UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

TÂNIA BASSO

PRIVAPP: A COMPREHENSIVE APPROACH TO GUIDE THE DESIGN OF PRIVACY-AWARE WEB APPLICATIONS

PRIVAPP: UMA ABORDAGEM EXTENSIVA PARA ORIENTAR O PROJETO DE APLICAÇÕES WEB COM PROTEÇÃO DE PRIVACIDADE

CAMPINAS
2015

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

TÂNIA BASSO

PRIVAPP: A COMPREHENSIVE APPROACH TO GUIDE THE DESIGN OF PRIVACY-AWARE WEB APPLICATIONS

Orientador: Prof. Dr. Mario Jino
Coorientador: Prof.ª Dr.ª Regina Lúcia de Oliveira Moraes

PRIVAPP: UMA ABORDAGEM EXTENSIVA PARA ORIENTAR O PROJETO DE APLICAÇÕES WEB COM PROTEÇÃO DE PRIVACIDADE

Doctorate thesis presented to the Electrical Engineering Graduate Program of the School of Electrical Engineering of the University of Campinas to obtain the Doctor grade in Electrical Engineering, in the field of Computer Engineering.

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas para obtenção do título de Doutora em Engenharia Elétrica, na área de Engenharia de Computação.

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELA ALUNA TÂNIA BASSO E ORIENTADA PELO PROF. DR. MARIO JINO

CAMPINAS
2015

Agência(s) de fomento e nº(s) de processo(s): CNPq, 141054/2011-5

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Elizangela Aparecida dos Santos Souza - CRB 8/8098

B295p

Basso, Tânia, 1981-
PrivAPP : a comprehensive approach to guide the design of privacy-aware web applications / Tânia Basso. – Campinas, SP : [s.n.], 2015.

Orientador: Mario Jino.
Coorientador: Regina Lúcia de Oliveira Moraes.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Privacidade. 2. Serviço na web. 3. UML (Computação). 4. UML (Linguagem de modelagem padrão). I. Jino, Mario,1943-. II. Moraes, Regina Lúcia de Oliveira,1956-. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: PrivAPP : uma abordagem extensiva para orientar o projeto de aplicações web com proteção de privacidade
Palavras-chave em inglês:
Privacy
Web service
UML (Computer)
UML (Standard modeling language)
Área de concentração: Engenharia de Computação
Titulação: Doutora em Engenharia Elétrica
Banca examinadora:
Mario Jino [Orientador]
Elisa Yumi Nakagawa
Ana Cristina Vieira de Melo
Ariadne Maria Brito Rizzoni Carvalho
Ivan Luiz Marques Ricarte
Data de defesa: 04-12-2015
Programa de Pós-Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE DOUTORADO

Candidato: Tânia Basso RA: 009935
Data da Defesa: 04 de Dezembro de 2015

Título da Tese: "PRIVAPP: A COMPREHENSIVE APPROACH TO GUIDE THE DESIGN OF PRIVACY-AWARE WEB APPLICATIONS" (*PRIVAPP: UMA ABORDAGEM EXTENSIVA PARA ORIENTAR O PROJETO DE APLICAÇÕES WEB COM PROTEÇÃO DE PRIVACIDADE*)

Prof. Dr. Mario Jino (Presidente, FEEC/UNICAMP)
Prof. Dr. Elisa Yumi Nakagawa (ICMC/USP)
Prof. Dr. Ana Cristina Vieira de Melo (IME/USP)
Prof. Dr. Ariadne Maria Brito Rizzoni Carvalho (IC/UNICAMP)
Prof. Dr. Ivan Luiz Marques Ricarte (FT/UNICAMP)

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no processo de vida acadêmica do aluno.

To my husband, Leandro Piardi, for the support, incentive and patience.

ACKNOWLEDGMENTS

To God, for the blessings and opportunities that allowed me to carry out this work.

To my advisor, Professor Dr. Mario Jino, for allowing me to perform this dreamed work. For his patience, for his valuable advices and for his understanding and generosity.

To my co-advisor, Professor Dr. Regina Lúcia de Oliveira Moraes, to whom I have no words to thank. Anything I say would be very little given the immense help she gave me to perform this work and for my professional and personal development. I also thank her for being more than a co-advisor – role that, by the way, has been performed exceptionally well – but for doing so much for me during this walk, believing in my effort and work. I thank her for the patience and availability, incentives, advices, friendship. To her I owe much, much more than she can imagine.

To Professor Dr. Marco Vieira, for the valuable advice, which helped me to better understand several aspects involved in this research and for the incentive to proceed with this work. Also thank to the reception and support during my stay in Coimbra.

To Professor Dr. Andrea Bondavalli, for the generous reception in Florence and for his criticism, always thorough and constructive, contributing to many improvements in this work.

To my husband, Leandro Piardi, for the support and understanding in the moments he was deprived of my company. Thank for his companionship and for helping me during the most stressful and difficult moments, which always motivated me to overcome, often believing in me more than myself.

To my mother and family, for giving me encouragement and support.

To Nuno Antunes for his time and expertise in building applications and test environments. His help was essential to the development of this dissertation.

To Leonardo Montecchi, for helping me, so patiently, with his experience in UML. He is a great working partner, which I hope to count on even after I have completed this work.

To Andreia Rossi, who is so kind giving up his own tasks to help a friend. To Suelen Mapa, always helping me to get emotional balance in such a hard period.

To other friends, who greatly contributed with their incentive, and who have provided fun moments in the course of this work.

To the University of Campinas (UNICAMP), which provided the opportunity for my intellectual and professional development.

To the National Council for Scientific and Technological Development – CNPQ – and to Higher Education Personnel Improvement Coordination – CAPES – for the financial support.

To the Universities of Coimbra and Florence for the welcome and for putting at my disposal the necessary resources for the development of this work.

And finally, I thank the people who challenged me and who often let me down, because they have taught me to recover from each stumble on the road of life.

RESUMO

Aplicações e Serviços Web são tecnologias extremamente relevantes atualmente, uma vez que proveem uma grande variedade de serviços online tais como vendas, operações bancárias e financeiras, entre outras. Normalmente, para utilizar esses serviços, os usuários, clientes e parceiros de negócio precisam fornecer informações de identificação pessoal como, por exemplo, endereço, número do cartão de crédito ou do seguro social. Além disso, aplicações mais recentes são capazes de coletar, automaticamente, informações relacionadas a atividades dos seus usuários como, por exemplo, padrões de utilização ou localização aproximada. Uma vez que essas informações são disponibilizadas elas não estão mais sob o controle de seus proprietários no que diz respeito a como elas são realmente manipuladas, e isso causa preocupações com a privacidade. Se por um lado empresas e organizações desejam obter, minerar e compartilhar informações de identificação pessoal, por outro lado elas também estão interessadas em manter essas informações privadas pois precisam atender as leis de privacidade e obter credibilidade. O presente trabalho apresenta uma abordagem que auxilia a análise, projeto e desenvolvimento de aplicações e serviços Web que incluem proteção de privacidade. A abordagem é composta por um Modelo Conceitual de Privacidade (sistematiza conceitos de privacidade, exibindo os elementos de privacidade e suas relações, de maneira organizada), uma Arquitetura de Referência (arquitetura abstrata que descreve as funcionalidades que devem ser implementadas para proteger a privacidade de usuários em aplicações Web) e uma extensão da Linguagem de Modelagem Unificada (Profile UML – Unified Modeling Language Profile)(extensão para incorporar conceitos de privacidade). Ela permite que as pessoas envolvidas no projeto entendam melhor o domínio de privacidade e desenvolvam modelos e aplicações Web consistentes com políticas de privacidade a fim de garantir que elas sejam de fato aplicadas. Com isso, informações de identificação pessoal podem ser gerenciadas de maneira mais segura e protegida de diferentes fontes de violação de privacidade. Um estudo de caso foi desenvolvido, aplicando a abordagem para melhorar a proteção de privacidade de uma livraria virtual. A abordagem foi avaliada em relação a dois atributos de qualidade importantes: aplicabilidade e completude. Resultados mostram que a abordagem agrega valor ao projeto como um todo e é uma importante contribuição para melhorias no processo de desenvolvimento de aplicações no domínio de privacidade.

***Palavras-chave**: Privacidade. Serviço Web. UML. Arquitetura de Referência.*

ABSTRACT

Web applications and web services are relevant technologies nowadays, supporting a wide range of services, such as e-commerce, e-banking, e-government, and others. Usually, to access these services, users, customers, and business partners need to provide personally identifiable information (PII), such as addresses, social security IDs, and credit card numbers. Furthermore, modern applications can automatically gather information related to users' activities, such as, for example, usage pattern or approximate location. Once this information is made available, it is no longer under the control of their owner regarding how it is actually handled, which raises privacy concerns. If on one hand companies and organizations want to be able to gather, data mine and share PII information, on the other hand they are interested in keeping such information private due to privacy laws and their credibility with respect to how able they are to protect the privacy of their users. This work presents a comprehensive approach to support the analysis, design, and development of web applications and services with privacy concerns. The approach is composed of a Privacy Conceptual Model (systematizes privacy concepts, showing privacy elements and their relations in an organized way), a Privacy Reference Architecture (abstract architecture which describes functionalities that must be addressed during the development of web applications to protect the privacy of the users) and a Privacy UML Profile (extension of the UML language to incorporate privacy concepts) and it allows stakeholders to better understand the privacy domain, as well as modeling and developing web applications consistently with the privacy policies enabling their enforcement. This way, PII can be managed in a more secure manner and protected from different sources of privacy violation. A case study was developed applying the approach to improve privacy protection for an online bookstore. The approach was evaluated considering two important key attributes: applicability and completeness. Results show that the approach adds value to the stakeholders and is an important contribution towards improving the process of designing web applications in the privacy domain.

Keywords: Privacy. Web Service. UML Profile. Reference Architecture.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ACM	Association for Computing Machinery
APEC	Asia-Pacific Economic Cooperation
COPPA	Children's Online Privacy Protection Act
ENISA	European Union Agency for Network and Information Security
EPA	Enterprise Privacy Architecture
EPAL	Enterprise Privacy Authorization Language
EU	European Union
FERPA	Family Educational Right to Privacy Act
GLBA	Gramm-Leach-Bliley Act
HIPAA	Health Insurance Portability and Accountability Act
IAPP	International Association of Privacy Professionals
ICT	Information and Communication Technology
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
NSA	National Security Agency
OASIS	Organization for the Advancement of Structured Information Standards
OECD	Organization for Economic Co-operation and Development
OMG	Object Management Group
OMT	Object-Modeling Technique
OOSE	Object-Oriented Software Engineering
OSI	Open Systems Interconnection
P3P	Platform for Privacy Preferences Project
PGP	Pretty Good Privacy
PII	Personally Identifiable Information
PIPEDA	Personal Information Protection and Electronic Documents Act
PRBAC	Privacy-Aware Role Based Access Control
QoS	Quality of Software
RA	Reference Architecture
RBAC	Role Based Access Control
RSA	Rivest-Shamir-Adleman (cryptography algorithm)
SMPC	Secure Multi-Party Computation
TPC	Transaction Processing Performance Council

UML	Unified Modeling Language
US	United States
USA	United States of America
XACML	eXtensible Access Control Markup Language

CONTENTS

1 . INTRODUCTION

Applications and services provided via web such as e-commerce, banking and financial services are essential and widely used by modern society. The provision of these services, provided with the aid of a browser, is possible only due to the support of computer networks that allow communication between suppliers and users. Many times, to use these services, users and customers need to provide Personally Identifiable Information (PII), i.e., any data that could potentially identify a specific individual (e.g. driver's license number, home address, telephone number, digital identity, credit card numbers, etc.). Furthermore, cookies, web beacons and similar technologies can record user's data, actions and preferences (e.g., search strings, visited links, approximate location, etc.), often without their knowledge and consent.

Once sent through the network, the collected information becomes known to the service provider and, often, to other business partners that do not even have any interaction or involvement with the users. It means that once personal information is made available it is no longer under control of its owner with respect to how they are used and the consequences. Companies and organizations want to be able to gather, data mine and share this information efficiently, but without putting their reputation at risk. Customers want choices regarding the way their personal information is used and ease of access to these information. Thus, relevant privacy concerns arise from both sides; and by privacy, for now, we mean *the right of an entity to be secure from unauthorized disclosure of personally identifiable information that is contained in an electronic repository* (Bertino *et al*., 2008).

Protecting the privacy of information manipulated by web applications and services is essential, due to privacy laws (the companies and organizations that hold private data must comply with them) and competitiveness differentials (the more a company protects the privacy of its customers and business partners, the more credibility it gets). The need to ensure the privacy of personal information handled by web services and applications has led to a significant development of technology (e.g., Ni *et al*. (2007) (presented a privacy-aware role based access control); Tbahriti et al. (2011) (presented a framework for privacy management in web services interactions); Giffin et al. (2012) (presented a mandatory access control and a declarative policy language to the Model-View-Contoller architecture)). However, a recurring problem in constructing web applications and services with privacy requirements is the insufficient resources for modeling and documenting them (Hoepman,

2014). In practice, the lack of integration of privacy requirements in the application design and development makes privacy protection difficult, since privacy mechanisms have to be devised based on both the application and the privacy policy. Also, the lack of privacy reference models hampers the standardization and evolution of systems (Nakagawa *et al.*, 2012). Approaches for understanding the privacy domain and modeling privacy views of system applications through a structured model are needed and can help to better describe how an application must deal with personally identifiable information in order to protect the privacy of the users.

In this work we propose PrivAPP, which is a novel and comprehensive approach to guide the design of privacy-aware applications. The goal is to systematize the privacy concepts that are related to the scope of web applications and, consequently, provide a better understanding of the privacy domain and ease the modeling and development of privacy-aware web applications and services.

The approach includes a Privacy Conceptual Model, a Privacy Reference Architecture, and a UML Profile for privacy aware modeling. Briefly, the Conceptual Model is composed of elements that represent privacy concepts and their relations. Its goal is to specify and organize the privacy domain knowledge. The Privacy Reference Architecture describes the features and functionalities that must be addressed during development to protect the privacy of the users. The elements of the conceptual model are distributed through layers (application layers based on three-tier architecture pattern) where they can be implemented. The goal of the Reference Architecture is to allow deriving concrete architectural models that facilitate the development of privacy-aware technology. Last but not least, the UML Profile extends the UML language (OMG, 2011) to incorporate privacy concepts. UML diagrams can then be improved and better applied to the development process of privacy-aware applications and services. It documents the elements of the conceptual model in order to reduce ambiguities in the solution.

PrivAPP can be used in a modular way, i.e., it is possible to use, for example, only the Reference Architecture or only the UML Profile, depending on requirements or needs. We performed an evaluation process, based on a set of privacy policies, through which we analyze the applicability and completeness of the approach. We also developed a case study, applying the approach to improve privacy protection in a web application of an online bookstore. The results of both activities (approach evaluation and case study) gave an indication that PrivAPP has high levels of completeness and applicability and it is an

important contribution towards improving the process of designing web applications in the privacy domain.

In practice, we want to answer these two main research questions:

Q1 - How to enhance the construction of privacy-aware web applications and services, i.e., how to help stakeholders to pay attention in constructing web applications and services with privacy protection?

Q2 - Reference models and specific UML resources can help in this task?

1.1 MOTIVATION

Personal information is quite valuable. The notion of big data has certainly gained momentum in recent years, essentially dealing with the efficient management of large volume, complex, and growing datasets from multiple sources, as well as the extraction of useful knowledge from these datasets (Wu *et al.*, 2014). Usually, companies and organizations use this knowledge to increase profitability. This can be done, for example, through statistical research, identification of profiles, sending customized advertisements or even selling information to other companies. However, at the same time that companies and organizations are interested in big data, they have also great interest in protecting the privacy of information manipulated by their web applications and services. The two main reasons, as already mentioned, are privacy laws and competitive differentials.

In the European Union (EU) (EU, 1995), Canada (PIPEDA, 2000) and Australia (Privacy Act, 1988), for example, regulations for the protection of personally identifiable information have been created, and some of them cross industry sectors. The United States of America (USA) has taken a sectorial approach, enacting separate regulations for health care (HIPAA, 1996), finance (Gramm-Leach-Bliley, 1999) and protection of children's data (COPPA, 1998). In either case the objectives are clear: to protect personal information, i.e., the companies and organizations that hold these data have the obligation and responsibility to protect them.

Regarding the competitive differentials, the reputation of a company can be strictly dependent on privacy protection, i.e., the more a company protects the privacy of its customers and business partners, the more credibility it gets. Proof of this is the last research undertaken by Truste (Truste 2015), where 91% of the internet users interviewed said they avoid doing business with companies they do not believe protect their privacy online.

There are many security and privacy violation factors over the multiple components of a system, including web applications. Even with the raising concern about credibility and reputation, there are numerous occurrences of privacy violation involving web applications, ranging from small to big companies, such as Carrier IQ (Adweek, 2011) and Facebook (NYTimes, 2012). Given this scenario, it is of utmost importance to construct web applications and services that protect privacy. To the best of our knowledge there is not a well-established and consolidated guide to help developing applications and services with privacy protection. This is a complex and difficult task that involves many factors such as privacy laws, technology support and user preferences.

1.2 OBJECTIVES

This work proposes an approach that aims to help improving the scenario of lack of privacy protection in the construction of web applications and services. The main objectives of the approach are: (i) systematize privacy concepts within the scope of web applications; (ii) facilitate the understanding of the privacy domain by the stakeholders; (iii) serve as a guideline for the design of concrete architectures that support web applications and services with privacy protection features; (iv) provide resources for the documentation of privacy specifications of web applications, helping to structure particular concepts of privacy; and (v) improve privacy protection definition and enforcement.

To establish the approach, secondary objectives must be addressed. To systematize the privacy concepts it is necessary (i) to understand how web applications should handle privacy; (ii) to outline the privacy elements required for this task. This calls for a model of the domain concepts that are required for modeling views of the system where privacy management and protection are applied.

Based on this conceptual model, we need to understand how privacy decisions can be made at the architectural level and provide an architecture model, which can serve as guidance for the development, standardization, and evolution of systems in the privacy domain. Another objective is to integrate privacy protection features in the design process – and, consequently, in the development – process, helping service providers to fulfill privacy requirements.

Furthermore, we want to investigate if the proposed approach can be applied in practice. To do this, case studies must be performed. The idea is to use the approach to

construct an application with privacy protection features. Also, we want to evaluate the approach and its elements according to quality attributes (e.g., completeness, applicability). The goal of this process is to identify weaknesses and to add improvements on PrivAPP.

1.3 CONTRIBUTIONS

Research in the field of privacy is relatively recent and involves many aspects such as privacy laws, user's perceptions, privacy policies, etc.. Some effort has been spent by the privacy community to address the widespread concern that indiscriminately collects and manages personally identifiable information pose to Internet connected systems.

In this work the focus is on the design of web application and services with privacy protection features and on what can be done concerning the understanding of privacy domain and the documentation of these features. Thus, we provide a model for the use of privacy elements in the web application context, considering even user's privacy preferences. The main contributions of this dissertation are:

- A discussion on the current privacy context, regarding: (i) different privacy definitions; (ii) the value of the personal information nowadays; (iii) how privacy is a competitive differential to companies and organizations; (iv) privacy laws and regulations across the world; (v) relevant and recent cases of privacy violation;
- A discussion on the current relationship between web applications and privacy, i.e., how these applications deal with privacy concerns. Also, we introduce the privacy policies (documents that explain how an organization handles any customer, client or employee information gathered through its operations) and the problems addressed by them nowadays;
- A Privacy Conceptual Model, which defines and organizes privacy concepts and their relationships;
- A Privacy Reference Architecture, describing privacy features that should be considered during the development of an application across the different application layers (presentation, application and persistence);
- A Privacy UML Profile, for modeling web applications with privacy protection features through UML diagrams;

- A Database Framework, with a set of independent tables that allows users to express their privacy preferences in detail and a mechanism that, based on predefined policies and user preferences, allows or denies access to personally identifiable information;
- A set of processes and adapted techniques that can be used to evaluate Reference Architecture quality attributes (*completeness, applicability, usability*, and *feasibility).*

Furthermore, the work concerning this dissertation is not limited to the contents of this document: the Ph.D. candidate has participated in two international projects (*MENON-WS – Methodologies for the Development of Non-Vulnerable Web Services* – and *DEVASSES – DEsign, Verification and VAlidation of large scale, dynamic Service SystEmS*), spending a period at the University of Coimbra (Portugal) and a period at University of Florence (Italy), for research purposes. This allowed exchanging research experience with Ph.D. and Post Doc students from these universities. The candidate also co-advised an undergraduate student in Information Systems. As a result of this effort, articles and technical production have been generated, and skills in research and advising students were gained.

1.4 TERMINOLOGY

The understanding of principles concerning privacy and data protection has evolved over the years, at the international and national levels. According to Solove (Solove, 2006), the term "privacy" is an umbrella term, referring to a wide and disparate group of related things. Various definitions for privacy, data protection and related concepts have been proposed (Pfitzmann and Hansen, 2009; ISO, 2011; IAPP, 2012). However, a standard has not yet been adopted. This affects the understanding of requirements to be realized and supported in legal, organizational and technical systems. The terminology even diverges in different communities dealing with "privacy by design" or "data protection by design" (ENISA, 2014).

For a better understanding of the content of this work, we introduce terms relevant for the scope of this dissertation. The following related terms and correspondent meanings are based mostly on ISO/IEC 29100 (ISO, 2011) and the glossary of privacy terms proposed by the International Association of Privacy Professionals (IAPP, 2012). These sources were selected due to their international acceptance. Terms are presented in alphabetical order.

- **Activity Tracking**: the act of recording user's activity on the computer as, for example, visited websites, search strings, online purchases, or any other activities that leave a digital trail.
- **Affiliate**: an entity that controls, is controlled by, or is under common control with the entity that is the subject of the privacy notice. Affiliates include, for example, the entity's sister companies, parent or subsidiaries.
- **Anonymity:** situation in which someone's personally identifiable information cannot be identified by the recipient, in order to protect the identity of the data subject.
- **Data Breach:** The unauthorized acquisition of computerized data that compromises the security, privacy, confidentiality, or integrity of personal information maintained by a data collector.
- **Consent:** the individuals' way of giving permission for the organization to collect, use or disclosure his/her personally identifiable information. Consent may be affirmative (e.g., opt in), negative (e.g., opt-out) or implied (e.g., the individual didn't opt out).
- **Consumer:** individual who is the subject of a personal data record. In this dissertation it can be also referred as **User** or **Data Subject.**
- **Cookie:** small text files that are stored on a client machine and which may be later retrieved by a Web server from a client machine. Cookie files allow the Web server to keep track of the end user's Web browser activities.
- **Data lifecycle:** the period of time defined from the originating point at which an organization acquires personal information to the time when the information is removed from the organization.
- **Data Subject:** term used in some data protection legislation to describe an individual who is the subject of a personal data record. In this dissertation it can be also referred as **Consumer** or **User.**
- **Directive:** formal and usually mandatory executive order or official pronouncement on a policy or procedure which needs to be attained.
- **Disclosure:** the release, transfer, provision of access to, or divulging in any other manner of information outside the entity holding the information.
- **Identity Theft:** the use of an individual's personally identifiable information (PII) in order to fraudulently appropriate their identity.

- **Information Security:** the act of safeguarding an organization's data from unauthorized access or modification to ensure its availability, confidentiality, and integrity.
- **Opt-in:** a user's expression of affirmative consent based upon a privacy policy statement.
- **Opt-out:** A user's exercise of choice through a negative consent based upon a privacy policy statement.
- **Personal Information:** any information that (i) relates to an individual and (ii) identifies or can be used to identify the individual (see **Personally Identifiable Information**).
- **Personally Identifiable Information:** any information that can be traced to a particular individual, such as a name, phone number, social security number, or e-mail address. Personal user preferences tracked by a Website via a cookie are also considered personally identifiable. In this dissertation it can be also referred as **Personal Information.**
- **Policy enforcement:** the act of fulfilling the privacy promises described in the privacy policy, i.e., to guarantee that the system in fact implements resources to accomplish the privacy policies statements.
- **Privacy Policy Statement:** part of the text described in privacy policies, with full meaning. It can describe, for example, which personal information is collected; how it will be used; with whom it will be shared; etc.
- **Privacy policy:** an internal statement that governs an organization or entity's handling practices of personal information. It is directed at the users of the personal information. A privacy policy instructs employees on the collection and the use of the data, as well as any specific rights the data subjects may have.
- **Privacy preferences:** specific choices made by a data subject about how their personally identifiable information should be processed for a particular purpose.
- **Privacy Principles:** set of shared values governing the privacy protection of personally identifiable information when processed in information and communication technology systems.
- **Privacy Stakeholder:** individual executives within an organization who lead and own the responsibility of privacy activities.
- **Privacy**: the appropriate use of personal information under the circumstances. What is appropriate will depend on context, law, and the individual's expectations; also, the right of an individual to control the collection, use, and disclosure of personal information.

- **Private Data:** any and all data that relates to an identifiable individual (see **Personal Information** and **Personally Identifiable Information**).
- **Recipient:** Any person or organization to whom data is disclosed, whether a third-party or not.
- **Sensitive information:** category of personally identifiable information whose nature is sensitive, related to data subject's most intimate sphere, as, for example, racial or ethnic origin, political opinions, religious beliefs, trade union membership, health, sex life, and criminal convictions.
- **Third-party:** privacy stakeholder other than the data subject, the company which detains the personally identifiable information and the persons who are authorized to process the data under the direct authority of the referred company.
- **User profile:** information about actions, preferences, interests and other characteristics that companies track and compile about their users.
- **User:** individual who is the subject of a personal data record (see **Consumer, Data Subject**).
- **Web beacon:** Also called a *Web bug* or a *pixel tag* or a *clear GIF*. Used in combination with cookies, a Web beacon is an often-transparent graphic image, usually no larger than 1 pixel x 1 pixel, that is placed on a Web site or in an e-mail and used to monitor the behavior of the user visiting the Web site.

1.5 DISSERTATION OUTLINE

In **Chapter 2** there is a brief contextualization of privacy, starting with a discussion about its several definitions. Then we discuss the value of the personal information nowadays and how important privacy protection is for companies and organizations. Also, we present some relevant privacy laws and regulations across the world – including Brazil – as well as some cases of privacy violation reported recently. The relationship between privacy and security is briefly discussed. Finally, the chapter contains a privacy contextualization in the scope of web applications and services, regarding the lack of privacy protection and resources currently used to address some privacy concerns, especially privacy policies.

In **Chapter 3** background and related work are presented. It describes relevant previous work (identified through literature reviews) which presents, respectively, reference architectures and UML profiles that served as basis for the construction of PrivAPP. Also, we

discuss the importance of these models and how they can contribute to the scenario of lack of privacy protection.

In **Chapter 4** PrivAPP is presented. We described in detail its three components: the Privacy Conceptual Model, the Reference Architecture and the UML Profile. The Privacy Conceptual Model addresses, within the scope of web applications and services, the privacy concepts and their relationships. The Reference Architecture addresses abstract software components that represent functionalities related to privacy protection. Finally, the UML Profile addresses an extension of the UML metamodel to allow using privacy protection features in UML diagrams.

Chapter 5 describes a case study using PrivAPP. The proposed approach is applied in the design (and, consequently, implementation) of data privacy protection features for a web application that represents an online bookstore. The concrete architectures, UML diagrams, details of implementation of components and experimental results are shown in this chapter.

In **Chapter 6** an evaluation process for the proposed approach is presented. This process evaluates the Reference Architecture regarding some quality attributes. Also it evaluates the *completeness* and *applicability* of PrivAPP by means of an empirical study, where a set of privacy policies from relevant companies were analyzed and the elements from PrivAPP were employed to help enforcing these policies. Then, we finish with a discussion of the limitations of the proposed approach.

In **Chapter 7**, we conclude by summarizing the results from this dissertation and topics for future work that can be explored to advance the research in the field.

In **Appendix A** we describe details of the literature review we performed to this work, including research questions and search strings. In **Appendix B** we present complementary details of the Privacy Reference Architecture evaluation process, as questionnaires and mappings of the Reference Architecture with privacy taxonomies and concrete architectures. **Appendix C** shows statements and UML diagrams used as a case study to the evaluation of the UML Profile. **Appendix D** presents a detailed description of the database framework and policy model used to implement the access control model in the case study. Finally, **Appendix E** presents the results of the activity we performed to evaluate the applicability of the approach, i.e., the requirements identified by professionals to construct a web application using the PrivApp.

2 . CONTEXTUALIZING PRIVACY

Privacy is a very abstract concept whose values and extensions vary from person to person. What one person considers an invasion of privacy, another person might consider as something completely normal and acceptable. Also, privacy has a broad and comprehensive context: Leino-Kilpi *et al.* (2001), for example, describe the privacy concept in four dimensions: social, physical, informational and psychological.

There is no single accepted understanding of privacy, but a set of intertwined notions. Definitions of privacy and levels of protection of the privacy sphere are in constant flux across nations and cultures and historical periods (Venier, 2010). In a broad sense, privacy is strongly connected with the idea that there are some things that other people should not see or know (Elgesem, 1996).

A privacy concept which is widely spread is the one discussed by Warren and Brandeis (1890), where they state that "*privacy is the right to be left alone*". Also in this work the authors state that "*the right to privacy ceases upon the publication of the facts by the individual, or with his consent*". From this statement it is possible to identify that individuals should be careful about their data because once these data are available there should be no way to protect their privacy anymore.

The technological advances have brought many facilities to the modern life, but, on the other hand, privacy concerns arise. Digital records have been rapidly growing over the last twenty years, as more and more business processes have been computerized. In the last few years it has been easy to share this information with others, either inside or outside an organization. New *big data* and *data mining* tools can manage large data sets and let researchers and analysts view the information in new ways, and determine trends or patterns that have important commercial applications. Last, we cannot forget to mention the social networks, which allow users to share any information, ranging from ideas, activities, events, and particular interests.

In this new scenario, a new sense of privacy can be assumed, in which it is more like a property, namely, "*the property of having control over the flow of personal information*" (Elgesem, 1996). Fried (1984) states that "*privacy is not simply an absence of information about us in the mind of others; rather is the control we have over information about ourselves*".

In the Web applications and services context, privacy refers to privacy of information (Clarke, 1999). A known definition for privacy of information is presented by Westin (1987): "*privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others*". Furthermore, Wang *et al.* (1998) state that "*privacy usually refers to personal information and the invasion of privacy is usually interpreted as the unauthorized collection, disclosure, or other use of personal information as a direct result of electronic commerce transactions*". More recently, Bertino *et al*. (2008) says that privacy is "*the right of an entity to be secure from unauthorized disclosure of sensible information that is contained in an electronic repository"*. For the scope of this dissertation, we will adopt these last two definitions (Wang *et al.*, 1998; Bertino *et al.*, 2008), since our work considers whole data lifecycle, ranging from collection, management, storage and disclosure.

2.1 THE VALUE OF PERSONAL INFORMATION

Companies and organizations, usually the medium-size and large-size ones, use sophisticated technologies to gather personal information from their customers or possible customers. Personal information is valuable from the business point of view because with it is possible, for example, to identify consumer profiles and to send specific advertisement directly to the consumers, according to their specific interests. Also, it is common to sell personal information to third parties. In addition, the sale of personal information to third parties is a practice that can cause dissatisfaction and annoyance to the owner of the information, who can be the target of unwanted advertisements.

The profile identification could even be advantageous to consumers, once they would receive only advertisements of major interest (and, consequently, reducing the amount of received advertisement). However, the personal information collected by the companies can contain extremely confidential information, such as financial or health-related data. When used in an inappropriate way or even stolen, this information may be targeted for criminal purposes such as identity theft or credit card fraud. In the case of health-related information, information about symptoms, examination results, diagnosis, treatment and course of diseases can provide new knowledge as, for example, the relationship between certain diseases and certain occupational, social and cultural profiles, local habits and living places. This new knowledge can be used, on the one hand, for commendable purposes as, for example, a better

understanding of diseases causes and treatments. However, on the other hand, they can be used for questionable purposes and cause, for example, the identification of patients who, as consequence, can suffer some kind of retaliation or prejudice.

Research has shown that although people approve the practice of collecting their personal information when they have given permission to do so and when the company uses their information only for the purpose to which they agreed, they deeply resent when a company is not forthcoming about how it plans to use personal information or when it uses personal information for other purposes. To these people, these actions constitute a violation of their privacy (Perkins and Markel, 2004; Reay *et al.*, 2009).

A survey conducted by Forrester Research Inc. (Reitsma *et al.*, 2011) showed that most Internet users in the United States are concerned about the security and privacy of their online data. From the more than 31,000 adults surveyed, 39% believe that no company keep their personal information protected enough; 43% do not want their data to be stored or made public; 23% do not want their behaviour and online information tracked and shared with third parties.

The Truste Company performed, in 2013, a survey to understand consumer concerns, consumer trust, and business impact related to online privacy of adults in the United States (Truste, 2013). More than two thousand Internet users were interviewed and 89% revealed to be concerned about the privacy of their information when using the Internet. Shopping online, using social networks and using internet banking are the activities that mostly generate this concern: 89% of the respondents are concerned about privacy when shopping on the internet; 87% when using social networks and 86% when using Internet banking. Recently, in 2015, the same company released a Consumer Confidence Edition (Truste, 2015), which shows that consumers concerns about their privacy is rising: this time 92% of the interviewed stated that worry about their privacy online (see Figure 2-1).

Figure 2-1. Truste 2015 Privacy Survey: Consumer Concern, Consumer Trust and Business Impact (Truste, 2015).

Figure 2-1 shows that consumer trust is declining: in 2013, 57% stated that they trust the companies with their personal information online, while in 2015 55% stated the same. This mistrust implies a direct impact on the company's business: in 2015, 91% of US internet users say that they avoid companies that do not protect their privacy online. And this impact is also rising when comparing with 2013, when 89% of the interviewed said the same.

In fact, consumer concerns are well founded: a survey performed by Accenture Company in 2009 shows that among the 5,500 business leaders from companies located in 19 countries, 58% revealed cases of loss of personal information of its customers. From these 58%, 19% said that the loss of personal information occurred more than five times. In addition, 55% of companies declared to provide personal information to third parties or even to outsource the collection or management of such data (Accenture, 2009).

In Brazil, a recent work of Silva (2015) identified which are the personal information the Brazilians are most concerned about privacy protection. The author interviewed 1.104 Brazilian citizens from regions South, Southeast, Midwest, North and Northeast and results showed that they are most concerned about protecting passwords (74%), credit card numbers (73%), agency and bank accounts (67%) and expenses with credit card (63%).

2.2 PRIVACY AS A COMPETITIVE DIFFERENTIAL

Based on the numbers presented in the previous section, there is no doubt that investing in privacy protection provides a competitive differential to companies and organizations. Perkins and Markel (2004) state that companies should favor rigorous protection not only because it is the right thing to do, but also because it is in their best interests. Ironically, the pragmatic argument is based on utility, not on rights. Low privacy protection can be good for business in the short run; a robust privacy protection is good for business in the long run. The reason for this is that people avoid companies with low reputation and credibility.

To evaluate the impact of privacy concerns on the use of online behavioral advertising, the Ponemom Institute interviewed, in 2010, 90 organizations, all located in the United States, which are consumer-facing companies using Internet advertising as a primary marketing channel. More than 70% of the respondents agree that online behavioral advertisement increases their company's marketing and sales performance. However, the same survey states that due to privacy-related concerns, the $2.4 million currently spent on online behavioral advertisement would increase to $8.96 million if respondents did not have privacy concerns, i.e., the privacy concerns impair the growth of this segment. Also, for all 90 benchmarked companies taken as a whole, the amount not spent on online behavioral advertisement due to privacy fear amounts to $604.9 million. This translates into $2.8 billion dollars not earned in that year (Ponemon, 2010).

An important factor that makes companies and organizations be concerned about privacy is their reputation. A survey on Reputation Impact of a Data Breach (Ponemon, 2011), performed in 2011, surveyed 843 senior-level individuals with deep expertise and knowledge about their organization's brand and reputation management objectives. According to the survey, 92% of the respondents believe that privacy and data protection is important in protecting the organization's reputation and brand value; 65% rate privacy and data protection practices as a most important factor contributing to their organization's brand and reputation.

Although there is interest by companies and organizations in protecting privacy, the challenge is too big to believe that these companies alone will be able to create the right incentives or outcomes for this purpose: the threats are too broad, the actors too numerous, the knowledge levels too unequal, the risks too easy to avoid internalizing, the privacy problem

too prevalent. So, incentives as government intervention in the form of privacy laws are necessary (Cate, 2009). Where companies fail to produce appropriate incentives for privacy protection, people usually look to law.

2.3 PRIVACY LAWS AND REGULATIONS

According to Glass and Gresko (2012), the way that privacy is handled, and to what extent it is legislated, is largely dictated by the dominant cultural values. So, it is natural that legislative approaches differ from country to country. Currently, many countries have regulations for privacy of information as, for example, China, India, Australia, countries of the European Union, United States, Canada, etc. Especially when dealing with international business interactions, it is important to know and understand the involved privacy legislation, because differences between countries must be considered when developing and implementing global web-based applications (Ives and Jarvenpaa, 1991). In particular, the different approaches to legislation between countries are significant and can impact system design (Hinde, 2003).

2.3.1 Privacy Legislation

The two most relevant legislations in the literature are the ones from European Union (EU) and the United States (US) (Hinde, 2003; Perkins and Markel, 2004). They are addressed in this section, as well as a recent privacy law from Brazil, since this work was developed in the scope of a Brazilian graduate program.

The European Union applied, in 1998, the European Privacy Directive, a legal instrument that supports the exercise of a right to privacy and rules for personal data protection (EU, 1995). By forcing EU member states to create legislation that ensures a minimum level of data protection, the directive is intended to remove trade barriers between countries within the EU. It is intended to ensure that data may be collected and possessed only for a specified and legitimate purpose, may be kept only long enough to fulfill that purpose, must be kept accurate and up-to-date, and may not be transferred to a third country that does not have an adequate level of protection. To guarantee that these principles are enforced, each EU member must create an independent authority to supervise data protection; organizations

within each EU member must appoint and register with government authorities a "data controller" to be responsible for collected data; and individuals must have the right to access their personal information, correct inaccuracies, and arise against any unauthorized use of the information.

In the US, the privacy legislation is stated through acts. The Privacy Act of 1974 (USPrivacyAct, 1974) has been encoded at different contexts. In a more general context, the act establishes a code of practice for the collection, maintenance, use, and disclosure of information about individuals that are stored by federal agencies. It provides three fundamental rights: the right to see the records about themselves (there are exceptions), the right to alter records that are inaccurate, and the right to sue the government if it violates the act. In addition, the Congress enacted the Freedom of Information Act in 1966 (FOIA, 1996). The basic provisions allow individuals not only to access paper documents, but also to access electronically created documents and information, such as electronic databases, electronic documents, word-processing documents, and e-mail.

