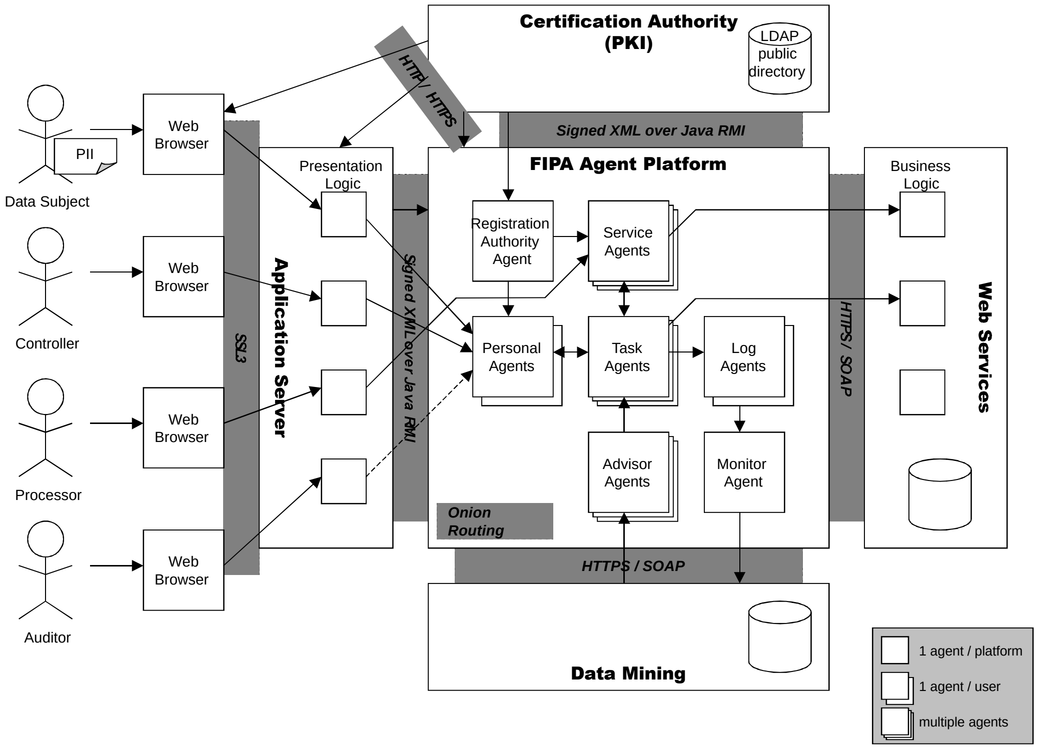
Handbook of Privacy and Privacy-Enhancing Technologies

The case of Intelligent Software Agents



Editors:

G.W. van Blarkom
J.J. Borking
J.G.E. Olk

PISA

Privacy Incorporated Software Agent

Handbook of Privacy and Privacy-Enhancing Technologies

The case of Intelligent Software Agents

Handbook of Privacy and Privacy-Enhancing Technologies

The case of Intelligent Software Agents

PISA Consortium

Editors:
G.W. van Blarkom RE
drs. J.J. Borking
dr.ir. J.G.E. Olk

Project coordinator:
ir. J. Huizenga

Handbook of Privacy and Privacy-Enhancing Technologies - The case of Intelligent Software Agents
The Hague, 2003

ISBN 90 74087 33 7

Print: TNO-FEL, P.O. Box 96864, 2509 JG, The Hague, The Netherlands
Publisher: College bescherming persoonsgegevens, copyright ©2003
P.O. Box 93374, 2509 AJ, The Hague, The Netherlands
email: info@cbpweb.nl

Preface

Houston, we have a problem! That's what we thought when we saw the 1986 video clip of Apple Corp. about The Knowledge Navigator issued at the introduction of its personal digital assistant the Newton in 1987. In this video clip visualizing the work of a professor in 2012 his intelligent software agent Phil is acting autonomously in order to accomplish tasks for his master in a complex network environment. The lack of supervision of the intelligent software agents (ISAs) from its masters, could lead to undesirable actions, such as violation of the privacy. In January 1999 the Dutch Data Protection Authority (NDPA, College bescherming persoonsgegevens) together with TNO published a technology assessment study about intelligent software agents and privacy. ISAs can appear in the role of a benevolent digital butler or a malicious digital criminal or as a means for constant surveillance. It became quite clear that fundamental research was needed about the risks of intrusion into privacy and the solutions that were needed to minimize privacy violations by agents.

An international research consortium was set up to investigate whether an ISA can be built that respects and protects the privacy of its users and that of other users while acting on the Internet. The consortium with the acronym PISA (Privacy Incorporated Software Agent) started its work on 10 January 2001 researching for and subsequently building a privacy guardian for the electronic age and not without success! A working job-acquiring agent has been designed and built as a demonstrator system showing how privacy can be safeguarded.

The PISA research has been funded by the EU under its IST 5th Framework program and this Handbook that is lying before you is one of the milestones of this project. The Handbook reports what results have been achieved and what needs to be done to create privacy safe agents. It shows what architecture has been developed by the consortium for information and collaboration systems for telecommunications, information technology and media, what EU privacy law rules apply to ISAs and what Privacy-Enhancing Technologies (PETs) are needed to create a Privacy Incorporated Software Agent (PISA). Especially the fundamental research on privacy ontologies in this project translating privacy law into object code for automatic execution of the law while processing personal data, will be continued in a the new EU research project PRIME (Privacy and Identity Management) that will start in December 2003. Without the research grant of the EU under the IST fifth framework program this timely research would never happened. The researchers are most grateful for this.

The consortium acknowledgements are to Mr. Andrea Servida, project officer of the European Commission and the reviewers Prof. Piereangela Samarati (UniMilan), Dr. Alberto Escudero-Pascual (UniStockholm) and Helena Lindskog (Ericsson) for their encouragements and positive remarks during the review meetings of the PISA project without whom this project would never yielded the important results for the protection of privacy of the EU and world citizens.

We would also like to thank all participants in this project for the extra miles that were necessary to achieve the results of this project and our PISA friends especially Unisys

Netherlands whose support was crucial for the start of our work.

We hope that this Handbook will be of assistance for software agents developers as well for those that will order for such intelligent systems, the users of these agents and the Privacy Commissioners worldwide.

The Hague, 28 November 2003

Jan Huizenga
PISA coordinator
TNO

John Borking
PISA dissemination
NDPA

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This introduction chapter deals with the concepts: privacy, Privacy-Enhancing Technologies (PET), intelligent software agents and the relations between these subjects. Furthermore this chapter provides a roadmap of this handbook, i.e. about this book, how to use it, the objectives and an outline of this book. Do not try to read this book at one sitting: there is no need to. Instead use this book as a dictionary or an encyclopedia. The subjects discussed in this book are meant to be assistance for all those that are involved in the construction or use of intelligent software agents and that are concerned about the privacy requirements.

1.1 Background

Privacy is defined by Westin [Wes67] as:

> *the claim of individuals...to determine for themselves when, how and to what extent information about them is communicated to others.*

Privacy is a fundamental human right, defined in Article 8 of the 1950 European Convention of Human rights. It is one of the most important human right issues of our evolving information age.

Technological developments have led to more and more automatic processing of information and storage of such information in computer databases. Part of this information is personal information. Individuals valueing their privacy are concerned about the fact that so much personal information is routinely stored in computer databases over which they have no control.

In a number of countries in the world, notably the democratic ones, informational privacy of people is acknowledged and laid down in treaties, EU directives, national laws and other (non binding) regulations. For example: OECD guidelines on, Council of Europe, EU Directives 95/46, 99/93, 2002/58 and Canadian law. These regulations give the background of informational privacy and adhered rights, restrictions and obligations of the parties during exchange and storage of personal data[1]. Especially the increasing automatic processing of personal data via software systems, databases, electronic communication etc. has made the protection of privacy a pressing issue. Unfortunately, despite all the regulations that

[1] See article 2 EU directive definition.

are put on paper, there is no well established and world wide accepted view on the way privacy protection and the consolidation thereof can be build into software. So, because software products are almost by definition international, there is a need to develop such a view in the international arena, to assure that all privacy respecting countries follows ’good practice’.

On the one hand this necessitates the introduction of a methodology for the privacy consolidation, including privacy threat analysis; on the other hand this will necessitate a classification of software products, dealing with personal data, to ensure that no lightweight privacy products are used for heavyweight personal data. Such methodology that follows software development through every single stage is mandatory; like software quality and other similar aspects like reliability, security or system safety: it is no paint that can be added after everything else has been dealt with. To put it straightforwardly: postponement of the handling of personal data implications ’until a later phase’, may easily lead to an information system that is adversarial to privacy adaptations: some measures may have been necessary very early on in system development, before much of it is cast in the concrete of structure. See [HB98] for conceptual models for several types of information systems ensuring the prevention of privacy violations.

1.2 The case of agents

The tremendous growth of the Internet has made the Internet an ’information infrastructure’ for virtually every subject and application domain. For instance, e-commerce and e-government are becoming part of citizens’ everyday life. Unfortunately, the enormous and fast-growing amount of data available has led to an ’information overload’. For users it is becoming more and more difficult to find the information they need. The information might be out there but finding it simply takes too much effort and time. Also, information might be available that could be of interest to a user without the user actually specifically searching for this information.

One way of dealing with the large amounts of available information is to alter the way a user tries to cope with the information. Instead of having a user sit behind its computer and directly navigating through the information, the user could have the computer do this autonomously through so-called (intelligent) software agents. A software agent is basically a piece of software that can execute tasks with minimal interference by their user. A software agent could run on the computer of its user but could also move itself around on the Internet or some other network infrastructure.

1.2.1 Intelligent Software Agents

There is no general agreement on a definition of the word ‘agent’, just as there is no consensus within the artificial intelligence community on a definition of the term ‘artificial intelligence’. In general, one can define an agent as a piece of software and/or hardware capable of acting in order to accomplish a task on behalf of its user.

A definition close to present-day reality is that of Ted Selker from the IBM Almaden Research Center:

> *An agent is a software thing that knows how to do things that you could probably do yourself if you had the time.*

Agents come in many different flavours. Depending on their intended use, agents are referred to by an enormous variety of names, e.g., knowbot, softbot, taskbot, userbot, robot, personal (digital) assistant, transport agent, mobile agent, cyber agent, search agent, report agent, presentation agent, navigation agent, role agent, management agent, search and retrieval agent, domain-specific agent, packaging agent.

The word ‘agent’ is an umbrella term that covers a wide range of specific agent types. Most popular names used for different agents are highly non-descriptive. It is therefore preferable to describe and classify agents according to the specific properties they exhibit.

Intelligent software agents are software agents that while executing their assigned task are capable of reasoning and actually learning during the execution of the task. They thus seem to behave and act in an intelligent manner. For example, an agent with the task of keeping track of the latest movies can learn what movies the user likes based on for instance the actors featured in the movies or the type of movie (action, romantic, etc.). Over time the agent will learn what its user likes and dislikes. The agent might then actually buy tickets (electronically) for a movie with the user’s favourite actor when it is released.

1.2.2 Privacy and Software Agents

While executing its task, an agent can collect, process, store and distribute data. Some of these data could be personal data about individuals. For instance, to delegate tasks to an agent, a user needs to provide the agent with a user-profile containing personal data about the user like mail addresses, habits and preferences. Part of these personal data might be privacy-sensitive or might become privacy-sensitive when the agent processes the personal data or combines it with other data. As long as the agent does not process the personal data, the privacy of the individuals involved will not be violated. A potential violation of the privacy might occur when the agent communicates or exchanges (sensitive) personal data with its environment. Because agents are both collectors and processors of (personal) data, and therefore form a possible threat to the privacy of those involved, they need to meet the requirements specified in (national and international) regulations to ensure privacy. Various governments and international governmental organisations have drawn up privacy regulations and privacy guidelines.

