems.

An important aspect of EPA's privacy enforcement system is the management of the data subject's consent on a per-person basis. Consent management includes the management of the consented policy as well as the management of the users opt-in and opt-out choices. The core of EPA's notion of consent management is the sticky policy paradigm: When submitting data to an enterprise, the user implicitly consents to the applicable policy and to the selected opt-in and opt-out choices. Opt-in and opt-out choices as well as the consented policy are associated with the collected data. This holds even if the data is sent to another enterprise. Note that policy management on a per-user-basis is useful once consent and different sources need to be considered [4]. Examples are managing data of different policy versions (e.g., due to different collection times), different user roles (e.g., premium and users funded by advertising), or users from different legislation (e.g., Europe and US).

Acknowledgments

We would like to thank Kathy Bohrer, Nigel Brown, Jan Camenisch, Calvin Powers, and Els Van Herreweghen for valuable comments. In addition, we would like to thank all members of the EPA team for the productive cooperation.

References

[1] Mark S. Ackerman, Lorrie Faith Cranor, and Joseph Reagle. Privacy in e-commerce: examining user scenarios and privacy preferences. In *1st ACM Conference on Electronic Commerce*, pages 1–8. ACM Press, 1999.

[2] Grady Booch. *Object Oriented Analysis and Design.* Benjamin Cummings, 1994.

[3] Jan Camenisch and Anna Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In *Advances in Cryptology - EUROCRYPT '2001*, Lecture Notes in Computer Science 2045, pp.93-117. Springer, 2001.

[1]Of course, this trust can be reduced by external audits and on-line auditing systems that are maintained by independent parties.

[4] Rüdiger Grimm and Alexander Roßnagel. Can P3P help to protect privacy worldwide? In *ACM Multimedia Workshop*, pages 157–160. ACM Press, 2000.

[5] Günter Karjoth, Matthias Schunter, and Michael Waidner. The Platform for Enterprise Privacy Practices — Privacy-enabled Management of Customer Data. To appear in *2nd Workshop on Privacy Enhancing Technologies (PET2002).* April 2002.

[6] Günter Karjoth, Matthias Schunter, and Michael Waidner. A Privacy Model for Enterprises. To appear in *Computer Security Foundations Workshop (CSFW2002).* IEEE Press, 2002.

[7] Sarah Spiekermann, Jens Grossklags, and Bettina Berendt. E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior. In *Electronic Commerce (EC'01)*, pages 38–47. ACM Press, 2001.

[8] The Platform for Privacy Preferences (P3P), W3C Candidate Recommendation, http://www.w3.org/TR/2000/CR-P3P-20001215, 2000.