When the EU directive went into effect, they ruled that because the US lacks a comprehensive privacy law, it lacks an "adequate" level of protection. In response to the EU directive, a US directive known as the Safe Harbor Program was created (SafeHarbor, 2000). Its purpose was to assure EU member states that those US companies that subscribed to Safe Harbor provisions were demonstrating an "adequate" level of protection. Safe Harbor includes principles, a set of frequently asked questions, and a description of enforcement measures. A company that wishes to be Safe Harbor compliant must send to the Department of Commerce a written statement that it agrees to adhere to these requirements.

In more specific contexts, there are several other laws. For example, the act called Gramm-Leach-Bliley Act (GLBA, 1999) establishes that financial institutions are obliged to respect the privacy of their customers and protect the security and confidentiality of personal information that is not public. The Children's Online Privacy Protection Action (COPPA, 1998) states that an online service provider may not collect personal information from a child, unless with the consent of parents. Also, Family Educational Right to Privacy Act (FERPA, 1974) gives students the right to access and change their school records and some control over the disclosure of information from the records. Finally, the Health Insurance Portability & Accountability Act (HIPAA, 1996) defines policies, procedures and guidelines for privacy and security of individually identifiable health information.

In Brazil, recently (April, 2014), the president sanctioned, after almost four years of discussions, the *Marco Civil da Internet*, a Brazilian Internet law (MarcoCivil, 2014). This

law establishes principles, guarantees, rights and duties for the use of the Internet in Brazil, including the protection of personal information and user privacy. Internet companies which work with their user's personal information for advertising purposes cannot transfer this information to third parties without explicit and free consent of the personal information owners. Also, the contents of private communications in electronic media must have the same privacy protection that was already guaranteed in traditional media, such as letters or telephone conversations.

Another advance promoted by the Brazilian Internet Law Framework is what they called "guarantee of neutrality in the network", which means that service providers should treat all data circulating on the Internet in the same way, without distinguishing content, origin, destination or service (a provider cannot, for example, to benefit the flow of traffic from a site or a service over the other). Lastly, the law also addresses better protection of freedom of expression on the Internet. The removal of content from the Internet will be made only by judicial order, except in cases of "revenge porn".

2.3.2 Privacy Principles

Privacy Principles are the fundamental rules about how organizations should handle personal information. The Fair Information Practices (FTC, 2000) is a legislation recommended by the Federal Trade Commission, from the USA, which set forth a basic level of privacy protection for consumer-oriented commercial Web sites. It establishes basic standards of practice for the collection of information online, and provides an implementing agency with the authority to promulgate more detailed standards. Consumer-oriented commercial Web sites that collect personal identifying information from or about consumers online would be required to comply with the four fair information practices: (i) **Notice** (web sites must provide consumers clear and conspicuous notice of their information practices); (ii) **Choice** (web sites must offer consumers choices as to how their personal identifying information is used); (iii) **Access** (web sites must offer consumers reasonable access to the information which they have collected about these consumers) (iv) **Security** (web sites must take reasonable steps to protect the security of the information they collect from consumers).

Similarly, The OECD Privacy Principles (OECD, 2010) provide a privacy framework, which is reflected in existing and emerging privacy and data protection laws, and serve as the basis for the creation of leading practice privacy programs and additional

principles. The OECD (Organization for Economic Co-operation and Development) is an international economy organization of 34 countries, founded in 1961 to stimulate economic progress and world trade. Currently, the list of countries includes 21 of the 28 European Union members and countries such as Japan, Mexico, Chile, USA, Australia, Israel, and others. Brazil is not part of this group (OECD, 2015). The principles are eight: (i) **Collection Limitation** (there should be limits to the collection of personal data); (ii) **Data Quality** (personal data should be relevant to the purposes for which they are to be used); (iii) **Purpose Specification** (the purposes for which personal data are collected should be specified); (iv) **Use Limitation** (personal data should not be disclosed); (v) **Security Safeguards** (personal data should be protected by reasonable security safeguards); (vi) **Openness** (there should be a general policy of openness about developments, practices and policies with respect to personal data) (vii) **Individual Participation** (an individual should have the right to obtain confirmation of whether or not the data controller has data relating to him); (viii) **Accountability** (a data controller should be accountable for complying with measures which give effect to the principles stated).

There are other privacy principles from other organizations. The APEC (Asia-Pacific Economic Cooperation), for example, establish a document with 9 principles (Preventing Harm; Notice; Collection Limitations; Uses of Personal Information; Choice; Integrity of Personal Information; Security Safeguards; Access and Correction; Accountability) (APEC, 2005). A Working Group in the Information and Privacy Commissioner of Ontario, Canada, also defined a Global Privacy Standard, with 10 privacy principles (Consent; Accountability; Purposes; Collection Limitation; Use, Retention, and Disclosure Limitation; Accuracy; Security; Openness; Access; Compliance) (Cavoukian, 2006). Recently, in 2013 (and amended in 2014), the Australian Government established the Australian Privacy Principles, which applies to Australian and Norfolk Island government agencies and also to private sector organizations with an annual turnover of $3 million or more. It is composed of 13 principles (Open and transparent management of personal information; Anonymity and pseudonymity; Collection of solicited personal information; Dealing with unsolicited personal information; Notification of the collection of personal information; Use or disclosure of personal information; Direct marketing; Cross-border disclosure of personal information; Adoption, use or disclosure of government related identifiers; Quality of personal information; Security of personal information; Access to personal information; Correction of personal information) (APP, 2014).

2.4 CASES OF PRIVACY VIOLATION

Once technological advances are faster than privacy laws and privacy control, there are innumerable cases of privacy violation, every day. Undoubtedly, the most recent case that received considerable attention all over the world is the collection of information from the population – without their knowledge or consent – by the US government (TheGuardian, 2013). The discovery of this action has happened in 2013, through complaints made by Edward Snowden, an outsourced former employee of the US National Security Agency (NSA). The NSA was collecting the telephone records of millions of US customers of Verizon, one of America's largest telecoms providers, under a top secret court order. The White House defended the position of the NSA, stating that the data collection is "an essential tool to protect the country from terrorist threats". Even heads of state such as Dilma Roussef, the president of Brazil, and Angela Merkel, the chancellor of Germany, were victims of espionage and violation of privacy by the NSA (G1, 2013). The impact of these complaints remains until nowadays. Reports from June 2015 says that the NSA has also been spying the last three French presidents: Jacques Chirac, Nicolas Sarkozy and François Hollande (BBC, 2015).

In 2012, a case of privacy violation by Google and other giants of the Internet network was reported. These companies were able to track user's activities through cookies that were automatically recorded in the Safari browser used in iPhones, without the users' knowledge (GloboNews, 2012). Also in 2012, a group of Austrian students pressed Facebook for defining and presenting stricter privacy rules. That's because a law student in Vienna noticed that all his information that he removed from his Facebook account was not actually removed from the social network (YouTube, 2012).

Other relevant case about privacy violation was reported in 2011, where the American government investigated an application from the Carrier IQ company. This application was present in mobile phones with Android system and recorded every key pressed, as well as the content of text messages sent to and received by the mobile. The information is collected without the user knowledge, and even becoming aware of the application, it was not possible to avoid sending these data (Veja, 2011).

A curious case that also deserves to be mentioned happened in 2010. DVDs containing personally identifiable information (such as social security number, address, full name, phone, commercial activity, sex and marital status) from more than 7,6 million

Brazilian tax paying citizens – which must be kept private by the government – were being sold by street vendors in downtown São Paulo (R7, 2010).

2.5 RELATIONSHIP BETWEEN SECURITY AND PRIVACY

The notions of privacy and security frequently appear closely connected in the literature on software development but, here we interpret these two terms distinctly, albeit they are clearly related.

Both privacy and security deal with essential protections, but they vary widely in what's protected and why (Hurlburt *et al.*, 2009). According to Peltier (2001), information security encompasses the use of physical and logical data access controls to ensure the proper use of data and to prohibit unauthorized or accidental modification, destruction, disclosure, loss or access to automated or manual records and files as well as loss, damage or misuse of information assets. Privacy strives to protect an individual's sensitive information from unwarranted exposure. It is defined as a legal right.

From this viewpoint, we assume that privacy goes beyond security. Some security methods have a direct effect on privacy but might not be deployed primarily to protect it, i.e., security is not the only resource to be used in order to protect privacy, because privacy goes beyond. Nonetheless, in this context where privacy isn't the absence of personal information, but the control of it, security mechanisms (although they are not the only resource) play a very important role in privacy protection: it is possible to have poor privacy and good security practices. However, it is difficult to have good privacy practices without a sound, comprehensive data security program (Heather, 2010).

2.6 WEB APPLICATIONS AND PRIVACY: KEY PROBLEMS

Nowadays, to acquire some product or service through companies' web sites, the most common approach we face is a presentation of a privacy policy, usually before sending our personal information. The focus of privacy policies is to describe the organization's practices, including, most of the time, the collection, usage, storage and disclosure of personally identifiable information from their users and customers. The policies intend to protect the organization and to signal integrity commitment to site visitors. To guide browsing

and transaction decisions, consumers adhere (or should) to the stated website policies. They are so important that can influence the organization's credibility: if the privacy policies are clearly and explicitly stated, then the visitor/consumer perceives the organization as more trustworthy (Han and Maclaurin, 2002).

Usually, the only options users have are to agree with the whole policy and continue the purchase or disagree with the whole policy and do not acquire the referred product or service. Most of the times it is not possible for the users to express their own privacy preferences. Moreover, most of the users do not read the privacy policies: a survey performed by Microsoft in 2013 interviewed more than 1,000 users in USA and Europe Union and 76% say they skip this information or simply accept the terms and conditions without reading the details (Microsoft, 2013).

In this scenario, we identify four main problems, which are discussed in the following:

1 –Privacy policies are not always written to protect customer privacy. Even with all the already mentioned reasons to write privacy policies in accordance to the laws and to respect the right to privacy of their customers, some companies still write privacy policies that tend to appropriate their customer's personal information, i.e., they do not explicitly state that they are compiling and selling them.

Consider the example from the privacy statement on Ford Motor Company's site: "*There are instances where Ford Motor Company requests personally identifiable information to provide site visitors with a service. This information, such as name, mailing address, email address, and type of request, is collected and stored in a manner appropriate to the nature of the request, as determined by Ford Motor Company, to fulfill your needs*." (Ford, 2003). The company will collect and store personal information in a manner that it deems appropriate: "to fulfill your needs." The Ford statement is, unfortunately, typical.

In the following policy statement from the General Electric site, some words are underlined: "*When other information is collected from you, such as your name and e-mail address, we <u>generally</u> let you know at the time of collection how we will use the personal information. <u>Usually</u>, we use the personal information you provide only to respond to your inquiry or to process your request (such as to receive electronic annual reports or to be added to our supplier diversity database). This information may be shared with other GE businesses, but only if necessary to fulfill your request or for <u>related purposes.</u>"* (GE, 2003).

The underlined words create a significant ethical gap. The company *generally* (not always) informs the visitor about how it intends to use the information; the company *usually*

(not always) uses the information to provide a requested service; the company shares the information to provide a requested service or for *related purposes*. Related according to whom?

The examples above were based on the analysis of Perkins and Markel (2004) and shows that, as privacy policies are textual, written in natural language, some information can be implicit or subject to different interpretations. So, a lot of semantics is involved. This leads to the next problem.

2 - Privacy policies are difficult to be machine-readable. It is known that natural language is the more adequate manner of communicating users about the privacy policy. In our view, privacy-related policies can be organized in a hierarchy: highest-level policies are described in natural language; lowest-level policies are specified in machine-readable format, and used by the application itself to, e.g., perform access control. In principle, lower-level policies describe a refinement of higher-level policies. This is because reproducing high-level statements in machine-readable statements is a very difficult task due to the semantics involved. The lower the level, the greater is the loss of semantics.

A lot of research has focused on low-level approaches (Mont *et al.*, 2011). Such works aim at producing machine-readable specifications, which can be directly used as input for software enforcement frameworks. Most of the work in this category addresses access control as, for example, XACML (OASIS, 2013), PRBAC (Ni *et al*. 2007), or Ponder (Damianou *et al.*, 2001).

3 - Expressing user's preferences is still a limited task. As previously mentioned, most of the companies websites present the privacy policy and give the user only the options to agree or disagree with this whole policy, not allowing them to express their own privacy preferences. This leads to the possibility of the private data be accessed with purposes different from the ones intended by their owners.

Two notable contribution within this lack of preference expression are the P3P (Cranor et al., 2006) and EPAL (Ashley et al., 2003), which are technologies for specifying user privacy preferences in the web domain. P3P (Platform for Privacy Preferences Project) is a protocol that allows websites to declare, in a standard format, privacy policies with the intended use of the information they collect about users, such as what data is collected, who can access those data and for what purposes, and for how long the data will be stored. This information can be retrieved automatically and is easily interpreted. EPAL (Enterprise Privacy Authorization Language) allows enterprises to formalize their privacy promises into policies. These policies can define the categories of users and data, the actions being

performed on the data, the business purposes associated with the access requests, and obligations incurred on access.

Although both these technologies allow users to express their preferences, it is done in a general way, defining, for example, for which goal their information can be used or who can view their information. Privacy protection based only in these definitions is frequently limited or insufficient. More rules and elements are necessary to describe protection information decisions. For example, to allow users to express their preferences for each piece of personal information individually, it is necessary to define rules that address this purpose.

Another limitation of P3P and EPAL is that they do not provide any mechanism to enforce the privacy policies and the user's preferences. Privacy enforcement is a huge gap in web applications and services privacy protection and will be addressed in the following.

4 – The enforcement of privacy policies is not guaranteed. Besides the privacy policies definition, mechanisms to enforce them are necessary to make sure companies keep their privacy promises to consumers and business partners.

According to the European Union Agency for Network and Information Security (ENISA, 2014), privacy and data protection features are, on the whole, ignored by traditional engineering approaches when implementing the desired functionality. This ignorance is caused and supported by limitations of awareness and understanding of developers and data controllers as well as by the lack of tools to realize privacy design and implementations. Also, the integration of privacy requirements in the design of a system is not a simple task. First, privacy in itself is a complex, multifaceted and contextual notion. In addition, generally it is not the primary requirement of a system and it may even be in conflict with other (functional or non-functional) requirements.

2.7 THE OVERALL PRIVACY CONTEXT

Regarding the four main privacy problems in web applications and services we presented in the previous subsection, the present work remains in the scope of the enforcement of privacy policies. In this dissertation, we address the two main reasons for privacy being ignored when implementing web applications and services (limitations of awareness and understanding of stakeholders and lack of tools to realize privacy design and implementations). Referring to the limitations of awareness and understanding of privacy

domain by developers and other privacy stakeholders, we provide models, in a comprehensive approach, which contribute to ease this problem, by treating the privacy domain in a comprehensive way. Also, referring to the lack of tools to realize privacy design and implementations, we provide an extension to the UML language, which contributes with the task of modeling privacy features, by making them closer to the implementation and making easier the task of the developers.

A major part of the work presented in this dissertation is centered on topics related to the design and modeling of web applications with privacy protection features. However, we consider, in our solution, the expression of privacy preferences. The focus is not to provide a complete model to this task, but just give users a way to opt-in or opt-out related privacy policy statements, i.e., an approach that enables people to decide whether to authorize the company to use their personal information in certain ways. It would prevent the company from using that information in those ways until given explicit permission.

2.8 CONCLUDING REMARKS

As Privacy is a very abstract concept and there is no single accepted understanding of it, in this chapter we discussed the definitions which are related to our work. Also, we highlighted the value of personal information and the necessity of keep them private, supported by recent surveys which evaluates the perception of internet users about their privacy. We also presented, summarily, some important privacy laws for different countries and the most important privacy principles, whose rules describe how organizations should handle personal information. We also presented some recent cases of privacy violation with great repercussion and discussed the relationship between privacy and security. Finally, we described some key problems regarding the lack of privacy protection in the scope of web applications and services, with focus on privacy policies.

3. RELATED WORK ON PRIVACY AND WEB APPLICATIONS MODELS

To establish the background for privacy and web applications according to the scope of this dissertation, we searched for related work proposing privacy models and UML extensions in this context. The identification of related work was done through a literature review process, which is described in APPENDIX A. The main sources we used for search are the digital libraries from ACM (Association for Computing Machinery), IEEE (Institute of Electrical and Electronics Engineers) and ScienceDirect (Elsevier), because they are considered quite relevant by the computer science community. As we want relatively recent information, we selected only work from 2010 on. The review was divided into four parts, to investigate: (i) privacy approaches and reference models; (ii) privacy reference architectures; (iii) privacy UML profiles; and (iv) tools and solutions to help guaranteeing data privacy. See APPENDIX A for details.

According to our literature review, there is no other approach to systematize privacy in web applications in a comprehensive way, which provides enforcement elements and UML resources for documentation. Most of the works provide solutions for specific parts of the whole context we want to address. The work of Chakaravarthi *et al*. (2014), for example, addresses information communication. They propose a mechanism for protecting privacy in the communication between applications through the internet. They called the solution as HTTPI protocol, which satisfies the QoS (Quality of Software) requirements, such as authentication, authorization, integrity and confidentiality at various levels of the OSI (Open Systems Interconnection) model layers. The work of Ghazinour and Barker (2013) concerns access control. They propose a privacy-preserving model, called Lattice-based Privacy Aware Access Control (LPAAC) Model, which considers privacy preferences of both the data provider and the collector, facilitating the customization of privacy agreements. Ghazinour *et al*. (2014) follow this same line, presenting a generic framework to capture privacy preferences from data providers. The framework captures weights (weights describe the importance of data items for individuals if that particular private data item is exposed) from data providers, and can be considered as a mediator to quantify privacy commitment. The privacy commitment value is defined by using the notion of formal concept analysis and weighted concept lattice structure.

In an effort more focused on web services, Meziane and Benbernou (2010) propose a formal model for privacy, called Privacy Agreement, which establishes that both service customer and service provider must agree before any running process. This agreement is done through privacy policy evolution primitives and an agreement negotiation protocol. This protocol should preserve privacy-agreement and avoids conflicts between the parties when events happen during the running process. With this, the authors aim to guarantee the compatibility of privacy policies between web services and, consequently, the privacy protection of personal information.

Still in these specific contexts of privacy models for web applications, Jiang *et al*. (2012) propose a randomized response model (k-shuffle) and a statistical information recovery procedure to protect privacy of patient records, guaranteeing that a data receiver cannot reconstruct the record-to-identity mapping for each individual. Instead of using standard operations (e.g., generalization, suppression or additive noise), the k-shuffle introduces plausible deniability using a mixture of distributions, followed by a statistical information recovery procedure. Similarly, Gkoulalas-Divanis and Coupe (2011) provide a publication process model for data consortia that allow users to extract the maximum amount of information from heterogeneous databases in a privacy-aware manner. Typically, data consortia provide subscribing members of specific industry sectors (such as financial services and healthcare) with access to a wide array of information (such as financial performance records, risk data, patient health statistics). In cases where the shared information is business sensitive, data consortia can provide the needed guarantees of data privacy that enables firms to agree to disclose their data. The model is called Operational Riskdata eXchange (ORX) and is composed by a set of preprocessing steps that are applied to the data records from each source prior to publication. These steps include linking records across various databases, compressing the data descriptions by applying homogeneity and scaling analysis, and generalizing or suppressing data items that may lead to inadvertent disclosure, violating privacy.

However, few efforts for designing reference models for data privacy in a comprehensive context exist. Cherdantseva and Hilton (2013) present a Reference Model of Information Assurance & Security (IAS), which endeavors to address the recent trends in the IAS evolution. The model incorporates four dimensions: Information System Security Life Cycle, Information Taxonomy, Security Goals and Security Countermeasures. The goal is to provide the understanding and communication among stakeholders through informal visual

representation. Although security is strictly related to privacy, the focus of the paper is in data security and considers few privacy aspects.

The work of Sathiyamurthy (2011) defined a conceptual model named “holistic privacy archetype”, which provides a pragmatic approach for business to manage and stay abreast of growing regulatory and fiduciary requirements. The model is divided into three main layers (business process layer, strategy and governance layer, and operational layer) and was applied to a financial business model to describe its capabilities. However, due to being more enterprise-focused (business-processes oriented), the model neglects more specific characteristics of web services and applications as, for example, privacy policies definition and management.

The Privacy Management Reference Model and Methodology (OASIS, 2012) is an OASIS (Organization for the Advancement of Structured Information Standards) specification that provides a conceptual model and a methodology for understanding and analyzing privacy policies and their privacy management requirements. It allows selecting technical services that must be implemented to support privacy controls. The model is based on a non-normative working set of operational privacy definitions and the privacy requirements are defined through use cases. Although this is a recent privacy reference model, it considers only intrinsic characteristics (core) of privacy, i.e., it does not directly incorporate privacy requirements related to different sources of privacy violation, in a broader privacy context. Also, it is generic and do not specify resources for enforcing privacy policies.

3.1 REFERENCE ARCHITECTURES AND SOFTWARE ARCHITECTURES

Software architecture constitutes the backbone of any successful software system. In practice, decisions made at the architectural level directly enable, facilitate, or interfere with the achievement of business goals as well as functional and quality requirements. By software architecture, we mean “*the structure of the components of a program/system, their interrelationships, and principles and guidelines governing their design and evolution over time*.” (Garlan and Perry, 1995).

A reference architecture refers to a special type of software architecture that captures the essence of the architectures of a set of software systems of a given domain, i.e., they have emerged as abstractions of concrete architectures. The purpose of a reference architecture is to serve as guidance for the development, standardization, and evolution of

systems in that domain, as well as to guarantee the interoperability between systems and between components of systems (Nakagawa *et al*., 2012; Muller, 2008).

Nowadays, the increasing complexity of software, the need for efficient and effective software design processes and for high levels of system interoperability has led to an increasing role of reference architectures in the software design process (Angelov *et al.*, 2009). According to Angelov *et al.* (2009), a fully accepted definition for software reference architectures does not exist. In this dissertation, we use the definition provided by Bass *et al.* (2003), which states that a reference model is "*a division of functionality together with data flow between the pieces*", and a reference architecture is "*a reference model mapped onto software elements (that cooperatively implement the functionality defined in the reference model) and the data flows between them*".

According to the goals and scope of this dissertation, we want to provide a privacy reference architecture that can serve as a foundation for the analysis, design and development of web applications with privacy concerns. Using the reference architecture, these applications can manage personal information in a more secure manner, protecting such information from different sources of privacy violation. The review process (APPENDIX A) was conducted in order to find work related to reference architectures in the privacy domain.

3.1.1 Abstract Architectures

Abstract architectures are generic architectures for classes of information systems or particular domains. They are used as a foundation for the design of concrete architectures from these classes and domains.

Works regarding abstract architectures cover specific parts of the privacy and web application context. The work by Sangani and Vithani (2012), for example, is focused on web applications hosted in a cloud and the security challenges in this environment. They propose an architecture that can be used by cloud providers, cloud security providers and consumers, which are small and medium enterprises. As the authors state, small and medium enterprises do not have the knowledge to protect such applications due to the lack of security technical expertise or financial budgets. They identified six security components that need to be plugged during the deployment of the web applications in a cloud. These components are meant to assist these enterprises to mitigate the disruption in business caused by hackers and provide the knowledge to understand such key security features. Heitmann *et al*. (2010)

focused on technologies from the emerging Web of Data (Friend-Of-a-Friend, WebIDs and the Web Access Control vocabulary). They present an architecture that describes how to combine existing infrastructure of the Web of Data and existing standards for decentralized identity management in order to achieve privacy-enabled user profile portability. User profiles and activity stream data can then be securely shared with any third party that supports the architecture.

Still in a specific context, Barcellona *et al.* (2014) are more focused on communication. They provide a solution that performs users' profiling and keeps sensitive information private. They used an iterative clustering algorithm for data mining called Fuzzy C-Means in a privacy preserving way. They used techniques drawn from the subfield of cryptography known as Secure Multi-Party Computation (SMPC). The SMPC main idea is to perform a computation among different parties, where each party contributes with an input value that remains private, even if the final output is public. In the architecture, the users are organized in ring structures to communicate with services providers. Osawa *et al.* (2010) focus on the web services exchange information, between users, services providers and identity providers. In their architecture, service providers require private information of users in order to provide their services. The identity provider has the function to manage users' private information. The application server of the identity provider interacts with user clients. The context handler of the identity provider negotiates to disclose private information considering the service providers' disclosure requirement and the user's preferences. A privacy management policy is defined in terms of logic and requests are sent to users to maintain their preference of information disclosure also in the form of logic.

To the best of our knowledge, the three main contributions regarding privacy reference architectures are the standards ISO/IEC 29100 (ISO, 2011), ISO/IEC 29101 (ISO, 2013) and the work of Shin *et al.* (2011). The ISO/IEC 29100 is a privacy framework that allows defining privacy safeguarding requirements as they relate to personally identifiable information (PII) processed by any information and communication system in any jurisdiction. It is applicable at an international scale and sets a common privacy terminology, defines privacy principles when processing PII, categorizes privacy features and relates all described privacy aspects to existing security guidelines. However, this information is not organized and provided at an architectural level. The ISO/IEC 29101 (ISO, 2013) describes best practices for the technical implementation of privacy requirements. The standard covers the various stages of data life cycle management and the required privacy functionalities for protecting data, as well as the definition of the roles and responsibilities of all the involved

parties. Similarly, Shin *et al.* (2011) present a privacy reference architecture as a security model for the management of personal information in its lifecycle. They divide the lifecycle of personal information into four stages and introduce the steps of the personal information processing performed at each stage. The architecture is based on the three types of actors involved in PII processing: principal, controller and processing. The authors state that the architecture is generic and should be implemented while considering the specific properties of Information and Communication Technology (ICT) systems. However, to cope with this requirement, significantly more users and roles should be considered.

Although the goal is to integrate privacy considerations into the technical design, the two works mentioned above (ISO, 2013; Shin *et al*., 2011) contextualize the privacy concerns only through the personal information life cycle. They do not consider the different sources of potential privacy violation of the system, which can be critical if certain features or functionalities are not considered.

3.1.2 Concrete Architectures

Concrete software architectures represent structures of a software system, comprising software elements, relations among them, and properties of both elements and relations. They are designed on the basis of required functionalities and system, business, and architecture qualities defined by the stakeholders (Bass, 2003). These functionalities and qualities reflect a specific context and the business goals of the stakeholders. Concrete architectures can be derived from abstract architectures.

There are two important concrete software architectures worth mentioning here: the IBM Tivoli Privacy Manager (Bücker *et al.*, 2003) and the HP Privacy-Aware Access Control architecture (Mont *et al.*, 2005). IBM Tivoli Privacy Manager (Bücker *et al.*, 2003) provides an enterprise-wide system that enables a company to use the personally identifiable information (PII) it collects according to the principles of Fair Information Practices (FTC, 2000) and to monitor and enforce its compliance with those principles. The goal is to help privacy and security officers as well as their staff to understand and implement the referred architecture in an enterprise environment. The work covers the design of the Enterprise Privacy Architecture (EPA) and considers the impact of privacy issues on enterprise policy, standards, and procedures. EPA shows the technical component architecture details and also describes the actors and data involved in respect to privacy.

More focused on access control context, Mont *et al.* (2005) propose a concrete software architecture to help the enforcement of privacy policies for personal data stored by enterprises. The goal is to demonstrate how privacy policies, dictating constraints and conditions on personal data, can be integrated with enterprise access control policies by leveraging a common authoring, deployment and enforcement framework. The authors describe a privacy enforcement model and a technical approach to model personal data, author privacy policies and customers' consent, to deploy and to enforce them in an integrated framework. The management of access control policies is integrated with the management of privacy policies and this process is made by the Policy Builder component, i.e., a graphical tool to author and manage access control policies on resources at different levels of granularity.

These both aforementioned architectures (Bücker *et al*., 2003; Mont *et al.*, 2005) were used in the evaluation process of the Privacy Reference Architecture we proposed as part of this dissertation. This evaluation process is described in Section 4.2.3.

3.2 UML PROFILES

The Unified Modeling Language (UML) is a resource in the development of modern software systems. Historically, UML was born as the result of the joint effort of important personalities in object-oriented software development, who were working together in the 90's with the aim to fuse their leading design methods (Grady Booch, Object-Oriented Software Engineering – OOSE – James Rumbaugh, Object-Modeling Technique – OMT – Ivar Jacobson) in a single standard language. The initial version of UML (1.0) was proposed to the Object Management Group (OMG) in 1996, and was officially adopted as a standard in November 1997. The most recent version released by OMG is UML 2.4.1 (OMG, 2011).

The objective of UML is to provide system architects, software engineers, and software developers with tools for analysis, design, and implementation of software-based systems and for modeling business and similar processes. UML consists of thirteen diagram types, each one addressing a different aspect of a system or providing a different way for organizing system concepts. They can be classified as structural diagrams (which are used to model the structure of the system) and behavioral diagram (which focus on modeling the behavior of the system and its components).

UML's rich modeling capabilities result in a language with a very broad scope that covers a large and diverse set of application domains. Therefore, not all of its modeling capabilities are necessarily useful in all domains or applications. For the same reason, it may be difficult to specify precise concepts belonging to the domain of interest in a convenient way. To a certain extent, this limitation can be in part overcome through the profiling mechanism, i.e., the extension mechanism provided by the UML standard.

A UML profile is an extension of the UML metamodel containing specializations for a specific domain, platform, or purpose. Profiles are defined using stereotypes, attributes, and constraints. Stereotypes are the main construct in a profile and help identifying elements of interest in a model. A stereotype is an extension of an existing UML metaclass, possibly defining a set of additional attributes (i.e., properties). When a stereotype is applied to an instance of a UML metaclass, values can be specified for its attributes. Finally, a UML profile may define additional constraints, i.e., statements that need to be satisfied for the model to be well-formed according to the profile. However, it should be noted that the profiling mechanism does not allow to directly modify the existing UML metamodel, and it is not possible to take away any of the existing constraints: the source metamodel is considered as "read-only", and profiles can only extend it.

There are several reasons for extending the existing UML metamodel, e.g., providing a specific terminology for a certain domain; providing a different notation for existing elements; adding constraints on the usage of the metamodel; adding information that can be used for model-transformation or code generation purposes. Currently, one of the most researched directions for UML profiles is in improving UML ability to describe non-functional information. The Object Management Group itself has published as OMG standards several UML profiles related to non-functional system properties, e.g., the SPT (OMG, 2005), QoS&FT (OMG, 2008), and MARTE (OMG, 2011-b) profiles.

According to our literature review (APPENDIX A) and to the best of our knowledge, no UML Profile for privacy domain has been defined. Most of related work defines UML Profiles for web applications and services. However, they address specificities of web services context. Scheithauer and Wirtz (2010) propose a metamodel for business-oriented service descriptions and develop a corresponding modeling notation on the basis of a UML Profile, which supports documenting, communication, and reasoning about descriptions at a strategic level. The resulting document is an input for service descriptions at a conceptual level. Two possible fields of application have been outlined: service-oriented modeling and service engineering. Other work related to web services is that by Li *et al.* (2013). The authors

present an approach to integrate a formal method, the refinement for Component and Object Systems (rCOS), into UML. A UML profile was created to represent the main concepts of rCOS in UML and support the development methodology of rCOS. The focus is on how the full development process can be supported in an incremental and interactive manner by applying the rCOS refinement rules, and how the object-oriented and component-based techniques can be seamlessly combined and used in the development process.

In a more general context of web applications, the work by Mubin and Jantan (2014) proposes a web application design model which is based on the UML 2.0 profile and stereotypes. The Profile provides user interaction diagram to capture users' information from use case diagram and specific UML modeling elements to model conceptual, navigational and user interface features of complex web applications. By complex web applications they mean applications that present interrelated processes, dependent links and time dependent processes. Also, the work by Domínguez *et al.* (2013) defines a UML profile for statecharts that specifies system behavior. The idea is to automatically generate a stereotyped UML class diagram containing information for tracing the system behavior without losing the statechart dynamic semantics. According to the authors, their solution has been transferred technologically to a company which has implemented it in a commercial product that manages biobank data. The profile is used to help register a system trace in a complete way, easing enhanced auditing processes and trace processing for process improvement.

Most of the related work we found are the ones which proposes UML profiles for the security domain, even though they are far from representing the privacy domain (Cirit and Buzluca, 2009; Jürjens, 2002). Cirit and Buzluca (2009) propose a UML Profile for Role-Based Access Control (RBAC). It provides access control specifications that can be modeled graphically from the beginning of the design phase, making it possible to extend security integration over the entire development process. The profile is based on the four model components of the RBAC standard and additional RBAC constraints, to represent: (i) RBAC Core Components; (ii) Hierarchical RBAC; and (iii) Constrained RBAC. Although this work represents part of security (and privacy) concerns, it is mainly limited to describe access control policies, and possibly for deriving code to enforce such rules.

The work of Jürjens (2002) is the closest to the one we propose. It presents a UML extension (profile) that allows expressing security relevant information within UML diagrams. The goal is to assist in the difficult task of developing security-critical systems in an approach based on the notation of the Unified Modeling Language. The profile encapsulates the knowledge of recurring security requirements of distributed object-oriented

systems, such as secrecy, fair exchange, and secure communication link. It is useful for expressing security related information and for doing security evaluations. The defined security requirements are high level and general. However, although security and privacy are strictly related concepts, this work addresses only security concerns. We propose to extend UML also for addressing privacy concerns (in Section 2.5 we have discussed the relationship between security and privacy).

3.3 PRIVACY PROTECTION TOOLS

Although the focus of this work is on abstract models and the systematization of privacy protection in web applications and services, we consider it is important to investigate tools that can help in this protection. The goal is to identify tools that could be employed to contemplate privacy requirements defined by the model. Obviously, there are lots of tools available in the literature and we restricted our research according to the keywords and search strings described in the literature review (APPENDIX A).

Meziane *et al.* (2010) present a PaM (Privacy Agreement Monitoring) system, a tool for controlling the private data usage flow dynamically in the area of web services. PaM allows to make analysis, diagnosis and provides reasoning services on violations; for instance, why violations happen, what improvement in the agreement makes the compliance of the agreement happen, etc. The tool uses the concept of system data flow views, expressed through state machines. These state machines represent all the operations that involve private data from the initial state (activation of the privacy agreement) up to the final state (end of the privacy agreement). In this same line of privacy violation analysis, Gao *et al.* (2010) present a collaborative method which identifies web services that disclose user's private data. This identification is done through a protocol based on secure computing multipart, which identifies e-mail addresses disclosed to third parties without the consent of the e-mail owners and used as spams. It is important to mention that both works (Meziane *et al.*, 2010, Gao *et al.*, 2010) are used only for analysis, i.e., they do not interfere with the system to make privacy to be enforced when the privacy agreement is violated.

Tbahriti *et al.* (2011) present a framework, called Meerkat, for privacy management in web services interactions. The framework implements a protocol to evaluate the compatibility between clients and services provider's privacy policies, as well as a negotiation model to conciliate them in case of incompatibility. The solution considers an

important characteristic of web services, which is the dynamic information exchange between them. It does not allow that information to be exchanged between client and service if they present incompatible privacy policies (otherwise, the client's privacy would be easily violated). However, the enforcement of the privacy promises still is needed.

An approach to enforce privacy protection in web applications and services is proposed by Hewett and Kijsanayothin (2009). The goal is to establish a minimum service composition and to guarantee that customers' privacy preferences will be fulfilled. It is composed of three stages, where the first one looks for the minimum service composition, the second one removes unnecessary services and the third one verifies the data conformity with the customer's privacy preferences. This verification is done based on provided information, e.g., the security social number can be used in a government office, but cannot be used by a credit card company.

In a closer security context, Kim *et al.* (2010) propose an extension to RBAC (Role-Based Access Control), whose goal is to restrict the access to users, based on role definitions. The solution manages the system security policies and includes modules to prevent intrusions, which is done through monitoring, auditing and alerting processes. It also includes modules to compare the database access queries with authorization policies, in order to manage private data and generate hierarchical structure with semantic web information.

The last work (Kim *et al.*, 2010), raises an important observation. As we have already stated, security and privacy are approaches strictly related. However, other approaches (or sub approaches) can also be related to privacy, even helping in privacy protection, as, for example, access control, anonymization, identity management, activity tracking. Although we do not intend to include all of these approaches in our literature review, we think it is important to have a very brief background and to cite relevant work in these areas. Given that our proposed solution, PrivAPP, is comprehensive enough, these works are also an indication of possible applications that contemplate some privacy requirements defined through the PrivAPP's models. They are described in the following paragraphs.

Anonymization consists in techniques that can be applied to prohibit the recovery of individual information. For example, do not allow that the result of a statistical query to be shown when the number of records retrieved falls below some threshold. Also, to enter deliberately small inaccuracies or "noise" in the results of statistical queries make the deduction of individual information difficult (Elmasri and Navathe, 2011). A well-known technique for anonymization is called k-anonymity, which consists of a protection model

proposed by Samarati and Sweeney (1998) where an algorithm is applied in information releases to generalize and/or suppress part of the data to be disclosed.

Regarding attacks by malicious users in web applications and services, which can also violate privacy, several efforts aim to contribute to this scenario, identifying vulnerabilities or attacks that exploit them. Examples are the Sign-WS tool (Antunes and Vieira, 2011), which is based on attack signatures and interface monitoring for detection of injection vulnerabilities, and the J-Attack (Fernandes *et al.*, 2011), a tool to perform attacks looking for XSS, SQL Injection and CSRF vulnerabilities.

Activity tracking consists of identifying the user's activities in the Web and build his or her profile, many times without his or her consent. To avoid being tracked, users can perform some web browsers security configurations or use tools as the one called TrackingTracker, developed by Roesner *et al*. (2012). This tool detects and classifies trackers' activities automatically, alerting users about these trackers.

For access control, many well-known technologies can be used, as the already mentioned RBAC (Role-Based Access Control) (Sandhu, 1998) or XACML (eXtensible Access Control Markup Language) (OASIS, 2013), which permits to create and enforce access control policies. The same happens for cryptography, with the famous PGP (Pretty Good Privacy) (PGP, 1999), which is a public key encryption program based on the Rivest-Shamir-Adleman (RSA) algorithm.

Solutions for identity management can also be useful for protecting privacy. A widespread solution is Shibboleth (Shibboleth, 2014), whose emphasis is on the privacy of user attributes, based on privacy policies and the user's personal preferences. Finally, regarding auditing process, Biswas and Niemi (2011) propose a solution to streamline the log generation process by deriving the auditing specifications directly from the policies to be audited.

3.4 CONCLUDING REMARKS

In this chapter we presented the related work on privacy and web applications models. We start presenting some privacy reference models available in the literature. Then we present some abstract architecture, which cover specific parts of the privacy in the web application context. The two most important concrete software architectures for privacy protection, from large multinational companies were presented. Although no UML Profile for

privacy domain has been defined yet, we presented some UML Profiles for web applications and services, especially a UML Profile for developing security-critical systems. Finally, we describe some tools that could be employed to contemplate privacy requirements in the scope of web applications and services.