1.2.3 Privacy-Enhancing Technologies (PET)

In order to help agents to protect internally stored personal data (that might be privacy–sensitive), Privacy-Enhancing Technologies (PET) have been developed over the years. PET stands for a coherent system of ICT measures that protects privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data, all without losing the functionality of the information system. PET try to manage the privacy threats that software agents face. Ultimately PET might replace inspections for enforcing the privacy regulations. Enforcing by means of inspections or audits to verify whether all organisations that collect personal data are complying with the privacy regulations is rather time-consuming and thus expensive. Goldberg [GWB97] gives an overview of existing and potential Privacy-Enhancing Technologies for the Internet.

PET help to tackle the two main types of privacy threats that are posed by the use of ISATs: threats caused by agents acting on behalf of a user (through the disclosure of the user’s personal information) and threats caused by foreign agents that act on behalf of others (via traffic flow monitoring, data mining, and even covert attempts to obtain personal information directly from the user’s agent).

There are four ways of using PET to protect an individual in an agent-based environment:

1. wrapping PET around the individual's agent;
2. integration of PET in the individual's agent;
3. combining wrapping and integration of PET;
4. using PET to create an infrastructure of trusted components.

Unfortunately the application of PET to ISAs in order to meet privacy regulations is not straightforward. Yet, this handbook illuminates the use of PET in software agents.

1.3 Objectives of the handbook

The objectives of the handbook on Privacy-Enhancing Technologies (PET) and Intelligent Software Agents (ISAs) are:

- to describe the privacy related issues (privacy preferences, privacy threats and possible solutions) to the use of software agents in general and illuminate these aspects by describing in detail several cases;
- to describe a solution (Privacy Ontology and Privacy rules and policies) for a software agent such that it can act according to the EC-Directive;
- to describe method for Privacy by Design;
- to describe technical solutions for Privacy protection for networks, human computer interfaces, public key infrastructures, cryptography and data mining and matching;
- to describe other initiatives and standards for privacy protection ands to discuss opportunities and area's of new research;
- to act as a guideline for designers of software agents to meet privacy regulations;
- to provide customers with information that allows them to draw up privacy aware system specification involving software agents.

1.4 About the Handbook

The handbook was established within the PISA (Privacy Incorporated Software Agent) project supported by the European Union (project RTD IST-2000-26038), see Appendix E and F. This handbook presents the results of the joint study by the Dutch Data Protection Authority, TNO, Delft University of Technology, Sentient Machine Research, FINSA Consulting, National Research Centre Canada and GlobalSign, as partners in the PISA project. With this study the partners have attempted to identify possible threats to the privacy of individuals resulting from the use of agent technology. Secondly, the study sought to identify and demonstrate ways of applying Privacy-Enhancing Technologies (PET) to agent technology in such a way as to eliminate the impact of these threats.

1.5 Outline of the Handbook

The handbook starts in Chapter 2 by generally describing and defining privacy and discussing the aspects involved with privacy like identity and personal data, legal interception, privacy preferences and policies, privacy threat analysis, privacy legislation. Chapter 3 discusses in detail Privacy-Enhancing Technologies (definition, history, legal ground, concept, and developments) and compares the Common Criteria used for information technology security evaluation with the principles of PET. PET is further illuminated by a description of several PET projects. Chapter 4 describes the development of agent technology and the the link between software agents and PET. Chapter 5 discusses the security and privacy protection aspects. An infrastructure, PKI, for providing security is described in Chapter 6. Chapter 7 discusses evaluation criteria to determine whether a software agent meets privacy regulations. In Chapter 8 a general architecture model is established for discussing privacy related issues of software agents. It is shown how measures to protect privacy can be incorporated. Other important architecture issues like a trust model and network aspects are discussed in Chapter 9. A design method keeping in mind privacy aspects is presented in Chapter 10. The technique of data mining is illuminated in Chapter 11 both as a concern for personal privacy and as an opportunity for privacy protection. Chapter 12 focusses on the Human Computer Interaction. Finally, Chapter 13 concludes the handbook and gives a view on future developments with respect to privacy and software agents that can be expected.

1.6 Reading the handbook

This handbook is targeted at designers and developers of Intelligent Software Agents (ISAs) as well as customers and people responsible for the use of ISAs. This handbook gives designers starting points to integrate privacy in a structured way in information and communication systems, specifically ISAs. Additionally, the handbook provides customers information required for drawing up system specifications and negotiate with designers and developers. The handbook tries to answer the following questions:

- What is privacy?
- What are software agents?
- What threats to privacy can be attributed to software agents?
- What (technical) measures are there to eliminate or reduce the impact of these threats?
- How to design a privacy incorporated system?
- How to audit and evaluate a privacy incorporated system?

To help the reader, guidelines are given for reading the handbook based on the reader's background and interests. The following types of readers are identified:

Developer. A developer of agent technology applications.

Manager. The person responsible for checking the design documents.

User. The user of agent technology applications.

Table 1.1: Overview of the chapters in the handbook of specific interest to the various types of readers.

type of	**handbook chapter**											
reader	2	3	4	5	6	7	8	9	10	11	12	13
developer	√	√		√	√		√	√	√	√	√	√
manager	√						√	√				√
user	√	√	√				√	√			√	√
decision maker				√		√						√

Decision maker. The person who decides to build a Multiagent system (MAS) and who is aware of all aspects, technical and legal, of such an environment.

Table 1.1 gives an overview of the chapters of the handbook of most interest to each different type of reader.

1.7 Acknowledgements

This handbook is the result of the collaboration of all partners of the PISA consortium. The following people contributed to this handbook (in no particular order): J.J. Borking (CBP), P. Verhaar (TNO-FEL), B.M.A. van Eck (CBP), P. Siepel (CBP), G.W. van Blarkom (CBP), R. Coolen (TNO-FEL), M. den Uyl (Sentient Machine Research), J. Holleman (Sentient Machine Research), P. Bison (Sentient Machine Research), R. van der Veer (Sentient Machine Research), J. Giezen (TNO-TPD), A.S. Patrick (NRC), C. Holmes (Carleton University), J.C.A. van der Lubbe (TUD), R. Lachman (TNO-FEL), S. Kenny (CBP), R. Song (NRC), K. Cartrysse (TUD), J. Huizenga (TNO-FEL), A. Youssouf (GlobalSign), L. Korba (NRC), J-M. Kabasele (GlobalSign), M. van Breukelen (TNO-TPD), A.P. Meyer (TNO-TPD), A.K. Ahluwalia (TNO-TPD), J.A. Hamers (TNO-TPD), J.G.E. Olk (TNO-FEL), A. Ricchi (FINSA), G. Manzo (FINSA), and X. Zhou (NRC).

Chapter 2

Privacy

G.W. van Blarkom J.J. Borking
gbl@cbpweb.nl jborking@euronet.nl
CBP, The Netherlands

J. Giezen
giezen@tpd.tno.nl
TNO-TPD, The Netherlands

R. Coolen P. Verhaar
coolen@fel.tno.nl verhaar@fel.tno.nl
TNO-FEL, The Netherlands

This chapter explains the concepts of privacy and data protection, the European Directives that rule the protection of personal data and the relevant definitions. It focuses on the question of when personal data items become non-identifiable, the sensitivity of data, automated decisions, privacy preferences and policies and the nine condensed privacy rules applicable to intelligent software agents. The reader shall also find in this chapter the legal security requirements and the method of assessing the privacy threats as a tool for privacy risk analysis. The development privacy threat analysis has been based on the Code of Practice for Risk Analysis and Management Method, Information Security Handbook for the Central Computers and Telecommunications Agency (CCTA).

2.1 Introduction to the Issue

In a number of countries in the world, notably the democratic ones, information privacy related to people is acknowledged and laid down in treaties, EU Directives, national laws and other (non-binding) regulations[1]. These regulations give the background to informational privacy and adhered rights, restrictions and obligations of the parties during the exchange and storage of personal data[2]. Personal data, however, shall not only be stored and read by the human eye. This use has privacy implications in itself. The advent of automatic processing via software systems, databases, electronic communication, etc., has also made the protection of privacy a pressing issue. Unfortunately, despite all the regulations that are put on paper, there is no well established and worldwide accepted view on the way privacy protection and the consolidation thereof can be built into software. There

[1] For example: OECD guidelines, the Council of Europe, EU Directives 95/46, 99/93, 2002/58 and/or Canadian law on privacy.

[2] See Article 2, EU Directive definition.

is, therefore, a need to develop such a view in the international arena to ensure that all privacy respecting countries follow ’good practices’ because software products are almost by definition international.

On the one hand, this necessitates the introduction of a methodology for privacy consolidation, including privacy threat analysis, and, on the other hand, this also necessitates a classification of software products, dealing with personal data, to ensure that no lightweight privacy products are used for heavyweight personal data. Such methodology that follows software development through every single stage is mandatory; such as software quality and other related aspects of which reliability, security and system safety are just but a few examples. You cannot just add it as an afterthought when everything else has been dealt with. In a nutshell, we could say that the postponement of dealing with personal data implications ’until a later phase’, may easily lead to an information system that is contrary to privacy adaptations. Certain measures may have been necessary very early on when developing the system before much of this system has been ’cast in stone’. See [HB98] for conceptual models for several types of information systems that ensure privacy violation prevention.

Legislative drafting seems to mirror this observation. For instance, Directive 2002/58/EC (Directive on privacy and electronic communications) states in Article 4 (Security), Paragraph 1, that ‘(preventive) measures shall ensure a level of security appropriate to the risk presented’ and is reinforced by the following addition to Recital 13: ‘the requirement to inform subscribers of particular security risks does not discharge a service provider from the obligation to take, at his or her own costs, appropriate and immediate measures to remedy any new, unforeseen security risks’[3].

Methodology is, of course, the first concern because this shall also produce insight on how to classify software systems, since, by necessity, it shall function as a benchmark for the quality of the privacy consolidation process. The more stringent the methodology that is applied, the better the software is expected to be, that is, from the privacy protection point of view; strict adherence to software quality standards and other non-functional restrictions still need to be ascertained. This paper presents a methodology on how this can be achieved to evoke a discussion on the subject of privacy consolidation as a step on the road to international consensus concerning the construction of privacy-safe information systems.

To facilitate the application of the privacy consolidation methodology, and of any methodology during the various stages of software development, qualified staff, budget and time are necessary resources. If, however, no thorough methodology is to be followed, the outcome can easily be that inadequate resources shall be allocated. The development of such a methodology shall, therefore, not just improve the awareness of the privacy issue for system developers, it shall also contribute to the allocation of the necessary facilities.

2.2 General Description of Privacy

Privacy is defined by Westin [Wes67] as:

> *the claim of individuals...to determine for themselves when, how and to what extent information about them is communicated to others.*

[3] Though it may be going too far to state the possibility that risk management shall be put on a similar regulatory footing to Auditing, it is certainly apparent that powerful forces are gathering strength around an *ex ante* as opposed to an *ex post* comprehension and implementation of measures protecting against moral and other damages arising from privacy violations.

Privacy is a fundamental human right as defined in Article 8 of the 1950 European Convention of Human Rights. It is one of the most important human right issues of our evolving information age [Ban00]. Informational privacy has two distinct characteristics:

1. The right to be left alone;
2. The right to decide oneself what to reveal about oneself.

The above means that, although it is a situation that is wanted by an individual, it primarily comprises a set of rules of conduct between the individual person and the person's environment with respect to personal data processing[4] in the Internet environment. A major step towards privacy protection in Europe was the adoption of Convention 108 of the Council of Europe. Today, informational privacy protection for individuals is expressed through different European Union Directives such as:

- 95/46/EC, Data Protection Directive hereafter DPD;
- 2002/58/EC, Directive on Telecommunications 2002;
- 99/93/EC, Digital Signature Directive and non-EU legislation[5].

This type of legislation defines a set of rights concerning personal data accruing to individuals irrespective of sector of application and creates obligations concerning the processing of data by third parties. The EU Privacy Directive 95/46/EC has two objectives:

1. Creating a high level of protection of personal data;
2. Enabling the free movement of data within the EU.

The EU privacy legislation has, furthermore, two main functions:

1. Empowering the individual to manage his or her own personal data and protecting these data within the limits as defined in the Directive;
2. Creating a protective legal regime when personal data are processed[6].

In PISA (see Appendix A, B, and E) we use the following definition for privacy:

> The claim of individuals to be left alone, free from surveillance or interference
> from other individuals, organisations or the state.

It comprises a set of rules of conduct between the individual person and the person's environment with respect to the manipulation of personal identifiable information even though it is a situation that is wanted by an individual. Personal data can be defined as the collection of all data that are or can be related to the individual. This includes factual data including their identification, physical data, behavioural data, social data, financial data and any other personal data that may be involved.

[4] This document uses the EU privacy legislation as the legal framework. This legislation uses the term personal data. The term Personal Identifiable Information is used instead of personal data in some environments. The terms are interchangeable, but this document shall use the personal data standard term.

[5] See the 2000 Canadian Personal Information Protection and Electronic Documents Act [Can00].

[6] Summary, Conclusions and Recommendations of the TNO Report, STB-02-13a, Privacy-Enhancing Technologies and Information Systems of the Central Administration, The Hague, 28 February 2002, commissioned by the Ministry of the Interior and Kingdom Relations (BZK), handed out 23 May 2002 in The Hague, after the symposium Privacy by Design organised by the Dutch Data Protection Authority (CBP).

2.3 Definitions

> "Directive 95/46/EC[7] of the European Parliament and of the Council of 4 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data."

The following definitions[8] apply and shall be used in this handbook throughout:

(a) *Personal data*[9] shall mean any information relating to an identified or identifiable natural person (*data subject*); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors that are specific to his or her physical, physiological, mental, economic, cultural or social identity;

(b) *Processing of personal data (processing)* shall mean any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organisation, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination, or otherwise making available, alignment or combination, blocking, deletion or destruction;

(c) *Personal data filing system (Filing system)* shall mean any structured set of personal data items that are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis;

(d) *Controller* shall mean the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data; where the purposes and means of processing are determined by national or Community laws or regulations, the controller or the specific criteria used for his or her nomination may be designated by national or Community law;

(e) *Processor* shall mean a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller;

(f) *Third party* shall mean any natural or legal person, public authority, agency or any other body than the data subject, the controller, the processor and the persons who, under the direct authority of the controller or the processor, are authorised to process the data;

(g) *Recipient* shall mean a natural or legal person, public authority, agency or any other body to whom data is disclosed, whether a third party or not; however, authorities which may receive data in the framework of a particular inquiry shall not be regarded as recipients;

(h) *The data subject's consent* shall mean any freely given specific and informed indication of his or her wishes by which the data subject makes it known that he or she does not object to personal data relating to him or her being processed.

[7] This Directive is hereafter referred to as the Data Protection Directive (DPD).

[8] See Article 2 "Definitions" of EU Directive 95/46/EC [EC95].

[9] PII (Personally Identifiable Information) is used instead of PD (Personal Data) in certain regions (e.g. the United States). This handbook consistently uses PD.

Note: the term *data subject* is implicitly defined. To simplify reading the term, *data subject*, once used in context, shall occasionally be replaced by the term *person*, without change of definition. Additionally, the term *Data Protection Authority* (or *Privacy Commissioner*) is used to indicate the legal authority in charge of providing privacy regulation support (through reporting, reprimands, enforcement or imposition).

2.4 Data Subject and Personal Data

This paragraph discusses several issues from the Directive applicable to the processing of personal data in general.

2.4.1 Identified or Identifiable

The legislator gives the definitions of the entities that relate to privacy legislation in Article 2 of the Data Protection Directive. In Article 2 (a) the definitions are given for "**data subject**" and "**personal data**":

> (a) "personal data" shall mean any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors that are specific to his or her physical, physiological, mental, economic, cultural or social identity;

The definition of data subject and of personal data distinguishes between two very important situations, namely whether a natural person is 'identified' by means of personal data items such as name, date of birth and gender or if a natural person is 'identifiable' by means of the available personal data items other than the type of items mentioned before. In legal terms, personal data means any piece of information regarding an identified or identifiable natural person[10]. Whether we can talk of 'personal data' depends on a number of elements of which, within the scope of this document, 'identification' is the only significant element. According to Article 2 of the EC Directive 95/46, a natural person can be identified 'directly or indirectly'. Direct identification requires basic details (e.g., name, address, etc.), plus a personal number, a widely known pseudo-identity, a biometric characteristic such as a fingerprint, PD, etc. Indirect identification requires other unique characteristics or attributes or a combination of both, to provide sufficiently identifying information (see [BR01]).

Non-identification is assumed if the amount and the nature of the indirectly identifying data are such that identification of the individual is only possible with the application of disproportionate effort, or if assistance by a third party outside the power and authority of the person responsible is necessary[11]. Whether we can talk of disproportionate effort depends, on the one hand, on the nature of the data and the size of the population and, on the other hand, the resources of time and money one is willing to spend in order to be able to identify the person. We can only note in passing here that the concepts of 'identification' and 'identity' are essentially contested in theory and are often arbitrary and ambiguous in practice (see [Raa99]).

[10] See Article 2 of EU Directive 95/46/EC; 'personal data' is defined in Article 1 of the UK Data Protection Act 1998 and there is a lot of confusion about identification and verification.

[11] See Recital 26 of the EC Directive 95/46 [EC95].

Internet identifiers such as an IP address, browsing activities of a user, session login details and the listing of websites visited by an Internet user are classified as personal data.

A data subject is identifiable if sufficient facts about a data subject are known. Example: an information system on a loyalty system database registers all details related to items sold for analysis purposes. It also registers the following demographical data items, in order to perform sales analyses, such as, gender, date of birth, loyalty card number and postal code. The DPD categorises this database to contain personal data because the controller can easily identify the data subject by matching the loyalty card number with the customer database. See recital 26: *whereas, to determine whether a person is identifiable, account must be taken of all the means that may reasonably be expected that shall be used, either by the controller or by any other person to identify the said person;*

The Common Criteria for Information Technology Security Evaluation (CC)[12] define the anonymity family as: the family to ensure that a user may use a resource or service without disclosing the user's identity. The requirements for Anonymity provide protection for the user identity. Anonymity is not intended to protect the subject identity. In order to anonymise personal data according to the CC, it is enough to remove those data items that directly identify the data subject. Anonymity is, however, more than that from a data protection point of view. Anonymity is intended to protect the data subject identity in relation to privacy legislation. The definition also categorises data as personal data as long as the data subject can be identified.

Given an information system that anonymises personal data according to the CC definition, it is very likely that the system still processes personal data according to the DPD. Example: date of birth must be transformed into an age category, the last positions of the postal code must be removed and the loyalty card number must be encrypted using a one-way hashing algorithm (this ensures that the items bought by one data subject can be linked without the ability to identify the data subject).

2.4.2 Special Categories of Data

The Privacy Directive defines several special categories of data in Article 8. Paragraph 1 defines the basic set of categories of special data and, subsequently, forbids the processing of such data.

1. Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and the processing of the data concerning health or sex life.

Paragraph 5 of the same Article adds data relating to offences, criminal convictions or security measures to this list. Finally, data relating to administrative sanctions or judgments in civil cases concludes this list. Member States are obliged to determine conditions under which national identification numbers or any other identifier of general application may be processed in the last paragraph of this Article.

Paragraph 2 of the same Article defines a number of well-defined conditions under which processing of these special categories could be lawful.

[12] See Section 3.2 for a more detailed description of the CC.

2.4.3 Automated Individual Decisions

The legislator has drawn special attention to automated decisions. In a MAS (Multiagent System) environment, some of the 'decisions' of an agent may well be classified as an automated decision. Article 15 of Directive 95/46/EC defines the automated decision as:

1. Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.

When it comes to the rights of the data subject, the Directive states, in Article 12 (a) third dash, that the data subject has the right to obtain knowledge of the logic involved in any automatic processing of data concerning said data subject at least in the case of the automated decisions referred to in Article 15 (1). This part of the privacy legislation may become a real issue if an agent somewhere down the line makes a 'decision' as to the facilities provided by the next agent to pass personal data onto.

2.5 Controller and Processor

The definitions are given for "**controller**" and "**processor**" in Article 2 (d) and (e):

(d) "Controller" shall mean a natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data. Where the purposes and means of processing are determined by national or Community laws or regulations, the controller or the specific criteria for his or her nomination may be designated by a national or Community law;

(e) "Processor" shall mean the natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller.

The most important parts of these definitions are that both the controller and the processor shall be natural or legal person that can be held responsible by the data subject for lawful processing of personal data within the context of a handbook for Internet based agents.

2.6 Security

2.6.1 Legal Framework

Article 17 Security of processing

1. Member States shall provide that the controller must implement appropriate technical and organisational measures to protect personal data against accidental or unlawful destruction or accidental loss and against unauthorised alteration, disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing.

   Having regard to the state of the art and the costs of their implementation, such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected.

2. The Member States shall provide that the controller must, where processing is carried out on his or her behalf, choose a processor who provides sufficient guarantees in respect of the technical security measures and organisational measures governing the processing to be carried out and must ensure compliance with those measures.

3. The carrying out of processing by way of a processor must be governed by a contract or legal act binding the processor to the controller and stipulating in particular that:

   - the processor shall act only on instructions from the controller;
   - the obligations set out in Paragraph 1, as defined by the law of the Member State in which the processor is established, shall also be incumbent on the processor.

4. For the purposes of keeping proof, the parts of the contract or legal act relating to data protection and the requirements relating to the measures referred to in Paragraph 1 shall be in writing or in another equivalent form.

Recitals

(15) Whereas the processing of such data are covered by this Directive only if it is automated or if the data processed are contained or are intended to be contained in a filing system structured according to specific criteria relating to individuals, so as to permit easy access to the personal data in question;

(26) Whereas the principles of protection must apply to any information concerning an identified or identifiable person;
whereas, to determine whether a person is identifiable, account must be taken of all the means that can reasonably be expected that shall be used either by the controller or by any other person to identify the said person;
whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable;
whereas codes of conduct within the meaning of Article 27 may be a useful instrument in providing guidance as to the way in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible;

(47) Whereas where a message containing personal data is transmitted by means of a telecommunications or electronic mail service, the sole purpose of which is the transmission of such messages, the controller in respect of the personal data contained in the message shall normally be considered to be the person from whom the message originates, rather than the person offering the transmission services; whereas, nevertheless, those offering such services shall normally be considered controllers in respect of the processing of the additional personal data necessary for the operation of the service;

2.6.2 The Security of Personal Data

This section contains a summary of the AV23 study: Security of Personal Data [vBB01].

Introduction

Our privacy legislation defines standards for the exercise of a fair and lawful processing of personal data. It is very important, for example, that appropriate technical and organisational measures are taken to protect personal data against loss and unlawful processing. This document considers how the obligation to provide protection must be fulfilled in practice, that is, the requirements that personal data protection measures must meet. The document sets out the Dutch Data Protection Authority's (DPA) practical guidance for controllers in the context of the statutory framework.

A controller's duty to protect personal data is a corollary to the individual's right to privacy. This right is established in international treaties, in European legislation, in the Dutch constitution and in legislation. The *Wet bescherming persoonsgegevens* (WBP; Dutch Personal Data Protection Act) has provided the general basis for this field of law since it came into force on 1 September 2001 in the Netherlands. Responsibility for supervising compliance with the WBP lies with the Dutch DPA. The WBP regulates the processing of personal data, i.e. any procedure involving such data, from its collection to its destruction. Before collecting personal data, a controller, that is, the person responsible for the processing of the data, must look into the question of appropriate security measures. The Act covers both automated and manual data processing. Measures and procedures already in place for the protection and processing of data need to be tested against the requirements of the WBP and revised as necessary.