4 . THE PROPOSED APPROACH: PRIVAPP

The approach we propose systematizes the privacy concepts within the scope of web applications. The greatest contribution of the approach is a set of reference documents that aims to provide a better understanding of the privacy domain and, consequently, to facilitate research, modeling and development of privacy-aware applications. The approach can be applied in the whole software lifecycle, supporting the different tasks. However, it was performed more focused to be used on the requirement analysis. If the requirements are defined considering privacy protection, the other phases will take privacy protection into consideration too. Figure 4-1 outlines the PRIVAPP, which is composed of a *Privacy Conceptual Model*, a *Reference Architecture* and a *UML Profile*.

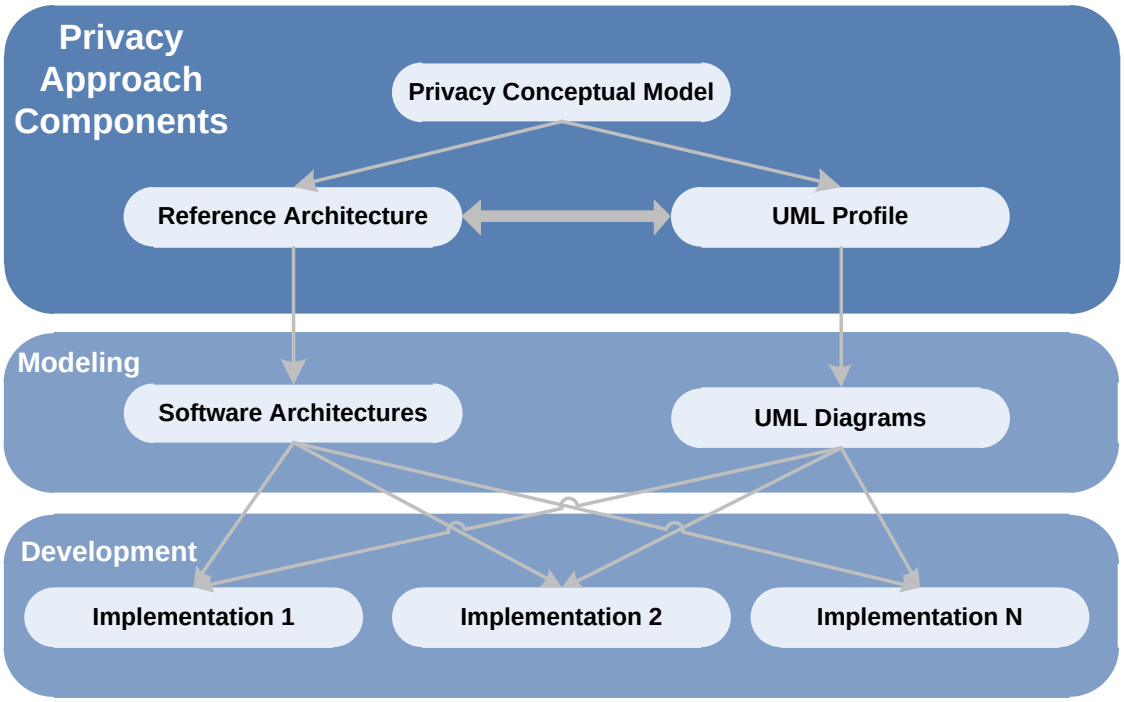

Figure 4-1. The proposed privacy approach and its application.

In Figure 4-1, the **privacy conceptual model** is composed of elements that represent privacy concepts and their relationships, in an organized way. The goal is to specify how applications should handle privacy. In practice, the model represents the privacy policies and their statements, as well as the related services and the resources to be used for enforcing

these statements. It includes the use of privacy preferences, through which users can agree or not with the policy statements.

Based on the conceptual model we defined a **reference architecture**, which describes the features and functionalities that must be addressed during the development to protect the privacy of the users. The reference architecture makes possible to design concrete architecture models which can help in providing a better understanding of the privacy domain and, consequently, facilitating the development of privacy-aware technology.

Also based on the conceptual model we created the **UML profile**, which allows extending the UML language to incorporate privacy concepts. The profile is useful to describe the privacy policies and how they are enforced, taking into consideration user's preferences. The description is done through UML diagrams that support the development process of privacy-aware applications and services.

As the approach provides reference documents that can be used to support different tasks, there is not an established step by step to apply it. However, a suggestion of use is given in section 5.2. Briefly, PrivAPP was applied following the steps: (i) understand the privacy policy statements; (ii) define the resources that can be used to enforce these statements (based on the privacy reference architecture); (iii) define UML models (based on the UML Profile); (iv) define a software architecture (based on the UML models); (v) implement the privacy-aware software or component.

The components of the approach are detailed next.

4.1 PRIVACY CONCEPTUAL MODEL

The Privacy Conceptual Model is a model of the domain concepts required for modeling views of the system where privacy management and protection are applied. The model is within the scope of web applications and, obviously, focused on the concepts in this domain. It is based on the current characteristics of these applications as well as on our extensive study about related work, privacy laws and principles. It comprises elements as privacy policy and its statements, as well as users and their privacy preferences concerning their personal information (they can agree or not with the policies statements). It comprises resources that can be used to enforce the privacy policy statements, taking into consideration users' preferences. These features allow users to make more thoughtful online choices on the use of their personal information and help to guarantee the protection of their privacy. Privacy

elements and their relationships are organized in a conceptual model, presented in Figure 4-2 and the description of the model follows. A summary of this description is given in Table 4-1.

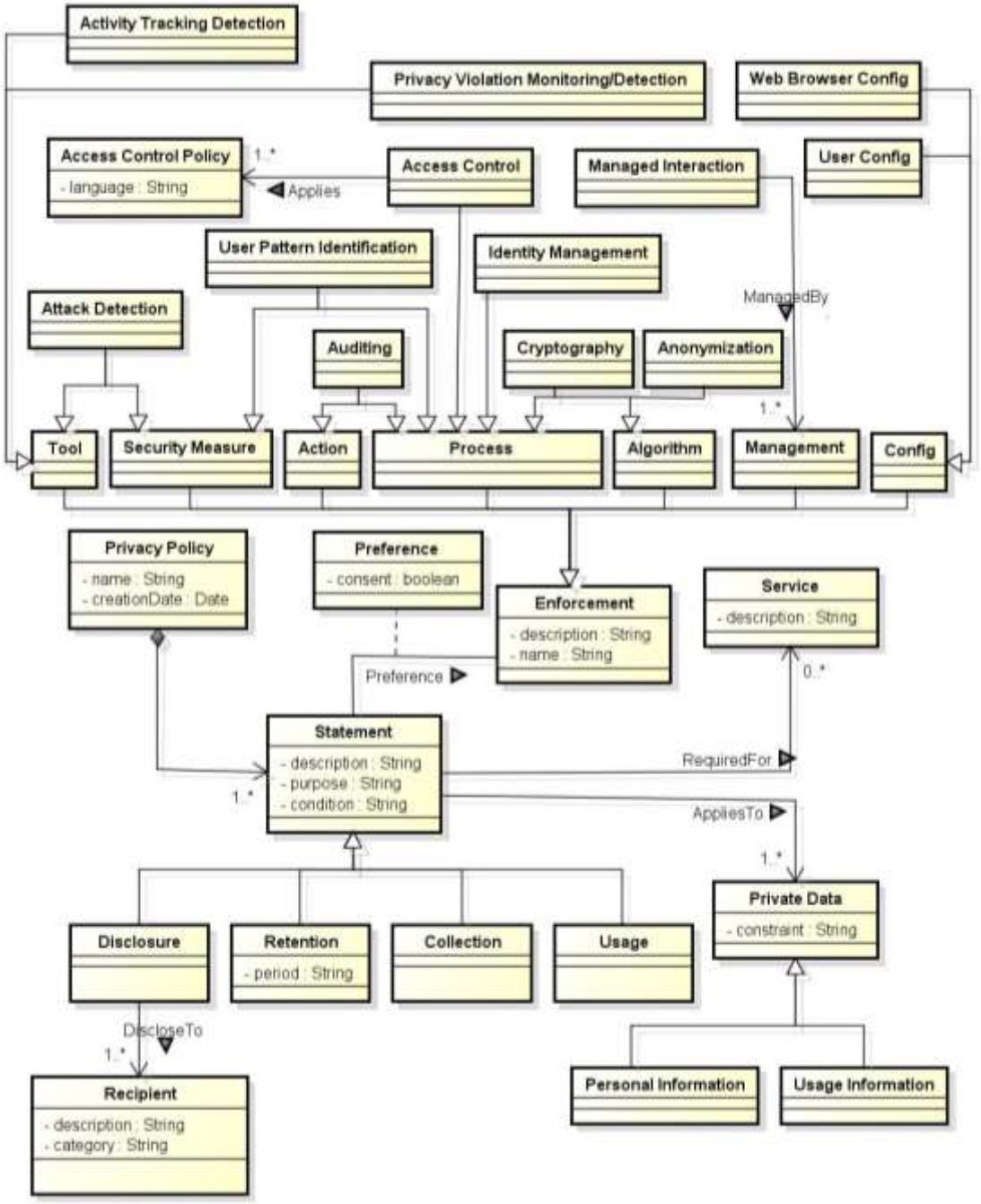

Figure 4-2. The Privacy Conceptual Model.

In Figure 4-2, the *Privacy Policy* element represents the artifact that must be defined and presented to the user. A *Privacy Policy* element can be defined by means of its attributes: *name* (name of the policy) and *creationDate* (the date the policy was created).

The *Privacy Policy* element is composed of one or more *Statements*. A *Statement* represents the description of one of the rules that are specified in the privacy policy. The attributes that identify a statement are: *description* (description of the rule), *purpose* (the purpose for which the data is collected or managed, e.g. research and development, or contacting visitors for marketing of services or products); *condition* (e.g.: "before collecting, using or disclosing personal information from a child, an operator must obtain verifiable parental consent from the child's parent").

In addition to the generic *Statement*, there are four specialized types: *Disclosure* (specifies which data will be disclosed and to whom), *Retention* (specifies the period the data will be retained), *Collection* (specifies which information, i.e., which private data will be collected) and *Usage* (specifies how the private data will be used). Based on the statements, users inform their privacy preferences.

Disclosure is related to *Recipient*. *Recipient* represents who will access the data to be disclosed. Its attributes are *description* (a textual description) and *category* (used to classify the recipient according to a given taxonomy, e.g., internal or external groups, individual or organization, etc.).

Still in Figure 4-2, *Service* represents a service offered by the company, i.e., a service a person can use if he/she provides his/her private information to the company' applications. A *Service* attribute is the *description* (description of the actions and results provided by the service). There is an association between a *Statement st* and a *Service sv* if the utilization of *sv* is dependent on acceptance of *st* by the user.

Private Data, also related to *Statement*, represent data to be collected and managed by the application according to the privacy policies statements. Its *constraint* attribute can be used to narrow down the kind of private data it represents; this feature is useful for modeling statements that apply only to data having specific characteristics (e.g., "only *anonymous* data is collected"). *Private Data* can be of two types: *Personal Information* (information the user provides to the system) and *Usage Information* (data the system collects, e.g. links accessed, user's actual location, search strings, etc.). The association between a *Statement* and a *Private Data* keeps track of private data on which each statement is applied.

Besides *Statement*, another key element of the privacy conceptual model is *Enforcement*. This element represents the resources that can be used to specify how to enforce the privacy policy statements, considering the data subjects' preferences. The attributes of *Enforcement* are *description* and *name*, which represent, respectively, the identification, the

description and the name of the resource to be used. *Statements* can be associated with the resources that are adopted for their enforcement (*Enforcement* elements). The association is performed through the *Preference* relationship; such a relation has an attribute, *consent*, which is used to specify the user's preference (*true* or *false*, meaning the user's consent or not to the statement) that resource relates to. A statement may be associated to one or more *Enforcement* elements.

Continuing in Figure 4-2, *Enforcement* can be represented as: *Tool* (e.g., tracking activities tool, intrusion detection tool); *Security Measure* (e.g., security packages updates, use of antiviruses and firewalls); *Action* (e.g., allow access, deny access, anonymize data, remove from storage devices, logging actions, encrypt data); *Algorithm* (e.g., k-anonymity – for anonymizing data, RSA – for encrypting data); *Process* (e.g., identity management, access control, auditing); *Management* (management of privacy policies); and *Config* (e.g., web browser security configurations, changes in default configurations). We specified *Enforcements* with elements we consider very relevant to the purpose (obviously, the model is highly representative but not exhaustive. Yet it is extensible enough to include more *Enforcement* elements as necessary).

Activity Tracking Detection is a tool that verifies if a system's user has his/her activities tracked. *Privacy Violation Monitoring/Detection* verifies if the user's privacy has been violated. *Attack Detection* verifies if the system suffers an attack and, as it is related to the security of the system, it can also be considered a *Security Measure*.

User Pattern Identification is a process that analyzes users' stored behaviors and uses them as a security resource against fake users. Usually it consists of observing and collecting data over time periods and then applying analysis methods to identify deviate user patterns.

Auditing refers to auditing resources that web application must use to monitor and identify possible privacy violation sources. These resources should monitor all the system elements, such as databases, servers, application, services calls, etc. This can be done automatically, through processes, or with the support of auditing actions.

Identity Management is a set of processes and technologies to manage and protect against unauthorized access. *Access Control* is a process with a set of rules by which users are authenticated and by which the access to applications and other information services is granted or denied. *Access Control Policy* represents the document that specifies roles and the information each role can access. The *language* attribute refers to the language used to define the access control policy (e.g., XML).

Cryptography represents the process used to cypher information and to avoid unauthorized access. It can be done by using algorithms such as, for example, RSA. *Anonymization* represents the process used to avoid disclosure of confidential stored information retrieved even by means of data analysis. k-anonymity is a representative algorithm which supports this process.

Management refers to the management of privacy policies in a ubiquitous environment, where data are transferred to different third-parties components or services. When transfer happens it is necessary to assure that policies are compatible so that they do not violate the main privacy policy (the privacy policy of the main application, i.e., the one the user has agreed to with the statements). It is necessary to verify this compatibility and, if there is no compatibility, measures must be taken (for example, adaptations in the policies). *Managed Interaction* represents interfaces between parts of the system using different privacy policies. As such interfaces may involve violation of privacy policies, they should be correctly managed. Thus, a managed interface has a relation with a *Management* element which is in charge of managing/protecting the communication, by verifying the policies of entities interacting with the system.

Web Browser Config represents configurations outside the system (i.e., users must configure their own web browser to protect their privacy, especially when they do not want to be tracked). *User Config* represents the configurations users can define in the system itself or in a web page to refuse services such as, for example, advertisements or cookies. Both elements (*Web Browser Config* and *User Config*) were added to the conceptual model as a result of the evaluation of the approach (Section 6).

Table 4-1 presents a summary of the elements described above, in alphabetical order.

Table 4-1. Summary of the Conceptual Elements descriptions.

Element	Description
Access Control	Process with a set of rules by which users are authenticated and information services is granted or denied.
Access Control Policy	Document that specifies roles and the information each role can access.
Action	Action to enforce statements of the privacy policy (e.g., allow access, deny access, anonymize data, remove from storage devices, logging actions, encrypt data).
Activity Tracking Detection	Tool that verifies if a system's user has his/her activities tracked.
Algorithm	Algorithm to enforce statements of the privacy policy (e.g., k-anonymity – for anonymizing data, RSA – for encrypting data).
Anonymization	Process used to avoid disclosure of confidential stored information retrieved even by means of data analysis.
Attack Detection	Tool that verifies if the system suffers an attack.
Auditing	Auditing resources that web application must use to monitor and identify possible privacy violation sources.
Collection	Statement that specifies which information, i.e., which private data will be collected.
Config	Configurations to enforce statements of the privacy policy (e.g., web browser security configurations, changes in default configurations).
Cryptography	Process used to cypher information and to avoid unauthorized access.
Disclosure	Statement that specifies which data will be disclosed and to whom.
Enforcement	Resources that can be used to specify how to enforce the privacy policy statements, considering the data subjects' preferences.
Identity Management	Set of processes and technologies to manage and protect against unauthorized access.
Managed Interaction	Interfaces between parts of the system using different privacy policies
Management	management of privacy policies in a ubiquitous environment, where data are transferred to different third-parties components or services (policies must be compatible)
Personal Information	Information the user provides to the system.
Preference	Specify the user's preference (consent or not to the statement).
Privacy Policy	Artifact that must be defined and presented to the user.
Privacy Violation Monitoring/Detection	Tool that verifies if the user's privacy has been violated.
Private Data	Data to be collected and managed by the application according to the privacy policies statements.
Process	Process to enforce statements of the privacy policy (e.g., identity management, access control, auditing);
Recipient	Who will access the data to be disclosed.
Retention	Statement that specifies the period the data will be retained.
Security Measure	Security measure to enforce statements of the privacy policy (e.g., security packages updates, use of antiviruses and firewalls).
Service	A service offered by the company.
Statement	Description of one of the rules that are specified in the privacy policy.
Tool	Tool to enforce statements of the privacy policy (e.g., tracking activities tool, intrusion detection tool).
Usage	Statement that specifies how the private data will be used.
Usage Information	Data the system collects (e.g. links accessed, user's location, search strings, etc.).
User Config	Configurations users can define in the system itself or in a web page to refuse services such as, for example, advertisements or cookies.
User Pattern Identification	Process that analyzes users' stored behaviors and uses them as a security resource against fake users.
Web Browser Config	Configurations outside the system (i.e., users must configure their own web browser to protect their privacy, especially when they do not want to be tracked).

As the main elements of the conceptual model were also conceived from the privacy principles (section 2.3.2), we describe in Table 4-2 some of these relations. It is important to mention that different countries or group of countries have different privacy principles (although some principles are similar). So, we will use, just for illustration, the two we found most relevant in the literature: The Fair Information Practices (FTC, 2000) and The OECD Privacy Principles (OECD, 2010).

Table 4-2. Relation between privacy principles and elements from the conceptual model.

Privacy Principle		**Conceptual Model**
Fair Information Practices	**OECD Privacy Principles**	**Element**
Notice	Openness	Privacy Policy
Choice	Individual Participation	Preference
Security	Security Safeguards	Security Measures (Enforcement)
	Use Limitation	Enforcement
	Purpose Specification	Statement (purpose)
	Accountability	Management (Enforcement)
	Collection Limitation	Collection

4.2 THE PRIVACY REFERENCE ARCHITECTURE

The Privacy Reference Architecture (PRA) is based on the Privacy Conceptual Model, i.e., the elements of the conceptual model are distributed through the layers where they can be implemented. Elements are presented in a higher level of abstraction; described features and functionalities must be addressed during the development of web applications to protect the privacy of the users' information. From PRA concrete architecture models can be derived to help in a better understanding of the privacy domain and, consequently, to facilitate the development of privacy-aware technology.

PRA was built using ProSA-RA (Process based on Software Architecture - Reference Architecture) (Nakagawa *et al.*, 2014), a systematic and iterative process for specification, design and evaluation of reference architectures. This process is based on 4 stages: (i) investigation and selection of the sources of information to be used; (ii) establishment of privacy architectural requirements; (iii) design of the reference architecture; (iv) evaluation of the constructed architecture. The next subsections describe the stages in the specification of the privacy reference architecture.

4.2.1 Architectural Requirements

According to the ProSA-RA method (Nakagawa *et al.*, 2014), before establishing the architectural requirements it is necessary to investigate the sources of information to be used. The privacy architectural requirements were established based on the domain information represented in the conceptual model, which, as already mentioned, was conceived through our extensive study about related work, privacy laws and principles. These multiple sources of information have been considered, namely: (i) reference architectures for privacy available in the literature (e.g. Shin *et al.*, 2011); (ii) legislation, standards and norms for developing applications that protect privacy information (e.g. ISO, 2013); (iii) solutions, frameworks and tools for privacy information protection (e.g. Cranor *et al*., 2006; Ashley *et al.*, 2003); and (iv) privacy violation taxonomies (e.g. Antón and Earp, 2004; Solove, 2006). These sources were selected as they present, in a broad context, current privacy problems and possible resources to protect information against those problems. They were investigated through literature reviews, where search strings and exclusion criteria were defined and applied, resulting in a set of consistent related works (APPENDIX A). From this study, we identified 12 main requirements (named as PAR-(number)), by which the reference architecture must:

PAR-1. Permit the development of applications that ensure the privacy of data during its collection, management and storage.

PAR-2. Allow the use of resources to protect the users' activities against tracking.

PAR-3. Permit the use of resources to protect personal information against attacks to web applications (given that certain types of attacks, when successful, can access personal private information and, consequently, violate the users' privacy).

PAR-4. Support the use of resources to protect personal information against security violation at different application layers. Also, it must provide guidelines regarding security configurations that support this requirement.

PAR-5. Permit the use of signatures and digital certificates because they provide higher levels of data privacy and security in electronic transactions, allowing unambiguous identification of the parties involved and the integrity and confidentiality of data.

PAR-6. Support the use of cryptography to protect the information during network traffic against non-authorized visualization or modification.

PAR-7. Permit the use of information anonymization techniques to prevent recovery of personal information, especially when dealing with statistical databases. The access to a statistical database should not enable one to learn, even through inferences, anything about an individual that should not be learned.

PAR-8. Provide resources for the definition, enforcement and management of privacy policies in web applications, with special attention to information in transit between different web applications and services.

PAR-9. Provide resources to create and maintain digital identities, especially in the context of collaborative networks.

PAR-10. Provide access control resources that, based on predefined rules, permit or deny access by users to applications and other services information.

PAR-11. Provide resources that allow auditing web applications to perform evaluative analysis of the data privacy or sources of privacy violation.

PAR-12. Enable the owners of the information to express their privacy preferences.

The requirements (and, consequently, the architecture) proposed here are generic and should be instantiated considering the specific properties of the target web application. Our goal is to provide a general view of the elements an application can adopt to avoid the violation of the privacy of personal information.

The relationship between the requirements and the elements from the Reference Architecture is shown in Table 6-4.

4.2.2 Reference Architecture Design

This section describes the design of the reference architecture, shown in Figure 4-3. It is based on a three-layer architectural style: *Presentation*, *Application* and *Persistence*. For privacy protection we introduced a *Privacy* layer between the *Application* and the *Persistence* layers. Each of these layers and their elements regarding the privacy domain are explained next. The relationships between these elements and the architectural requirements are shown in Table 6-4.

Due to the higher level of abstraction, the concepts represented in the privacy reference architecture are independent from the development approach. So, to cope with the

architecture, diverse techniques and tools can be adopted or developed using different technologies.

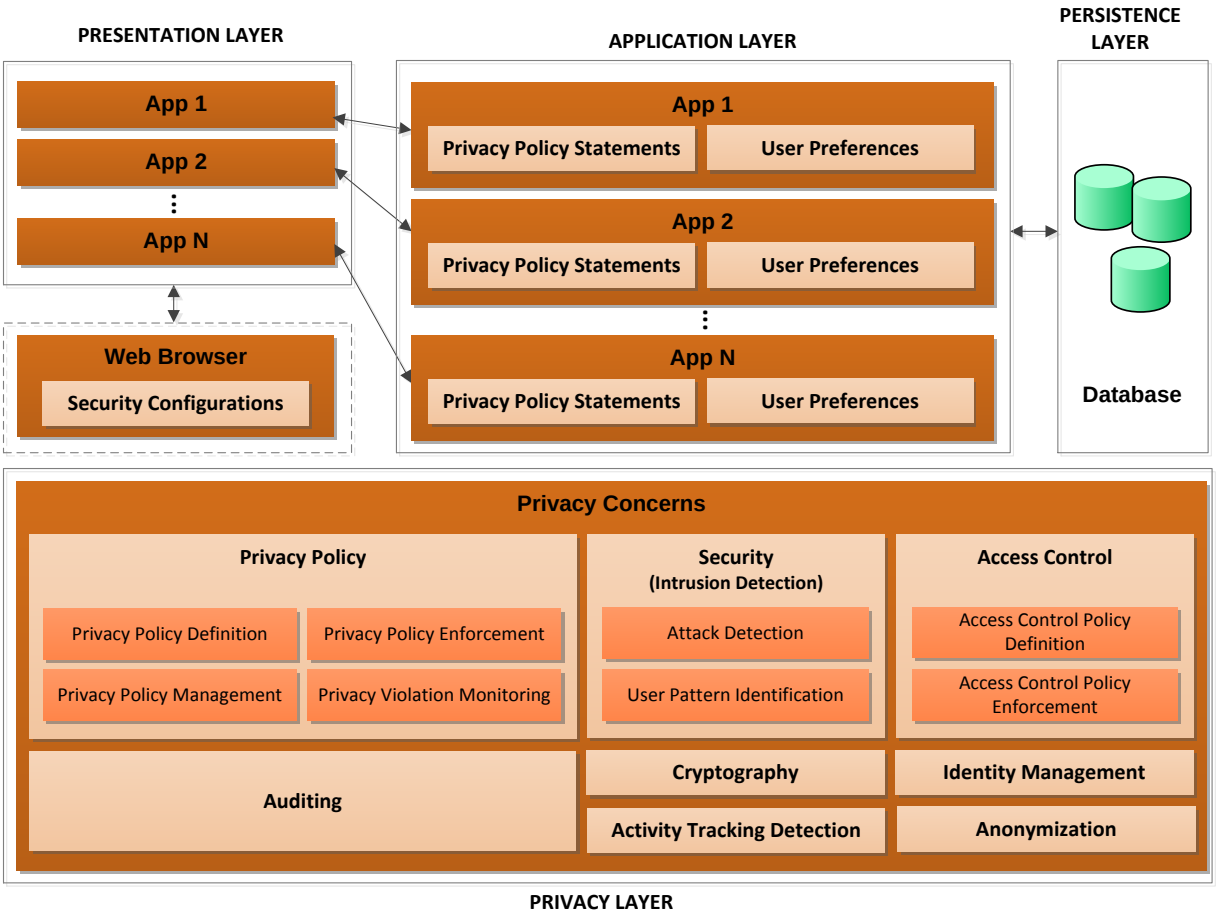

Figure 4-3. General view of the Privacy Reference Architecture.

In Figure 4-3, the **Presentation Layer** refers to the user interface. It allows the user to interact with the application. In practice, it is possible to have different applications with different interfaces (i.e., different sets of functionalities). The *Web Browser* element refers to security configurations that must be implemented to protect the personal information from activity tracking. This tracking consists of identifying user activities on the Web without his/her consent and may represent privacy violation.

The **Application Layer** represents the application logic, with functionalities inherent to the organization's business model. For each application there are two elements: *Privacy Policy Presentation* and *User Preferences.* The *Privacy Policy Presentation* element refers to the fact that the web application must provide this document for its customers and business partners. The *User Preferences* element refers to the need for the web application to

permit users to state their privacy preferences regarding personal information, agreeing or not with the presented policies (or part of them: the statements).

The **Persistence Layer** is responsible for information storage. The *Database* element represents the storage resources and the functionalities the web application may use, such as the DBMS (Database Management System) and other technologies that support data management and recovery. Different applications can access the same database. Data sharing requires ways of ensuring that private information of one application is not accessed by other non-related application.

Still in Figure 4-3, the **Privacy Layer** includes most of the concepts directly related to the privacy domain. It contains the *Privacy Concerns* element, which refers to orthogonal services for personal information privacy protection. They represent functionalities that are independent of the application and may be encapsulated as transverse elements or aspects (i.e., this layer is a logical organization and its elements can be implemented in different parts of the system). It has a set of eight elements: (i) *Privacy Policy*, (ii) *Security/Intrusion Detection*, (iii) *Activity Tracking Detection*, (iv) *Access Control*, (v) *Cryptography*, (vi) *Identity Management*, (vii) *Auditing*, and *(viii) Anonymization*. A short description of each is provided next.

(i) **Privacy Policy**. This element is responsible for defining, enforcing and managing privacy policies. The *Privacy Policy Definition* element is responsible for privacy policies to be defined and presented to the user. Also, based on the policies, users should be able to state their privacy preferences.

Besides defining privacy policies, web applications must ensure that such policies are enforced, i.e., that the agreement signed in the privacy policy is fulfilled. This important requirement is assured by the *Privacy Policy Enforcement* element.

The *Privacy Policy Management* element represents the management of privacy policies between third-parties (i.e., independent web applications or services that interact with the main application). This is an important issue because different applications and services can have different privacy policies and the information exchanged between them must agree with these policies. This element is also responsible for updates in the privacy policies. The updates in the privacy policy must be informed to the users and new preferences about these updates must be considered.

The *Privacy Violation Monitoring* element refers to mechanisms that can be used to detect privacy violation. These mechanisms continuously monitor access to personal data and detect misuse or abnormal behavior.

(ii) **Security/Intrusion Detection**. Security and Privacy in web applications are closely related because security breaches can result, in some cases, in misappropriation and misuse of information by malicious users, leading to violation of information privacy. Attack detection is extremely important as it allows actions to be taken to avoid privacy violation. This feature is represented by the *Attack Detection* element.

Another way for malicious users to access the application is by using valid credentials, usually obtained through identity theft. To help avoiding these fake users, behavioral tendency resources can be used, if the web application collects the users' behaviors. The application must analyze stored behaviors and use them as a security resource. Fake users potentially show a behavior potentially different from that of authentic users and when such difference is detected, the application may ask for new identification to confirm the user identity and, thus, reinforce the security. This resource is represented by the *User Pattern Identification* element.

(iii) **Activity Tracking Detection**. Activity tracking consists of identifying user activities without consent and building a profile, which in practice may represent a privacy violation. To protect against such violation, the web application must use resources to detect tracking to make possible actions to be taken when needed. The *Activity Tracking Detection* is associated with the *Web Browser* element.

(iv) **Access Control**. Access Control is a set of rules by which users are authenticated and by which the access to applications and other information services is granted or denied. The web application must allow access control policies to be defined, specifying roles and information each role can access. This policy definition is represented in the *Access Control Policy Definition* element.

Besides defining access control policies, the application must enforce these policies, ensuring that authorized users only will access particular private information. The *Access Control Policy Enforcement* element refers to these enforcement resources.

(v) **Cryptography**. Personal private information must be protected during the traffic through the web. Cryptography can be used for this purpose, as it makes possible the transmission of incomprehensible messages that are harmless even if third entities eventually intercept them. Cryptography provides: (i) confidentiality (only the authorized receiver can read the message); (ii) integrity (the receiver will be able to identify whether the message has been changed along the way); (iii) authentication (the receiver can identify if the sender is the person who is supposed to send the information); and (iv) non-repudiation (it should not be possible for the sender to deny that was he/she did send the message).

(vi) **Identity Management**. If the web application uses some digital representation of the known information about a specific individual or organization, it must use a digital identity management resource. This resource consists of a set of processes, tools, social contracts and supporting infrastructure to create, maintain, and terminate a digital identity. It enables secure access to an expanding set of systems and applications.

(vii) **Auditing**. Auditing is used to evaluate internal controls in an automated information system and to verify the results of phases and processing systems. This element refers to auditing resources the web application must use to monitor and identify possible sources of privacy violation. These resources should monitor all the system elements, such as databases, servers, application, etc.

(viii) **Anonymization**. This element specifies that the web application must provide techniques to avoid disclosure of confidential information that is retrieved even by means of data analysis. This must be done especially when dealing with statistical databases, which are used mainly to produce statistics on various populations and may contain confidential data regarding individuals.

Another view of the Privacy Reference Architecture is presented in Figure 4-4. It represents the module view, which shows the interactions between the privacy reference architecture packages.

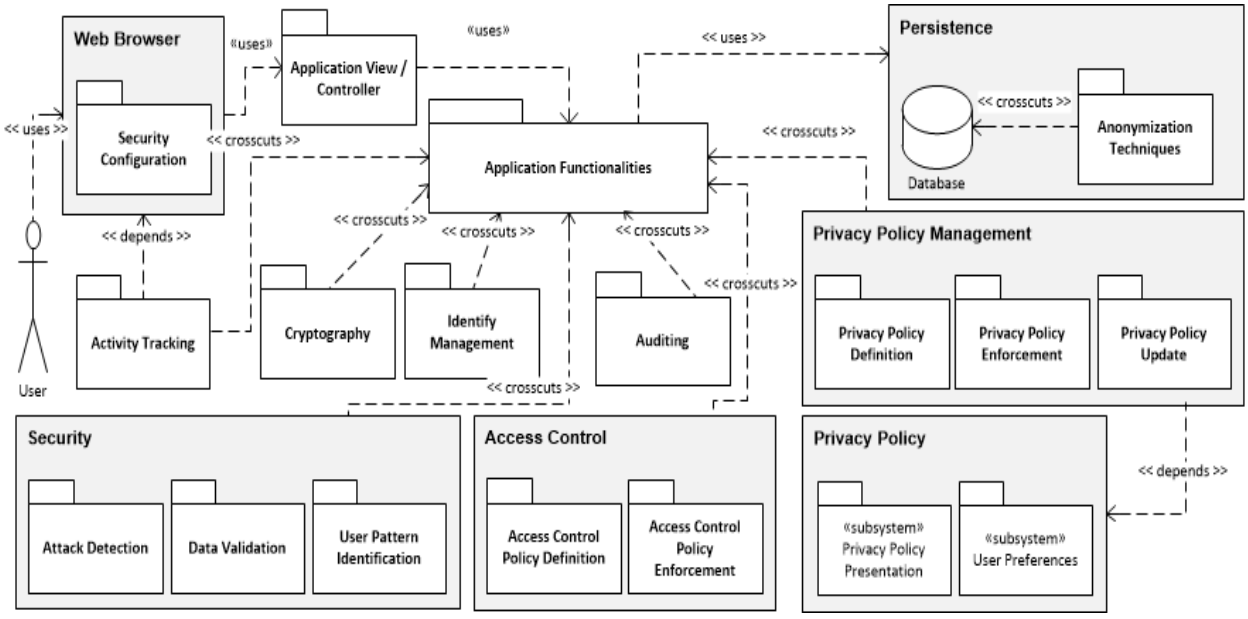

Figure 4-4. Modules of the Privacy Reference Architecture for web applications.

In Figure 4-4, code units, packages and dependency relations are used to represent the module view. While the web application functionalities (application functionalities package) can be implemented using classes, components, or subsystems, the modules that implement crosscutting concerns use aspects in their structures. The dependency relationships labeled <<crosscut>> mean that these aspects crosscut other modules and change the execution flow of these modules by inserting functionalities related to a crosscutting concern.

4.2.3 Architecture Evaluation

Evaluation of the reference architecture helps to identify strong and weak aspects of the architecture and provides an indication of the potential success of the system development process. To the best of our knowledge, there are no methods aimed at holistic evaluation of reference architectures. Angelov and Grefen (2008) present general qualities required for their reference architecture. As done by Angelov and Grefen (2008), we considered ***completeness, usability*** and ***applicability*** as the most important quality attributes of the proposed reference architecture. For evaluation of these qualities, we used adapted techniques similar to those used by Angelov and Grefen (2008). This evaluation process is described in Section 6.1, as part of the PrivAPP quality attributes evaluation.

4.3 THE PRIVACY UML PROFILE

The Privacy UML Profile was constructed based on the Privacy Conceptual Model, i.e., it documents the elements of the conceptual model in order to reduce ambiguities in the solution. It defines new modeling elements, bringing specific concepts related to privacy protection to the UML language. Models created using the profile are meant to be used both during the development phase of a web application as well as after its deployment. During the development, models created using the profile help developers to keep track of privacy requirements and of how they are implemented. After the deployment, the same model can provide the users with a structured description of how the application will handle its private information.

4.3.1 UML Profile Overview

As mentioned in Section 3.2, UML profiles are defined using *stereotypes*, *attributes*, and *constraints*. The elements of our extended Privacy UML Profile are listed in Table 4-3.

Table 4-3. The Privacy UML Profile.

Stereotype	Base Metaclass or Stereotype	Attributes
<<PrivateData>> *(abstract)*	Property, Class	
<<PersonalInformation>>	*PrivateData*	
<<UsageInformation>>	*PrivateData*	
<<PrivacyPolicy>>	Artifact	name (string), creationDate (date), constraint (string)
<<Statement>>	Class	description (string), purpose (string), condition (string)
<<Disclosure>>	*Statement*	
<<Retention>>	*Statement*	period (string)
<<Collection>>	*Statement*	
<<Usage>>	*Statement*	
<<Recipient>>	Actor	description (string), category (string)
<<Enforcement>> *(abstract)*	Class	name (string), description (string)
<<Tool>>	*Enforcement*	
<<Action>>	*Enforcement*	
<<Algorithm>>	*Enforcement*	
<<Process>>	*Enforcement*	
<<Config>>	*Enforcement*	
<<Service>>	Component	description (string)
<<Preference>>	Association	consent (Boolean)
<<SecutiyMeasure>>	*Enforcement*	
<<ActivityTrackingDetection>>	Tool	
<<PrivacyViolationMonitoring/Detection>>	Tool	
<<AttackDetection>>	SecurityMeasure, Tool	
<<UserPatternIdentification>>	SecurityMeasure, Process	
<<Auditing>>	Action, Process	
<<AccessControl>>	Process	
<<AccessControlPolicy>>	Artifact	
<<IdentityManagement>>	Process	
<<Cryptography>>	Algorithm, Process	
<<Anonymization>>	Algorithm, Process	
<<Management>>	*Enforcement*	
<<ManagedInteraction>>	Port	

In Table 4-3, the conceptual elements from the privacy conceptual model (see Figure 4-2) are mapped to UML *stereotypes*, and listed in the first column; for completeness, also abstract stereotypes are included in the table. As stereotypes extend UML *metaclasses* or other stereotypes, the base element of each stereotype is listed in the second column. It should

be noted that a stereotype may also extend another newly introduced stereotype. Finally, stereotype *attributes* are listed on the last column.

The <<PrivacyPolicy>> stereotype extends the *Artifact* metaclass, which represents the specification of a physical piece of information that is used or produced by a software development process, or by deployment and operation of a system (OMG, 2011). It also extends the *Class* metaclass. In UML, a *Class* describes a set of objects that share the same specifications of features, constraints, and semantics (OMG, 2011).

The <<Statement>> stereotype extends the *Class* metaclass. In UML profiling, *Class* is often selected as a "default" base metaclass, and it is typically adopted for stereotypes that do not represent software elements as well. The <<Statement>> stereotype is further extended by stereotypes that characterize the nature of the statement of the privacy policy: <<Disclosure>>, <<Retention>>, <<Collection>> and <<Usage>>.

The <<PrivateData>> abstract stereotype extends both the *Property* and the *Class* metaclasses. The *Property* metaclass is a structural feature which represents an attribute (OMG, 2011), i.e., a portion of data; the *Class* in this context is seen as an aggregation of multiple elements of information. <<PersonalInformation>> and <<UsageInformation>> are used to mark data that is regarded as personal information or usage information, respectively, and they extend <<PrivateData>>.