The security that must be provided goes beyond information security; all matters relevant to the processing of personal data must be addressed; not just those that fall within the ICT domain of information security. This document describes the additional measures to be taken to supplement those needed for compliance with general security requirements. This is because the WBP stipulates that extra measures must be taken to ensure the security of personal data processing activities. The measures put in place must, therefore, be broader than those needed to satisfy general data security requirements.

The protection of personal data is concerned with three quality aspects: exclusivity, integrity and continuity. The guidance set out in this document is particularly concerned with measures and procedures for protecting the exclusivity of personal data. The other two aspects are covered by the general system of security measures.

The legal basis for personal data protection measures is formed by Article 13 of the WBP. The relevant passage of the Act states: "The controller shall implement appropriate technical and organisational measures to secure personal data against loss or against any form of unlawful processing. These measures shall guarantee an appropriate level of security, taking into account the state of the art and the costs of implementation, and having regard to the risks associated with the processing and the nature of the data to be protected. These measures shall also aim at preventing unnecessary collection and further processing of personal data." The level of security that a controller must provide shall depend on the risk class. Article 13 of the WBP forms the basis for the use of Privacy-Enhancing Technologies (PETs). PETs are a coherent system of ICT measures protecting informational privacy (in accordance with European Directive 95/46/EC and the WBP) by eliminating or minimising personal data or by preventing the unnecessary or unwanted processing of such data, without compromising the functionality of the information system. The use of PETs is more than an appropriate technical measure; it is a means of systematically ensuring compliance with the WBP.

Protection Levels for Personal Data

It is important that the measures taken when protecting personal data address threats that are realistic, given the nature of the data concerned and the scale of the processing activities. The risk may be regarded as the product of the likelihood of an undesirable event and the seriousness of the implications of that event. The greater the risk, the stricter the protection requirements that must be met. As a guide to the measures that are appropriate, data processing procedures are divided into a number of predefined risk classes. Each class is linked to a particular level of protection. The main factors influencing the level of protection required include:

- The significance attached by society to the personal data to be processed;
- The level of awareness within the processing organisation regarding information security and the protection of personal data and subjects' privacy;
- The nature of the ICT infrastructure within which the personal data is to be processed.

The controller must perform a thorough analysis in each case. On the basis of the findings, the controller can decide which risk class the intended procedure falls into and what level of protection is, therefore, required. The analysis must be verifiable and it must be possible to give an account of the analysis if necessary. Four risk classes are recognised:

Risk class 0: Public level risk. The personal data to be processed is already in the public domain. It is generally accepted that use of the data for the intended purpose represents no risk to the subjects. This document, therefore, proposes no special protection measures.

Risk class I: Basic level risk. The consequences for the subjects of the loss or unauthorised or inappropriate use of their personal data are such that standard (information) protection measures are sufficient.

Risk class II: Increased risk. The loss or unauthorised or inappropriate use of the personal data would have additional consequences for the subjects. Certain types of personal data referred to in Article 16 of the WBP enjoy special legal protection and, therefore, require at least the level of protection associated with this risk class. The types of personal data in question are data concerning a data subject's religion or philosophical beliefs, race, political opinions, health, sex life, trade union membership, criminal record or record of unlawful or antisocial behaviour following the imposition of an injunction.

Risk class III: High risk. Where several collections of special categories of personal data are to be processed, the potential consequences of processing can be sufficiently serious for the data subjects involved that the procedure warrants inclusion in risk class III. The measures taken to protect data processed in a class III procedure must meet the highest standards.

The interrelationships between the various risk classes are summarised in Table 2.1.

The Security of Personal Data in Practical Situations

The requirements that apply to the protection of personal data are presented below, divided into fourteen categories. The measures that must be taken in a particular case depend on the risk class in which the processing procedure is placed on the basis of risk analysis.

Table 2.1: Risk Classification of Personal Data.

Nature of personal data:		Personal data	Sensitive personal data In accordance with Article 16 WBP	Personal data of a financial and/or economic nature
Quantity of personal data (nature and volume)	Nature of the processing			
Small quantity of personal data	Simple processing	Risk class 0	Risk class II	Risk class II
Large quantity of personal data	Complex processing	Risk class I	Risk class III	

Security policy, protection plan and implementation of the system of measures and procedures Management formulates a policy setting out general information security requirements and specific personal data protection requirements. On the basis of this policy, a plan is drawn up for the creation of a system of protection measures. The responsibilities of staff members with regard to the implementation of the protection policy must be defined. Compliance must be checked regularly and validity in the current circumstances must be reviewed regularly.

Administrative organisation The term ‘administrative organisation’ covers the entire system of measures relevant to the systematic processing of data in order to provide information to facilitate the management and operation of the organisation, as well as to facilitate the process of accounting for such activities. The measures and procedures must be defined in a structured manner. It is, furthermore, important that they are reviewed whenever changing circumstances make this appropriate. To this end, regular checks must carried out to ascertain whether the procedures and measures are consistent with current practice. It is also important that the responsibilities with regard to information security and data processing are properly defined.

Privacy awareness Practical application of protection measures is normally down to an organisation’s staff, all of which need to play their part. It is, therefore, necessary to build up or maintain an adequate level of privacy awareness within the organisation. Checks on application and compliance are important.

Personnel requirements In order to minimise the risk of inappropriate processing, it is important that the need for care and attention in this area is taken into account during the recruitment and selection of personnel. Responsibilities with regard to the protection of personal data must be set out in job descriptions.

Organisation of the workplace One of the most important aspects of information security is ensuring that data (in general) do not come into the possession of unauthorised individuals. Appropriate measures, which often do not need to be at all complicated, are required to minimise the risk of unauthorised access to personal data. Personal data are sometimes transported on portable media. PCs and data carriers must be properly secured and unauthorised access prevented.

Management and classification of the ICT infrastructure An organisation needs to have an up-to-date overview of its ICT facilities for operational purposes. Proper management of the ICT infrastructure is also necessary for the protection of personal data.

Access control An organisation is required to define measures and procedures to prevent unauthorised access to locations and information systems at or in which personal data is held. This implies being able to close off and control access to relevant areas and ensuring that only authorised personnel can access information systems. It is also important that records are kept indicating who is authorised to do what, and that regular authorisation checks are carried out.

Networks and external interfaces The transmission of personal data via a network involves a significant security risk. It is, therefore, strongly recommended that personal data be encrypted for transmission. In this way, it is at least possible to ensure that messages containing personal data are not read by unauthorised persons without explicit, deliberate intent.

Use of third-party software In order to minimise security risks, illegal or unapproved third-party software must never be used for data processing. Regular checks are important to ensure that this principle is adhered to. All software modifications must be documented and managed using a problem and change management system. Procedures for the modification and replacement of software must also be in place.

Bulk processing of personal data Processes involving personal data on numerous subjects are frequently automated or semi-automated. Integrated or non-integrated automated bulk processes are initiated and then allowed to complete without further interruption. The controller, nevertheless, remains responsible for the processing at all stages; responsibility cannot be transferred to the system manager.

Storage of personal data Management of data carriers is important for the protection of personal data. System backups must be made at regular intervals. Clear procedures must be defined for the production of backup files containing personal data, and adherence to these procedures must be checked. Data carriers bearing personal data must be carefully stored and transported to ensure that unauthorised persons cannot remove them or view the data. Data carriers bearing personal data in risk classes II or III must be stored in lockable areas with appropriate break-in protection, even if the data is stored in encrypted form.

Destruction of personal data Personal data and data carriers that are no longer required must be destroyed in an appropriate manner once any applicable statutory or other retention period has elapsed. Responsibilities with regard to the destruction of personal data and data carriers must be clearly defined and personnel must be aware of who is responsible for doing what.

Contingency plan Every organisation must have a contingency plan indicating exactly what is to happen in the event of an emergency. Such a plan, however, is useful only if personnel are familiar with it and regular drills have been held to practise its implementation. An organisation's contingency plan must set out a procedure for the resumption of data processing following an emergency.

Contracting out and contracts for the processing of personal data An organisation shall sometimes choose to contract out some or all of its personal data processing activities, rather than perform them in-house. Under such circumstances, the external processor must guarantee the same level of protection as that provided by the controller. A contract made with an external processor must make provisions for the protection of personal data. The external processor, furthermore, must sign a confidentiality contract. If an external processor handles personal data, the controller must supervise the processor's protection arrangements by, for example, carrying out periodic checks.

2.7 Privacy Preferences and Policies

Many of the data protection issues are linked with the consent of the data subject. A data subject can give his or her consent to a form of processing his or her personal data if he knows how the controller shall process his or her personal data; the transparency principle. The controller shall inform the data subject of his or her intentions. This can be published in a privacy policy statement. The data subject may have written down the conditions under which he is prepared to disclose his or her personal data to a controller. Such a statement may be called the privacy preferences of the data subject. In theory, there can now be a broker of personal data. Assuming that this broker can be trusted, the data subject may give privacy preferences and his or her personal data. The broker, on receiving a request for disclosure, shall ask for the privacy policy of the controller in question. If privacy preferences and privacy policies can be defined in a binary system, the decision whether to disclose or not to disclose could be an automated decision. It is one of the objectives of a PISA that each agent in the environment is equipped with the privacy policy of its controller. Before personal data is transferred to an agent, directly by the data subject or by an agent already loaded with personal data, the agent discloses its privacy policy, which is then matched against the preferences of the current holder of the data, data subject or agent. The preferences are compared with the policy and, if they match, this can be viewed as the consent required by the Data Protection Directive.

2.8 Legal Interception

Agents do operate in the environment of the Internet. This means that the Privacy and Electronic Communications Directive of 2002 also applies to the MAS environment as well as data protection legislation. Recital 11 and Article 15 of the 2002 Privacy and Electronic Communications Directive may restrict the scope of the rights and obligations provided for in Article 5, Article 6, Article 8(1), (2), (3) and (4), and Article 9 of the DPD when such restrictions constitutes a necessary, appropriate and proportionate measure within a democratic society to safeguard national security (i.e. state security), defence, public security, and the prevention, investigation, detection and prosecution of criminal offences or of unauthorised use of the electronic communication system, as referred to in Article 13(1) of the Data Protection Directive. To this end, Member States may, inter alia, adopt legislative measures providing for the retention of data for a limited period as long as the criteria of Article 8 of the European Convention of Human Rights are respected. This provision in the Privacy and Electronic Communications Directive means that the privacy of the user, the data subject, is not guaranteed in all circumstances. National implementations of this Directive may foresee in judicial powers to have conditional access to log files to be created by providers of public telecommunication services. This has lead to the following provisions in the telecommunications law in the Netherlands. Article 1 of

Chapter 13 states that public telecommunication networks and public telecommunication services may only be made available to the public if the data processed can be intercepted. (This provision does not arise from the Directive but from the Council recommendations of 1995 concerning interception ability of communications.) Because there is no legal obligation for all EU Member States to implement similar regulations, not all Member States have this obligation. This legislation may lead to a very confusing situation in the following situation. A controller offers a telecommunication service where all the (personal) data processed are encrypted. This controller uses a telecommunication provider that has no commercial links with the controller. The encryption process is secured using a Trusted Third Party (TTP). The service provider shall be denied access to the TTP for the decryption. Even if the service provider discloses the log files, the contents cannot be made readable to the investigative authorities. The same situation may also arise even when the controller and the provider are one and the same organisation. The questions to be asked now are twofold:

- The service provider renders an unlawful service;
- The TTP can be forced to disclose the encryption key to the investigative authorities. There must be a provision in the contract between the controller and the TTP regarding this matter.

2.9 Nine Condensed Rules from the EU Viewpoint on Privacy

The interpretation on how privacy has to be approached is not anywhere near uniform, not even in the countries with a privacy protection regime. The rigor with which the legislation is applied also varies from country to country. The most stringent The EU regime recognises nine different basic guiding rules. Eight of these rules deal with privacy protection directly while the remaining rule is a consequence of the global differences in the privacy regimes. In order to protect its citizens against violations of privacy in other non-EU countries where more permissive interpretations exist after (electronically) receiving personal data, it stipulates that, if at all possible, this data traffic must be prevented. Each rule shall be indicated by a brief title, a short explication of its meaning and a number of paragraphs providing further clarification.

Reporting the processing (Data Protection Authority) The processing of personal data must be reported in advance to the Data Protection Authority (DPA) or a privacy officer, unless processing has been exempted. This rule stipulates that the collector of personal data has to inform the DPA of its collection and its intentions with this collection. The DPA is the main guardian of the rules of the privacy regime. Its authority and competence must be expressed in national legislation[13]. It must also, naturally, be equipped with an appropriate budget, staff and other required facilities and it must be backed by national parliament.

Transparent processing The person involved must be able to see who is processing his or her personal data and for what purpose. The main issues of this rule is that (1) the data subject is informed about all the controllers of his or her personal data and (2) that these controllers inform the data subject about the purposes of their collections and further processing.

[13] The USA have not appointed a Federal Privacy Commissioner.

’As required’ processing Personal data may only be collected for specific, explicit and legitimate purposes and **not** further processed in a way incompatible with those purposes. This rule is meant to make perfectly clear that use and processing for purposes that are not known to the data subject are forbidden, not to say illegal.

Lawful basis for data processing The processing of personal data must be based on a foundation referred to in legislation, such as permission, agreement, legal obligation, justified interest, and such. For special categories of data, such as medical information, stricter limits prevail. The legality of the processing of personal data collected must be ensured. Despite eventual consent from the data subject, any processing that is in violation of the law remains illegal.

Data quality Personal data must be as correct and as accurate as possible, sufficient, to-the-point and not excessive. This rule supports the possible volatile nature of personal data and charges the collector with the obligation to keep all the items of personal data in good order, including accuracy, adequacy and timeliness, and is an incentive to minimise data as much as possible.

Rights of the parties involved The parties involved have the right to take cognisance of and to update their data as well as the right to raise objections. This, in fact, addresses two different issues: (1) the right of the data subject to have the ability to update personal data through a certain mechanism made known to him or her with the obligation, at the controller’s side, to implement the amendments in relation to all the duplications of the personal data at its collection or its processor’s collection and (2) the right of the data subject to raise objections in case the mechanism remains obscure or the amendments are not implemented in due time.

Processing personal data by a processor If processing is outsourced to a processor, it must be ensured that he shall observe the instructions of the person responsible. This rule states that, with the transfer of personal data to a processor, the rights of the data subject remain unaffected and that all restrictions and obligations of the controller equally apply to the processor. There is no erosion of personal data constraints in the course of this process.

Protection against loss and unlawful processing of personal data Suitable measures of a technical and organisational nature make up the necessary tailpiece of lawful processing [vBB01]. This rule puts the controller under the obligation to take all meaningful and possible measures to keep the collection of personal data, and any item related to the data, in good order and also to take precautions that use of the collection outside the legal and purposeful limitations are virtually impossible.

Data traffic with countries outside the EU In principle, the traffic of personal data to a country outside the EU is permitted only if that country offers adequate protection. In the real world, this implies that the intended country must have a privacy protection regime that is at least as stringent as the one applied in the EU and preferably a Privacy Commissioner.

2.10 The Privacy Threat Analysis

Informational privacy is gradually becoming more accepted in many democratic countries as well as by the citizen and organisations. This change in recognition, however, is not reflected in the software systems that in some way or another deal with privacy-related information; the majority of these systems can only be described as being privacy naïve. A universal method for privacy threat analysis of intelligent software agents is necessary. A common hurdle to thorough privacy threat or risk analysis is the lack of the necessary budgetary and time resources and qualified staff to do the job, which may be due to the drive of technical and commercial management alike to strike the market with the product.

2.10.1 General Approach Risk Analysis

The general approach for risk analysis and subsequent requirement determination as shown in Figure 2.1 is derived from a comparable domain: risk assessment for information security in British Standards 7799, the Code of Practice for Risk Analysis and Management Method [BS], and Information Security Handbook of the Central Computers and Telecommunications Agency (CCTA) [Cen].

The presented approach highlights the way privacy consolidation and, subsequently, privacy protection can be handled, but by the same token, after due adaptation, the same line of thinking can be applied to other topics where vital elements can be violated, such as (physical) safety, security, etc. The privacy consolidation approach starts from the privacy regulations, which mandates the Privacy Consolidation Ordination, or the obligation to implement such measures in the system to ensure that the privacy regulations are met to the satisfaction of the regulators. The first layer of analysis has two branches. The "assets" represent the objects or subjects that bear personal data and hence are vulnerable to privacy threats. The identification of such assets is necessary because their appraisal is a factor in risk assessment. The threats are identified in parallel to this. These threats can be linked to an asset but may also be a direct consequence of the regulations. Later, they can be linked to specific assets.

The second layer of results establishes the two factors relevant for risk assessment. The first factor has two inputs. The first input is the appraisal of the assets: how important is it that the assets are protected by possibly expressing the assets in an ordinal list to produce a sequence, in some monetary unit, to make it financially tangible, or in some other unit; the purpose is to rank the assets in value. The second input is the result of the assessment of the severity of the threats: what is the gravity of the consequences of a privacy violation. The second factor represents the likelihood of occurrence, expressed as the likelihood that it happens in a certain timeframe. This is determined by expert judgment or, preferably, by statistical analysis. The third layer is the result of risk assessment. The approximate formula is that a risk is the product of the appraisal of the consequences of the threat times the likelihood of occurrence of such a threat. This must at least give some insight for the prioritisation of the countermeasures (or risk controls) and to determine where and how effort and money must be invested. The fourth layer is the definition of requirements that are deemed adequate to meet the safety regulations. Because the countermeasures may, however, out of themselves, also create secondary threats, eventually the whole privacy protection approach has to be repeated, until no further threats are expected. The last layer is the implementation of the requirements for the countermeasures, which is the practical result of the activities. In actual practice, it can be that privacy loopholes are still present, in which case a new privacy protection ordination arises, but this is outside the scope of the current considerations.

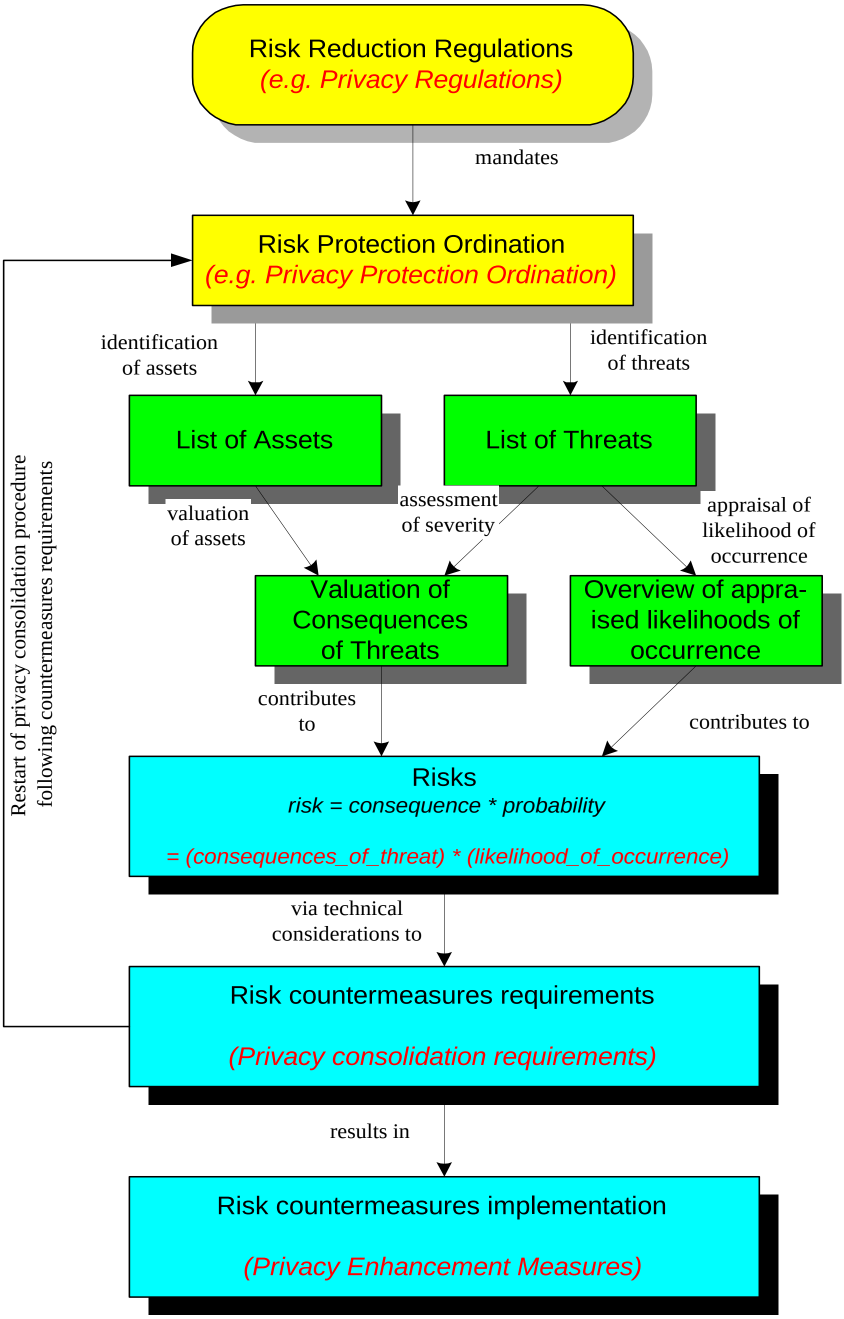

Risk Reduction Regulations
(e.g. Privacy Regulations)
mandates
Risk Protection Ordination
(e.g. Privacy Protection Ordination)
identification of assets
identification of threats
List of Assets
List of Threats
valuation of assets
assessment of severity
appraisal of likelihood of occurrence
Valuation of Consequences of Threats
Overview of appra-ised likelihoods of occurrence
contributes to
contributes to
Risks
risk = consequence * probability
= (consequences_of_threat) * (likelihood_of_occurrence)
via technical considerations to
Risk countermeasures requirements
(Privacy consolidation requirements)
results in
Risk countermeasures implementation
(Privacy Enhancement Measures)
Restart of privacy consolidation procedure following countermeasures requirements

Figure 2.1: Model Risk Assessment.

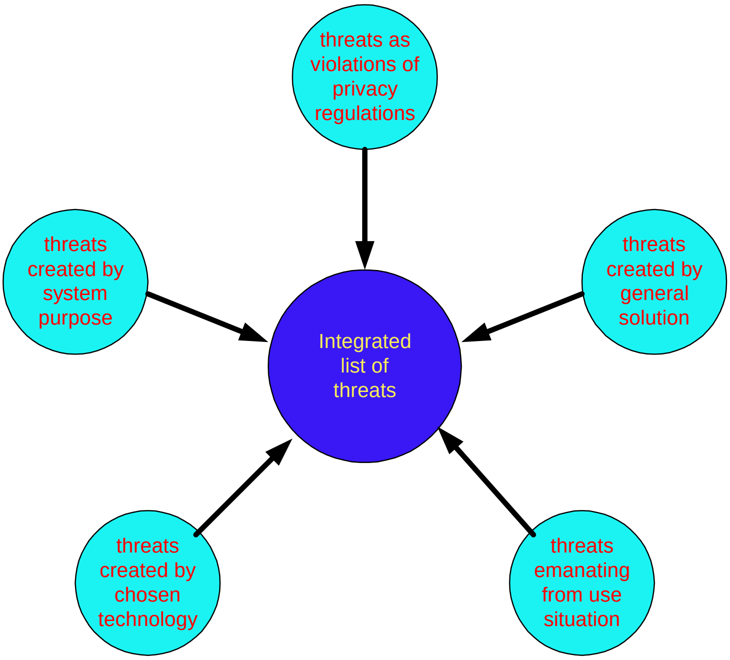

threats as violations of privacy regulations
threats created by system purpose
threats created by general solution
Integrated list of threats
threats created by chosen technology
threats emanating from use situation

Figure 2.2: Five-pronged Approach to Privacy Threat Analysis.