The <<Enforcement>> stereotype and its descendants represent resources and solutions that are used to enforce the statements described in the privacy policy. Ideally, the profile should allow the modeler to relate enforcement solutions directly to elements in the model of the software architecture. Depending on the context, an enforcement solution (e.g., an algorithm) may be described by either a structural (e.g., a *Component*) or a behavioral feature (e.g., an *Activity*). In order to be able to cover both cases, our <<Enforcement>> stereotype extends the *Class* metaclass, which is a common ancestor of both the *Component* and *Behavior* UML metaclasses (OMG, 2011). The <<Enforcement>> stereotype is then extended to better categorize the nature of the enforcement solution: <<Tool>>, <<SecurityMeasure>>, <<Action>>, <<Algorithm>>, <<Process>>, <<Config>>, <<Management>>. Then, we extended these stereotypes slightly beyond in order to specify more concrete statement enforcement resources: <<ActivityTrackingDetection>> and <<PrivacyViolationMonitoring/Detection>> (extend <<Tool>>); <<Attack Detection>> (extends <<Tool>> and <<SecurityMeasure>>) ; <<UserPatternIdentification>> (extends <<SecurityMeasure>> and <<Process>>); <<Auditing>> (extends <<Action>> and <<Process>>); <<AccessControl>> and <<IdentityManagement>> (extend <<Process>>);

<<Cryptography>> and <<Anonymization>> (extend <<Process>> and <<Algorithm>>). Here it should be mentioned the related <<AccessControlPolicy>> stereotype, which extends the *Artifact* metaclass, and the <<ManagedInteraction>> stereotype, which extends the *Port* metaclass. A *Port* may be used to specify in more detail the services a classifier provides (requires) to (from) its environment.

The <<Preference>> stereotype extends the *Association* metaclass, which specifies a semantic relationship that can occur between typed instances, in our case elements of the <<Statement>> and <<Enforcement>> elements. Such association relates an enforcement solution with a statement for which it is needed, also detailing for which kind of user preference (*opt-in*, *opt-out*) is actually needed.

The <<Service>> stereotype extends the *Component* and *Port* metaclasses. A *Component* describes a modular part of a system that encapsulates its contents, i.e., without focusing on its internal implementation, but only on the service(s) it provides.

Finally, the <<Recipient>> stereotype extends the *Actor* metaclass. This metaclass specifies a role played by a user or any other system that interacts with the subject.

The *constraints* needed to express our domain concepts are limited to relationship multiplicities (see Figure 4-2); no additional constraints are included in the profile. The constraints expressing the multiplicities are instead summarized in Table 4-4.

Table 4-4. Privacy UML Profile constraints

For each *Statement* element there must be at least one association with a *Private Data* element.
For each *Disclosure* element there must be at least one asssociation with a *Recipient* element
For each *Privacy Policy* element there must be at least one containement association with a *Statement* element

4.3.2 Illustrative Example

In order to evaluate the applicability of the UML Profile, we modeled some real privacy policy statements and possible enforcement resources for them. We adopted the privacy policy of Google services (Google, 2014) and, although there is some discussion about Google's privacy policy when describing how it uses personal data gathered from its web services and products (BBC, 2015-b; ICO, 2015), we decided to model this policy because it is very popular and used by many users and client applications around the world. To be sure that a privacy policy is well constructed, i.e., if it complies with the established

laws and principles, it would be necessary to perform a thorough analysis, which is out of the scope of this work. Therefore, we assume that Google's policy is acceptable and we try to deal with its problems (like fuzzy or unclear statements), and focus on our goal to apply the proposed profile to a real use case.

For the purpose of this evaluation, we focus on specific parts of Google's privacy policy, among those that are more informative to the user (i.e., discarding general statements such as "*we collect information to provide better services to our users*"). The Google privacy policy (Google, 2014) is described in natural language and consists of 8 pages and more than three thousand words. From all the material we isolated a set of 9 statements of interest for our evaluation process. Then, we modeled three of them in UML diagrams. The statements we selected are described in APPENDIX C.

It should be noted that, although we have complete access to the Google privacy policy (Google, 2014), we have no information on the means that are applied by the company to actually enforce its privacy policy. Consequently, while most of the elements in the following models are derived from the real world, *Enforcement* elements used in the example are fictional, and serve only for the purpose of describing the application of the profile.

Figure 4-5 details the model for statement we called ST9 ("*We restrict access to personal information to Google employees, contractor and agents who need to know that information in order to process it for us.*"). The statement describes that the access to personal information is restricted to Google employees who need to know that information for processing it; in case they fail to respect privacy obligations, disciplinary measures may be undertaken.

Statement ST9 is a <<Disclosure>> statement, since it describes the disclosure of information to internal Google members. The purpose for which data is disclosed is for data processing. The statement is associated to the *PrivateData* elements, which represents the generic data that will be disclosed. Since the statement does not detail which kind of data will be disclosed, we assume that both personal information and usage information are involved; for this reason both the <<PersonalInformation>> and <<UsageInformation>> stereotypes are applied to the same *PrivateData* element.

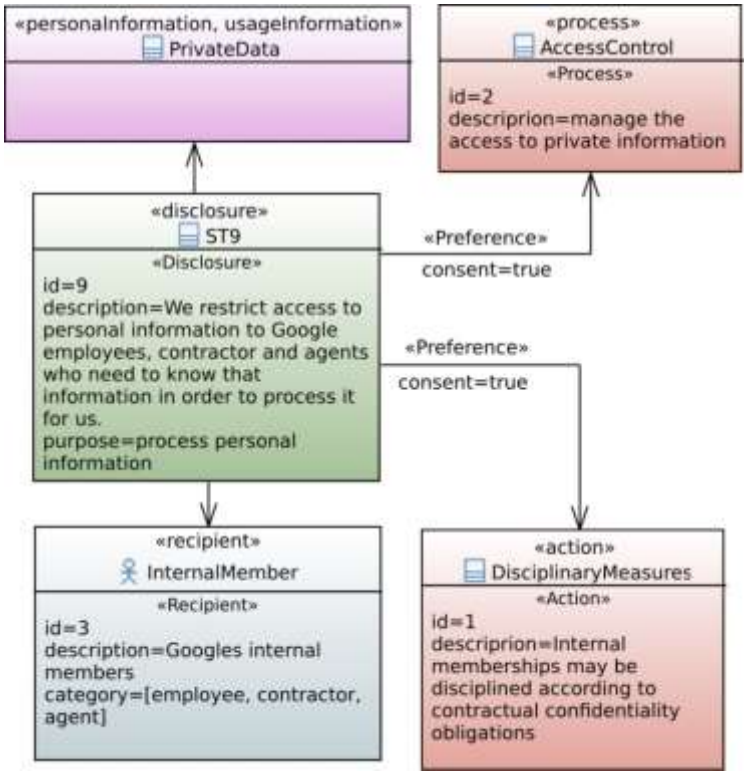

Figure 4-5. Representation of Google's statement using the privacy profile.

The *InternalMember* element is a <<Recipient>>, and it represents the recipient of data disclosure. The categories associated with this recipient are *employee*, *contractor*, and *agent*. Among the enforcement solutions, we find the *AccessControl* element and the *DisciplinaryMeasures* element. The *AccessControl* is a <<Process>> element, and it has been already described in the previous section; however, in this case it is applied to enforce an *opt-in* preference of the user. This is modeled by the <<Preference>> relation with the *consent* attribute having the *true* value.

The other enforcement solution, *DisciplinaryMeasures*, is modeled with an <<Action>> element and describes the application of disciplinary measures, e.g., revocation of contracts in case of privacy obligations are not met by the recipient. Also in this case, the enforcement solution is applied in case of an *opt-in* preference of the user.

The other statements we selected and correspondent UML diagrams are described also in APPENDIX C. Although we did not exercise all the elements of the Profile, we have an indication of the applicability of the proposed Profile. It allowed to model privacy policy statements effectively and through these diagrams is possible keep track of them and their requirements during the development of applications.

Due to the close relationship between security and privacy, here we highlight the contribution of the Privacy Profile we propose with respect to the security profile UMLSec (Jürjens, 2002). Due to the nature of the privacy and security, the requirements for both UML profiles are high level and general. While the UMLSec is more focused on security on communication met by physical layer (providing secrecy and integrity of data considering different threat scenarios and incorporating security mechanisms (e.g. access control)), the Privacy UML Profile is more focused on the privacy policy (providing resources to web applications and services manage private data in order to not violate the privacy of their customers and business partners). So, both UML Profiles can be used in complementary manner.

4.4 CONCLUDING REMARKS

In this chapter we presented the proposed approach: PrivApp. We described in detail its three components: the Privacy Conceptual Model, the Reference Architecture and the UML Profile. The Privacy Conceptual Model addresses, within the scope of web applications and services, the privacy concepts and their relationships in order to systematize these concepts. The Reference Architecture addresses abstract software components that represent functionalities related to privacy protection. We describe the whole process to the conception of the Privacy Reference Architecture, since the architectural requirements establishment until the evaluation process regarding some quality attributes. Finally, the UML Profile, which is an extension of the UML metamodel to allow using privacy protection features in UML diagrams, was presented.

5 . CASE STUDY

A case study was developed to provide data privacy protection in a web application. The goal is to evaluate the practical application of the proposed approach. The following steps were carried out: (i) an application without privacy protection resources (an online book store) was selected; (ii) a privacy policy was established for this application (based on the Amazon's privacy policy (Amazon, 2014)); (iii) a software architecture including the privacy protection elements was created for the original web application; (iv) the UML diagrams were created, based on the main statements of the application's privacy policy.- these diagrams are derived from the UML Profile and show how the application must enforce the statements, including the technologies that must be used; (v) the solution designed by the diagrams and the architecture was implemented.

The procedure above led us to the implementation of a solution which consists in an access control mechanism that allows users to express their privacy preferences and requested information are permitted/denied according to these preferences. This mechanism is integrated into the relational database system, providing security against possible attacks to the web application or the network.

5.1 THE BOOKSTORE APPLICATION

The web application we used in the case study is a Java implementation of a TPC-W (TPCW, 2015). TPC-W is a benchmark for web-based transactional systems where several clients access the website to browse, search, and process orders. The typical workload that it supports consists of shopping sessions. Each session emulates the behavior of a customer connected to the server and generally consists of a sequence of interactions: search, browse, add to shopping cart, make purchases, and so on. In this study, we adapted the TPC-W through an implementation of a retail online book store, which simulates the sale of books in the Internet. On purpose, the application is devoid of any data privacy protection. Hence, for the sake of security and privacy, we did not use real data. Figure 5-1 shows the main page of our application.

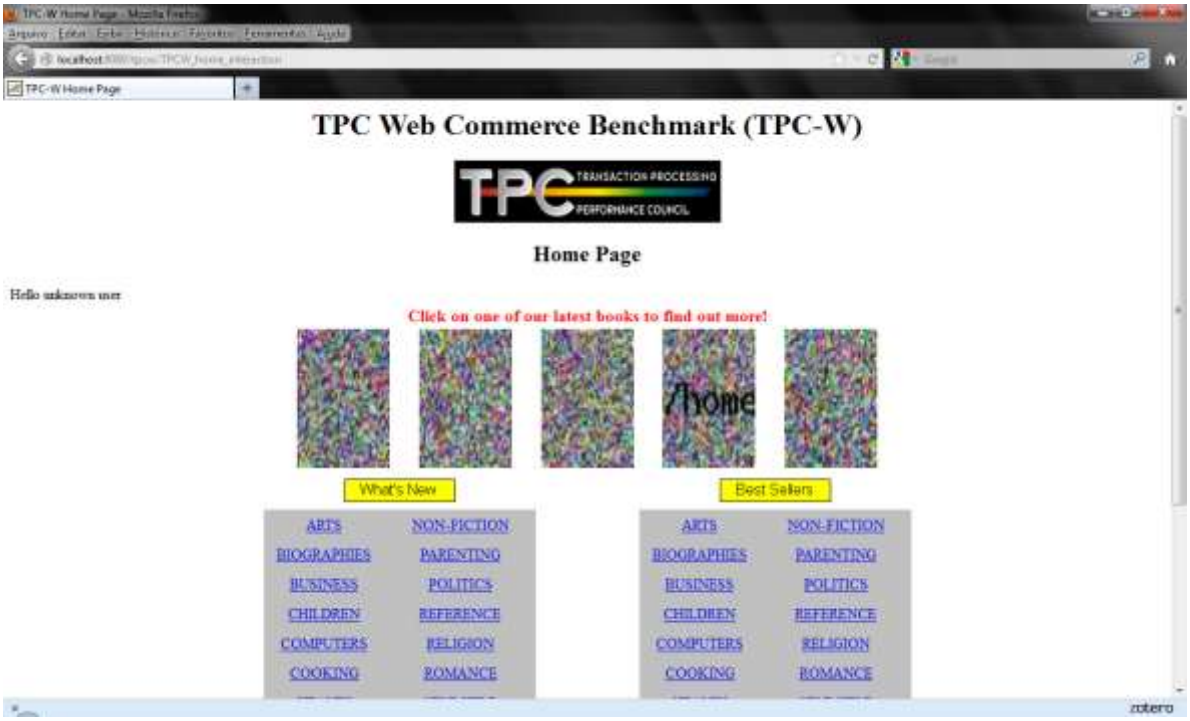

Figure 5-1. Main page of our adaptation of TPC-W application.

The diagram in Figure 5-2 shows a high-level view of the TPC-W architecture.

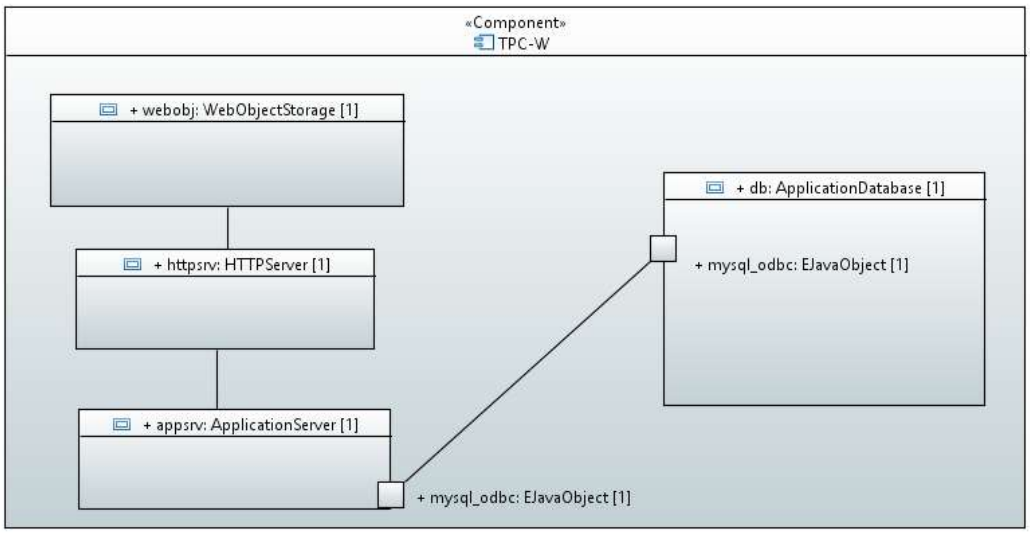

Figure 5-2. TPC-W's architecture diagram.

As in all e-commerce benchmarks, TPC-W has a client-server architecture. The client computers work as remote browser emulators to simulate the workload that real customers would generate. In Figure 5-2, the system includes an HTTP server with web

object storage, an application server, and an application database. This system communicates with the clients through a dedicated network.

The TPC-W component of our major interest is the Application Server. The bookstore implementation operates in this server. Figure 5-3 details this server, showing its components.

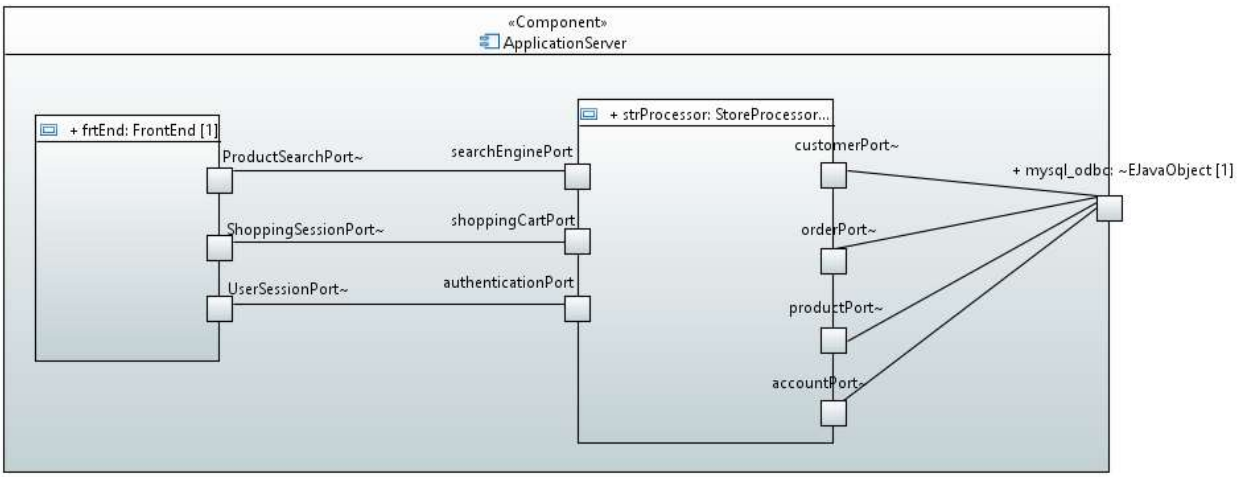

Figure 5-3. TPC-W's Application Server detailing.

In Figure 5-3, the *FrontEnd* component is the Front End of our implementation. It corresponds to the presentation layer, the interface between the user and the application. The *StoreProcessor* component is the Store Processor of the implementation, i.e., the procedures used to purchase the books online. They exchange information through their interface (ports), where the symbol ~ means that the interface is required. Next we detail each of these components.

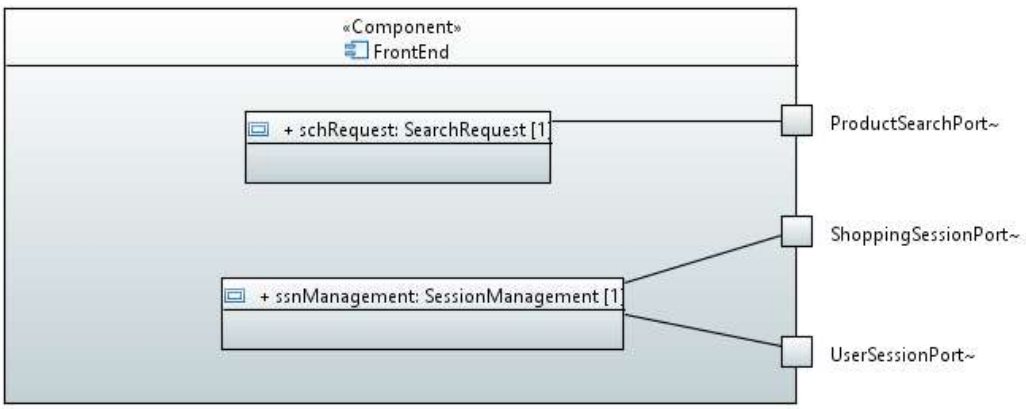

Figure 5-4. *FrontEnd* component detailing.

In Figure 5-4, the *FrontEnd* component is composed by the *SearchRequest* and the *SessionManagement* components. The *SearchRequest* is responsible for the interface through which customers and visitors can search books in the bookstore. The *ProductSearchPort* is the interface through which components process search queries. The *SessionManagement* is the interface where sessions can be created in two ways: (i) the visitor adds books to the shopping cart without registering (shopping session); (ii) a registered user authenticates in the application to use it (user session). The *ShoppingSessionPort* and *UserSessionPort* are, respectively, the ports through which information related to the sessions is exchanged.

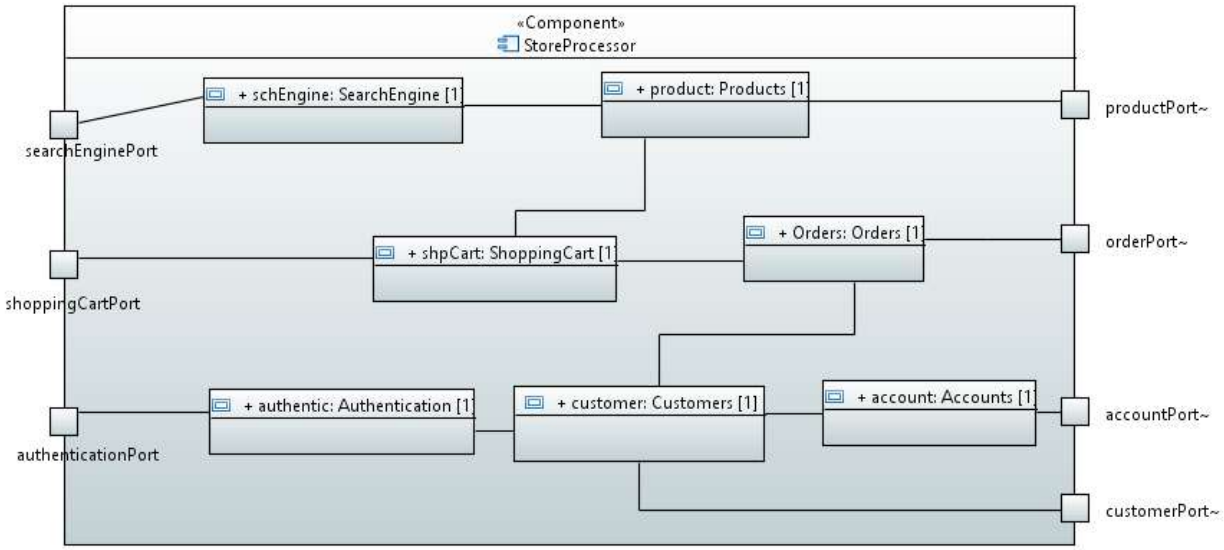

Figure 5-5. The *StoreProcessor* component.

Figure 5-5 shows the *StoreProcessor* and its seven components. The *SearchEngine* is responsible for processing the search strings the user requested. It is related to the *Products* component, which manages the inclusion, exclusion and updates of books. The *ShoppingCart* component is responsible for the management of items to be bought, while the *Orders* component process the orders and the payments. To make a purchase, the visitor must register first. Thus, the *Customers* component is related to *Orders* and is responsible for managing the customers' records. Also, a customer's account is managed by the *Accounts* component. Once the visitor is registered and an account is associated to him/her, an authentication process is necessary. This is done by the *Authentication* component.

The ports *productPort, orderPort, accountPort* and *customerPort* are used by the corresponding components to interact with the database, i.e., they make use of the interface provided by the database.

5.2 THE PRIVACY POLICY

To implement privacy protection in this application it is necessary to define a privacy policy to guide all the privacy control process. Since our focus is not policy definition, we adopted Amazon's privacy policy (Amazon, 2014) because Amazon is a well-known and successful online book store, with a policy that is certainly representative of this segment. Obviously, we cannot use the whole policy because our application is simpler than Amazon's application. We selected 5 statements that are closely related to the operation of our application and which require enforcement resources to enforce them. The statements are described in Table 5-1. It is important to mention that, to guarantee privacy protection, we interpreted fuzzy statements in a conservative (worst-case) meaning, e.g., if a statement says "*we usually keep the copy*" we interpreted it as "*we do keep the copy*".

Table 5-1. Selected statements from the privacy policy for enforcement of privacy protection (Amazon, 2014).

Statement	Description
ST1	"*We work to protect the security of your information during transmission by using Secure Sockets Layer (SSL) software, which encrypts information you input*."
ST2	"*You can add or update certain information on pages such as those referenced in the "Which Information Can I Access?" section. When you update information, we usually keep a copy of the prior version for our records."*
ST3	"*Cookies are unique identifiers that we transfer to your device to enable our systems to recognize your device and to provide features such as 1-Click purchasing, Recommended for You, personalized advertisements on other Web sites (e.g., Amazon Associates with content served by Amazon.com and Web sites using Checkout by Amazon payment service), and storage of items in your Shopping Cart between visits"*
ST4	"***Affiliated Businesses We Do Not Control:*** *We work closely with affiliated businesses. In some cases, such as Marketplace sellers, these businesses operate stores at Amazon.com or sell offerings to you at Amazon.com. In other cases, we operate stores, provide services, or sell product lines jointly with these businesses. Click here for some examples of co-branded and joint offerings. You can tell when a third party is involved in your transactions, and we share customer information related to those transactions with that third party*."
ST5	"***Third-Party Service Providers:*** *We employ other companies and individuals to perform functions on our behalf. Examples include fulfilling orders, delivering packages, sending postal mail and e-mail, removing repetitive information from customer lists, analyzing data, providing marketing assistance, providing search results and links (including paid listings and links), processing credit card payments, and providing customer service. They have access to personal information needed to perform their functions, but may not use it for other purposes*."

5.3 APPLYING THE APPROACH

As already mentioned, the online book store is, on purpose, devoid of any data privacy protection. Our goal is to incorporate privacy protection into this application, by modeling privacy concerns and the elements needed to enforce privacy.

We first created in the architecture a logical group of measures to help in privacy protection. This logical group is defined as <<aspect>> because Aspect Oriented technology is rooted back to the separation of concerns by which different concerns of the software system can be designed and reasoned about in isolation from each other (Aldawud *et al*., 2003). These aspects can be used in (i.e., crosscut) different components of the application; thus, they can be used in both *ApplicationServer* and *DatabaseServer* of the original bookstore application. This logical group is represented by the *PrivacyManagement* component, in Figure 5-6. The next subsections shows how we applied the PrivAPP: first, we created the UML diagrams, based on the Privacy UML Profile. In parallel, we used the Reference Architecture to identify the Enforcement elements that are part of the UML diagrams. In sequence, also based on the Privacy Reference Architecture, the software architecture that represents the application with privacy protection was defined. Based on the UML diagrams and the software architecture, we implemented a database framework for access control. During the development, models created using the profile help developers to keep track of privacy requirements and how they are implemented. An experimental evaluation was performed to assess the scalability and performance impact of the proposed solution.

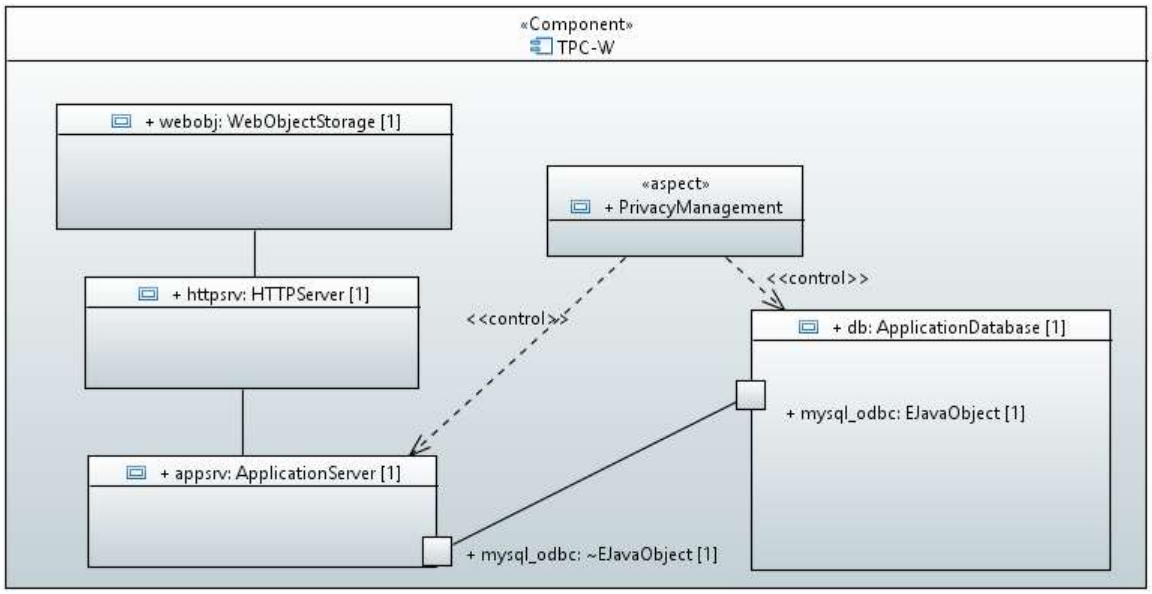

Figure 5-6. Inclusion of the *PrivacyManagement* component, responsible for privacy protection.

5.3.1 Applying the Privacy UML Profile

We start implementing privacy protection by defining UML models because they help to better understand the privacy policy statements and the resources that can be used to enforce these statements. The Privacy UML Profile proposed in Chapter 4 was used to create the UML diagrams and to document the system.

The Privacy UML Profile and the Privacy Reference Architecture are elements from our approach that can be used individually or in parallel, in a complementary manner (see Figure 4-1). In this case study, we used them in a complementary manner. While constructing the UML diagrams, the Privacy Reference Architecture was used to identify the enforcement elements that are more adequate for each statement.

For the sake of organization, we split the UML diagrams into 2 parts, represented in Figures 5-7 and 5-8. The privacy policy statements are described in Table 5-1.

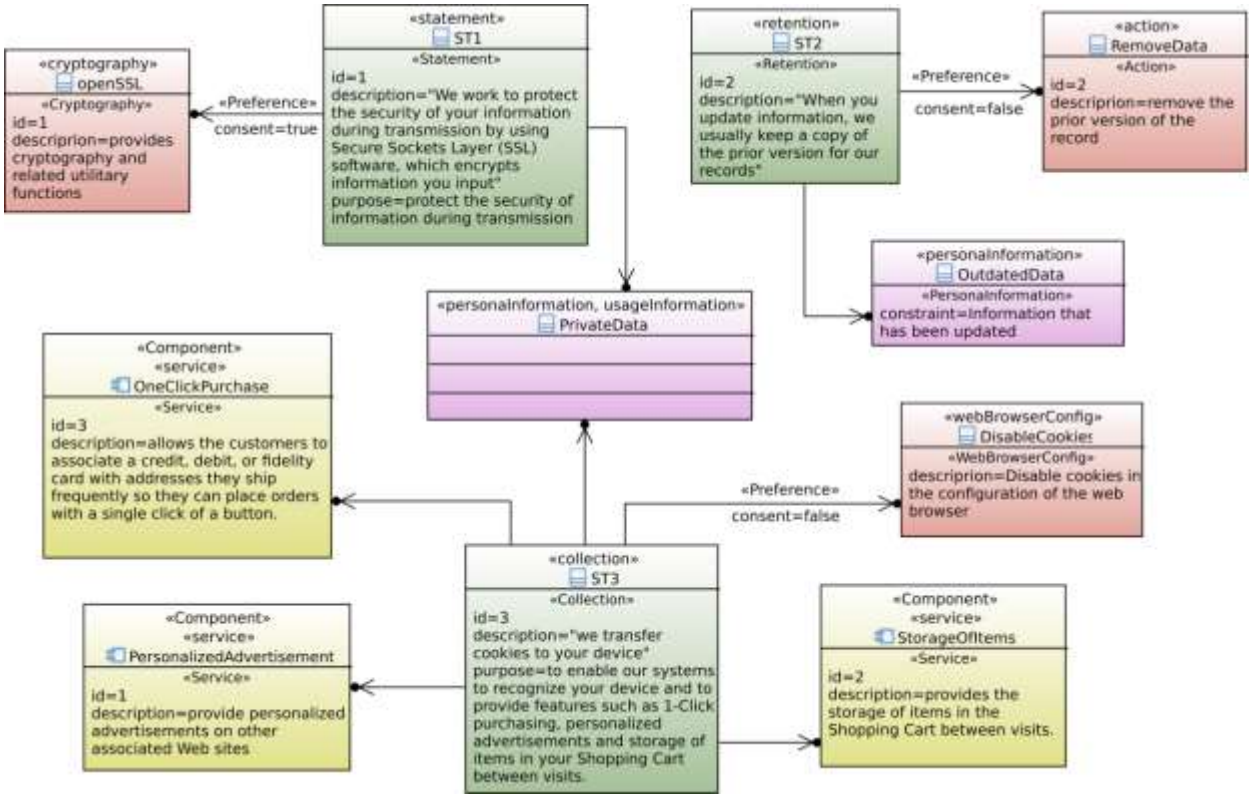

Figure 5-7. Representation of Statements ST1, ST2 and ST3, with their related elements.

In Figure 5-7, Statements *ST1*, *ST2* and *ST3* are shown, representing, respectively, the <<*Statement*>>, <<*Retentiton*>> and <<*Collection*>> elements. *ST1* and *ST3* are related to *PrivateData*, which represents both <<*PersonalInformation*>> and <<*UsageInformation*>>. Statement *ST3* is related to 3 types of services provided by the application: *OneClickPurchase, PersonalizedAdvertisement* and *StorageOfItems*. *ST2* is related to *OutdatedData*, which represents<<*PersonalInformation*>>. Furthermore, each statement is related to a user preference (<<*Preference*>>, *consent = true* or *consent=false*) and to an enforcement element according to this preference. *ST1* is related to*SSL*, which is a <<*Cryptography*>> element; *ST2* is related to *RemoveData*, which is an <<*Action*>> element; *ST3* is related to *DisableCookies*, which is a <<*WebBrowserConfig*>> element. The same happens in Figure 5-8: Statements *ST4* and *ST5* represent the <<*Disclosure*>> element. They are related, respectively, with the *AffiliateBusinessOperations* and *BasicFunctions* services (<<*Service*>>) and the *AffiliatedBusinesses* and *ThirdPartyServiceProviders* recipients (<<*Recipient*>>), as well as the *PrivateData* (represented by <<*PersonalInformation, UsageInformaiton*>>). Both statements are related to their own user preference (<<*Preference*>>, *consent = false*) and the enforcement is given by the *AccessControlMechanism*, an <<*AccessControl*>> element, related to its

<<*AcccessControlPolicy*>>, which we called *ACPolicy1*. Also for Statement ST5, if the user agrees with the policy (<<*Preference*>>, *consent = true),* two enforcement elements can be applied: <<*Management*>> and <<*Auditing*>>. The <<*Management*>> element, which we called *CheckThirdPartiesPolicies,* is responsible for checking the compatibility of the original application's privacy policy and the third party service provider's privacy policy. If they are compatible, the services (*BasicFunctions*) can be provided. The <<*Auditing*>> element (*AuditThridPartiesPurposes*) is responsible for periodically verifying if the third parties are using the personal information shared with them according to the specified purposes.

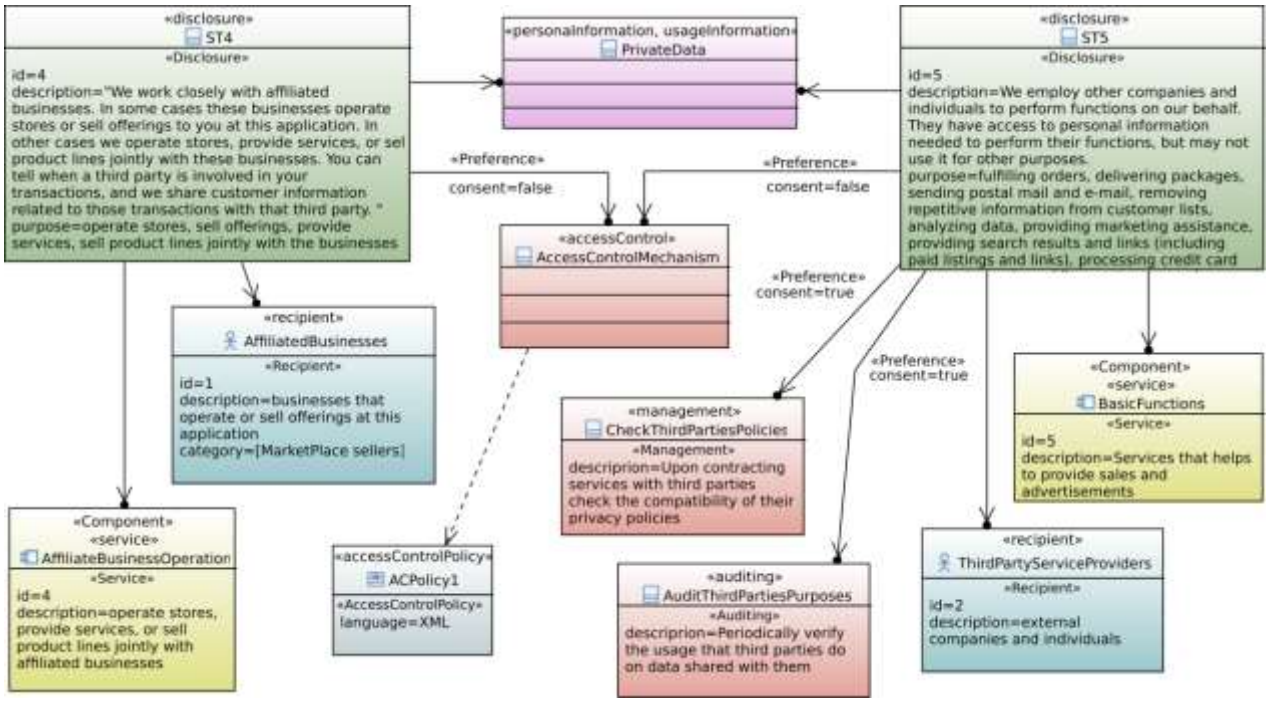

Figure 5-8. Representation of Statements ST4 and ST5, with their related elements.

According to the Privacy UML Profile, a complete model should also include a <<*PrivacyPolicy*>> element, having a containment relation with all the statement elements included in the model. For this case study, <<*PrivacyPolicy*>> contains the 5 statements shown in Figures 5-7 and 5-8. To simplify the presentation of both diagrams, the <<*PrivacyPolicy*>> element is not shown. A complete diagram representing all the statements aggregated to the <<*PrivacyPolicy*>> is presented in Figure 5-9.

Figure 5-9. Privacy policy element and associated statements.

5.3.2 Applying the Privacy Reference Architecture

After defining the UML diagrams for the privacy policy, we created a software architecture that represents the application with privacy protection; thus, we identified the components corresponding to enforcement elements adopted in the UML diagrams in the reference architecture. They are: *User Preferences, Web Browser (Security Configurations)*, *Privacy Policy Enforcement, Cryptography*, *Access Control Policy Definition*, and *Access Control Policy Enforcement* (see Figure 4-3). From these elements, the *Enforcement* ones can be grouped into the *PrivacyManagement* component (see Figure 5-6) as subcomponents to be used in the software architecture. Figure 5-10 shows this group.