The focus, in relation to privacy threat analysis, shall be on the right-hand side which shall be representing threat identification and assessment of severity of consequences of such threats.

2.10.2 Five-pronged Approach to the Privacy Threat Analysis

A methodology is based on a *’way of thinking’* about a certain area of investigation. Next, the ’way of thinking’ is strengthened by the *’way of working”* that is expressed in the methodology itself. The final part is the *’way of writing’* which describes, using any desired level of formality, how the results have to be put on paper[14]. A five-pronged approach has been chosen for privacy threat analysis, which is depicted in Figure 2.2.

That five different perspectives on the issue of personal data are applied, finds its rationale in the observation that one single line of thought may fail to illuminate all the aspects of privacy that are relevant for the ultimate implementation. Five different perspectives, and hence five different lines of thinking, is more likely to reveal everything that there is to know about privacy threats. That this may result in the same threat identified more than once is not an issue.

The five perspectives chosen are:

[14] Examples of this are abundant in the arena of software system development. A case in point is object-oriented development. Object-oriented thinking and object-oriented design methodologies are available and also description languages such as UML.

- Privacy regulations, as defined in a certain country or country union: these regulations inherently list a number of privacy threats, if these regulations are not adhered to;
- Purpose of the system, which creates its own threats: because the user (private person) wants to achieve something, that person creates privacy threats;
- Solution adopted, which may or may not create threats of its own;
- Technology used: because of the way a certain system is implemented, certain threats may ensue which are not necessarily a consequence of the intended purpose. Meanwhile, the technology shall harbour some of the privacy enhancement measures;
- Situation in which the ultimate system shall be used: which, although not necessarily creating threats of its own, may or may not aggravate (or alleviate) previously identified threats and hence may incur more demanding technological measures. This part is especially needed when a commercial off-the-shelf (COTS) product shall be used in an unforeseen situation; the previous four types can be followed whether or not the system is a COTS or dedicated to a certain problem and environment.

Risk Analysis Branch 1: Threat Identification and Appraisal of Likelihood of Occurrence

Following the five-pronged approach, the sequence of steps in the Threat Identification branch, the right-hand ramification of the risk analysis diagram can be elaborated in the diagram shown in Figure 2.3.

As can be derived from the figure, the subsequent steps result in separate threat survey documents. The sequence of steps also shows the feature of following the system design lifecycle fairly closely, which has the obvious advantage of easy incorporation.

Risk Analysis Branch 2: Asset Identification and Appraisal of Consequences of Threats

The investigation of the left-hand branch of the risk analysis diagram is depicted in Figure 2.4.

Risk Analysis Branch 3: Risk Identification, Prioritisation, Privacy Requirement Determination and Implementation

The list of risks is also specified in Figure 2.4. This list can be prioritised because the higher the risk, the more stringent the privacy-enhancing measures to counter the threat. This results in requirements for the intended system and this, in turn, shall be implemented in the ultimate system.

2.10.3 Definition of the Primary Threats

The nine EU privacy rules are in fact the countermeasures against a certain number of threats. These underlying threats, unfortunately, are not explicitly formulated in Directive

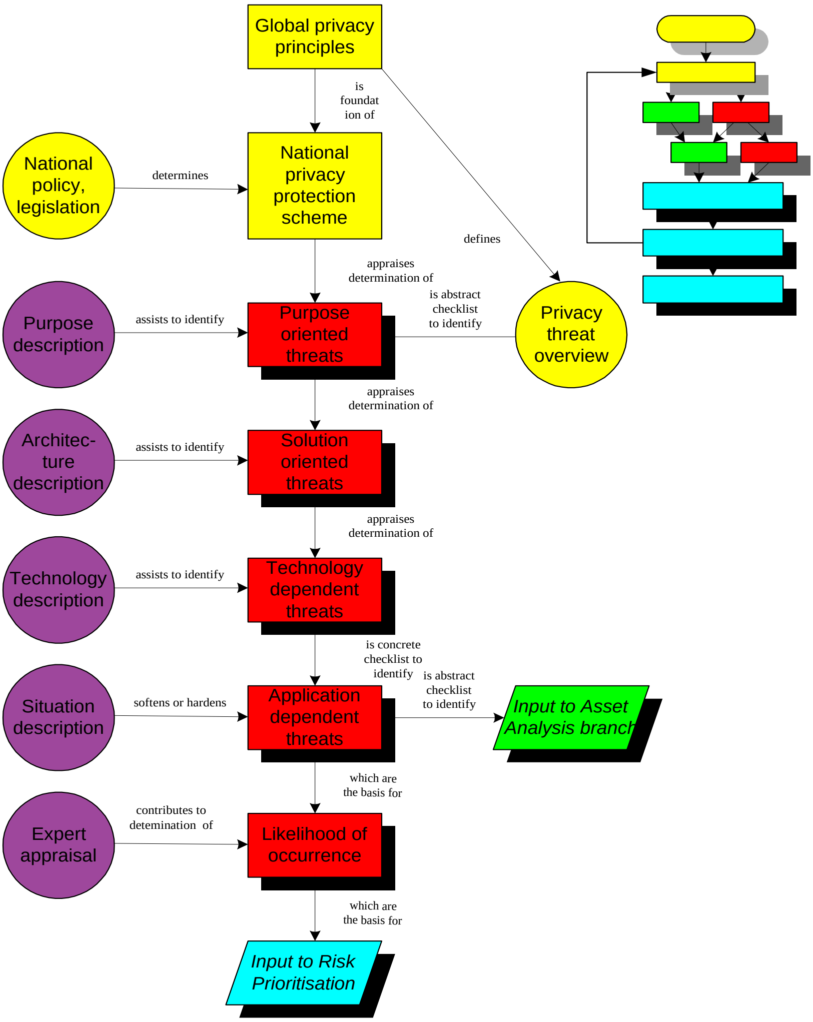

Global privacy principles
is foundat ion of
National policy, legislation
determines
National privacy protection scheme
defines
appraises determination of
Purpose description
assists to identify
Purpose oriented threats
is abstract checklist to identify
Privacy threat overview
appraises determination of
Architec-ture description
assists to identify
Solution oriented threats
appraises determination of
Technology description
assists to identify
Technology dependent threats
is concrete checklist to identify
is abstract checklist to identify
Situation description
softens or hardens
Application dependent threats
Input to Asset Analysis branch
which are the basis for
Expert appraisal
contributes to detemination of
Likelihood of occurrence
which are the basis for
Input to Risk Prioritisation

Figure 2.3: Risk Analysis Branch 1.

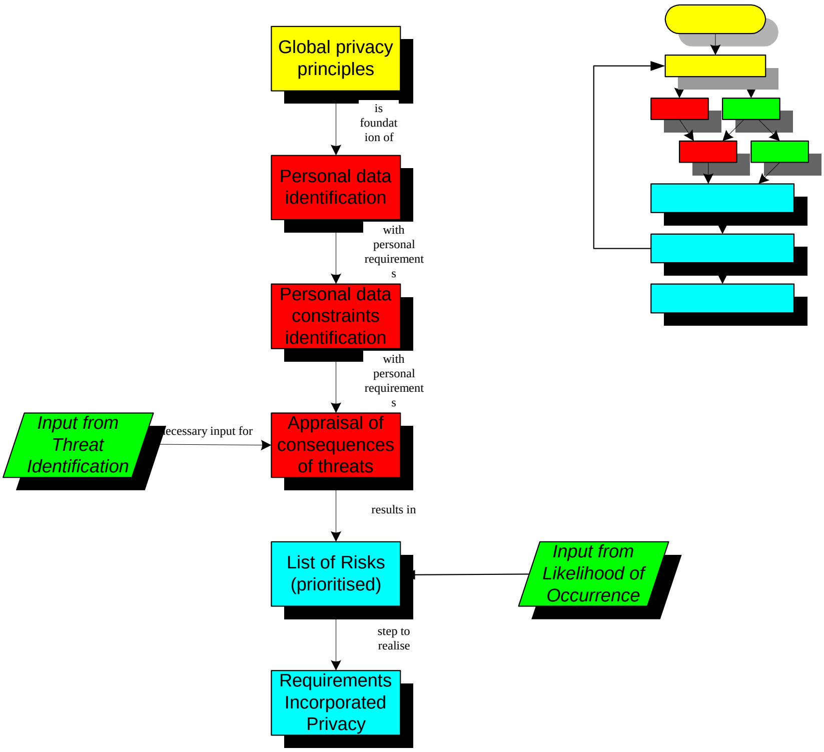

Global privacy principles
is foundat ion of
Personal data identification
with personal requirement s
Personal data constraints identification
with personal requirement s
Input from Threat Identification
ecessary input for
Appraisal of consequences of threats
results in
List of Risks (prioritised)
Input from Likelihood of Occurrence
step to realise
Requirements Incorporated Privacy

Figure 2.4: Risk Analysis Branch 2.

95/46/EC. They are, however, needed as a starting point for the threat identification process. A list of fundamental threats must, therefore, first be formulated which mirrors the EU rules on privacy in the threats domain[15]. If so desired, a different set of threats can be compiled for another privacy regime and such a change does by no means affect the methodology as such.

The threats identified are given below and include a brief explanation of each. First, the EU rule is stated and then the resulting privacy threat.

- *Reporting the Processing:* **Threat***: Secret possession of personal data [files]*: a controller or processor has the control over personal data which is not reported to the Data Protection Authority or Privacy Commissioner;
- *Transparent Processing:* **Threat***: Secret processing of personal data*: a controller or processor has legally justified control over personal data but processes this data without official consent from the data subjects;
- *As Required Processing:* **Threat:** *Out of bound processing (violation of personal data processing restrictions)*: the controller or processor has the permission to pro-

[15] A note of caution, however, is appropriate since the list presented here is just the product of the methodology and has only the role to demonstrate the feasibility of the methodology; scrutiny by privacy experts may lead to a reformulation, although the authors are confident that the presented list does not, by any means fundamentally deviate from the intentions of the privacy regulations discussed earlier.

cess or handle the personal data according to some set of personal data constraints, but violates the rules formulated;

- *Processing Personal Data by a Processor:* **Threat:** *Out of law processing*: if the data subject has not formulated any constraints but has not provided permission or given consent for dedicated processing either, the collector or processor shall be bound by the rules of law.

- *Data Traffic with Countries Outside the EU:* **Threat:** *Out of jurisdiction processing*: the transmission of data outside the EU or processing of data outside the area of jurisdiction on a location where privacy rules are less strict or non-existent;

- *Rights of parties involved:* **Threat:** *Irresponsiveness to discontent of the data subject*: the controller or processor is not responsive to complaints about possession, processing or correctness of the personal data or violations of the personal data constraints and does not provide facilities to implement changes to the set of personal data related to the data subject;

- *Data Quality, Personal Data and Personal Deterioration*: **Threat:** *Wrong decisions based on wrong data:* the controller or processor does not keep these data current or in good order in time and place;

- *Personal Data Management Violation Constraints:* **Threat:** *Serious privacy intrusion and liability*: the controller or processor does not attend to the restrictions and obligations formulated in the constraints (other than the processing restrictions).

2.10.4 Purpose of the System to be Devised

Purpose of the system, to perform the *purpose-oriented threat identification*:

> *The activity of searching for yet unknown data from as yet unknown sources, to expand, complement and match the acquired data with known data, so as to find an answer to a request or question.*

Note that the purpose of the system has to be defined explicitly, because otherwise legal approval shall be denied by default. In addition, a sloppy defined purpose needs to be rejected to ensure that only well-defined purpose declarations shall remain. Such a declaration can be strengthened by a description of the personal data to be collected and processed and the personal data constraints that shall be accepted by the controller. The pinnacle of personal data description is a data dictionary of such data according to the ISO standard on data administration[16].

The *Contributing Purposeful Factors* are the threat-bearing factors that can be derived directly from the purpose declaration: they must appear as such in the declaration.

These factors can be found in the left-hand column of Table 2.2. The column headers are the threats formulated earlier as being the basis for the *Privacy Facilitation Rules*. (PD in the table stands for personal data and PDC stands for Personal Data Constraints.)

As can be seen, the column header, containing the basic threats, can be used as a tool to ensure that the threats of the *contributing purposeful factors* are more specific.

[16] See ISO 11179 series.

Table 2.2: Overview of contributing purposeful factors.

Contributing Purposeful Factors	**Secret possession of personal data**	**Secret processing of personal data**	**Out of bounds processing**	**Out of law processing**	**Out of jurisdiction processing**	**Irresponsiveness to discontent**	**Personal data, constraints deterioration**	**PD constraints management part violation**
Nature of Question/ request								
Unknown data (receiver act as collector)			PD received, but no PDC received (processing part)				Corruption of PD Acquisition of unreliable data	
Unknown sources of data (collector)	Data subject plus DPA not informed	Data subject plus DPA not informed			Jurisdiction unknown	Impossibility to communicate	Impossibility to correct PD, PDC	
Activity of Searching								PD constraints not received

2.10.5 Description of the Solution to be Adopted

The statement of the solution is:

> *The solution to the stated purpose shall be to use intelligent agents in order to perform autonomous actions in transactions with other such agents by means of roaming over an independent transmission network.*

Similar to the purpose declaration, the solution description (definition) must shed light on the threat-bearing factors; the participants in the approach to address the issue. A full solution description must contain figures on how the solution shall work, shall be organised, what the functions, tasks and responsibilities of the main actors shall be, how communication shall take place, how personal data shall be involved, etc. The example in this chapter sketches the contours of this to demonstrate that such an approach works and delivers results. From the description of the solution, the *Contributing Solution-oriented Factors* can be extracted, which can be found in the left-hand column of Table 2.3.

This example can be completed with technology-oriented analysis and, when more is known about the situation in which the system is going to be used, by situational analysis. Currently, unfortunately, too little is known of both sides and, therefore, it has to be omitted but it stands to reason that the threat identification process can be continued in a similar vein.

2.10.6 Conclusion

The conclusion is that technological privacy consolidation and protection is certainly conceivable, and that a haphazard approach is not necessary: a structured, repeatable and verifiable approach with a sequence of intermediate documents, in close harmony with the system development lifecycle, can be followed. The presented approach, however, is just the first try and improvements and refinements are bound to follow.

The PET concepts of anonymity, pseudonymity, unlinkability and unobservability are also described in ISO 15408, Common Criteria for Technology Security Evaluation [ISO99].

Table 2.3: Overview of solution-oriented factors.

Participant	**Secret possession of personal data**	**Secret processing of personal data**	**Out of bounds processing**	**Out of law processing**	**Out of jurisdiction processing**	**Irresponsiveness to discontent**	**Personal data, PD constraints deterioration**	**PD constraints management part violation**
Transmission network					Wrong routing over network		Corruption of PD/PDC (own) Involuntary disclosure of PD/PDC-O, software or instructions.	
Agent as harbourer of personal data (including acquired foreign PD/PDC)							Corruption of PD-F, PDC-F - Unreliable PD	
Agent as collector of personal data	No notification of DPA, data subject	No notification of data subject	Foreign PD: PDC not well received					Unplanned actions. Erroneous execution of instructions
Agent as autonomous actor							Erroneous loading of PD, PDC or instructions	
Initiation sequence							Loss of control Involuntary disclosure	
Autonomous transactions								Violation of PDC, management part
Foreign agent, foreign site	No notification of DPA		Processing for unintended purposes. Unintended integration	Processing for forbidden purposes. Unlawful integration	Unknown or deceptive location of foreign site			
Data Protection Authority (DPA)	Inadequate administration. No notification of data subject	Inadequate administration. No notification of data subject						

2.11 Product Liability

Product liability is not to be confused with legal liability[17] deriving from the DPD. This section discusses product liability[18] if a manufacturer of intelligent software agents defaults. Several legal entities are involved in the development, creation and usage of ISAs. This section introduces the roles of these entities and their responsibilities.

The first role is that of the (ISA) developer. The developer is the (legal) entity that develops the software code for the ISA. The developer determines how the ISA shall act. For example, he determines what actions can be performed, what can be learned, when it communicates with other ISAs or legal entities, how it can be managed and what it does with PD. If software products like ISAs are defectively manufactured or designed, the developer and/or the distributor is directly liable for injuries and damage caused by his or her software product, especially if the design defects are foreseeable. This is the case with regard to privacy violations. We have carried out a privacy threat analysis. We know the risks. We cannot ignore these. The rationale is that the costs of psychological injuries and damages must be borne by the developer and not by the user. The burden of psychological or financial losses consequent upon use of defective software must and is, therefore, borne by those who are in a position to control the danger.

The second role is the role of the (ISA) provider. The provider is the legal entity that markets the ISA for usage. This could be done by selling or hiring the ISA software or by supplying the ISA software to the market without costs being involved. ISPs offer Internet services free of charge and request more precise in-depth information of the user in return. Such a provider has been qualified in the privacy threat analysis, as a threat and any detailed information of the user of the ISA must be hidden for this provider. Generally speaking, it is the responsibility of the provider to give instructions about the ISA operation and to provide all relevant information about the usage of the ISA, including any possible privacy risks. The user shall, nevertheless, be entitled to claim compensation for damages caused by a malfunctioning ISA, from a product liability point of view, concerning privacy violations and other damages that the user may have suffered due to the ISA or other ISAs or individuals. The third role is that of the (ISA) user. The user uses and manages the ISA in such a manner that the ISA performs tasks for the user. The ISA shall perform tasks in certain cases that require PD from its own user. In other cases, however, it may however require PD from the users of other ISAs. This is, for example, the case if the ISA provides a service to other ISAs. It is, in our viewpoint, the responsibility of the user to determine whether or not to use the ISA. Knowingly using an ISA that potentially might violate the privacy of others shall lead to heavy liabilities. The user can assume that the ISA is a safe product that shall do no harm to others' interests unless the developer or the provider has explicitly given warnings about privacy violations (hazardous to your privacy) when using its ISA. Legal theorists however defend that when product liability applies to software agent, a disclaimer (Caution) shall be cumbersome, as the instructions have to be of such a nature that the instructions render the use of the ISA actually safe. The warning contained in the software of the ISA must, therefore, be designed with internal user warnings that shall warn the user of a privacy violation whether he wants it or not and that shall even, ideally, prevent misuse.

[17] Article 23 1 DPD states that the controller is liable to the consequences of any form of unlawful processing.

[18] Article 23 2 DPD, the controller may be exempted from this liability, in whole or in part, if he proves that he is not responsible for the event giving rise to the damage.

Chapter 3

PET

G.W. van Blarkom J.J. Borking P. Verhaar
gbl@cbpweb.nl jborking@euronet.nl verhaar@fel.tno.nl
CBP, The Netherlands TNO-FEL, The Netherlands

Developments in Information and Communication Technology (ICT) are providing ever more possibilities to collect, store, process and distribute personal data. The potential violations to consumers' and citizens' privacy online and offline increase consequentially. The same technologies that appear to threaten citizens' privacy can, however, also be used to help protect it. These technological solutions that offer agent user privacy protection, extensively known under the name Privacy-Enhancing Technologies (PETs), and the legal basis for PET shall be discussed in this chapter. These techniques depend on a set of underlying technologies of which perhaps the most fundamental is that of encryption. Seven principles shall, furthermore, be clarified. The Common Criteria for Information Technology Security Evaluation (CC) supplies important information for building privacy secure agents.

3.1 Privacy-Enhancing Technologies (PET)

3.1.1 Definition

> *Privacy-Enhancing Technologies is a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system.*

3.1.2 History

The Dutch and Canadian Data Protection Authorities together with the Dutch TNO jointly published the study 'Privacy-Enhancing Technologies – A Path to Anonymity' [vRGB$^+$95] in 1995. This report proved that technology could be used as a means of protecting individuals against misuse of their personal data by reducing the processing of personal data.

ICT offers solutions in the shape of privacy protection for users, consumers and citizens. The application of ICT to protect privacy has become extensively known under the name

Privacy-Enhancing Technologies (PET or PETs); see [HB98]. PET stands for a coherent system of ICT measures that protects privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data, all without losing the functionality of the information system [Bor96]. PET incorporated systems use Identity Protectors and divide systems into identity, pseudo-identity and anonymity domains.

PETs have already achieved an important place in the practical and theoretical repertoire of privacy protection instruments in most, if not all, countries where data protection laws and systems are already in place or are being created. (see [Tet00], page 28).

PETs have been at the centre of attention in countries, especially the USA, where the legislative provisions for data protection are patchy or deficient. Raab [Raa02] writes: PETs are sometimes thought of as *substitutes* for other instruments of privacy protection, such as laws and the regulatory bodies that enforce and implement legislation. This conception is not helpful. PETs are better thought of as *complementary* to other instruments with which they must work together to provide a robust form of privacy protection. Law is, first and foremost, the instrument to which PETs must relate, incorporating legal principles into technical specifications [Raa02].

To implement matters technically, a system element called the 'identity protector' [RF95] is used within the data system to convert the identity of the 'data subject' involved (the person whose data are being processed) into one or more pseudo-identities[1].

The placement of the identity protector provides for at least two different domains within the data system; one domain where the identity of the person involved is known or accessible (the identity domain) and at least one domain where this is not the case (the pseudo-identity domain). The aim of the pseudo-identity domain is to ensure the person involved cannot be identified on the basis of previously obtained personal data, and vice versa, to ensure that personal data cannot be obtained on the basis of the obtained identity.

The identity protector in an information system can take several forms, for example:

- A separate function implemented in the data system;
- A separate data system supervised by the individual (for instance, the smart card for biometrics identification);
- A data system supervised by a party entrusted by a service provider and consumer ('Trusted Third Party' (TTP))[2]

The use of an identity protector, therefore, makes it possible to intervene preventively within the data system to hide the identity of the person involved. Other possible techniques are digital signatures, blind digital signatures, digital pseudonyms, digital certificates [Ver01] and MIX nodes. A MIX node is a processor that takes as input a certain number of messages that it modifies and outputs in a random order. The messages are modified and reordered in such a way that it is nearly impossible to correlate a message that comes in with a message that goes out. The MIX nodes can be used to prevent traffic analysis. See [BPS00, STRL00, BFK00].

One of the most important principles of privacy protection is that no more data must be collected or processed than are strictly necessary for the specified purpose. If an assessment indicates that the use of PET would reduce the amount of personal data items being

[1] Pseudonymity as well as anonymity are key concepts in the design of information systems and privacy-protecting tools, and have received considerable attention in recent years. See [Cla96].

[2] TTP services are for: a) authenticity and integrity and b) for confidentiality. See [Ver01].

processed, such use is effectively a legal requirement under the Dutch, German and Portuguese data protection legislation[3]. There are, of course, others means of offering protection but technical measures are very effective because you cannot avoid them. PET can, furthermore, be employed to block certain forms of processing thereby helping to ensure that data is used only for a specified purpose. PET is also valuable in the context of information security. The choice of PET techniques depends on the level of security needed to match the level of risks represented by the personal data.

The PISA project uses PET to protect all personal data with traditional PET components such as identity protectors and the creation of anonymity and pseudo-identity domains. When all personal data has been modified and the data subject is no longer identifiable, the DPD shall not be applicable any more.

3.1.3 Legal grounds for PET

EU Privacy Directive 95/46/EC

Article 17 Security of Processing

1. *Member States shall provide that the controller must implement appropriate technical and organisational measures to protect personal data against accidental or unlawful destruction or accidental loss and against unauthorised alteration, disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing.*

   Having regard to the state of the art and the costs of their implementation, such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected.