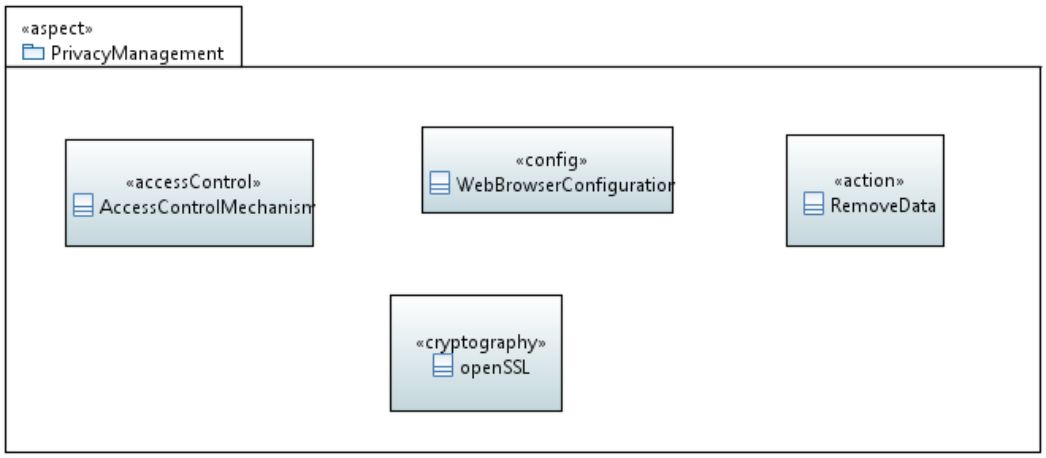

Figure 5-10. Enforcement components.

From the selected enforcement elements, we have detailed, in this case study, the implementation of the <<accessControl>> with the *AccessControlMechanism* component,

because of the following: (i) <<cryptography>>, with *OpenSSL*, is off-the-shelf; (ii) <<config>>, with *WebBrowserConfiguration*, does not belong to the application, but rather to the user environment; (iii) <<action>>, with *RemoveData* is a simple implementation.

Basically, an access control mechanism includes access control policies and a mechanism that, based on these policies, get the requested information and allows or denies access to these information to the requester. This is explained with more detail further. We created a software architecture where we include the access control mechanism in the original TPC-W architecture. As this component could be used in both TPC-W's *ApplicationServer* and *ApplicationDatabase* components, we represented the two situations, respectively, in Figures 5-11 and 5-12.

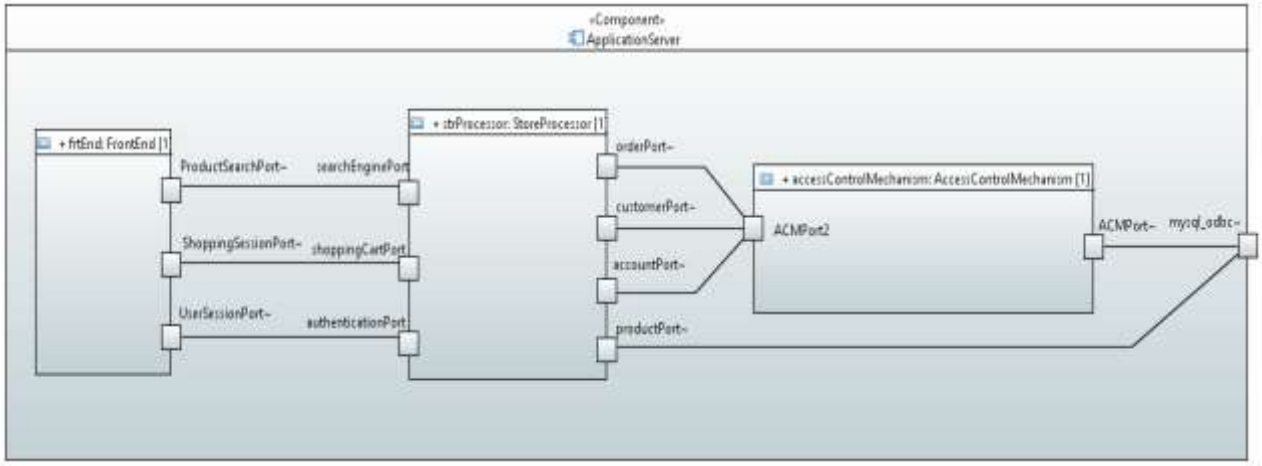

Figure 5-11. TPC-W's Application Server with the addition of the Access Control Mechanism

In Figure 5-11 the *AccessControlMechanism* component was added to the original TPC-W's *ApplicationServer* in order to help protecting privacy according to the privacy policy (statements *ST4* and *ST5*). The *ApplicationServer* is presented in Figure 5-3 and again, in Figure 5-11; the *orderPort, customerPort* and *accountPort* ports are connected to the *ACMPort2*, which is the interface provided by the access control component. The idea is to control the access of third parties to information that includes orders, customers and account data. *productPort*is not connected to the access control because the products to be sold in the bookstore have free access to customers and visitors, i.e., the access control is not necessary for this information.

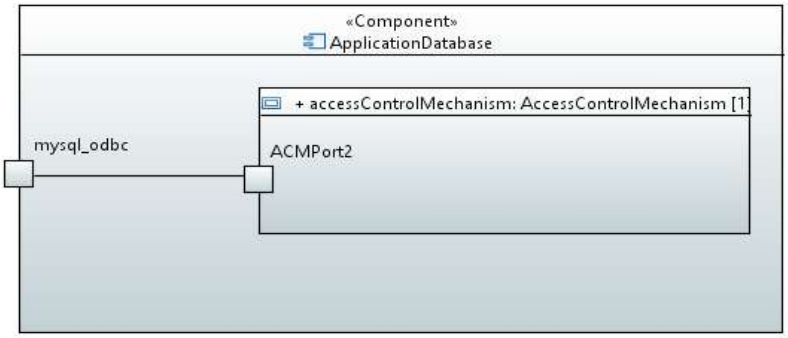

Figure 5-12. TPC-W's Database Server with the addition of the Access Control Mechanism.

In Figure 5-12 the *AccessControlMechanism* component was added to the original TPC-W's *ApplicationDatabase.* In this case, the *ACMPort2* port is connected to the provided interface of the correspondent TPC-W's component. Implementing the access control inside the database server has the advantage of filtering the data directly in the database, which helps to protect against possible attacks to the web application or the network. That is the reason we decided to implement the mechanism in this server.

One advantage of the access control mechanism that we represented in the software architecture is the users' privacy preferences management. The mechanism must allow users to express their privacy preferences, concerning each piece of their personal information, and this must be taken into account, i.e., the access to private information must be controlled according to these preferences. Thus, still defining the software architecture, we detailed the components of the access control, shown in Figure 5-13.

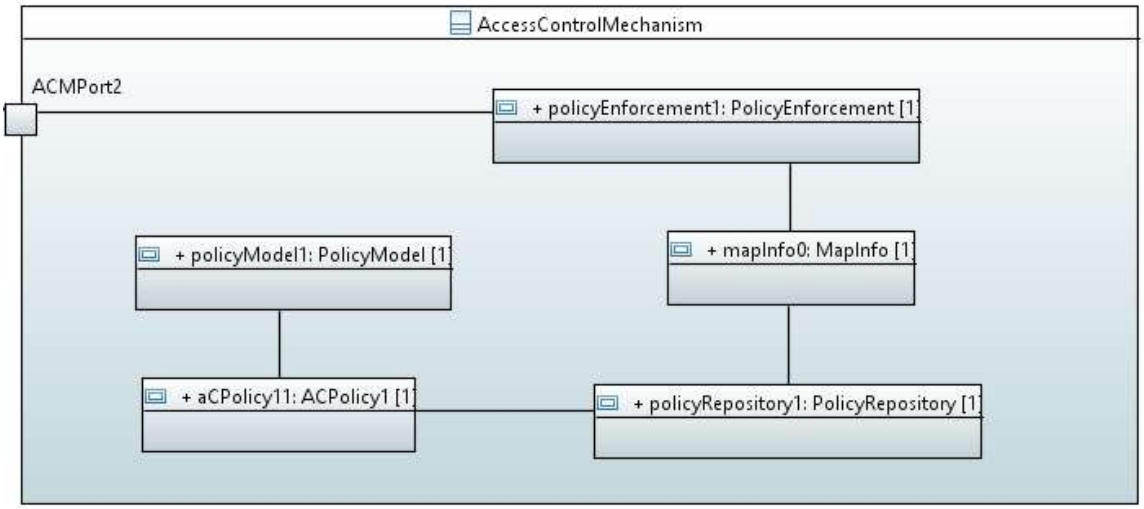

Figure 5-13. Access control mechanism detailing.

Briefly, the components of the access control mechanisms are: (i) *policyModel1*: represents the model or set of models to be used in order to create the access control policies; (ii) *aCPolicy11:* represents the access control policies. This component is responsible for helping to create these policies; (iii) *PolicyRepository1:* responsible for maintaining the access control policies; (iv) *mapInfo0*: as the access control mechanism we represent is to be implemented in the database application, the policies information must be managed so that their information can be manipulated by it. The users' preferences must be considered and also managed by this component; (v) *policyEnforcement1*: this component analyzes the access control policy information and the users' preferences and, according to them, it allows or denies access to the requested information.

5.3.3 Implementing and Evaluating the Access Control Mechanism

Based on the UML diagrams and the software architecture, we implemented a database framework for access control. Basically, this framework consists of a set of independent tables that can be added to the application's database. These tables will contain the information necessary for performing the access control according to predefined access control policies, which consider users preferences concerning each piece of personal information to be collected, stored or managed. An access control is performed by implemented database packages and has the advantage of filtering the data directly in the database, providing better data protection against malicious attacks that occurs in the web application. A detailed description of the database framework is presented in APPENDIX D.

The access control policies are based on the policy model we proposed. This model is simple (has the simplicity as its main advantage) and was constructed through an extensive study based on the literature and interviews with IT professionals. Thus, it is based on real problems a policy model should tackle considering the requirements which are relevant for the implementation of a good access control mechanism. The description of this policy model is presented in the APPENDIX D. For the sake of simplicity, we focused on using information on the profiles allowed to access data and on the criticality levels of data, disregarding other types of information defined by the model (e.g., *from where* the required information can be accessed).

The policies are defined as XML files and a job (i.e., a combination of a schedule and a program, along with any additional arguments required by the program) is executed

periodically to verify the input of policies in a specific application directory (policy repository). Then, the job maps these policies to the set of tables of the framework.

Users' preferences are also mapped into the framework. To allow users and visitors to express their preferences concerning the privacy of each piece of their personal information, the application must implement controls over the user interaction with the application during the collection of this personal information. We adapted the applications interface with simple text box or combo box informing, for each piece of data, the criticality levels that can be chosen. These criticality levels are based on the work of Vieira and Madeira (2005) and are represented as numbers from 1 to 5, which represent different requirements in terms of security, ranging from non-critical data (level 1) to data that has to be very strongly protected against unauthorized access (level 5). For more details about these criticality levels, see APPENDIX D.

With all these information in the framework, when a query is executed from the application, the mechanism (*policyEnforcement)* obtains the policies information and the criticality levels of particular fields returned by this query. Then, the data are masked (or not) and presented to the requestor. To mask the data and, consequently, enforce the policies, the mechanism implements an algorithm, based on the following steps:

1. Obtain the identifier of the user that is requesting the private data, the role of this user, and the data that are being requested.

2. From this information (received in Step 1), the table that stores the privacy policies is queried to identify the criticality levels this user (requester), with respective role, can access.

3. Obtain the preferences of the data's owner, i.e., the criticality level the owner assigned the private data being requested.

4. Verify whether the criticality level the user (requester) can access is higher than the criticality level of the data and:

a. If true, the data are provided to the user who is requesting them.

b. If false, the data are masked in order to enforce the policies and abide by the preferences of the data owner.

A case study was performed to assess the scalability and performance impact of the proposed solution. By scalability we mean the number of records that can be processed by the mechanism without impairing seriously the application, i.e., the goal is to assess how much the number of records in the database application affects the performance. By

performance impact we mean the throughput and the average processing time, to determine if to use the mechanism can be a disadvantage to the user.

Experimental Setup. The database used in the experiments is the Oracle Database 10g Express Edition Release 10.2.0.1.0 (Oracle, 2015); the mechanism was implemented using PL-SQL (a procedural language extension for SQL). The metrics were collected using the JMeter tool (Jmeter, 2015). The tests apply a scenario of application use to simulating a third-party user (a user trying to access unduly data or even a potential attacker) trying to obtain data of a registered customer through a search process. Privacy policies were implemented and the criticality level of each piece of data of each customer was randomly generated through a database script.

For the experimental evaluation we used, respectively, 500, 5000 and 50000 records in the database. The simulations of threads, which simulate concurrent connections to the server application, ranged from 1 to 128 users for each set of records. Also, in order to understand the performance impact, the tests were performed without the database framework in place (to obtain baseline indicators). For each run of the experiment, the whole system is returned to its initial state in order to avoid cached data. Figure 5-14 presents the overall results of the study.

Overall result analysis. In Figures 5-14a, 5-14b and 5-14c the average processing time (in milliseconds) of all requests for the customer search scenario is shown. Figures 5-14d, 5-14e and 5-14f shows the throughputs, which are calculated as requests divided by unit of time. The time is calculated from the start of the first sample to the end of the last sample, including any intervals between them, as it is supposed to represent the load on the server. The samples are given by the number of users multiplied by the number of the requests of the scenario of application use.

In terms of performance impact, the proposed solution has very low impact when few users are using the web application (see Figures 5-14a, 5-14b and 5-14c). Although in some cases the increased time corresponds to a high percentage (for example response time increasing from 10 to 20 milliseconds represents a 100% increase), the difference in absolute values, considering that time is measured in milliseconds, is practically irrelevant. Hence, the average time without the access control mechanism is very similar to the other results up to around 16 threads. As the number of threads increases, differences arise. The inclusion of this privacy protection mechanism affects the performance for higher number of threads but the

system performance increases linearly for a high demand. The performance impact in the throughput analysis (see Figures 5-14d, 5-14e and 5-14f) is also similar to the average processing time results, i.e., for the 16 first samples the results with and without the access control are similar and the differences arise as it increases. We believe the differences between the throughputs (for different number of records) arise due to the randomness of criticality levels, which may not have a realistic distribution of the values. The criticality levels are generated randomly and exclude the level 1, once information tagged with this level do not require privacy protection. So, the less level 1 generated the more level are recorded and the more processing time is necessary to protect the data.

It is worthwhile to mention that many access control solutions are already being used and consolidated (e.g., XACML, P-RBAC). First of all, our solution is not meant to replace them, but to provide a lightweight solution that can be integrated into a database with reduced performance overhead. Second, the introduction of this mechanism also served the purpose to show how our comprehensive approach can be applied to a real case-study. This gives an indication that the approach is, at least for this particular study, feasible and applicable.

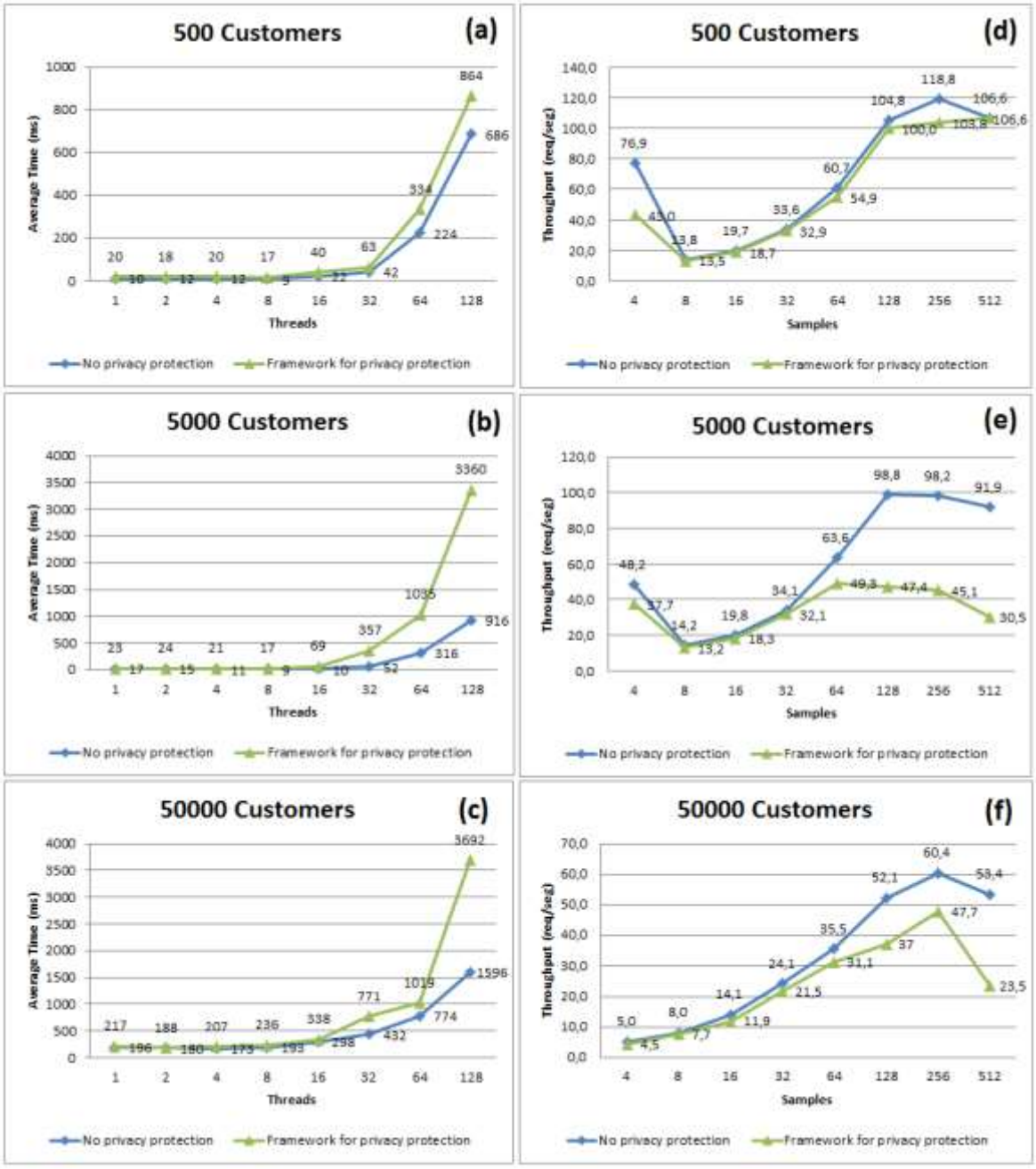

Figure 5-14. Access control experiments: average processing time and throughput.

5.4 CONCLUDING REMARKS

A case study applying the PrivAPP to provide data privacy protection in a web application was presented in this chapter. We describe the web application used in the case study (which is an online book store), its respective software architecture and privacy policy. Then the UML diagrams were created, based on the Privacy UML Profile. In sequence, the

software architecture that represents the application with privacy protection was defined, based on the Privacy Reference Architecture. Based on the UML diagrams and the software architecture, we implemented a database framework for access control. An experimental evaluation was performed to assess the scalability and performance impact of the proposed solution.

6 EVALUATION OF THE APPROACH

In addition to the case study described in Section 5, aimed at evaluating the applicability of the approach, we deemed useful to identify, concerning quality attributes, strengths and weaknesses of PrivAPP. This would give an indication of its potential success regarding privacy protection. We start presenting the evaluation process of our Privacy Reference Architecture. Following, we describe the evaluation process of the proposed approach PrivAPP as a whole. We used different approaches, as theoretical and empirical approaches, based on the work of Angelov and Grefen (2008).

6.1 PRIVACY REFERENCE ARCHITECTURE EVALUATION

To the best of our knowledge, there are no methods for holistic evaluation of this kind of approaches. Angelov and Grefen (2008) present general qualities required for their reference architecture. Based on their work, we used adapted techniques to evaluate our Privacy Reference Architecture ***completeness, usability*** and ***applicability***.

(i) Functional Completeness. For the evaluation of the proposed privacy reference architecture in terms of *functional completeness*, we have approached the problem with two different but complementary ways: theoretical and empirical (scenario-based).

As the **theoretical approach,** we used the Antón and Earp (2004) and Solove (2006) taxonomies to evaluate the completeness of the privacy reference architecture. They were selected because they are the only two privacy violation taxonomies that directly consider, or are immediately applicable, to information privacy (Reddy *et al*., 2008). Antón and Earp (2004) provide a taxonomy focused on the analysis of website privacy requirements, while Solove (2006) synthesizes American law and legal literature.

For both privacy violation taxonomies, we mapped each category to the architecture elements in the presentation, application and persistence layers. In order to exemplify and enhance understanding, we believe that, for now, presenting only one of the taxonomies is enough. We present here only the Antón-Earp's taxonomy and the correspondent privacy reference architecture mapping (Table 6-1). We decided to show this

taxonomy as it focuses on website privacy. Solove's taxonomy and correspondent mapping can be found in APPENDIX B, Table B-1.

Table 6-1. Antón-Earp's taxonomy and the privacy reference architecture correspondence.

Taxonomy Category	Category Description	Privacy Reference Architecture element
Protection goals	Safeguard the privacy of a customer's data	-
Notice / Awareness	Describe how a customer is informed about an organization's practices regarding their data	Privacy Policy Definition
Choice / Consent	Describe a customer's ability to choose how they want their data to be managed by organization	User Preferences
Access / Participation	Reflects a customer's ability to challenge, correct or modify their data as used by an organization	Privacy Policy Definition / Enforcement
Integrity / Security	Describe measures an organization takes to protect the accuracy and security of a customer's data	Security /Cryptography / Auditing
Enforcement / Redress	Describe the ways that organization approaches internal policy violations by their employees	Access Control
Vulnerabilities	Reflect potential privacy violation	-
Information Monitoring	Describes how an organization tracks customers' interaction with their website	Tracking Detection
Information Aggregation	Reflects the ways that an organization will combine customer data with third-party data sources	Privacy Policy Enforcement
Information Storage	Reflects practices regarding what/how customer records are stored in the organization's database	Privacy Policy Definition / Enforcement
Information Transfer	Describes how an organization may share their collected customer information with third-parties	Privacy Policy Definition / Enforcement
Information Collection	Shows what types of information an organization may collect and how to collect	Privacy Policy Definition / Enforcement
Information Personalization	Reflects the methods an organization uses to the presentation of their website to their customer	Privacy Policy Definition / Presentation
Solicitation	Shows the purposes and methods an organization would use to contact their customers	Privacy Policy Definition

The first and second columns of Table 6-1 represent the taxonomy category and their description, as extracted from Antón and Earp (2004). The third column represents the elements in our reference architecture that correspond to each taxonomy category. The correspondence was done based on the description of both taxonomy category and privacy reference architecture element. For example, Notice/Awareness is the taxonomy category that *describes how a customer is informed about an organization's practices regarding their data*. Similarly, the element Privacy Policy Definition in the Privacy Reference Architecture *is responsible for privacy policies to be defined and presented to the user.*

As can be observed, our privacy reference architecture provides support for each category of privacy violation in the referred taxonomy.

As the **empirical approach** to evaluate the completeness of our privacy reference architecture, we designed and conducted a workshop with potential users, with the goal of analyzing the architecture through the interests and experience of the participants. In the

workshop we presented the privacy reference architecture to 8 representative stakeholders interested in personal information privacy protection. Six of them are IT professionals and work in big IT companies in Brazil. Two of them are professors and academic researchers. All these professionals have, in average, five year of work experience. Obviously, this group is not representative of all the potential stakeholders of the proposed architecture, but we consider it sufficient to obtain some indication about its level of completeness (addressing all stakeholders would indeed be impossible). After an adequate introduction, we asked the participants to suggest possible scenarios for protecting privacy of personal information. The result was a set of 14 use-case scenarios, presented in Table 6-2.

Table 6-2. Scenarios defined by the workshop participants

Scenario	Votes	Privacy Reference Architecture element
1. Defining policies easy to read and understand	4	Privacy Policy Definition
2. Allow data subjects to express their preferences related to their personal information	7	User Preferences
3. Inform changes in the policy to data subjects	2	Privacy Policy Management
4. Allow data subjects to access/modify/exclude their data	5	Privacy Policy Enforcement
5. Allow data subjects to define the period their data will be stored	2	User Preferences
6. Use resources of user pattern identification to identify data subjects' unusual operations (prevent identity theft)	2	User Pattern Identification
7. Allow data subjects to opt in or opt out about receiving advertisements and respect the decision.	1	User Preferences / Privacy Policy Enforcement
8. Not make the data available to third parties (including apps) without data subjects' authorization	6	User Preferences / Privacy Policy Enforcement
9. Privacy Policies must comply the legislation	2	Privacy Policy Definition
10. Privacy Policies must be enforced, respecting data subject preferences	2	Privacy Policy Enforcement / User Preferences
11. Use access control resources to assist privacy policy enforcement	1	Access Control
12. Use resources of cryptography to protect personal information	3	Cryptography
13. Use resources of security to protect personal information	1	Security
14. Use auditing resources to identify the sources of privacy violation	1	Auditing

As shown, the scenarios were defined in the form of features a web application or service should implement. The votes in the second column represent the number of different stakeholders that proposed each scenario. The third column represents the elements of our reference architecture that correspond to each use-case scenario. The correspondence was done based on the description of both scenario and reference architecture element. For example, to the scenario 3 *(Inform changes in the policy to data subjects)*, the element Privacy Policy Management *is responsible for updates in the privacy policies, which should be managed. The updates in the privacy policy must be informed to the users and new preferences about these updates must be considered.*

For all the scenarios the architecture has a correspondent element. The most voted scenarios are number 2 (*allow data subjects to express their preferences relate to their personal information*) with 7 votes and number 8 (*do not make the data available to third parties (including apps) without data subjects' authorization*) with 6 votes. These scenarios are related to the control that the users want to have about their personal information, reinforcing the importance of this set of requirements.

Besides the workshop, we conducted a second meeting with all the stakeholders to validate the scenarios identification. The scenarios were prioritized in order of importance. In the majority of the responses (approximately 75%), the scenarios Number 1 (*Defining policies easy to read and understand*) and Number 2 (*Allow data subjects to express their preferences relate to their personal information*) are classified as the two most important ones. Again, they represent concerns about the user's control over their personal information.

The results from this workshop indicate that our privacy reference architecture addresses the functionalities that potential stakeholders believe to be necessary to protect personal information.

(ii) Usability. While presenting our privacy reference architecture to a set of IT professionals, we wanted to investigate if they could easily understand the major functionalities specified in the privacy reference architecture. A questionnaire was then elaborated based on the work of Padilha (2004). It has statements about the presentation and the contents of the privacy reference architecture. Examples of these statements are "*the reference architecture has a pleasant and readable graphical presentation*" and "*understanding each element is easy*". For each statement, the stakeholders could answer using the following levels: *strongly agree, agree, undecided, disagree* or *strongly disagree*. The complete questionnaire can be found in APPENDIX B, Table B-2.

From the analysis of the answers, 100% of the stakeholders agree or strongly agree that they liked the presentation of the Privacy Reference Architecture and that it is readable. Also, 100% agree or strongly agree that the presented architecture is quite relevant for understanding the privacy domain and to construct web applications and services with privacy protection requirements. About the content of the privacy Reference Architecture, 75% of the stakeholders agree that it is clear and consistent, while 25% are undecided.

This experiment suggests that the proposed privacy reference architecture uses naming conventions, notation, and a structure that are understandable by the stakeholders.

Clearly, as these results were obtained from a limited set of stakeholders, they can be considered only as an indication of the usability of the proposed privacy reference architecture.

(iii) Applicability. Regarding the applicability of the privacy reference architecture, we compared the functionalities it supports with the functionalities addressed/defined in existing commercial and academic concrete architectures. This comparison allows us to demonstrate the applicability of our privacy reference architecture as an analytical tool.

To support the comparison, we selected two commercial privacy concrete architectures: The IBM Enterprise Privacy Architecture (Bücker *et al.*, 2003) and HP Privacy-Aware Access Control architecture (Mont *et al.*, 2005). Both were selected due to their availability and the importance of these companies in the current IT market. Also, we selected an academic architecture (Bodorik and Jutla, 2008). This architecture is agent-based and support privacy in an environment based on web services.

We mapped the functionalities of each component of the three selected privacy architectures to the proposed privacy reference architecture elements. Again, in order to exemplify and enhance understanding, we present only one case: the IBM Enterprise Privacy Architecture (Bücker *et al.*, 2003) components and the privacy reference architecture mapping. Results are shown in Table 6-3. We decided to show this architecture as it is more complete than the other ones. However, the mapping to the remaining architectures (Mont *et al.*, 2005; Bodorik and Jutla, 2008) can be found in APPENDIX B, Table B-3 and Table B-4, respectively.

The first column of Table 6-3 shows the components of the IBM Enterprise Privacy Architecture, followed by their brief description in the second column. The third column represents the elements of our privacy reference architecture that correspond to the IBM's architecture components. The correspondence was done based on the description of both components from the concrete and reference privacy architectures. For example, the Policy presentation/negotiation from IBM architecture *displays the Privacy Policy and the user can opt-in or opt-out for certain procedures*. Similarly, in the Privacy Reference Architecture the element Policy Presentation *refers to the fact that the web application must provide this document to customers and business partners*. Also, the User Preferences element *refers to the need for the web application to allow users to state their privacy preferences regarding personal information, agreeing or not with the presented policies (or part of them: the statements).*

Table 6-3. IBM Enterprise Privacy Architecture and the privacy reference architecture correspondence

	Component	Description	Privacy Reference Architecture element
User Interaction Node	Policy presentation / negotiation	Display the Privacy Policy and the user can opt-in or opt-out for certain procedures	Policy Presentation / User Preferences
	Privacy Action Manager	Allows data subjects actions on their own PII (retrieve, update, etc.)	Policy Enforcement
	Privacy Contact Manager	Contacts the data subject in case of an enterprise-triggered event concerning privacy (for example, privacy policy changes).	Policy Management
Privacy Data Handling Node	Privacy-enabling Resource Manager	Contains PII and gives users access to it according to privacy rules.	Policy Enforcement /Access Control
	Deployment Engine (optional)	Translates specific technical terminology to generalized access requests.	--
	Privacy Policy Decision Point	Evaluates rules and returns a decision ("allowed" or "denied") based on policy.	Policy enforcement
	Privacy Data Transformation Engine	Applies privacy-related transformations to PII. (for example, depersonalized or totally anonymized way).	Anonymization
Privacy Service Node	Privacy Policy Manager	Stores privacy policies, including data-subject provided parts of them, ranging from simple consent over opt-in or opt-out choices	Policy Definition / User Preferences
	Privacy Action Audit Manager	Logs access to PII.	Auditing
	Privacy Obligation Event Services	Keep track of privacy-action obligations	Policy Enforcement
Directory and Security Node	Privacy-enabled authentication	Extension of security authentication to include one based on pseudonyms	Identity Management
	Identity mapping	Maintain information to link entries in privacy-enabling Resource Managers	---
	Attribute Exchange Engine	Supports the exchange of attributes between different organizations.	Policy Management
	Privacy-enabling Credential Service	Specific type of engine that supports the generation and verification of credentials.	Identity Management

Correspondences marked with a dash (-) means that this feature is not a reference architecture concern as it is too detailed for its abstraction level. Thus, it should not be represented. As shown, our privacy reference architecture provides support for all functionalities in the referred IBM's architecture.

A summary of all the mappings is shown in Table 6-4. The elements of our privacy reference architecture are listed in the first column and the cells marked with an "x" represent the cases where the component has a correspondent feature in the work (taxonomies, scenarios or concrete architectures) stated in the first row of the same column. Also, the relation between the elements of the architecture with the established architectural requirements is presented in the last column. From this summary, we observe that our reference architecture is more complete than any existing concrete architecture as it includes functionalities to protect personal information privacy that none of the analyzed architectures

consider: the *Web Browser Security Configurations* and the *Privacy Violation Monitoring* elements have no correspondence in any other work.

Table 6-4. Summary of Privacy Reference Architecture elements' correspondences.

Privacy Reference Architecture element	Antón-Earp's taxonomy	Solove's taxonomy	Scenarios by stakeholders	IBM's architecture	HP's architecture	Bodorik and Jutla's architecture	Architectural Requirement
Web Browser Security Configurations							PAR-4
Privacy Policy Presentation	x			x			PAR-8
User Preferences	x	x	x	x	x		PAR-12
Privacy Policy Definition	x	x	x	x	x	x	PAR-8
Privacy Policy Enforcement	x	x	x	x	x	x	PAR-8, PAR-1
Privacy Policy Management			x	x			PAR-8
Privacy Violation Monitoring							PAR-1
Activity Tracking Detection	x	x					PAR-2
Cryptography	x		x				PAR-6
Auditing	x		x	x	x	x	PAR-11
Attack Detection	x	x	x				PAR-3, PAR-4
User Pattern Identification	x	x	x				PAR-5, PAR-4
Access Control Policy Definition	x	x	x	x	x		PAR-10
Access Control Policy Enforcement	x	x	x	x	x		PAR-10
Identity Management				x			PAR-9, PAR-5
Anonymization		x		x			PAR-7

6.2 PRIVAPP EVALUATION

Also based on Angelov and Grefen (2008), we evaluated the approach as a whole. We considered ***completeness*** and ***applicability*** as the most important quality attributes to be evaluated for PrivAPP. For the evaluation of these qualities, we applied an empirical approach in a case study. The case study consists of selecting privacy policies from relevant companies and analyzing them, to check if elements from PrivAPP can help to enforce these policies. The idea is to split the privacy policy into statements and, for each statement evaluate: first, if the semantics allows users to express their preferences, by agreeing or not with the statement; second, to evaluate if PrivAPP provides elements to help the enforcement of the statement, abiding by the user preference (when it can be expressed). Details of the evaluation process are described in the next sections.

6.2.1 Evaluation Setup

As the number of e-commerce websites is very large, we established, empirically, 20 privacy policies as the target of the analysis. We cannot generalize the results of this case study to the universe of e-commerce companies for which privacy is very important. The study is meant as a "proof of concept" of the suitability of the approach for companies similar to the ones in the sample.

The two main criteria for selection of the companies and their corresponding privacy policies are:

(i) Laws and regulations. Legislative approaches differ from country to country and the privacy policies are (or should be) based on these approaches. Taking this matter into account we decided to select companies from different countries, including Brazil, USA, and European Union countries. The European Union and the United States are the two most prolific sources of laws and relevant statutes on privacy (Hinde, 2003; Perkins and Markel, 2004). Brazil was chosen because we want to investigate privacy policies from our country. The goal is to assess if the difference in the laws makes it necessary to add new elements for privacy protection in the model's approach.

(ii) Size and market segment. We selected companies that are "top-of-mind" with respect to volume of sales and consumer preference. To verify if the model can be applied to a diversity of companies, we decided to investigate different market segments, including electronics, tourism products, cosmetics, furniture, etc.. Results of ranking researches support this selection (G1, 2014; Webshoppers, 2015; R7, 2015). We did not include Amazon (Amazon, 2014) in our list as its privacy policy has already been used in the profile construction. The result of the selection is shown in Table 6-5.

The selected companies and their market segments are described in Table 6-5. As most of this research was performed in Brazil, we have adopted the perspective of a Brazilian user. Thus, the columns *Brazil* and *Other Countries* describe, respectively, where the companies operate. Five companies operate only in Brazil (Americanas, Casas Bahia, CVC, Cia Dos Livros, Submarino), nine companies operate in Brazil and other countries (Walmart, Dafiti, Decolar, Aliexpress, E-bay, DealeXtreme, Mercado Livre, OLX/Bom Negócio, Carrefour) and 6 companies **do not** operate in Brazil (Brigette's Boutique, Drugstore, Topshop, Media Markt, Worten, Selfridges).

Table 6-5. Companies (and privacy policies) selected to support the evaluation of PrivAPP.

Company	Market Segment	Brazil	Other Countries	Origin	Policy Size	Statements
Americanas	Wide variety of products such as books; games; Cine & Photo; Mobile Phones; Electronics, etc.	yes	no	Brazil	S	5
Casas Bahia	Home appliances, electronics, furniture and housewares.	yes	no	Brazil	M	15
CVC	Tourism products and services.	yes	no	Brazil	M	17
WallMart	Wide variety of products such as electronics, home appliances, computers, mobile phones; etc.	yes	yes	USA	L	22
Dafiti	Shoes, clothes, accessories, sports products, perfumes, beauty products and decorative items	yes	yes	Brazil	S	6
Cia dos Livros	Books	yes	no	Brazil	S	8
Decolar	Tourism products and services	yes	yes	USA	M	17
Aliexpress	Wide variety of products such as clothing, accessories, cars, motorcycles, cell phones, electronics, etc.	yes	yes	China	L	26
Brigette's Boutique	Cosmetics, makeup, hair products	no	yes	USA	M	16
E-bay	E-commerce solutions to help individuals and companies to buy and sell products via Internet	yes	yes	USA	L	45
Submarino	Wide variety of products such as books; games; mobile phones; electronics; watches, etc.	yes	no	Brazil	S	4
DealeXtreme	Wide variety of products such as electronics, phones, electrical tools, car accessories, etc.	yes	yes	China	L	32
Drugstore	Health, beauty, vision, and pharmacy products.	no	yes	USA	L	22
Mercado livre	E-commerce solutions to help individuals and companies to buy and sell products via Internet.	yes	yes	Argentina	L	43
OLX / Bom Negocio	E-commerce solutions to help individuals and companies to buy and sell products via Internet.	yes	yes	Argentina	M	20
Topshop	Clothes, shoes, bags and accessories, makeup.	no	Yes	United Kingdom	M	14
Media Markt	Wide variety of products such as flat-screen TVs, tablets, smartphones, coffee makers, etc.	no	yes	Germany	M	19
Worten	Home appliances, consumer electronics and entertainment.	no	yes	Portugal	S	7
Selfridges	Clothes, bags, makeup, cosmetics, perfumes, home appliances, mobile phones, tablets, wines, etc.	no	yes	United Kingdom	M	16
Carrefour	Supermarket, gas stations, drugstores and financial services.	yes	yes	France	S	9

In the *Origin* column we can observe that six companies were originated in Brazil (Americanas, Casas Bahia, CVC, Dafiti, Cia dos Livros, Submarino), five were originated in the USA (Walmart, Decolar, Brigette's Boutique, E-bay, Drugstore), two were originated in China (Aliexpress and DealeXtreme), two in Argentina (Mercado Livre and OLX/Bom Negócio), and five were originated in European Union countries (Topshop, Media Markt, Worten, Selfridges, Carrefour). This information was found in the *about us* links in the companies' websites.

The size of the privacy policy of each company (*Policy Size* column) is classified as: small (S), if the policy has 10 or less statements; medium (M), if the policy has from 10 to 20 statements; and large (L), if the policy has more than 20 statements. Thus, we have 6 small policies, 8 medium and 6 large ones. 351 statements were analyzed.

6.2.2 Analysis of PriVAPP's Elements Versus Policies Statements

Table 6-6 shows the results of the analysis of the companies' policy statements and the elements of our approach. Just for better organization, we have split the approach's elements into two groups: the *fundamental elements* and the *enforcement elements*. The numbers represent the frequency that each element is associated in a privacy policy. Although the privacy policies are publicly available, the companies are not explicitly identified to assure neutrality; furthermore, they usually do not allow the disclosure of results of this type of evaluation. Hence, the companies will be referred numerically, from this point on, from 1 to 20, in no particular order. We also assume that all of them comply with their privacy promises. Results are discussed in the next subsections.

Table 6-6. PrivAPP elements versus the companies' privacy policies.

	Element	Companies																				Total
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	
Fundamental elements	Privacy Policy Definition	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
	Statement	2	10	4	12	3	2	5	8	5	19	2	11	7	6	16	4	5	4	17	7	149
	Disclosure	1		1	1	1	2	4	8	3	13	1	1	2	3	1	1	2	1	7	3	56
	Retention		1	3	2			2			2		6	2		1	1		1	3		24
	Collection	2	4	4	5	1	3	5	9	6	6	1	7	10	3	2	1	2	9	8	2	90
	Usage		1	2	3	1	1	4	3	1	4		6	1	2	2	1	1	2	9	8	52
	Recipient	1		1	1	1	1	4	8	3	13	1	1	2	2	1	1	2	1	3	3	50
	Service	1	1	2	3	1	1	2	2	1	6	1		3	2	2		1	2	2	1	34
	Private Data		1	1	1		1	1	1		1	1	1	2			1	1		1		14
	Usage Information	1	1	1	1	1		1		1	1	1	2	1	2	1		1	5	1	1	23
	Personal Identifiable Information			1			1	1	3	1	2		1	1	1	1			1	1	1	16
	Preference	1	2	3	3	1	1	7	3	5	11	1	4	7	6	3	1	1	8	9	6	83
Enforcement elements	Enforcement												1									1
	Tool																					0
	Activity Tracking Detection	1	1	1	2	1	1	2	1	1	2	1	3	4	2	1				3	1	28
	Privacy Violation Monitoring/Detection	1			1					1	1	1							1	1	1	8
	Security Measure		1	2	3	1			2	2	2		1	1		1				2		18
	Attack Detection		1	1	2				1		1		1	1						2		10
	User Pattern Identification	1	1	1	1		1	1	1			1								2		10
	Auditing	1	2	2	3	1	2		1	2	7	1	6	4	4	2	2	1	1	11	2	55
	Action		1	1	2	1		2	1		9		1	4	3	4	3	1		8	2	43
	Process																					0
	Access Control	2		2	2			2	3	3	6	1	4	2	2				1	3	2	35
	Access Control Policy	2		2	2			2	3	3	6	1	4	2	2				1	3	2	35
	Identity Management												1							1		2
	Cryptography		1	1	1	1			1	1	1		1	1		1				2		12
	Anonymization		1		1	1			2									1				6
	Algorithm																					0
	Config	1	2	1	1	2	1	4	2	7	9	1	6	10	3	5	1	3	9	8	5	81
	Management	1	1		3		1	4	3	2	2	1	2	4	2	1		1				28
	Managed Interaction				2			3	2	1	2		1		1							12
	Total	20	34	38	59	19	20	57	69	50	127	18	73	72	47	46	18	24	48	108	48	

6.2.3 Fundamental Elements and the set of Policies

The most frequent element found in the set of *Fundamental Elements* is *Statement* (see total in last column of Table 6-6). *Statements* are generic and do not refer to any of their

specializations. Examples of *Statements* are: "*The User guarantees the truthfulness and accuracy of the personal data he/she provides to XXXX and assumes the corresponding responsibility*"; *"This online privacy policy applies only to information collected through our website and not to information collected offline.",* where XXXX is the name of the company, omitted here for reasons given previously. The high frequency of this element is due to the necessity of including more information in the privacy policy than that related specifically to private data (collection, retention, usage, disclosure).

The second more frequent element is *Collection*. All the policies we analyzed have at least one statement that refers to data collection. These statements can specify the collection of personal identifiable information, users' activities (e.g., links they click or sites they access), users' system information (IP address, operating system, web browser) or even generic data. Example of collection statements: *"Information including, but not limited to, user name, address, phone number, fax number and email address ("Registration Information") may be collected at the time of user registration on the XXXX."; "We record and retain details of users' activities on the XXXX. […]"*. 95% of the policies state that the collection of users' activities and system information is done through cookies, web beacons and similar technologies. From the point of view of policy evaluation, we believe this result is a good indication that the privacy policies are informing the users about processes that are transparent to them (i.e., they have no idea these cookies are being recorded in their web browser until they have already been recorded). With this information, it is possible for users to take actions to avoid these processes if they want to.

Preference is the third more frequent element found in the *Fundamental Elements* set. We consider as *Preferences* statements that are likely to offer the user the option of choice, i.e., to agree or disagree (opt-in or opt-out) with the referred statement. Examples: *"We may also send you from time to time (by email or post) information about products and services and details of promotions and special offers from XXXX"; "A Cookie is […]. We use Cookies to keep track of your current shopping session […]"*. For all these statements classified as *Preference*, users could say that they opt-out, i.e., they do not want to receive e-mails with advertisements or to have their activities tracked. In these cases, the companies need to take some actions to respect these preferences (the enforcement elements of our approach can help in this direction). *Preference* being a special kind of statement (i.e., it refers to the need for the web application to give the option to the user and take different actions according to this option), some statements are classified simultaneously as two elements of

the approach, e.g., a statement can be a *Collection* statement and a *Preference* statement at the same time.

6.2.4 Enforcement Elements

Config is the *Enforcement* element that has been most widely used in our analysis, with 81 occurrences. We associated the *Config* element with statements that allow two types of configurations to enforce the policy, even in the cases where users express their preferences: *web browser configurations* and *user configurations*. *Web browser configurations* belong to the Reference Architecture and, although it is a configuration outside the system (i.e., each user must configure its own web browser), this is an important resource that must be made explicit at least in the privacy policy. Guiding the user to configure their web browser would help to respect their privacy, especially when they do not want to receive cookies or to be tracked. *User configurations* is an instance of *Config* created to represent the enforcement of statements that allows users to refuse some services such as, for example, advertisements or cookies and similar tracking technologies. Example of statements for user configurations: *"If your personally identifiable information changes, or if you no longer desire our service, you may correct, update, request deletion, or deactivate it by making the change on the "your account" page or by e-mailing us at privacy@XXXX.com"*.

Auditing is the second more frequent enforcement element, with 55 occurrences. As the privacy policies must abide by the laws and regulations, statements refer often to the use of private data to comply with them. Thus, we must check if data are really being used for these legal purposes. A statement which needs to be audited to be enforced is, for example: "*Your Data may be retained beyond the expiry of its purpose if that is required by law, such as a provision of a statute, or a court order such as a search warrant or subpoena, or a warning by a law enforcement agency that delivery of a court order is imminent*". Auditing is also used for enforcement of statements that disclose private data with specific goals, e.g., *"As part of the customer data management, the data collected will be transmitted to third parties, the transport companies, for the exclusive purpose of the realization of the services or products purchased by the user"*. Auditing can be a complex and expensive resource, but it is necessary especially in the cases prescribed by law.

Action is an element representing mechanisms the system could implement to help to protect privacy. There is a wide variety of instances and some we propose are: *notify*

policy changes (to notify the users in case of changes in the privacy policy and, if necessary, to ask them to express their new preferences); *inform user about automatic collection* (to inform the user when cookies are sent or other mechanisms will track the activities, allowing the user to express their agreement or disagreement); *do not send text message* (when statements say text messages will be sent to the cell phone and the user disagrees).

Access Control and corresponding *Access Control Policy* elements have also been widely used. We adopt these elements in cases where statements determine that only qualified and authorized staffs are allowed to access personal data and in cases where they mention the disclosure of private data to third-parties. It is evident that controlling disclosure goes far beyond access control; the best we can do is to use the most adequate element from PrivAPP.

6.2.5 Quality Attributes and Improvement of the Approach

Regarding the *applicability* attribute, we want to assess if our approach can be applied to the design and analysis of privacy systems, specifically for privacy enforcement. Thus, we addressed the enforcement functionalities our approach supports with the statements in the commercial privacy policies. In Table 6-6 we can observe that, for the 351 statements analyzed, we used 384 enforcement suggestions (some statements require more than one enforcement). Also, the enforcements considered in our approach were applicable to all the companies, varying from the minimum of 6 (companies number 6 and 16) and to the maximum of 48 (company number 10) suggestions per company. This analysis indicates that, for the set of analyzed policies, the model can be applied as an analysis tool.

Regarding the *completeness* attribute, we want to verify if the proposed approach contains all the information necessary for helping privacy protection, according to the privacy policies we analyzed. In Table 6-6 we can observe, in the last column, that almost all elements were found in the policies, except *Tool*, *Process* and *Algorithm*. We did not find any specific correspondent element for these ones, but we used their specializations (*Activity Tracking Detection, Privacy Violation Monitoring/Detection, Access Control, Identity Management, Cryptography, Anonymization*). Thus, the approach still offers generic elements that could be used for statements referring to resources that we can associate to them and that are not too specific as their specializations. This is a good indication that the specializations of these elements have a good level of completeness.

The intention of this evaluation activity is also to determine if the approach and its models need improvement. We did not find any new element that could be added to the approach. This indicates that the model is fairly complete for this set of policies. However, we used frequently instances of the *Config* element and we could frequently identify two categories of configurations: *Web browser configurations* and *User configurations* (their descriptions are in Section 6.2.2). Hence, the conceptual model of PrivAPP was complemented by implementing both these specializations (see Figure 4-2. They are called *Web Browser Config* and *User Config*, respectively).

In spite of the high frequency of the *Statement* element in this study, we could not identify groups that could represent other specializations than the ones already present in the PrivAPP conceptual model.

6.2.6 Applicability Analysis

In addition to the elements' analysis, a case study was performed to get an indication of PrivAPP's applicability. The goal of this case study is to guide the requirement elicitation within the system development phase. Three IT professionals – who work in the same company and have the same position, i.e., with a similar background and knowledge – were asked to identify the functional and non-functional requirements for the development of an online bookstore. They work at a university's data center and are in charge of network support and software development.

Two of these professionals were given PrivAPP as a support for requirements identification. The third professional performed the same task without knowing PrivAPP. For all of them we provided the system description, the privacy policy and a requirements template and asked them to, freely, describe the requirements for developing the application. The results are shown in Table 6-7. Obviously, three professionals is a very small sample and the case study must be scaled to a much larger number of them; different backgrounds are also in order. The case study was performed to give just a very preliminary indication of PrivAPP's applicability.

Table 6-7. PrivAPP's elements used by the professionals for the applications requirements.

	Element	Professionals Interviewed			Total
		P1	P2	P3	
Fundamental elements	Privacy Policy Definition				
	Statement				
	Disclosure	1			1
	Retention	1			1
	Collection	1	1	1	3
	Usage				
	Recipient				
	Service				
	Private Data	2	1		3
	Usage Information				
	Personal Identifiable Information				
	Preference	1			1
Enforcement elements	Enforcement				
	Tool				
	Activity Tracking Detection				
	Privacy Violation Monitoring/Detection				
	Security Measure	1	5		6
	Attack Detection	1	1		2
	User Pattern Identification	2			2
	Auditing	1	2		3
	Action				
	Process				
	Access Control	2	2	1	5
	Access Control Policy				
	Identity Management				
	Cryptography	3	3		6
	Anonymization		1		1
	Algorithm				
	Config				
	Management				
	Managed Interaction				
	Total	16	16	2	34

The second column in Table 6-7 presents PrivAPP's elements; they are grouped into *Fundamental* and *Enforcement* elements (first column). For each participant (the three professionals) we reported the frequency each element of the approach was mentioned in the requirements (third, fourth and fifth columns, respectively). The professionals we called *P1* and *P2* are the ones who had access to PrivAPP; *P3* is the one who have no knowledge about the proposed approach.

Professional *P1* defined 13 functional requirements, 3 non-functional requirements and 4 business rules. In this specification, the elements of PrivAPP were

mentioned 16 times; the most frequent is *Cryptography*, followed by *Private Data*, *User Pattern Identification,* and *Access Control*. Professional *P2* defined 5 functional requirements, 8 non-functional requirement and zero business rule. In P2's specification, the elements of PrivAPP were also mentioned 16 times; *Security Measure* is the most frequent, followed by *Cryptography, Auditing,* and *Access Control*. Professional *P3* defined 18 requirements, zero non-functional requirements and zero business rules. In his specification, requirements are succinct and do not include privacy or security concerns explicitly. Nonetheless, we considered 2 mentions of concerns that can be represented by elements of the PrivAPP: *Collection* (related to the requirement he called "Register User") and *Access Control* (related to the requirement he called "Login"). All the requirements identified by the professionals are shown in APPENDIX E; since the participants are Brazilians and we want to keep the original answers, they are in Portuguese.

In the last column of Table 6-7 we present the number of times the elements were identified in the requirements. The most frequent are *Cryptography* and *Security Measures,* mentioned 6 times each, followed by *Access Control*, mentioned 5 times. It is notable that the two professionals who used PrivAPP were able to identify much more privacy concerns and include more privacy protection in their application requirements (Table 6-7, last line). Thus, these results allow us to infer that the proposed approach can be very useful to guide the requirement elicitation when privacy protection is mandatory in the system being developed, i.e., PrivAPP has a positive applicability in improving the quality of the requirement elicitation. Also, the experiment gives an indication that the proposed approach can be useful in the construction of web applications with privacy protection.

After the requirements definition, we applied a questionnaire to professionals P1 and P2. The goal is to evaluate their perceptions regarding the presentation and contents of the PrivAPP. The questionnaire has statements such as, for example, "*the content presented in the approach is clear and consistent*" and "*the approach is quite relevant and should be used to construct web applications and services with privacy protection*". For each statement, they could answer using the following levels: *strongly agree, agree, undecided*, *disagree* or *strongly disagree*. The complete questionnaire is also in APPENDIX E.

From the analysis of the answers, both professional agree or strongly agree that (i) the content of PrivAPP is clear and consistent; (ii) understanding the elements of the approach is easy; (iii) the approach is quite relevant and should be used to construct web applications and services with privacy protection; (iv) it is easy to add new elements, if necessary. Also, both consider that the approach helps developers and stakeholders to pay attention to privacy

protection in the beginning of the software construction process. This can avoid many consequences of privacy violation and reduce costs (to include privacy protection in an information system that is already in use can be very costly).

6.3 LIMITATIONS OF THE APPROACH

Although the effort of assessing PrivAPP showed relevant results, it is not enough to make the approach widely accepted by the scientific community. First, the sample is not large enough to be representative of the considerable amount of web applications and corresponding privacy policies throughout the world. Second, our case study addresses only part of the approach. Furthermore, an approach such as the one we propose can be only proven worthwhile after a long period of utilization, by several different stakeholders, with different skills, in a long and complete software development lifecycle, considering improvements and software maturity levels. It is not possible to have this type of evaluation in the time span of a dissertation. Hence, the results are a "proof of concept" as well as a preliminary indication of the approach's applicability, completeness and feasibility. Furthermore, although the approach has been shown to possess a good level of completeness, it cannot be considered exhaustive, especially concerning privacy enforcement technologies. Different or new technologies can be necessary but they can be added to the approach without much effort, due to its extensibility.

Other limitations identified are related to the characteristics inherent to privacy:

- Many UML Profiles provide automatic code generation to implement the solution from the models produced. However, due to the high level of abstraction inherent to privacy concepts, the models produced from our Profile cannot provide code generation.
- As discussed in Section 2.6, there is a huge difficulty in having machine-readable privacy policies. Splitting the privacy policy into statements in an automatic ways is not part of the scope of PrivAPP. This was done empirically in the evaluation process.
- Privacy is a relatively recent concern; it is affected by the emergence of new technologies. The tendency is the evolution of the concepts concerning privacy. The proposed approach is new and based on the most recent privacy concepts but, over time, it will certainly need to undergo changes and adaptations.

6.4 CONCLUDING REMARKS

In this chapter we performed an evaluation process for the proposed approach. We evaluated two quality attributes (completeness and applicability) through an empirical approach. To evaluate the completeness we selected privacy policies from relevant companies and analyzed them, checking if elements from PrivAPP can help to enforce these policies, abiding by the user preference (when it can be expressed). To evaluate the applicability we conducted a case study to guide the requirement elicitation within the system development phase (two groups of professionals were asked to define requirements for an online bookstore and, for only one group we provided the PrivAPP. Then we evaluated the differences in the requirements they defined). This evaluation process helped to improve the PrivApp. Finally, we described the limitations of the approach.

7 CONCLUSIONS

In this dissertation we discuss aspects concerning privacy within the scope of web applications and services. In times when digital information has immense value and privacy is a *must have*, the goal of this work is aimed at improving the scenario of lack of privacy protection in the construction of web applications and services. Our approach is a reference model which systematizes privacy concepts in the referred scope and serves as a guideline for modeling, designing and, ultimately, implementation of web applications and services with privacy protection features. By means of a case study and an evaluation process, we got an indication that this approach has a good level of completeness and applicability. We present below a review of the topics discussed in this dissertation, followed by the corresponding conclusions.

In Chapter 2, we provided a background on the importance of protecting the privacy of personal information. We started describing the value of personal information and the main reasons that motivate companies to protect it. We have also described relevant privacy laws and regulations. Interesting cases of privacy violation are reported, reinforcing the lack of privacy protection nowadays. We described the current scenario concerning privacy and web applications, addressing problems and challenges within this context. In Chapter 3 we have shown solutions related to our work (architectures, UML extensions, tools), which have the goal to help protecting privacy. Chapters 2 and 3 helped us to understand how web applications should handle privacy as well as to determine privacy elements required for this task.

In Chapter 4, we presented our approach, composed of the privacy conceptual model, the privacy reference architecture and the privacy UML profile. The privacy conceptual model defines the privacy elements and their relationships in an organized way, fulfilling our goal of systematizing privacy concepts within the scope of web applications. The model provides the domain concepts to represent views of the system wherein privacy management and protection are applied. The model is the basis of the approach.

The design of the Reference Architecture was guided by the well-established process ProSA-RA (Nakagawa *et al.*, 2014) and provided a detailed description of the functionalities to be addressed in the implementation of web applications and services, to protect personal information privacy. The process of evaluating the architecture was hindered by the lack of a generic method for the evaluation of reference architectures. Nonetheless, the

evaluation process we applied has shown important qualities of the architecture, such as functional completeness, usability and applicability.

The UML Profile is useful to describe the privacy policy applied by an application as well as to keep track of the elements in charge of enforcing it, e.g., to track privacy requirements or for documentation purposes. The direct relationship between the Reference Architecture and the UML profile allows privacy-related components to be immediately identified within the software architecture and their role to be highlighted with domain-specific stereotypes. We have found the visual support of UML diagrams, specialized with stereotypes indicating the roles of components, to be really useful for documentation purposes.

In Chapter 5, we discuss a case study performed by applying the proposed approach in the construction of a web application with privacy protection. We created a concrete architecture and UML diagrams for the design and implementation of data privacy protection features (more specifically, an access control component) for the web application of an online bookstore. Experimental results have shown the effectiveness of the solution and the applicability of the approach. In Chapter 6, we describe an evaluation of the *completeness* and *applicability* of PrivAPP through an empirical study, where a set of privacy policies from relevant companies were analyzed to determine the cases for which elements from PrivAPP can help to enforce these policies. Results have shown the approach to possess a good level of completeness and applicability. The evaluation process has helped to address improvements to the approach. Limitations of the approach were addressed in this chapter.

PrivAPP introduces a couple of benefits for software developers and business professionals: first, the elements of the approach serve as a guideline for the design of concrete architectures which support web applications and services with privacy protection features. The models derived from the approach – UML diagrams and software architectures – provide resources for the documentation of privacy specifications of web applications, helping to structure particular concepts of privacy. They ease the understanding of the privacy domain by stakeholders and are useful for communication and to support the discussions on the general analysis of privacy resources when dealing with web applications. Consequently, these models allow a faster development of privacy issues by leaving programmers free from the task of deciding which technology to use to enforce privacy policies. A partial observation of the approach's capability in practice could be attained with the implementation of a solution with a few of its elements. Furthermore, the levels achieved of completeness and applicability give an indication that PrivAPP is a fairly feasible solution.

So, answering the research question Q1 (presented in Section 1), PrivAPP can be a solution to enhance the construction of privacy-aware web applications and services, because our preliminary evaluation makes us confident that it is a significant contribution towards improving the process of designing web applications in the privacy domain, by integrating privacy-related information into the development process of a web application. Also, the applicability analysis performed in Section 6.3 can answer the research question Q2 (Section 1): the professionals who used the PrivAPP defined requirements attempting much more to privacy issues, so, privacy reference models and specific UML resources can help in constructing web applications and services with privacy protection.

7.1 FUTURE WORK

The study of privacy is a wide open research field requiring continuous monitoring as new technologies arise. Even within the context of privacy-aware web applications, there is still not a widely established and adopted standard to handle privacy, creating several opportunities for future work. We outline below possible future research work closely related to this dissertation.

Evaluation of Reference Architectures. Evaluation of architectures helps to determine strong and weak aspects of them and gives an indication of how successful the system development and implementation processes will be. A reference architecture serves as a guiding tool for many projects taking place in diverse contexts. Thus, its evaluation prior to its adoption by the stakeholders is of even greater importance. Furthermore, a strong positive evaluation of a reference architecture is an incentive for its wide adoption. The lack of methods dedicated to the holistic evaluation of reference architectures was a limitation to this work. We used reasoning techniques we found adequate, but additional work must still be done. As a sequel to our work we intend to investigate: quality attributes to be evaluated, the reasoning techniques that can be employed for this evaluation, and adaptations in methods proposed elsewhere for the evaluation of concrete architectures. Our aim is to establish a solid approach for evaluating reference architectures.

Evaluation of UML Profiles. The same reasons we stated concerning reference architectures remains valid for the evaluation of a UML Profile. The evaluation we performed herein is quite simple and limited; it did not address the whole profile. Furthermore, to the best of our knowledge, there are no methods dedicated to the holistic evaluation of UML

Profiles. Hence, as a follow-up to our work we intend to investigate quality attributes, methods and techniques that can be employed for this task, to establish an approach for the evaluation and improvement of UML extensions.

Privacy tests. We intend to investigate if our solution – PrivAPP – can be a step forward concerning tests activities in the privacy domain. The idea is to use concrete architectures and UML models annotated with our profile as a support for evaluating if applications and services enforce correctly privacy policies and protect user's privacy. It can also be useful to check if the statements of the privacy law of a country are being met.

Automatic code generation. The relationship between UML models and text code can be considered from a historical perspective as an evolution towards model-centric approaches. Structural code generation from a model can let the programmers free from writing code for the structures implied by the UML class diagrams. Although our Privacy UML Profile is already useful for documenting and helping the implementation of privacy-aware applications, we intend to make it more useful, by making feasible the automatic generation of code. The idea is to address specific privacy requirements and their relationships, and the corresponding implementation techniques that realize these requirements, improving thereby the development process. For this purpose we intend to investigate the reduction of the abstraction level of privacy, with focus on the enforcement elements.

Comparison of Privacy Policies. Comparing the compatibility of privacy policies in practice is still a big challenge. On one hand, translating privacy policies into machine-readable format causes loss of semantics. On the other hand, comparing them in natural language is a very difficult task due to the involved semantics. We intend to investigate how the models constructed through the approach can be applied; privacy policies would be compared at a high level description (very close to natural language). At first, we would propose a semi-automated approach, by which the user could interfere and take decisions in the most complex cases of comparison. Semantic Similarity approaches can help in this direction. Semantic measures are widely used today to compare units of language, concepts, instances or even resources indexed by them (e.g., documents, genes). They are central elements of a large variety of Natural Language Processing applications and they have been subjected to intensive and interdisciplinary research efforts in the last decades. Ontologies can also help in this direction and we intend to investigate how to apply this technology in the task of recognize and interpret terms in natural language texts.

7.2 PUBLICATIONS

The following list includes the published works that served as the basis for this dissertation. Most of them are ranked by Brazilian Qualis Ranking 2012-2014, as shown next.

BASSO, T.; MORAES, R.; JINO, M.; VIEIRA, M. (2015). "Requirements, Design and Evaluation of a Privacy Reference Architecture for Web Applications and Services". In: The 30th ACM/SIGAPP Symposium On Applied Computing, 2015, Salamanca, Spain, pp. 1425-1432. **Qualis A1.**

BASSO, T.; MONTECCHI, L.; MORAES, R.; JINO, M.; BONDAVALLI, A. "Towards a UML Profile for Privacy-Aware Applications". In: 15th IEEE International Conference on Computer and Information Technology (CIT-2015), 2015, Liverpool, UK, pp. 371 - 378. **Qualis B1.**

BASSO, T.; PIARDI, L.; MORAES, R.; JINO, M. ; ANTUNES, N. ; VIEIRA, M. "A Database Framework for Expressing and Enforcing Personal Privacy Preferences". In: XVI Workshop de Testes e Tolerância a Falhas (WTF2015), 2015, Vitória. XXXIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, 2015, pp. 1-14. **Qualis B4**.

Mello, V. ; BASSO, T. ; MORAES, R. "A Test Process Model to Evaluate Performance Impact of Privacy Protection Solutions". In: XV Workshop de Testes e Tolerância a Falhas (WTF 2014), 2014, Florianópolis. XXXII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, 2014, pp. 1-14. **Qualis B4**.

BASSO, T.; ANTUNES, N.; MORAES, R.; VIEIRA, M. "An XML-based Policy Model for Access Control in Web Applications". In: 24th International Conference on Database and Expert Systems Applications – DEXA 2013, Praga, pp. 274-288. **Qualis B1.**

Other published papers that are not directly related to this dissertation are:

BASSO, T.; MORAES, R.; JINO, M. “A Semi Automated Approach to Assess Web Vulnerability Scanner Tools Effectiveness”. In: XIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg), 2013, Manaus, pp 1-14. **Qualis B4.**

The following paper, submitted to a journal, is under review:

BASSO, T. ; MONTECCHI, L. ; MORAES, R. ; JINO, M. ; BONDAVALLI, A. “PrivAPP: A Comprehensive Approach to Guide the Design of Privacy-Aware Applications”. Submitted to the International Journal Information and Software Technology. **Qualis B1.**

REFERENCES

Accenture (2009). “How Global Organizations Approach the Challenge of Protecting Personal Data”. Available at: www.ponemon.org/local/upload/file/ATC_DPP%20report_FINAL.pdf. Last access on October, 2015.

Adweek (2011). Adweek. “Lawmakers Go After Carrier IQ Over Privacy Concerns”. December, 2, 2011. Available at www.adweek.com/news/technology/lawmakers-go-after-carrier-iq-over-privacy-concerns-136880. Last access on October, 2015.

Aldawud, O., Elrad, T., Bader, A. (2003). “UML Profile for Aspect-Oriented Software Development”. The Third International Workshop on Aspect Oriented Modeling, Boston, USA, pp.1-6.

Amazon (2014). “Amazon.com Privacy Notice” Available at www.amazon.com/gp/help/customer/display.html/ref=footer_privacy?ie=UTF8&nodeId=46849. Last access on November, 2015.

Angelov, S. and Grefen, P. (2008). “An e-contracting reference architecture.” J. Syst. Softw. vol. 81, issue 11 (November 2008), pp. 1816-1844.

Angelov, S., Grefen, P., Greefhorst, D. (2009). "A classification of software reference architectures: Analyzing their success and effectiveness," Joint Working IEEE/IFIP Conference on Software Architecture & European Conference on Software Architecture. (WICSA/ECSA), pp.141,150.

Antón, A. I. and Earp, J. B. (2004). “A requirements taxonomy for reducing web site privacy vulnerabilities”. Requirements Engineering, vol. 9, issue 3, pp.169–185.

Antunes, N. and Vieira, M. (2011). "Enhancing Penetration Testing with Attack Signatures and Interface Monitoring for the Detection of Injection Vulnerabilities in Web Services". IEEE In-ternational Conference on Services Computing (SCC), pp.104-111.

APEC (2005). “APEC Privacy Framework”. Available at www.apec.org/Groups/Committee-on-Trade-and-Investment/~/media/Files/Groups/ECSG/05_ecsg_privacyframewk.ashx. Last access on August, 2015.

APP (2014). “Australian Privacy Principles”. Available at www.oaic.gov.au/privacy/privacy-resources/privacy-guides/app-quick-reference-tool. Last access on October, 2015.

Ashley, P., Hada, S., Karjoth, G., Powers, C., Schunter, M. (2003). “Enterprise Privacy Authorization Language (EPAL 1.2)”. Available at www.zurich.ibm.com/security/enterprise-privacy/epal/Specification/index.html. Last access on November, 2015.

Barcellona, C., Tinnirello, I., Merani, M.L. (2014). "Rings for privacy: An architecture for privacy-preserving user profiling," IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp.199-200.

Bass, L., Clements, P., Kazman, R. (2003). “Software Architecture in Practice”. Addison-Wesley Prof.,2003.

BBC (2015). BBC News. “US 'spied on French presidents' – Wikileaks”. Available at www.bbc.com/news/33248484. Last access on October, 2015.

BBC (2015-b). BBC News. “Google agrees privacy policy changes with data watchdog”. Available at http://www.bbc.com/news/technology-31059874. Last access on October, 2015.

Bertino, E., Lin, D., and Jiang, W. (2008). “A Survey of Quantification of Privacy Preserving Data Mining Algorithms”. In *Privacy-Preserving Data Mining*, vol. 34, C. C. Aggarwal, P. S. Yu, and A. K. Elmagarmid, Orgs. Springer US, pp. 183–205.

Biswas, D. and Niemi, V. (2011). "Transforming Privacy Policies to Auditing Specifications," IEEE 13th International Symposium on High-Assurance Systems Engineering (HASE), pp. 368-375.

Bodorik, P. and Jutla, D. (2008). "Privacy with Web Services: Intelligence Gathering and Enforcement". IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, (WI-IAT), 2008, vol.3, pp.546-549.

Bücker, A., Haase, B., Moore, D., Keller, M., Kobinger, O., Wu, H-F. (2003). "IBM Tivoli Privacy Manager. Solution Design and Best Practices". IBM Redbooks, 2003.

Cate, F.H. (2009). "Security, Privacy, and the Role of Law". IEEE Security & Privacy, vol.7, no.5, pp.60-63.

Cavoukian, A. (2006). “Creation of a Global Privacy Standard”. Available at www.ipc.on.ca/images/resources/gps.pdf. Last access on October, 2015.

Chakaravarthi, S., Selvamani, K., Kanimozhi, S., Arya, P.K. (2014). "An intelligent agent based privacy preserving model for Web Service security". IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE), pp.1-5.

Cherdantseva, Y. and Hilton, J. (2013). "A Reference Model of Information Assurance & Security," Eighth International Conference on Availability, Reliability and Security (ARES), pp.546-555.

Cirit, Ç. and Buzluca, F. (2009). “A UML profile for role-based access control”, Proceedings of the 2nd International conference on Security of information and networks (SIN), 2009, ACM, New York, NY, USA, pp. 83-92.

Clarke, R. (1999). “Introduction to dataveillance and information privacy, and definitions of terms”. The Information Society, september 1999. Available at www.anu.edu.au/people/Roger.Clarke/DV/Intro.html. Last access on August, 2015.

COPPA (1998). United States. Children’s Online Privacy Protection Act of 1998 (COPPA), October 1998. Available at www.cdt.org/legislation/105th/privacy/coppa.html. Last access on August, 2015.

Cranor, L., Dobbs, B., Egelman, S., Hogben, G., Humphrey, J., Langheinrich, M., Marchiori, M., Presler-Marshall, M., Reagle, J. M., Schunter, M., Stampley, D. A., and Wenning R. (2006). “The Platform for Privacy Preferences 1.1 (P3P1.1) Specification”. World Wide Web Consortium NOTEP3P11-20061113.

Damianou, N., Dulay, N., Lupu, E., Sloman, M. (2001). “The Ponder Policy Specification Language”, Policies for Distributed Systems and Networks. Lecture Notes in Computer Science, Vol. 1995, pp 18-38.

Dominguez, E., Perez, B., Zapata, M.A. (2013). "A UML profile for dynamic execution persistence with monitoring purposes". 5th International Workshop on Modeling in Software Engineering (MiSE), pp.55-61.

Elgesem, D. (1996). “Privacy, respect for persons, and risk”. In Charles Ess, editor, Philosophical perspectives on computer-mediated communication, chapter 3, pp. 45–66. State University of New York Press, 1996.

Elmasri, R., Navathe, S.B. (2011). “Database Systems". Pearson - Addison Wesley, 6th. Edition.

ENISA (2014). European Union Agency for Network and Information Security. “Privacy and Data Protection by Design – from policy to engineering”. December, 2014. Available at www.huntonprivacyblog.com/files/2015/01/Privacy-and-Data-Protection-by-Design.pdf. Last access on August, 2015.

EU (1995). European Union. “The European Union Directive 95/46/EC: On The Protection Of Individuals With Regard To The Processing Of Personal Data And On The Free Movement Of Such Data”, February 20 1995. Available at www.privacy.org/pi/intl_orgs/ec/eudp.html. Last access on August, 2015.

Fernandes, P., Basso, T., Moraes, R. (2011). “J-Attack - Injetor de Ataques para Avaliação de Segurança de Aplicações Web”. XII Workshop of Test and Fault Tolerance, Campo Grande, Brazil, pp. 29-41.

FERPA (1974). “Family Educational Rights and Privacy Act”. *US Department of Education*. Available at www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html. Last access on August, 2015.

FOIA (1996). The Freedom Of Information Act 5 U.S.C. § 552, As Amended By Public Law No. 104-231, 110 Stat. 3048. The United States Department of Justice. Available at www.justice.gov/oip/blog/foia-update-freedom-information-act-5-usc-sect-552-amended-public-law-no-104-231-110-stat. Last access on August, 2015.

Ford (2003). Ford Motor Co. Privacy. Available at www.ford.com/en/support/privacystatement.htm?referrer=home. Last access on August, 2015.

Fried, C. (1984). “Philosophical dimensions of privacy”. In F. D. Schoeman, editor. Cambridge University Press, 1984.

FTC (2000). Federal Trade Commission. “Privacy Online: Fair Information Practices in the Electronic Marketplace: A Federal Trade Commission Report to Congress”. Available at www.ftc.gov/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission. Last access on August, 2015.

G1 (2013). “NSA Documents point Dilma Rousseff as espionage target”. G1 Politics. Available at www.g1.globo.com/politica/noticia/2013/09/documentos-da-nsa-apontam-dilma-rousseff-como-alvo-de-espionagem.html. Last access on August, 2015.

G1 (2014). “Brazil has two of the top 50 online retail, research shows”. G1 Economy and Bbusiness. Available at http://g1.globo.com/economia/negocios/noticia/2014/01/brasil-tem-2-entre-50-maiores-do-varejo-online-mostra-pesquisa.html. Last access on August, 2015.

Gao, F., He, j., Ma, S. (2010). "A collaborative approach for identifying privacy disclosure in Web-based services," IEEE International Conference on Service-Oriented Computing and Applications (SOCA), pp.1-4.

Garlan, D. and Perry, D. (1995). “Introduction to the Special Issue on Software Architecture”. IEEE Transactions Software Engineering Vol. 21, issue 4, pp. 269-274.

GE (2003). “General Electric Co. Privacy Policy”. Available at www.ge.com/en/ge/privacy.htm. Last access on July, 2015.

Ghazinour, K. and Barker, K. (2013). “A privacy preserving model bridging data provider and collector preferences.” Proceedings of the Joint EDBT/ICDT 2013 Workshops (EDBT), ACM, New York, NY, USA, pp. 174-178.

Ghazinour, K., Razavi, A.H., Barker, K. (2014). “A Model for Privacy Compromisation Value”. Procedia Computer Science, Vol. 37, pp. 143-152.

Giffin, D.B., Levy, A., Stefan, D., Terei, D., Mazières, D., Mitchell, J. C., and Russo, A. (2012). “Hails: Protecting Data Privacy in Untrusted Web Applications,”. Presented as part of the 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI), Hollywood, CA, pp. 47–60.

Glass, L., Gresko, R. (2012). "Legislation and Privacy across Borders". International Conference on Privacy, Security, Risk and Trust (PASSAT), 2012 and International Conference on Social Computing (SocialCom), pp.807-808.

GLBA (1999). “Gramm-Leach-Bliley Act”. US Federal Trade Comission. Available at www.ftc.gov/tips-advice/business-center/privacy-and-security/gramm-leach-bliley-act. Last access on July, 2015.

GloboNews (2012). “Google gets involved in invasion of privacy on the Internet”. Globo News. Available at www.g1.globo.com/globo-news/globo-news-em-pauta/videos/t/todos-os-videos/v/google-se-envolve-em-invasao-de-privacidade-na-internet/1818480/. Last access on August, 2015.

Google (2014). Google Privacy & Terms. “Privacy Policy”, release of march, 2014. Available at www.google.com/policies/privacy/. Last access on August, 2015.

Gramm-Leach-Bliley (1999). United States. Gramm-Leach-Bliley Act: Financial Privacy and Pretexting, November 12 1999. Available at www.ftc.gov/privacy/glbact/glboutline.htm. Last access on August, 2015.

Han, P., and Maclaurin, A. (2002). "Do consumers really care about online privacy?”. Marketing Manage, vol. 11, no. 1, pp. 35-38.

Heather, M. (2010). “What is the Difference Between Security and Privacy?” The Propay Perspective. Available at www.blog.propay.com/index.php/2010/09/15/what-is-the-difference-between-security-and-privacy/. Last access on August, 2015.

Heitmann, b., Kim, J. G., Passant, A, Hayes, C., Kim, H-G. (2010). “An architecture for privacy-enabled user profile portability on the web of data”. In Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec). ACM, New York, NY, USA, pp.16-23.

Hewett, R. and Kijsanayothin, P. (2009). "On securing privacy in composite web service transactions," International Conference for Internet Technology and Secured Transactions (ICITST), pp.1-6.

Hinde, S. (2003). “Privacy legislation: a comparison of the US and European approaches”. Computers & Security, 2003, vol. 22, no. 5, pp.378-387.

HIPAA (1996). United States. The Health Insurance Portability and Accountability Act of 1996 (HIPAA), October 1998. Available at www.hhs.gov/ocr/privacy/. Last access on September, 2015.

Hoepman, J-H. (2014). “Privacy Design Strategies”. In Proceedings of 29th IFIP TC 11 International Conference (SEC), Marrakech, Morocco, pp. 446-459.

Hurlburt, G. F., Miller, K.W., Voas, J. M., Day, J. M. (2009). "Privacy and/or Security: Take Your Pick", IT Professional , vol.11, no.4, pp.52-55.

IAPP (2012). International Association of Privacy Professionals. “Glossary of Privacy Terms”. IAPP Privacy Glossary, 2012. Available at www.privacyassociation.org/media/pdf/resource_center/IAPP_Privacy_Certification_Glossary_v2.0.0.2.pdf. Last access on September, 2015.

ICO (2015). Information Comissioner’s Office. “Google to change privacy policy after ICO investigation”, 2015. Available at www.ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2015/01/google-to-change-privacy-policy-after-ico-investigation/.

ISO (2011). ISO/IEC 29100. International Standard - Information technology - Security Techniques - Privacy Framework. First Edition, 2011-12-15.

ISO (2013). ISO/IEC 29101. International Standard - Information technology - Security Techniques - Privacy Architecture Framework. First Edition, 2013-10-15.

Ives B., Jarvenpaa S.L. (1991). “Applications of global information technology: key issues for management.” MIS Q 1991, March, pp. 33-48.

Jiang, X., Wang, S., Ji, Z., Ohno-Machado, L.; Xiong, L. (2012). "A Randomized Response Model for Privacy-Preserving Data Dissemination," IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB), pp.138-138.

Jmeter (2015). “Apache JMeter - Apache JMeterTM”. Available: www.jmeter.apache.org/.

Jürjens, J. (2002). “UMLsec: Extending UML for Secure Systems Development”. In Proceedings of the 5th International Conference on The Unified Modeling Language (UML '02), Jean-Marc Jézéquel, Heinrich Hußmann, and Stephen Cook (Eds.). Springer-Verlag, London, UK, pp. 412-425.

Kim, K.I., Kim, W.Y., Ryu, J.S., Ko, H.J., Kim, U.M. and Kang, W.J. (2010). “RBAC-based access control for privacy preserving in semantic web”. In Proceedings of the 4th International Conference on Uniquitous Information Management and Communication (ICUIMC '10). ACM, New York, NY, USA, pp. 1-5.

Leino-Kilpi, H.; Välimäki, M.; Dassen, T.; Gasull,M.; Lemonidou,C.; Scott,A.; Arndt,M. (2001). “Privacy: a review of the literature”. International Journal of Nursing Studies, Vol. 38, Issue 6, pp. 663-671.

Li, D., Li, X., Liu, Z., Stolz, V. (2013). "Support Formal Component-Based Development with UML Profile". 22nd Australian Software Engineering Conference (ASWEC), pp.191-200.

MarcoCivil (2014). “Marco Civil da Internet – LEI Nº 12.965, DE 23 DE ABRIL DE 2014”. Presidence of the Republic of Brazil. Available at www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Last access on August, 2015.

Meziane, H., Benbernou, S. (2010). “A dynamic privacy model for web services”. Computer Standards & Interfaces, Vol. 32, Issues 5–6, pp. 288-304.

Meziane, H., Benbernou, S., Zerdali, A.K., Hacid, M.-S., Papazoglou, M. (2010). "A view-based Monitoring for Privacy-aware Web services," IEEE 26th International Conference on Data Engineering (ICDE), pp.1129-1132.

Microsoft (2013). “Microsoft Trustworthy Computing. 2013 Privacy Survey Results. Protecting consumers’ online privacy is a shared responsibility”. Available at www.download.microsoft.com/download/A/A/9/AA96E580-E0F6-4015-B5BB-ECF9A85368A3/Microsoft-Trustworthy-Computing-2013-Privacy-Survey-Executive-Summary.pdf. Last access on August, 2015.

Mont, M. C., Thyne, R., Bramhall, P. (2005). "Privacy Enforcement with HP Select Access for Regulatory Compliance". Hewlett-Packard Company , 2005.

Mont, M.C., Pearson, S., Creese, S., Goldsmith,M., Papanikolaou, N. (2011). “A Conceptual Model for Privacy Policies with Consent and Revocation Requirements”. IFIP Advances in Information and Communication Technology, Vol. 352, pp. 258-270.

Mubin, S.A. and Jantan, A.H. (2014). "A UML 2.0 profile web design framework for modeling complex web application". 2014 International Conference on Information Technology and Multimedia (ICIMU), pp.324-329.

Muller, G. (2008). "A Reference Architecture Primer". Eindhoven Univ. of Techn., Eindhoven, White paper, 2008.

Nakagawa, E.Y., Guessi, M., Maldonado, J.C., Feitosa, D., Oquendo, F. (2014). “Consolidating a Process for the Design, Representation, and Evaluation of Reference Architectures”. In Proceedings of the 2014 IEEE/IFIP Conference on Software Architecture (WICSA). IEEE Computer Society, Washington, DC, USA, pp.143-152.

Nakagawa, E.Y., Oquendo, F., Becker, M., (2012). “RAModel: A Reference Model for Reference Architectures”. In 2012 Joint Working IEEE/IFIP Conference on Software Architecture (WICSA) and European Conference on Software Architecture (ECSA), pp. 297-301.

Ni, Q., Trombetta, A., Bertino, E., Lobo, J. (2007). “Privacy-aware Role Based Access Control”. In Proceedings of the 12th ACM Symposium on Access Control Models and Technologies (Sophia Antipolis, France, June 20-22, 2007). ACM, New York, pp. 41-50.

NYTimes (2012). The New York Times. “Austrian Law Student Faces Down Facebook”. Feburary, 5, 2012. Available at www.nytimes.com/2012/02/06/technology/06iht-rawdata06.html. Last access on August, 2015.

OASIS (2012). “Privacy Management Reference Model and Methodology (PMRM) Version 1.0”, 2012. Available at www.docs.oasis-open.org/pmrm/PMRM/v1.0/csd01/PMRM-v1.0-csd01.pdf. Last access on August, 2015.

OASIS (2013). “eXtensible Access Control Markup Language (XACML)”. OASIS Standard. Available at www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml. Last access on August, 2015.

OECD (2010). “OECD Privacy Principles”. Available at www.oecdprivacy.org/#principles. Last access on August, 2015.

OECD (2015). “List of OECD Member countries”. Available at www.oecd.org/about/membersandpartners/list-oecd-member-countries.htm. Last access on November, 2015.

OMG (2005). Object Management Group. UML Profile for Schedulability, Performance, and Time Specification (OMG SPT), Version 1.1. OMG Document formal/05-01-02. Jan. 2005

OMG (2008). Object Management Group. UML Profile for Modeling Quality of Service and Fault Tolerance Characteristics and Mechanisms (OMG QoS&FT), Version 1.1. OMG Document formal/2008-04-05. Apr. 2008

OMG (2011). Object Management Group. OMG Unified Modeling Language (OMG UML), Superstructure, Version 2.4.1. OMG Document formal/2011-08-06. Aug. 2011.

OMG(2011-b). Object Management Group. A UML Profile for MARTE: Modeling and Analysis of Real-Time Embedded systems, Version 1.1. OMG Document formal/2011-06-02. June 2011

Oracle (2015). “Oracle | Hardware and Software, Engineered to Work Together”. Available: http://www.oracle.com/index.html.

Osawa, Y., Imamura, S., Takeda, A., Kitagata, G., Shiratori, N., Hashimoto, K. (2010). "A Proposal of Privacy Management Architecture," 10th IEEE/IPSJ International Symposium on Applications and the Internet (SAINT), pp.161-164.

Padilha, A.V. (2004). "Web Usability: A Proposal of a Questionnaire for Evaluation of the Satisfaction Degree of E-commerce Users". Work of master's degree in Computer Science. Federal University of Santa Catarina. Florianópolis, Brazil.

Peltier, T. (2001) “Information Security Risk Analysis, Auerbach Publications”. Div. of CRC Press LLC, Boca Raton, FL, USA, p. 266.

Perkins, E., and Markel, M. (2004). "Multinational data-privacy laws: an introduction for IT managers". IEEE Transactions on Professional Communication, vol.47, no.2, pp.85-94.

Pfitzmann, A., Hansen, M. (2009). “A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management”. Available at www.dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf. Last access on August, 2015.

PGP (1999). “How PGP Works”. Available at www.pgpi.org/doc/pgpintro/ . Last access on August, 2015.

PIPEDA (2000). Canada. The Personal Information Protection and Electronics Document Act: Bill C6. Available at www.laws-lois.justice.gc.ca/eng/acts/p-8.6/. Last access on August, 2015.

Ponemon (2010). “Economic impact of privacy on online behavioral advertising”. Benchmark study of Internet marketers and advertisers. Available at www.ponemon.org/local/upload/file/2010_Economic_impact_of_privacy_on_OBA.pdf. Last access on August, 2015.

Ponemon (2011). “Reputation Impact of a Data Breach”. U.S. Study of Executives & Managers.” Available at www.experian.com/assets/data-breach/white-papers/reputation-study.pdf. Last access on August, 2015.

Privacy Act (1998). “Australia Privacy Act 1988”. Available at www.oaic.gov.au/privacy/privacy-act/the-privacy-act. Last access on August, 2015.

R7 (2010). “DVDs sold by street vendors in Brazil have private data from Dilma, Serra and Lula”. R7 Portal. Available at www.noticias.r7.com/eleicoes-2010/noticias/dvds-de-camelos-trazem-dados-cadastrais-de-dilma-serra-e-ate-de-lula-20100905.html. Last access on August, 2015.

R7 (2015). “10 reliable sites to buy travel packages and airline tickets”. Available at www.lista10.org/tech-web/pacotes-de-viagens-e-passagens-aereas/. Last access on August, 2015.

Reay, I., Dick, S., Miller. J. (2009). “A large-scale empirical study of P3P privacy policies: Stated actions vs. legal obligations”. ACM Transactions on the Web, vol. 3, nº. 2, Article 6, pp. 1-34.

Reddy, K., Venter, H. S., Olivier, M., Currie, I. (2008). "Towards Privacy Taxonomy-Based Attack Tree Analysis for the Protection of Consumer Information Privacy," Sixth Annual Conference on Privacy, Security and Trust (PST), pp.56-64.

Reitsma, R., O'Connell,J., Wise, J., Jaddou, S. (2011). "Consumers And Online Privacy: How Much Information Is Too Much?". Forrester Technographics Digital Consumer Community Report, August 2011.

Roesner, F., Kohno, T., Wetherall, D. (2012). "Detecting and defending against third-party tracking on the web". In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (NSDI). Berkeley, USA, pp.12.

SafeHarbor (2000). "U.S.-EU Safe Harbor Overview". Available at www.export.gov/safeharbor/eu/eg_main_018476.asp. Last access on August, 2015.

Samarati, P. and Sweeney, L. (1998). "Generalizing data to provide anonymity when disclosing information". In Proceedings of the seventeenth ACM symposium on Principles of database systems (PODS). ACM, New York, NY, USA, pp. 188.

Sandhu, R. S. (1998). "Role-based Access Control". In Advances in Computers, vol. 46, Elsevier, pp. 237–286.

Sangani, N.K., Vithani, T., Velmurugan, P., Madiajagan, M. (2012). "Security & Privacy Architecture as a service for Small and Medium Enterprises," 2012 International Conference on Cloud Computing Technologies, Applications and Management (ICCCTAM), pp.16-21.

Sathiyamurthy, S. (2011). "The Struggle for Privacy and the Survival of the Secured in the IT Ecosystem". ISACA Journal, 2011, vol. 2, no. 1, pp.1-7.

Scheithauer, G. and Wirtz, G. (2010). "Business modeling for service descriptions: a meta model and a UML profile". In Proceedings of the Seventh Asia-Pacific Conference on Conceptual Modeling - Volume 110 (APCCM '10), Sebastian Link and Aditya Ghose (Eds.), Vol. 110. Australian Computer Society, Inc., Darlinghurst, Australia, pp. 79-88.

Shibboleth (2014). "Shibboleth Identity Provider V3.0.0". Available: www.shibboleth.net/. Last access on July, 2015.

Shin, Y-N., Chun, W. B., Jung, H. S., Chun, M. G. (2011). “Privacy Reference Architecture for Personal Information Life Cycle”, in Advanced Communication and Networking, T. Kim, H. Adeli, R. J. Robles, e M. Balitanas, Orgs. Springer Berlin Heidelberg, pp. 76–85.

Silva, Vergilio Ricardo Britto da (2015). “Privacy Concerns on Internet: an Exploratory Research on Brazilian Scenario (Preocupação Com A Privacidade Na Internet: Uma Pesquisa Exploratória No Cenário Brasileiro)”. Work of master's degree in Administration and Business. Pontifical Catholic University of Rio Grande Do Sul. Porto Alegre, Brazil.

Solove, D. (2006). “A Taxonomy of Privacy,” University of Pennsylvania Law Review, Vol. 154, No. 3, pp. 477 - 560.

Sybase (2013). “Sybase XML Modeling PowerDesigner® 15.3”. Available: www.download.sybase.com/pdfdocs/pdd1100e/xmug.pdf . Last access on July, 2015.

Tbahriti S-E., Medjahed, B., Malik, Z., Ghedira, C., Mrissa, M. (2011). “Meerkat - A Dynamic Privacy Framework for Web Services”. In 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. vol.1, pp. 418-421.

TheGuardian (2013). “NSA collecting phone records of millions of Verizon customers daily”. *The Guardian website*. Available at www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order. Last access on August, 2015.

TPCW (2015). “Transaction Processing Performance Council”. Available at www.tpc.org/tpcw/. Last access on August, 2015.

Truste (2013). TRUSTe 2013 U.S. Consumer Privacy Confidence Privacy Report: What Consumers Think, Business Impact, and Recommended Actions. Available at www.truste.com. Last access on August, 2015.

Truste (2015). TRUSTe Privacy Index. 2015 Consumer Confidence Edition. Available at www.truste.com/resources/privacy-research/us-consumer-confidence-index-2015/. Last access on August, 2015.

USPrivacyAct (1974). “Overview of the Privacy Act of 1974”. The United States Department of Justice. Available at www.justice.gov/opcl/overview-privacy-act-1974-2012-edition. Last access on August, 2015.

Veja (2011). “Digital Life – US Congress will investigate application for smartphones that 'steals' personal information”. Veja Digital Magazine. Available at www.veja.abril.com.br/noticia/vida-digital/senador-americano-ira-examinar-escandalo-de-privacidade-envolvendo-smartphones. Last access on August, 2015.

Venier, S. (2010). “The respect to privacy in different cultural contexts”. *EACME Annual Meeting*, Oslo, September, 2010. Available at www.prescient-project.eu/prescient/inhalte/download/VenierEACME2010.pdf. Last access on August, 2015.

Vieira, M., Madeira, H. (2005). “Towards a security benchmark for database management systems”. In Proceedings of International Conference on Dependable Systems and Networks, DSN 2005. pp. 592–601.

Wang, H., Lee, M., and Wang, C. (1998). “Consumer privacy concerns about Internet marketing”. In Commun. ACM, vol. 41, no. 3, pp. 63-70.

Warren, S. D. and Brandeis, L.D. (1890). The right to privacy. Harvard Law Review, vol.4, no. 5, December 1890.

Webshoppers (2015). “Webshoppers 2015, 31st. Edition”. Available at www.webvenda.com/wp-content/uploads/2015/02/31_webshoppers.pdf . Last access on August, 2015.

Westin, A. (1987). “ Privacy and Freedom”. Bodley Head, 1987.

Wu, X., Zhu, X., Wu, G-Q., Ding, W. (2104). “Data mining with big data”. In IEEE Transactions on Knowledge and Data Engineering, vol.26, no.1, pp.97-107.

You Tube (2012). “Student processes Facebook”. You tube. Available at www.youtube.com/watch?v=ObbiBeXevkE. Last access on August, 2015.

APPENDICES

APPENDIX A - DETAILING THE LITERATURE REVIEW PROCESS

From the theme we have chosen to this research (which is about helping to improve the scenario of lack of privacy protection in the construction of web applications and services), this work has the goal of defining an approach that systematizes privacy concepts in the scope of web applications and, consequently, improving the construction of web applications and services with privacy protection definition and enforcement. As a first step of the proposal it was performed a literature review. We want to obtain, through this systematization, a general view from the research area to be considered, identifying the focus and quality of the related works. In addition, we want to analyze the solutions that have been conducted in the research and their corresponding results.

The following research questions were established for the literature review and conducted the study.

Question 1 – Which are the available conceptual or reference models for privacy in web applications?

Question 2 – Which are the available reference architectures for privacy in web applications?

Question 3 – Which are the available UML profiles for privacy?

Question 4 – Which are the solutions or tools available to help guaranteeing privacy in web applications and services? Are any of them related to security attacks?

The scope of the review was defined using the framework PICO, which structure the research question in four basic elements: population, intervention, comparison and outcome. So, to this work we have:

Population: web applications and services.

Intervention: approach for systematizing privacy concepts in the scope of web applications and services.

Comparison: the proposed solution shall be compared to the non-use of it.

Outcome: based on the result of the literature review we intend to establish the referred approach, evaluate its effectiveness and compare the results to other similar and relevant solutions available in the literature.

To the review, we selected the works from the following digital libraries. They were selected because are considered quite relevant by the computer community.

ACM Digital Library (http://dl.acm.org/)
IEEE Xplore Digital Library (http://ieeexplore.ieee.org/)
ScienceDirect (http://www.sciencedirect.com/)

Then, in the research context, the following keywords were defined:

Keywords: *Privacy, model, conceptual, approach, reference, web application, web service, architecture, UML Profile, UML extension, data privacy, policies, preserving, guarantee, warranty, violation, solution, mechanism, attack, security.*

We also verified the taxonomy or classification system from the referred digital libraries.

ACM Classification System: privacy, Web-based services.
IEEE Taxonomy: Data privacy, privacy, Web services.

The search mechanisms in these digital libraries allow searching stored works through many keywords, combined using Boolean operators (e.g., AND, OR), in the title or in the abstract. Thus, our search strategy was split in four parts in order to address the research questions individually. For the sake of space, we will present only the final search strings and corresponding results, after applying all the exclusion criteria.

PART I – Identifying Privacy Concepts and Reference Models.

Research Question: which are the available conceptual or reference models for privacy in web applications?

Criteria exclusions: it was excluded from the results the works that:

a) do not have the words “privacy” and “model” in the title.

b) are previous to the 2010 year.

c) do not attend the goal of the research question.

Search strings:

IEEExplorer

(("Document Title":privacy) AND (p_Title:model)) AND (conceptual OR reference OR approach OR Web Application)

Publication year: range from 2010 to 2015.

36 Results

ACM Digital Library

((Title:privacy and Title:model)) and (conceptual or reference or approach or web application)

Published since 2010

27 results

Science Direct

ttl(privacy and model) and (conceptual or reference or approach or web application)

pub_date > 2009

24 results

From the 87 results we selected 10 works that are the most related to our research. Through some references of these works we selected more 3 works that are quite relevant regarding privacy concepts and models for web applications. Table A-1 shows the references of the selected works, in alphabetical order.

Table A-1. Related work for privacy concepts and reference models**.**

	Selected Related Work For Privacy Concepts And Reference Models
Selected through the literature review process	Gkoulalas-Divanis, A. and Cope, E.W. (2011). "A publication process model to enable privacy-aware data sharing". IBM Journal of Research and Development, vol.55, no.5, pp.8:1, 8:10.
	Chakaravarthi, S.; Selvamani, K.; Kanimozhi, S.; Arya, P.K. (2014). "An intelligent agent based privacy preserving model for Web Service security". IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE), pp.1-5.
	Cheek, G.P., Shehab, M.(2014). "Human Effects of Enhanced Privacy Management Models," IEEE Transactions on Dependable and Secure Computing, vol.11, no.2, pp.142,154.
	Jiang, X., Wang, S., Ji, Z., Ohno-Machado, L.; Xiong, L. (2012). "A Randomized Response Model for Privacy-Preserving Data Dissemination," IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB), 2012, pp.138,138.
	Ghazinour, K. and Barker, K. (2013). "A privacy preserving model bridging data provider and collector preferences." Proceedings of the Joint EDBT/ICDT 2013 Workshops (EDBT '13). ACM, New York, NY, USA, pp. 174-178.
	Ghazinour, K., Razavi, A.H., Barker, K. (2014). "A Model for Privacy Compromisation Value". Procedia Computer Science, Vol. 37, 2014, pp. 143-152.
	Mahmood, S. and Desmedt, Y. (2013). "Two new economic models for privacy". SIGMETRICS Perform. Eval. Rev. 40, 4 (April 2013), pp. 84-89.
	Witt, S., Feja, S., Speck, A., Prietz, C. (2012). "Integrated privacy modeling and validation for business process models". In Proceedings of the 2012 Joint EDBT/ICDT Workshops (EDBT-ICDT '12), Divesh Srivastava and Ismail Ari (Eds.). ACM, New York, NY, USA, pp. 196-205.
	Ny, J. L. and Pappas, G. J. (2013). "Privacy-preserving release of aggregate dynamic models." In Proceedings of the 2nd ACM international conference on High confidence networked systems (HiCoNS '13). ACM, New York, NY, USA, pp. 49-56
	Meziane, H., Benbernou, S. (2010). "A dynamic privacy model for web services". Computer Standards & Interfaces, Vol. 32, Issues 5–6, October 2010, pp. 288-304.
Related references	Cherdantseva, Y. and Hilton, J. (2013). "A Reference Model of Information Assurance & Security," Eighth International Conference on Availability, Reliability and Security (ARES) 2013, pp.546,555.
	Sathiyamurthy, S. (2011). "The Struggle for Privacy and the Survival of the Secured in the IT Ecosystem". ISACA Journal, 2011, vol. 2, pp.1-7.
	OASIS (2012). "Privacy Management Reference Model and Methodology (PMRM) Version 1.0", 2012. Available at http://docs.oasis-open.org/pmrm/PMRM/v1.0/csd01/PMRM-v1.0-csd01.pdf

Figure A-1 shows the conceptual map that was constructed to organize the literature review. From this map it is possible to observe solutions, i.e., privacy models that are in the context of web applications and also in other contexts as, for example, cloud computing, social networks, data mining and RFID (Radio-Frequency Identification). We used the works from the offshoot *privacy model → for → web applications.*

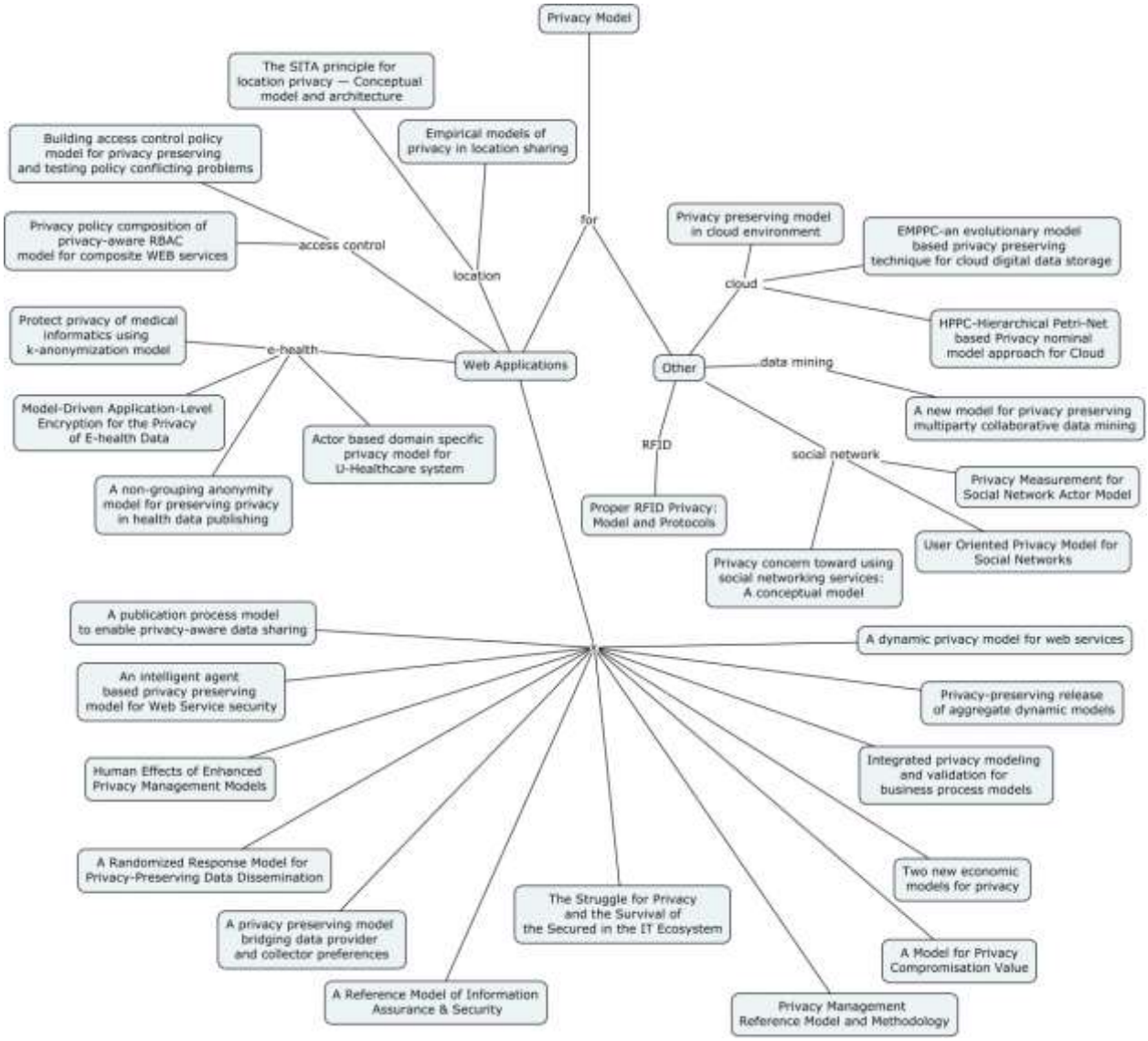

Figure A-1. Conceptual map for privacy concepts and reference models.

PART II – Identifying Privacy Architectures.

Research Question: which are the available reference architectures for privacy in web applications?

Criteria exclusions: it was excluded from the results the works that:

a) do not have the words “privacy” and “architecture” in the title.

b) are previous to the 2010 year.

c) do not attend the goal of the research question.

Search strings:

IEEExplorer
(("Document Title":privacy) AND "Document Title":architecture)
Publication year: range from 2010 to 2015.
48 Results

ACM Digital Library
((Title:privacy and Title:architecture))
Published since 2010
13 results

Science Direct
ttl(privacy) and ttl(architecture)
pub_date > 2009
4 results

From the 65 results we selected 5 works that are the most related to our research. Through some references of these works we selected more 5 works that are quite relevant regarding privacy architectures in the context of web applications. Table A-2 shows the result:

Table A-2. Related work for privacy architectures.

	Selected Related Works for Privacy Architectures
Selected through the literature review process	Sangani, N.K., Vithani, T., Velmurugan, P., Madiajagan, M. (2012). "Security & Privacy Architecture as a service for Small and Medium Enterprises," 2012 International Conference on Cloud Computing Technologies, Applications and Management (ICCCTAM), pp.16,21.
	Heitmann, b., Kim, J. G., Passant, A, Hayes, C., Kim, H-G. (2010). "An architecture for privacy-enabled user profile portability on the web of data". In Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec '10). ACM, New York, NY, USA, pp.16-23.
	Diaz-Tellez, Y., Bodanese, E.L., Nair, S.K., Dimitrakos, T. (2012). "An Architecture for the Enforcement of Privacy and Security Requirements in Internet-Centric Services," 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp.1024,1031.
	Osawa, Y., Imamura, S., Takeda, A., Kitagata, G., Shiratori, N., Hashimoto, K. (2010). "A Proposal of Privacy Management Architecture," 10th IEEE/IPSJ International Symposium on Applications and the Internet (SAINT), 2010 , pp.161-164.
	Barcellona, C., Tinnirello, I., Merani, M.L. (2014). "Rings for privacy: An architecture for privacy-preserving user profiling," IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2014, pp.199-200.
Related references	ISO/IEC 29100 (2011). International Standard - Information technology - Security Techniques - Privacy framework. First Edition.
	ISO/IEC 29101 (2013). International Standard - Information technology - Security Techniques - Privacy architecture framework. First Edition.
	Shin, Y-N., Chun, W. B., Jung, H. S., Chun, M. G. (2011). "Privacy Reference Architecture for Personal Information Life Cycle", in Advanced Communication and Networking, T. Kim, H. Adeli, R. J. Robles, e M. Balitanas, Orgs. Springer Berlin Heidelberg, pp. 76–85.
	Bücker, A., Haase, B., Moore, D., Keller, M., Kobinger, O., Wu, H-F. (2003). "IBM Tivoli Privacy Manager. Solution Design and Best Practices". IBM Redbooks.
	Mont, M. C., Thyne, R., Bramhall, P. (2005). "Privacy Enforcement with HP Select Access for Regulatory Compliance". Hewlett-Packard Company.

The conceptual map for this part of the literature review is presented in Figure A-2. We used the works from the offshoot *privacy architecture → for → web application.*

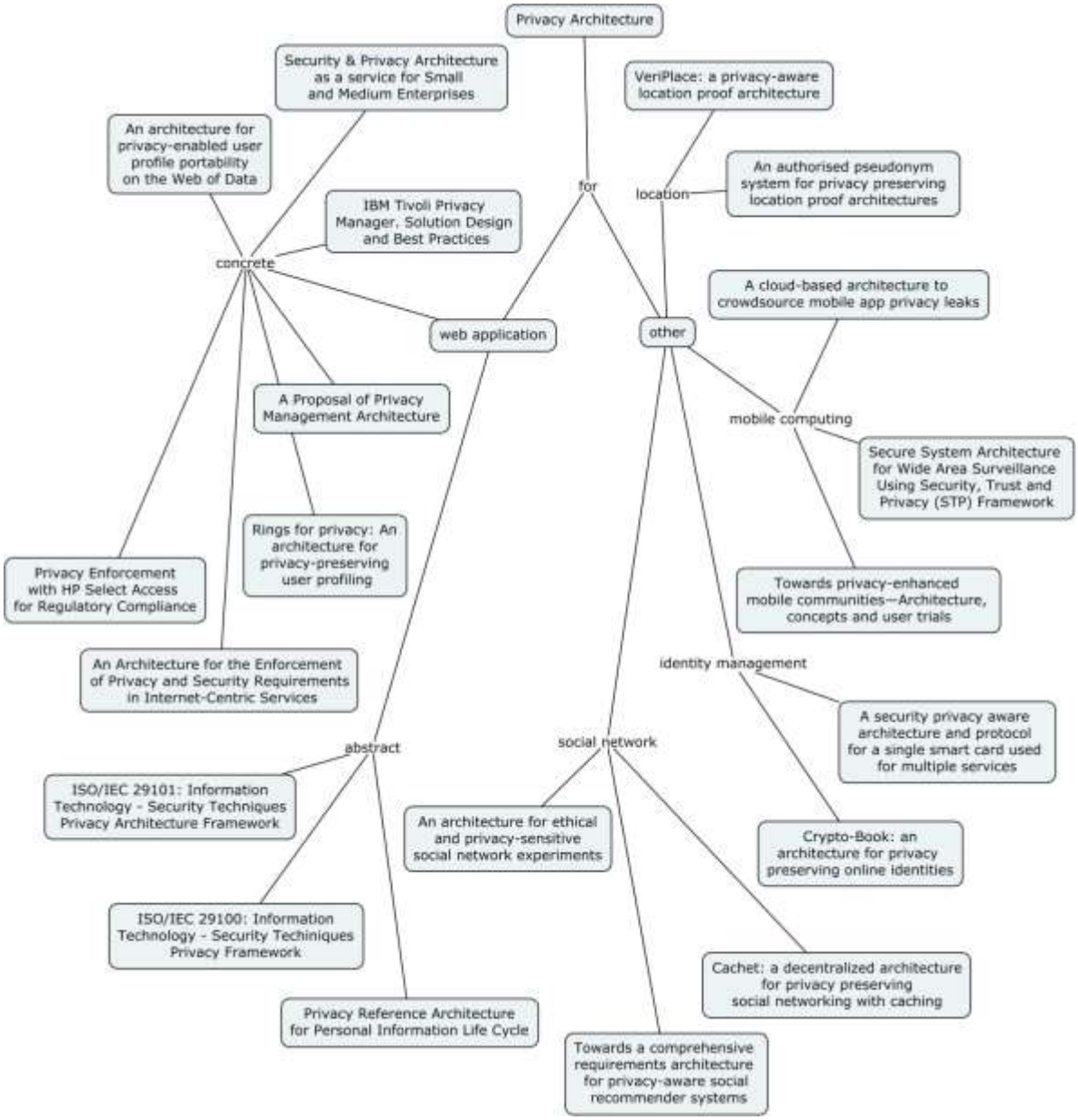

Figure A-2. Conceptual map for privacy architectures.

PART III – Identifying Privacy UML Profiles.

Research question: which are the available UML profiles for privacy?

Criteria exclusions: it was excluded from the results the works that:

a) do not have the words "UML Profile" in the title.

b) are previous to the 2010 year.

c) do not attend the goal of the research question.

Search strings:

IEEExplorer

("Document Title":uml profile)

Publication year: range from 2010 to 2015.

33 Results

ACM Digital Library

(Title:"uml profile")

Published since 2010

7 results

Science Direct

ttl("uml profile")

pub_date > 2009

6 results

From the 46 results we selected 4 works to this literature review process. Through some references of these works we selected more 2 works that are quite relevant regarding UML Profiles for privacy. Table A-3 shows the references. Figure A-3 shows the conceptual map. We used the *UML Profile* → *web applications* and *UML Profile* → *security* offshoots.

Table A-3. Related work for privacy UML Profiles.

	Selected Related Works for UML Profiles for Privacy
Selected through the literature review process	Scheithauer, G. and Wirtz, G. (2010). “Business modeling for service descriptions: a meta model and a UML profile”. In Proceedings of the Seventh Asia-Pacific Conference on Conceptual Modeling - Volume 110 (APCCM '10), Sebastian Link and Aditya Ghose (Eds.), Vol. 110. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, pp. 79-88.
	Li, D., Li, X., Liu, Z., Stolz, V. (2013). "Support Formal Component-Based Development with UML Profile". 22nd Australian Software Engineering Conference (ASWEC), 2013, pp.191-200.
	Dominguez, E., Perez, B., Zapata, M.A. (2013). "A UML profile for dynamic execution persistence with monitoring purposes". 5th International Workshop on Modeling in Software Engineering (MiSE), 2013, pp.55-61.
	Mubin, S.A. and Jantan, A.H. (2014). "A UML 2.0 profile web design framework for modeling complex web application". 2014 International Conference on Information Technology and Multimedia (ICIMU), pp.324-329.
Related references	Jürjens, J. (2002). “UMLsec: Extending UML for Secure Systems Development”. In Proceedings of the 5th International Conference on The Unified Modeling Language (UML '02), Jean-Marc Jézéquel, Heinrich Hußmann, and Stephen Cook (Eds.). Springer-Verlag, London, UK, UK, pp. 412-425.
	Cirit, Ç. and Buzluca, F. (2009). “A UML profile for role-based access control”, Proceedings of the 2nd International conference on Security of information and networks (SIN’09), 2009, ACM, New York, NY, USA, pp. 83-92.

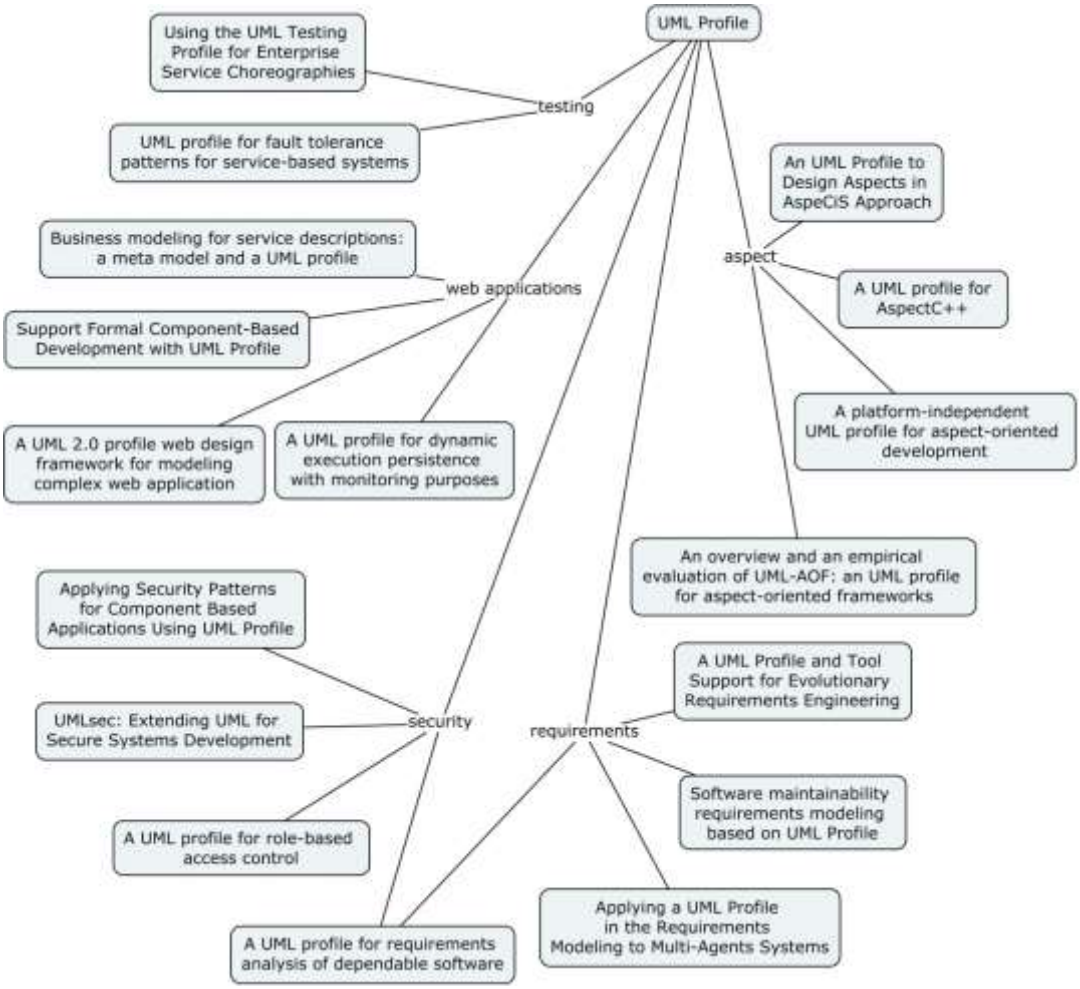

Figure A-3. Conceptual map for Privacy UML Profiles.

PART IV – Identifying Solutions and Tools for Privacy Protection.

Research question: Which are the solutions or tools available to help guaranteeing privacy in web applications and services? Are any of them related to security attacks?

Criteria exclusions: it was excluded from the results the works that:

a) do not have the keywords in the title or in the abstract.

b) do not have the words “privacy” and “web service” in the title.

c) are previous to the 2009 year

d) do not attend the goal of the research question.

Search strings:

IEEExplorer

(("Abstract":privacy) AND ("Abstract":web-based services OR web services) AND ("Abstract":policies OR policy OR preserving OR guarantee OR warranty OR violation OR solution OR data OR mechanism OR security OR attack OR injection OR sql OR xpath OR command))AND ("Document Title":privacy) AND ("Document Title":web services)

Publication year: range from 2009 to 2015.

16 results

ACM Digital Library

((privacy) AND (web-based services OR web services) AND (policies OR policy OR preserving OR guarantee OR warranty OR violation OR solution OR data OR mechanism OR security OR attack OR injection OR sql OR xpath OR command)) and (AbstractFlag:yes) AND (Title:privacy) AND (Title:web services)

Published since 2009

17 results

Science Direct

pub-date > 2008 AND abs((privacy) AND (web-based services OR web services) AND (policies OR policy OR preserving OR guarantee OR warranty OR violation OR solution OR data O

R mechanism OR security OR attack OR injection OR sql OR path OR command)) AND ttl(privacy) AND ttl(web services)

3 results

From the 46 results we selected 8 works that are the most related to our research. Their references are presented in Table 4-A. The conceptual map organizing the references is presented in Figure A-4. We used the offshoot *privacy → in → web applications and services.*

Table A-4. Related Work for Solutions and Tools for Privacy Protection.

Selected Related Works for Solutions and Tools For Privacy Protection
Gao, F., He, j., Ma, S. (2010). "A collaborative approach for identifying privacy disclosure in Web-based services," IEEE International Conference on Service-Oriented Computing and Applications (SOCA), 2010, pp.1-4.
Garcia, D., Allison D.S., Capretz, M., Toledo, M.B.F. (2010). "Privacy Protection Mechanisms for Web Service Technology". In 2010 Eighth ACIS International Conference on Software Engineering Research, Management and Applications, pp. 337–44.
Hewett, R. and Kijsanayothin, P. (2009). "On securing privacy in composite web service transactions," International Conference for Internet Technology and Secured Transactions ICITST 2009, pp.1-6.
Kim, K.I., Kim, W.Y., Ryu, J.S., Ko, H.J., Kim, U.M. and Kang, W.J. (2010). "RBAC-based access control for privacy preserving in semantic web". In Proceedings of the 4th International Conference on Uniquitous Information Management and Communication (ICUIMC '10). ACM, New York, NY, USA, Article 63, 5 pages.
Liu L., Huang Z., Xiao F., Shen G., Zhu H. (2012). "Verification of Privacy Requirements in Web Services Composition". In 2010 Second International Symposium on Data, Privacy, and E-Commerce, pp. 117–22.
Meziane H., Benbernou S. (2010). "A dynamic privacy model for web services". Computer Standards & Interfaces, vol. 32, issues 5–6, October 2010, pp. 288-304.
Meziane, H., Benbernou, S., Zerdali, A.K., Hacid, M.-S., Papazoglou, M. (2010). "A view-based Monitoring for Privacy-aware Web services," IEEE 26th International Conference on Data Engineering (ICDE), 2010 , pp.1129-1132.
Tbahriti S-E., Medjahed, B., Malik, Z., Ghedira, C., Mrissa, M. (2011). "Meerkat - A Dynamic Privacy Framework for Web Services". In 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. vol.1, pages 418-421.

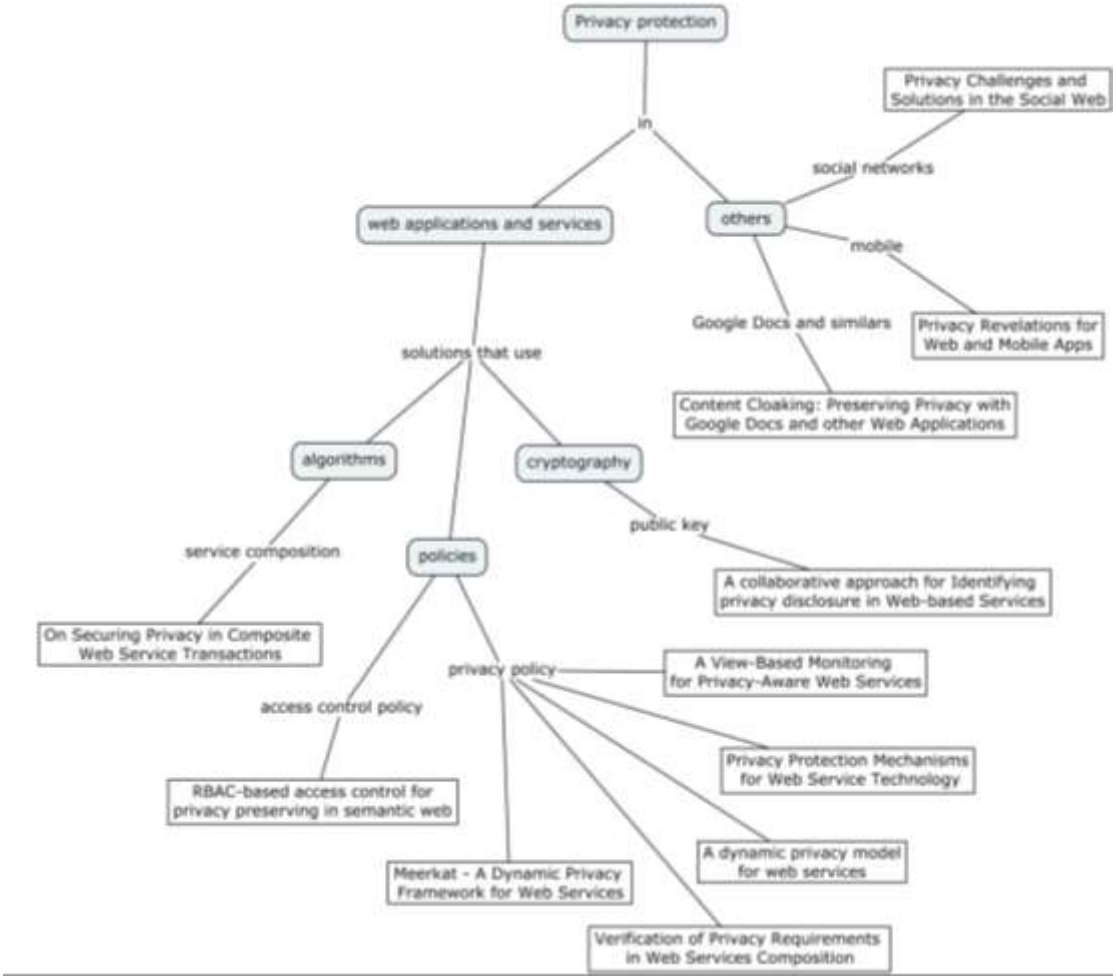

Figure A-4. Conceptual map for solutions and tools for privacy protection.

APPENDIX B - EVALUATION ELEMENTS OF THE REFERENCE ARCHITECTURE

Table B-1. Solove's taxonomy (Solove, 2006) and the privacy reference architecture correspondence.

Category	Description	Privacy Reference Architecture Element
Information Collection	Deals exclusively with privacy problems resulting from gathering information.	
Surveillance	Consists of methods of watching, listening and recording a subject's activities	Tracking detection
Interrogation	Describes methods an organization may use to ask or Elicit information from a subject	Privacy policy definition
Information Dissemination	Consists of privacy harms resulting from the release of information about a subject	
Breach of Confidentiality	Contains those harms based on the violation of a trust agreement to maintain confidentiality of a subject's information	Privacy policy enforcement
Disclosure	Describes harms related to the release of truthful information about a data subject.	Privacy policy enforcement
Exposure	Describes the dissemination of information about a subject's grief, body or bodily functions	Anonymization techniques
Increased Accessibility	Consists of the ways that a subject's public information may be made available to a wider audience than before.	Access Control
Blackmail	Involves a threat made to a data subject about a potential release of their information	Attack detection
Appropriation	Describes the use of a subject's identity or information to serve the purposes of the organization rather than the subject	User pattern identification
Distortion	Consists of harms related to the release of falsified information about a data subject.	Privacy policy enforcement
Information Processing	Describes methods to store, modify or manipulate a subject's information	
Aggregation	Combines individual and previously separate pieces of data about a subject	Anonymization techniques
Identification	Depicts an organization's methods for determining Which individual is described by a set of data	Access control Policy Definition
Insecurity	Is a failure to properly protect stored data	Attack Detection
Secondary use	Reflects the use of data for a purpose Other than that for which it was originally provided	User preferences/privacy policy enforcement
Exclusion	Describes the inability of a subject to have Knowledge of how their data is being used	Privacy policy enforcement
Invasion	Consists of the various intrusions on an individual's private life	
Intrusion	Is a form of invasion that describes all harms resulting from the disturbance of an individual's peace and solitude	Attack detection
Decisional interference	Is an invasion into a subject's decisions about their private affairs	User preferences

Table B-2. Questionnaire applied to stakeholders for the reference architecture evaluation process.

Statement	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
The RA has a pleasant and presentation tem uma apresentação gráfica agradável e legível					
As soon as I visualize the architecture I already know about what it refers					
The presented content is clear and consistent					
The understanding of each component of RA is easy					
All components are clearly classified and according to their goals					
I liked the presentation of RA					
I feel confident using this architecture to participate in the construction of applications and web services with aspects of privacy protection					
The RA is very relevant and should be used to understand the domain of privacy					
The RA is very relevant and should be used to build applications and web services with aspects of privacy protection					
Do you have any additional comments about the ease of use of RA? How do you think this architecture can be improved? How do you consider that the architecture would help in developing a software product?					

Table B-3. HP's architecture (Mont *et al*., 2005) and the privacy reference architecture correspondence.

Category	Description	Privacy Reference Architecture Element
HP Validator	Evaluate the policies to make decisions. Make "Yes & constraints" decisions, i.e. Decisions where access to data is allowed subject to the satisfaction of further privacy constraints -such as filtering out/obfuscating or statistically transforming part of these data;	Privacy Policy Enforcement/Access Control
Policy repository	Repository of privacy policies	Not inherent to the AR
HP Policy builder	Component to author access control policies and privacy policies. It Checks (at the enforcement time) the requestor's intent against the stated data storage purpose, take into account data subjects' consent & data retention policies and describe how the accessed personal data Has to be filtered, obfuscated or manipulated, etc.	Privacy policy definition/access control policy definition /user preferences
Audit	Log (among other things) Requests to access data and related decisions made by the enforcement system	Auditing
Data Enforcer	Is in charge of enforcing privacy decisions made by the Validator. It intercepts incoming calls to data resources, interacts with the Validator, performs fine grained manipulation of data resources and deals with the interpretation and enforcement of additional constraints as defined by the privacy policies	Privacy Policy Enforcement
Privacy Policy Enforcement	Enforces the privacy policies	Privacy Policy Enforcement
Privacy Policy Deployment	Framework for deploying both access control and privacy-based policies and making access decisions based on them	Privacy Policy Enforcement /Access control
Data Inventory & Privacy Policy Authoring	Allow administrators to express different kinds of privacy constraints.	Privacy policy definition

Table B-4. Academic architecture (Bodorik and Jutla, 2008) and the privacy reference architecture correspondence.

Component	Description	Privacy Reference Architecture
PI Monitor Agent	Intercepts and examines web services requests and replies for private information	Privacy Policy Enforcement
Enforcement/Monitoring Rules	Checks rules that determine how the request is handled from the privacy point of view	Privacy Policy Enforcement
Audit Log	Stores information about the request when a web service is invoked	Auditing
PI Agent	Mines the audit log and updates the KB when there is richer or different context to be gained due to users' and applications' actions	Privacy Policy Enforcement/Definition
Privacy Knowledge Base	Captures information on applications, web services they invoke, context of invocation, and private information stored, managed, and used by the enterprise	Privacy Policy Enforcement/Definition

APPENDIX C – EVALUATION ELEMENTS OF THE UML PROFILE

Table C-1. Selected statements from the Google's privacy policy**.**

ID	Statement
ST1	*Many of our services require you to sign up for a Google Account. When you do, we ask for personal information like your name, email address, telephone number, or credit card.*
ST2	*We may collect device-specific information, unique device identifiers and mobile network information.*
ST3	*When you use a location-enabled Google service, we may collect and process information about your actual location.*
ST4	*If other users already have your email or other information that identifies you, we may show them your publicly visible Google Profile information, such as your name and photo.*
ST5	*We will share personal information with companies, organizations or individuals outside Google when we have your consent to do so.*
ST6	*If your Google Account is managed for you by a domain administrator, then your domain administrator […] will have access to your Google Account information (including your emails and other data).*
ST7	*We provide information to our affiliates […] to process it for us, based on our instructions and in compliance with our Privacy Policy.*
ST8	*We may share aggregated, non-personally identifiable information publicly and with our partners […]. For example, […] to show trends about the general use of our services.*
ST9	*We restrict access to personal information to Google employees, contractor and agents who need to know that information in order to process it for us.*

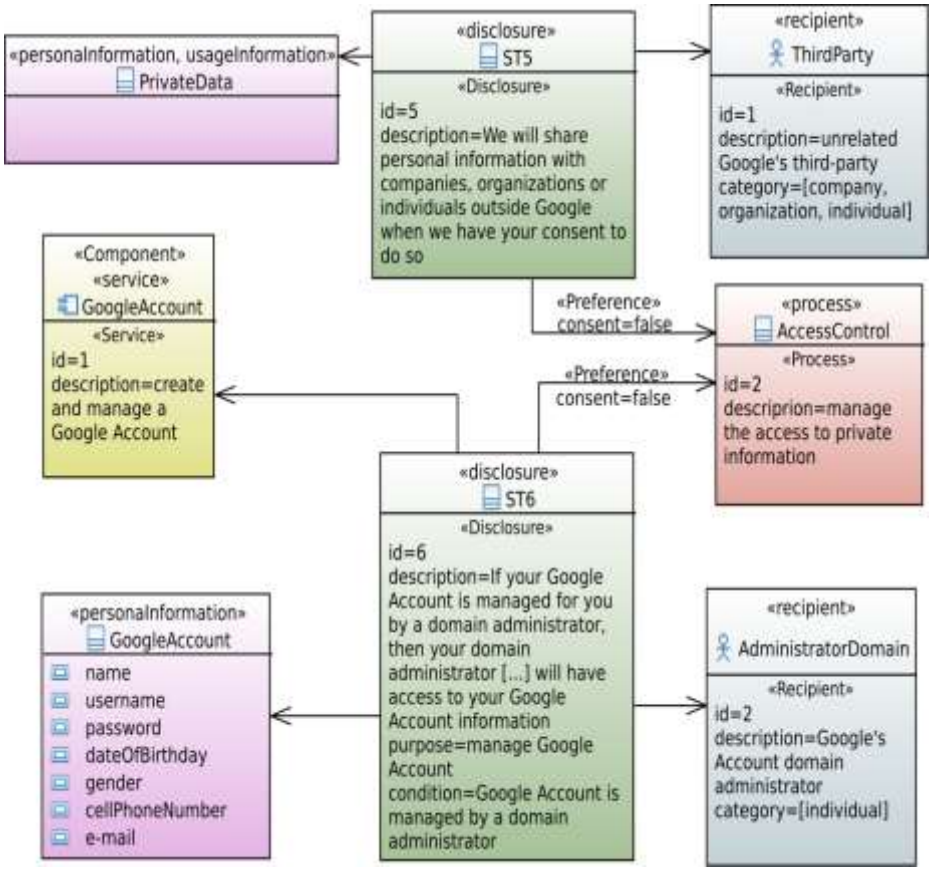

Figure C-1. Representation of statements ST5 and ST6 using the privacy profile.

In Figure C-1, both the statements are represented by <<Disclosure>> elements: ST5 describes a generic disclosure of data from Google to other third parties, while ST6 describes the possibility that some specific data of the Google Account will be available to the domain administrator (see Table C-1). In this case, the *purpose* for which data to be disclosed is the management of the Google account.

The two different types of recipients to which the statements refer are represented by <<Recipient>> elements. The *ThirdParty* recipient is connected with ST5, and represents a generic third-party unrelated to Google; the categories assigned to this recipient are derived from the statement itself: company, organization, individual. The *AdministratorDomain* recipient represents the domain administrator, and it is categorized as an individual. The *AdministratorDomain* is associated with ST6.

In the case of *opt-out* preference by the data subject, both the statements are enforced through access control. This is represented by the *AccessControl* <<Process>> element, since realizing access control requires the implementation of software and organizational measures. Both ST5 and ST6 are then connected with the *AccessControl*

element through a relation with the <<Preference>> stereotype. The *consent* attribute set to *false* reflects that access control is applied in case when the user disagrees with the statement.

The acceptance of statement *ST6* is mandatory for using the Google Account service; this is represented by the association between ST6 and the GoogleAccount <<Service>> element.

A complete model should also include a <<PrivacyPolicy>> element, having a containment relation with all the statement elements included in the model. For this case study, <<PrivacyPolicy>> contains the three statements represented in the diagrams (ST5, ST6, ST9), as well as the other six mentioned in Table C-1 (ST1 to ST4, ST7, ST8). To simplify the presentation, the <<PrivacyPolicy>> element was not shown. A complete diagram representing all the statements aggregated to the <<PrivacyPolicy>> is presented in Figure C-2.

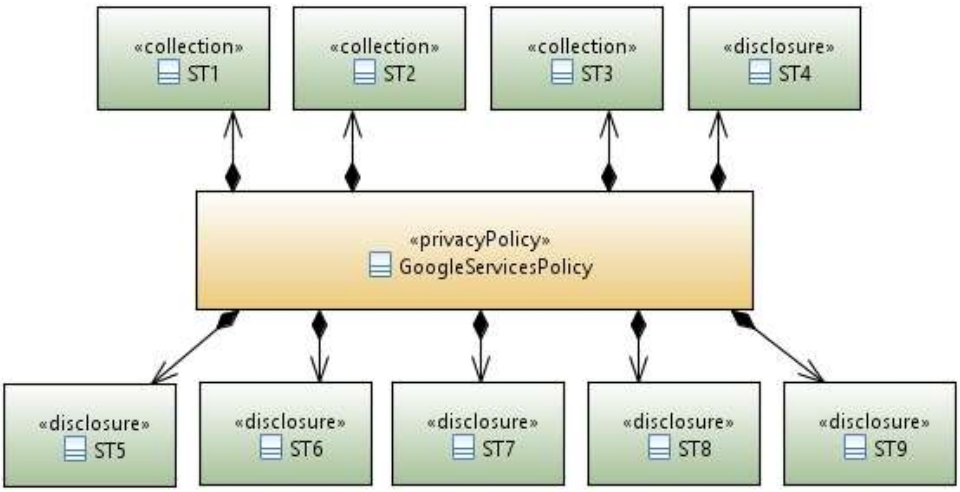

Figure C-2. Privacy Policy element and corresponding Statements

APPENDIX D – DETAILED DESCRIPTION OF THE POLICY MODEL, THE CRITICALITY LEVELS AND THE PRIVACY DATABASE FRAMEWORK

POLICY MODEL

The policy model is represented in an XML-schema file that contains the definition of the structure of an XML document and is used to derive different policies from the same structure. As illustrated in Figure 2, the model implements all the characteristics mentioned in the previous section. In a broad view, it defines *who* can access certain information, *when, from where, and how* the required information can be accessed.

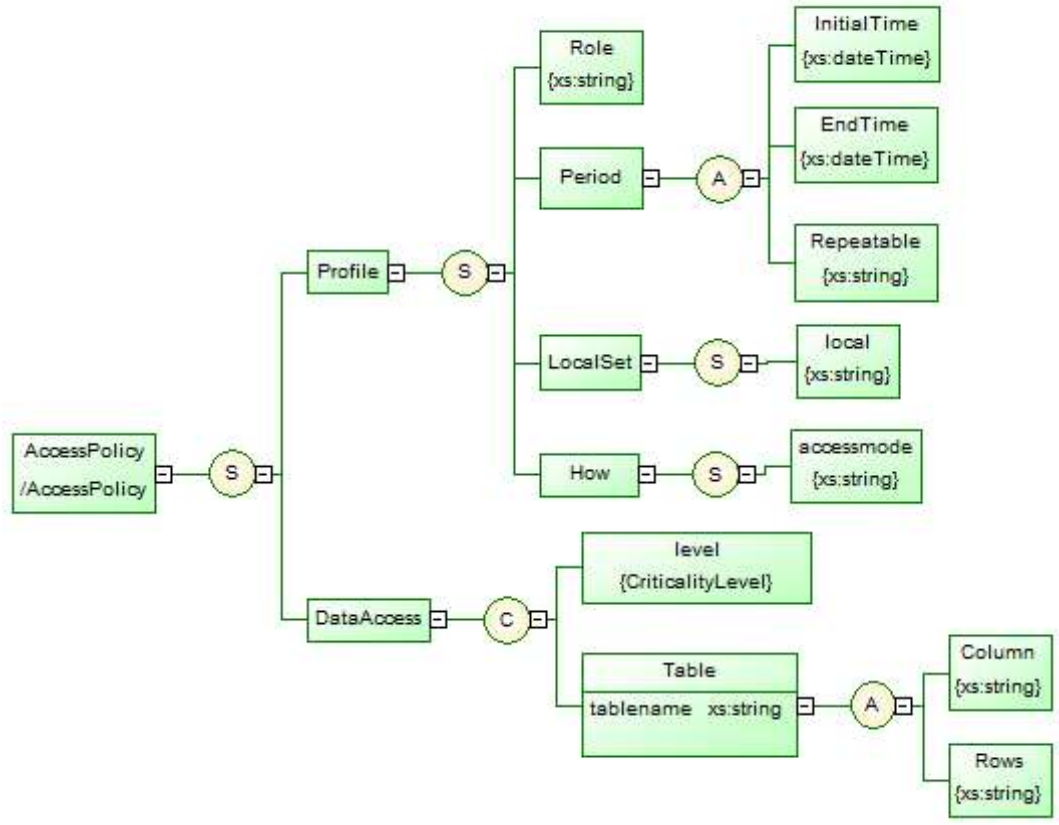

Figure D-1. Policy model.

The tag *AccessPolicy* in the model is the start point to define a policy. It is composed by two other tags: the *Profile* and *DataAccess*, both connected by a selector *S*. The selectors *A, S* and *C* mean, respectively, *All, Sequence* and *Choice*. *Sequence* means that all

the elements must appear at least once in the order of their declaration. *Choice* means that one element must be chosen. And *All* means that each element can appear or not, in any order (Sybase, 2013).

The *DataAccess* offshoot was designed to meet the requirements of restricting columns and rows information from tables, which is the major concern of IT professionals interviewed during the development of this work. From *DataAcess*, two tags were created and must be chosen: *Level* or *Table*. The *Level* tag represents an alternative way of restricting the information: according to their criticality level. The *Table*, C*olumns* and *Rows* tags represent the information in table columns and rows that will be restricted.

The *Profile* offshoot was designed to implement the groups and conditions requirements, described as very important in the literature and also pointed as necessary by almost half of IT professionals interviewed to this work. The *Role* tag represents the user groups. The *Period* tag restrict when the information can be accessed in terms of date and hour (through *InitialTime, EndTime* and *Repeatable* tags). *LocalSet* tag restricts one or more locations from which user access the information (e.g. local network or remote access). *How* tag represents the means through user can access information (e.g. from a web application, web service or a SQL console).

From the model presented in Figure C-1, it is very simple to derive policy files, which are represented as XML files.

CRITICALITY LEVELS

A typical database application manages data with different requirements in terms of security, ranging from non-critical data to data that has to be extremely protected against unauthorized access. These requirements can be represented through data criticality levels. These levels can be configured, added or even removed. In order to identify the different levels of criticality we established, for our study, the 4 levels described by Vieira and Madeira (2005). They are:

- Level 1: non-critical data, i.e., data that does not represent any confidential information.
- Level 2: data in this level must be protected against unauthorized modification (for this class of data unauthorized read is less critical than unauthorized modification). One typical example is the list of products in an online retail store. This information has to be

protected against modification (because it is used by the customers to perform orders) and should be open to all users.

- Level 3: data in this level must be protected against unauthorized read and modification. Most of the data in typical database applications is in this criticality level. Some examples are: clients' orders, costumers' information, and employees' information.
- Level 4: critical data that has to be extremely protected against unauthorized read and modification. This data must not be understandable even if someone is able to access the database using a valid username/password (i.e., this data has to be stored encrypted in the database). Typical examples are: usernames/passwords, credit card numbers, patients' files in hospitals, and bank accounts.

The levels we adopted are a good option to represent commercial online applications, but companies and organizations can establish their own levels according to their necessities.

PRIVACY FRAMEWORK DATABASE

The entity-relationship diagram in Figure C-2 represents a set of tables storing information about the privacy policies, including users and visitors preferences.

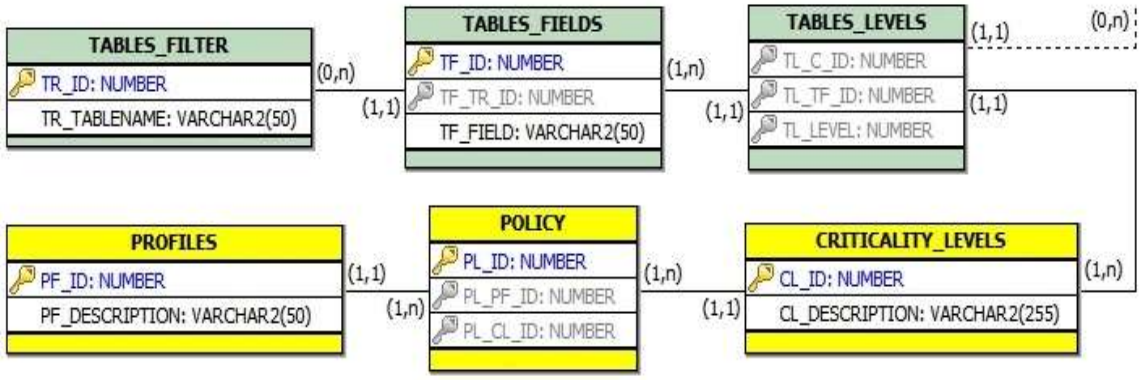

Figure D-2. Entity-Relationship Diagram of the database framework.

The information in the XML policies (constructed based on policy model described above in this appendix) is mapped to the *Policy* table, addressing the levels of criticality of the information that the profiles can access (e.g. system administrator can access information with levels 1 to 4; trainees can access information with level 1 and 2). The users

preferences are stored in *Tables_Levels*, addressing the criticality level for each information to be protected (e.g. if a user defines the phone number as level 3, this value can be accessed only by the system administrator). The functions of each of the tables are explained below.

- *Profiles*: stores all the different system profiles existing in the organization as, for example, administrator, customer, vendor, etc.
- *Criticality_Levels*: stores criticality levels, i.e., the default values adopted by the company or organization according to their needs. The definition of such criticality levels must be done in a thoughtful way because they will be associated to each user personal data.
- *Policy*: associates the profiles and criticality levels, specifying, through criticality levels, the information each profile can access. This table stores, in the form of data, the privacy policies defined through XML files and presented to users and visitors.
- *Tables_Filter*: stores the name of the tables whose fields will have the access controlled. Typically, these tables are the ones that stores data that pertains to profiles which express their privacy preferences (e.g. customers) and associated tables (e.g. address, country, phone numbers, etc.)
- *Tables_Fields*: stores the fields of the tables (specified at *Tables_Filter*) that will have restricted access.
- *Tables_Levels*: stores the users (data owners) preferences. Typically, *Tables_Levels* is associated to the table that stores the data of the person or profile subject to data privacy (e.g. customers and its associations).

The mapping from XML policy files to tables is done as follows: the *Profiles* and *Criticality_Levels* tables must be pre-fulfilled according to the company's criteria. As the policies establish the profiles and the levels of criticality of the information that these profiles can access, the content of the *Role* tag in XML is checked to exist in the *Profiles* table. The same verification is done to the criticality level, i.e., the job checks if the content of the *Criticality_Level* tag in XML exists in the table *Criticality_Levels*. If both informations are in their corresponding tables, the *Policy* table is fulfilled, characterizing the privacy policy. If they are not, a message is sent notifying the policy incompatibility.

For collection of users (data owner) preferences, tables *Tables_Filter* and *Tables_Levels* must also be pre-fulfilled according to the company's criteria. The user specifies his/her privacy preferences for each piece of data to be collected and managed through the criticality levels and these preferences are stored in the *Table_Levels* table. The records of this table specify the criticality level for each field of each table to be protected.

Default values can be set at first and then changed by users, via application, to express their preferences. In Figure C-2, *Table_Levels* has a relationship represented by a dashed line, not associated with another table. This line represents the relationship that *Table_Levels* has with the applications tables. These applications tables store the fields that should be protected.

APPENDIX E – REQUIREMENTS IDENTIFIED BY PROFESSIONALS USING PRIVAPP

ORIGINAL DOCUMENT FROM PROFESSIONAL P1(in Portuguese)

1. REQUISITOS FUNCIONAIS

ID	Descrição
RF1	O sistema deverá permitir o cadastro e armazenamento de informações de produtos (livros)
RF2	O sistema deve permitir que o usuário se cadastre, armazenando informações triviais para envio correto da mercadoria (Data de nascimento, nome, endereço, telefone e e-mail)
RF3	O sistema deve permitir com a permissão do usuário, o armazenamento de informações para futuros pagamentos utilizando a opção 1 clique (numero do cartão), caso o usuário não permita, está informação deve ser usada para gerar a ordem de pagamento e imediatamente "esquecida".
RF4	O usuário só poderá ter acesso há informações de seu cadastro, não podendo acessar informações pessoais de outros usuários
RF5	O sistema deve ser protegido contra ataques de quebra de senha por força bruta, e se detectado a tentativa deve-se bloquear o acesso do I.P atacante por um período de tempo (+-10 min)
RF6	Deve – se ser possível identificar através de analises de padrões não usuais de um usuário que a conta foi comprometida e assim que identificado enviar um e-mail ao usuário e bloquear a conta. Ex.: Usuário nunca gastou mais que R$ 200,00 em um mês e em um mesmo dia ele gasta mais de R$200,00
RF7	O armazenamento de senha do usuário não pode ser feito em texto claro, e sim com um mecanismo de criptografia que não possibilite a recuperação dessa mesma senha.
RF8	Deve ter um mecanismo de esquecimento de senha (link), onde o usuário receberá um link para troca de senha em caso esquecimento.
RF9	A comunicação entre o servidor web e o servidor de banco de dados deve ser criptografada utilizando SSL
RF10	O sistema deve armazenar informações de padrões de acesso de usuários, como tipo de livros que acessa, valores que gasta mensalmente, localização de onde acessa e seções do site que visita
RF11	Deve-se armazenar logs de transações e acesso por tempo mínimo de 1 ano, esse dados devem estar acessível para auditorias somente por usuário administrativos.
RF12	O sistema deve permitir diferentes níveis de controle de acesso, o usuário que só terá acesso aos seus dados podendo inclusive alterar em qualquer momento, o funcionário que terá acesso aos pedido realizados e cadastro de produtos, o administrativo que poderá ter acesso na base de usuários e produtos.
RF13	A sessão entre o cliente e o servidor (loja virtual) deve ser criptografada utilizando SSL

2. REQUISITOS NÃO FUNCIONAIS

ID	Descrição
RNF1	A resposta para qualquer operação (cadastro, compra...) não poderá ultrapassar 8 seg.
RNF2	O acesso simultâneo de usuário terá que ser expansível, aumentando conforme a demanda.
RNF3	A interface terá que ser amigável de fácil navegação.

3. REGRAS DE NEGÓCIO

ID	Descrição
RN1	Suas ações no nosso sistema são analisadas, para assim podermos melhor atendê-lo indicando produtos que mais lhe interessam.
RN2	Cookies são identificadores únicos que transferimos para o seu dispositivo para permitir que nossos sistemas reconheçam seu dispositivo e fornecer funções, como compra com apenas um clique e recomendações personalizadas.
RN3	Dados de nossos clientes não são repassados para terceiros.
RN4	Nós trabalhamos com a suposição de quê nossos clientes usam máquinas seguras para realizar as operações em nosso site

4. CONDIÇÕES DE TESTE

Será testado o comportamento de todos os sistemas, envolvendo os requisitos funcionais e não-funcionais, o ambiente de teste deve ser o mais próximo possível do ambiente de produção, observando informações como versão de programas usados no servidor, devem ser testados requisitos incompletos e não documentados, é necessário que o cliente realize testes de aceite do software

4. INFORMAÇÕES TÉCNICAS

Os testes deverão rodar em um servidor com as configurações mínimas abaixo:
Processador quad-core 2,6GHz
memória de 4GB 1333MHz
Disco Sata 300GB 7.2K RPM

ORIGINAL DOCUMENT FROM PROFESSIONAL P2(in Portuguese)

1. REQUISITOS FUNCIONAIS (NÃO ESTOU COLOCANDO UMA ORDEM DE PRIORIDADE)

ID	Descrição
01	Sugerir a instalação do plugin módulo de segurança ao iniciar a navegação (não sei se vale a pena obrigar essa instalação para uma web store. Eu sugeriria a instalação do plugin de segurança para ganhar um voucher de 5% na compra, por exemplo...).
02	Fazer login do usuário através de um username e uma senha.
03	Cadastrar dados do usuário (Username, Nome, Endereço, Cidade, Estado, CEP, País, Telefone e conta de e-mail, além de dados opcionais).
04	Cadastrar dados de pagamento (Tipo de cartão de crédito, Nome do titular do cartão – DESNECESSÁRIO por questões de segurança; eu usaria uma chave estrangeira que seria a chave primária da tabela que mantém os dados do usuário da linha anterior, só que criptografada para dificultar o cruzamento de dados, Número do cartão de crédito e Data de expiração do cartão).
05	Gravar os pedidos de compra com os devidos dados. Assim como no item anterior, eu vincularia todos os dados de pedidos de compras com um campo (chave estrangeira) que seria o código identificador do cliente (chave primária) de forma criptografada. Dessa forma, todo dado recuperado de formulário ou transação SQL efetuada protegeria a anonimidade do usuário.

2. REQUISITOS NÃO FUNCIONAIS (NÃO ESTOU COLOCANDO UMA ORDEM DE PRIORIDADE)

ID	Descrição
01	Utilização de protocolo de hipertexto seguro (HTTPS) com a configuração dos protocolos SSL ou TLS no servidor web e utilização de certificados digitais reconhecidos por uma ICP (Globalsign, Verisign, etc) para garantir a identidade do web site ao usuário final.
02	Criação de honeypots para monitorar possíveis atividades maliciosas. As honeypots podem ser hosts com a mesma configuração dos servidores em produção (web, DBMS, etc.), mas não precisam ter os dados reais. Servem apenas para identificar possíveis invasores e rastrear suas atividades para posteriormente corrigir falhas e fortalecer o sistema de firewall.
03	Utilização de uma infraestrutura de firewall consistente, com filtragem de tráfego, antivírus, proxies de acesso aos principais serviços, honeypots, monitoramento de logs de acesso, etc.
04	Atualização de sistemas operacionais e serviços, bem como correção de falhas de software.
05	Política de acesso aos dados em várias camadas, ou seja, implementação de vários níveis de autenticação ou autorização para obter acesso a dados gravados no DBMS. Isso evita que um funcionário ou empresa parceira tenha acesso de informação maior do que o necessário e também dificulta o roubo de dados por usuário maliciosos.
06	Criação de um plugin "módulo de segurança" que possa ser baixado e instalado pelo usuário, que faça verificações de segurança no computador e reforce a segurança do comprador, pois em algum momento ele digitará dados de cartão de crédito, por exemplo.
07	Gravação dos dados dos usuários de forma criptografada no DBMS para dificultar o uso de dados não autorizados em caso de roubo de informações e/ou intrusão.
08	Efetuar auditorias de acesso aos dados gravados no DBMS por funcionários e parceiros autorizados e monitorar constantemente (na maioria dos casos o maior inimigo está dentro da organização e não fora).

3. REGRAS DE NEGÓCIO

ID	Descrição
-	-

4. PROTÓTIPOS

N/A

5. CONDIÇÕES DE TESTE

N/A

6. INFORMAÇÕES TÉCNICAS

N/A

ORIGINAL DOCUMENT FROM PROFESSIONAL P3(in Portuguese)

1. REQUISITOS FUNCIONAIS

ID	Descrição
1	Exibir Catálogo de Produtos
2	Adicionar produto ao carrinho de compras
3	Remover produto do carrinho de compras
4	Alterar quantidade de produtos no carrinho de compras
5	Buscar produto no catálogo de produtos
6	Selecionar produto
7	Efetivar pedido
8	Registrar Usuário
9	Efetuar Login
10	Efetuar Pagamento
11	Exibir carrinho de compras
12	Cadastrar Categoria
13	Cadastrar Produto
14	Gerar usuário e senha
15	Gerar ordem de compra
16	Executar ordem de pagamento
17	Comprar com apenas um click
18	Gravar Cookies

2. REQUISITOS NÃO FUNCIONAIS

ID	Descrição
-	-

7.3

3. REGRAS DE NEGÓCIO

ID	Descrição
-	-

4. PROTÓTIPOS

N/A

5. CONDIÇÕES DE TESTE

N/A

6. INFORMAÇÕES TÉCNICAS

N/A

QUESTIONNAIRE APPLIED TO PROFESSIONALS P1 AND P2 (in Portuguese)

Questões	Concordo fortemente	Concordo	Indeciso	Discordo	Discordo fortemente
Os elementos da abordagem possuem uma apresentação gráfica agradável e legível.					
Logo que visualizo a abordagem e seus respectivos elementos já sei sobre o que ela se refere.					
O conteúdo apresentado da abordagem, como um todo, é claro e consistente.					
O entendimento de cada elemento da abordagem é fácil.					
Todos os elementos estão classificados claramente e de acordo com seus objetivos.					
Gostei da apresentação da abordagem.					
A abordagem possui todos os recursos necessários para proteger privacidade.					
É possível acrescentar facilmente novos elementos se necessário.					
Eu possuo todas as informações necessárias para entender o domínio de privacidade de aplicações e serviços web.					
Eu me sinto seguro utilizando essa abordagem para participar do processo de levantamento de requisitos de aplicações e serviços web com aspectos de proteção de privacidade.					
*Eu me sinto seguro utilizando essa abordagem para participar do processo de construção de aplicações e serviços web com aspectos de proteção de privacidade.					
A abordagem apresentada é bastante relevante e deve ser utilizada para entender o domínio de privacidade.					
A abordagem apresentada é bastante relevante e deve ser utilizada para construir aplicações e serviços web com aspectos de proteção de privacidade.					

Você tem algum comentário adicional sobre a facilidade de uso da abordagem?

Como você acha que essa abordagem pode ser melhorada?

Como você considera que a abordagem auxiliaria no desenvolvimento de um produto de software?