2. *The Member States shall provide that the controller must, where processing is carried out on his behalf, choose a processor who provides sufficient guarantees in respect of the technical security measures and organisational measures governing the processing to be carried out and must ensure compliance with those measures.*

3. *The carrying out of processing by way of a processor must be governed by a contract or legal act binding the processor to the controller and stipulating in particular that:*

   - *The processor shall act only on the instructions of the controller;*
   - *The obligations set out in paragraph 1, as defined by the law of the Member State in which the processor is established, shall also be incumbent on the processor.*

4. *For the purposes of keeping proof, the parts of the contract or legal act relating to data protection and the requirements relating to the measures referred to in Paragraph 1 shall be in writing or in another equivalent form.*

Paragraph 1 of this article gives the controller the possibility to choose between technical and organisational security measures. Theoretically, the system of measures in place can be defined in such a manner that there is effectively no difference between these two categories of measures. Namely, the controller is in a position to write all procedures, thus, describing the conditions under which the processing of personal data in his organisation can take place and is lawful. In order to enforce compliance to procedures, however, the

[3] Article 13, Dutch data protection legislation is derived from Article 17 of the DPD.

controller has to appoint quite a number of staff to monitor the work of staff members processing personal data under his responsibility. It is also very likely that severe sanctions must be in place if the rules laid down in the procedures are broken since the probability of detection is minimal. Technical measures, however, are more effective since special actions shall be taken to escape from the effect of such a measure.

Article 17 applies to all types of processing of personal data. Article 1 (b) defines 'processing of personal data' as any type of operation performed upon personal data whether or not by automatic means. It is safe to assume that the largest part of any processing of personal data shall be within the domain of Information and Communication Technology. It is, however, also safe to assume that some parts of the processing shall be non-automated. Technical measures cannot be put in place in a non-technical environment. It is, therefore, very unlikely that the security measures required to ensure lawful processing could be of a technical nature only. A number of organisational measures shall also be required; these must, however, be limited to an absolute minimum. It should go without saying that security measures are based on a security policy that has been adapted by the organisation. This shall probably involve paperwork, i.e. an organisational measure, which needs to be in place before any detailed measure, be it technical or organisational, can be defined and, subsequently, be implemented. Pursuant Article 32, EU Member States shall bring into force laws, regulations and administrative provisions necessary to comply with the Directive. During the parliamentary discussion in the Netherlands on the new privacy legislation based on the Directive the Minister of Justice explained that the text of the article corresponding to Article 17 must be explained in such a manner that an organisational measure is only then an acceptable alternative if a technical measure is not appropriate. The comparison of the pros and cons must be made taking into account the state of the art and the cost of their implementation where the measure required shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected. He then drew the conclusion that, in fact, there is a legal basis for the obligation to use Privacy-Enhancing Technologies as security measures. This legal basis is the ground for the statement that PET is the means to translate soft legal standards into hard system specifications.

3.1.4 Original PET Concept – Pseudo-identity

It was not until 1997 that the theoretical model published in the study was implemented for the first time in a commercial application. ICL in the Netherlands designed a working information system based on pseudo-identities in a Hospital Information System. In the Privacy-Incorporated Database® personal data is grouped into logically related domains, e.g. medication, appointments, treatment. None of these domains contain data items revealing the identity of the patient, the data subject. The patient is identified by a unique pseudo-identity in each of these domains. Next to these domains with medical data, one domain was created to hold the identity of the patient identified by a unique identifier. The pseudo-identity can be derived from the patient's identifier and vice versa. This process, the identity protector, is located in the application code. As the system is based on a client-server architecture, the identity protector is not available to (illegal intruder) users to the database. Only users of the application shall have the right to use the identity protector. The identity protector achieves a further level of protection, that is, the process uses an encryption protocol, the key for which is released to the application by a trusted third party only after successful authorisation of the user. Because the pseudo-identity is different for each of the pseudo-identity domains, a simple authorization scheme only permits access to those domains that a user needs to access based on his role that is defined by his user identity.

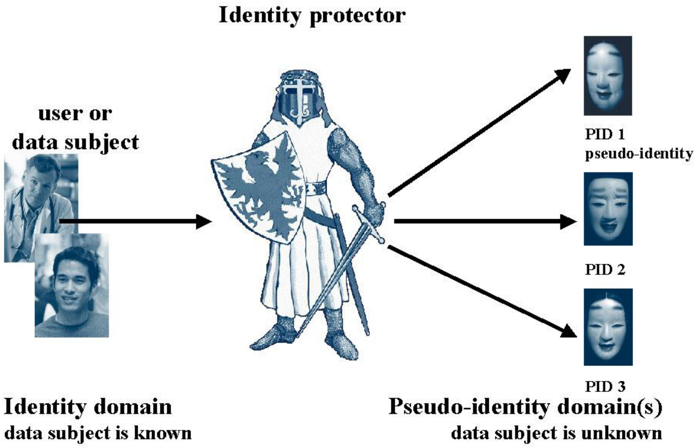


Figure 3.1: The identity protector.

3.1.5 Further PET Developments

Once the first implementation of a Privacy-Enhancing Technology was achieved, the ICT industry worldwide became interested in the possibilities of PET. The concept of the Privacy-Incorporated Database was being paraphrased on and, subsequently, used in a great variety of information systems.

After the pseudo-identity was established, newer PETs gave rise to a classification in seven principles:

The Seven PET Principles

1. Limitation in the collection of personal data
2. Identification/authentication/authorisation
3. Standard techniques used for privacy protection
4. Pseudo-identity
5. Encryption
6. Biometrics
7. Audit ability

The next seven sections shall describe these principles in some detail. You must keep in mind, however, that the possibilities of Privacy-Enhancing Technologies are not based on this limitative list of techniques. Designers of information systems that wish to comply with privacy legislation through technology must be open-minded. Firstly, they must identify those areas of the information system that represent the real risks in relation to privacy violation. Once identified, developing PETs is a creative process. The seven PET principles described below form a summary of PETs attained so far and, at the same time, could guide the mind of the designer in creating new PETs. The possibilities of PETs are endless.

A Privacy-Enhancing Technology is not necessarily something new, a new invention or something that has not been done or seen before. Existing technologies can be put in place in an information system and, subsequently, act as a PET. One must not forget the basic idea behind the obligation of using PETs, namely that it is very hard, if not impossible, to escape from the effect of a technical measure when compared to an organisational measure.

Limitation in the Collection of Personal Data

Privacy enhanced systems must first of all comply with Article 6, Paragraph 1 (c), of Directive 95/46/EC. Personal data collected must be adequate, relevant and not excessive in relation to the purpose for which they are collected and/or further processed. One does not require a vault, if one does not have anything to store. Personal data that have not been collected, do not need any measures to protect the data simply because one shall not be processing the data. When designing a database schema, data items must only be added if, and only if, their collection can be justified within the definition of the purpose of the information system.

Identification/Authentication/Authorisation

Pursuant to Article 16 of Directive 95/46/EC, any person acting under the authority of the controller, including the processor[4] himself or herself, who has access to personal data must not process them except on the instruction of the controller, unless required to do so by law. This article is often interpreted in such a manner that a user of an information system processing personal data must only process personal data on a need-to-know basis.

This information system functionality is a typical example where an organisational measure could be put in place using PETs. The controller must define an authorisation system. This system could have the following structure. Firstly, function profiles are created after which information system functionality is linked to organisational functions in this profile. Each processor within the organisation is given a unique identity to be used when being a user of the information system. As a last step, this user identity is allocated to one function profile. If a user is authorised to use a certain system functionality, that does not necessarily mean that this person has a need-to-know access to the data of all data subjects being processed. Within a hospital information system, fore example, a physician is only entitled to have access to the data related to the patients that the physician treats.

Taking the seventh PET principle into account, one has to be accountable for one's deed; it must be very obvious that each processor must be identified by a unique identifier; that is, a login name.

Standard Techniques Used for Privacy Protection

Article 6, paragraph 1 (b), of the Directive 95/46/EC makes the provisions that further processing of data for historical, statistical or scientific purposes shall not be considered

[4] See Directive 95/46/EC Article 2 (e) defining the 'processor' as a natural or legal person, public authority, agency, or any other body which processes personal data on behalf of the controller. A processor can thus be a natural person employed by the controller or by the organisation carrying out the processing on behalf of the controller, or the organisation carrying out that processing.

as incompatible provided that appropriate safeguards are put in place. These undertakings must guarantee that data kept must only be used for these purposes. The best guarantee is achieved by transforming the data in such a way that the data can no longer be used to identify the data subject. This process is often referred to as rendering anonymous or anonymisation.

Anonymous data as seen by the Data Protection Supervisory Authorities differs from the way the Common Criteria look at it. Within the Common Criteria, Anonymity is defined as follows:

> **9.1 Anonymity (FPR_ANO)**
> FPR_ANO Anonymity
> Family Behaviour
>
> This family ensures that a user may use a resource or service without disclosing the user's identity. The requirements for Anonymity provide protection of the user identity. *Anonymity is not intended to protect the subject identity.*

Paraphrasing on the above, this definition can also be applied to data subjects rather then to users. This means that, according to the Common Criteria, anonymity does not necessarily protect the identity of the data subject. The Supervisory Authorities, however, demand that in order to gain anonymity, the data is transformed in such a manner that the identity of the data subject can no longer be identified. In terms of the privacy legislation, the system no longer holds personal data. Rendering anonymous, according to this definition, is the process by which personal data is transferred into data.

According to the legislation, a data subject must give his consent regarding each processing of his personal data. The goal of this processing is to render his personal data into non-personal data; therefore, there is no need for the data subject to give his consent. This does, however, imply that it must be guaranteed that, after the process, the data subject is no longer identifiable.

The process of rendering anonymous can be split into two phases. Firstly, ensure that the filing system does not contain direct identifiable data items. This can be achieved by simply deleting all data items that contain the person's name, address, telephone number etc., which directly identify the data subject. The next phase is of a more complex nature. The data can no longer be used to indirectly identify the data subject. A data subject is indirectly identifiable if the filing system contains such a number of data items that, if combined, the data subject clearly can be identified. Recital 26 of Directive 95/46/EC states that when determining whether a person is identifiable, account must be taken of all the means that can be reasonably expected that may be used by the controller or by any other person to identify said person.

The process required is best explained by using a number of examples.

- There is no need to know the date of birth of the data subjects for statistical purposes. It suffices to know an age category. The date of birth must, therefore, be transformed to the code in question.

- There is no need to know the full postal code of the data subject for statistical purposes. It suffices to know the region where the data subject lives. The full postal code must, therefore, be truncated to the required length.

- There is no need to know the precise date of an incident for statistical purposes. It suffices to know the period in which the incident happened. All date items must, therefore, be truncated into the month/quarter/year of the occurrence of the incident.

- There is no need to know the exact profession of the data subject for statistical purposes. It suffices to know the industrial area in which the data subject earns his living. The job code must, therefore, be truncated to a code indicating the area.

Many more examples could be given. It must, however, also be clear that there is a real requirement to apply this type of PET on a great number of data items in order to really render data as being anonymous. The last remark on this subject is also a very important one. Filing systems for historical, statistical or scientific purposes are mostly created using a personal data filing system as the source. The same controller, very often, does not operate such a data warehouse. Disclosure of personal data is, therefore, possibly at stake. Rendering anonymous must take place under the responsibility of the controller of the personal data filing system. If that is not the case, personal data is disclosed, and the data subject must give his unambiguous consent.

Pseudo-identity

The concept of the original PET principle is the use of the pseudo-identity. A concise description of its first use was given in Section 3.1.4. Another example is a system where there is a real requirement for an anonymous linkability. Several personal data filing systems process data on the same group of data subject. For statistical purposes, there is a requirement to link these systems, but there is no requirement at all to keep the identity of the data subjects. The solution can be found in a combination of PETs. The first step is to create a pseudo-identity as an identifier. This identity is created in three steps. First, a string is created of the first three characters of the surname, the first initial, the date of birth and the code indicating the gender of the data subject. In the next step, a hash of his string is generated. Lastly, the original string is again removed. The pseudo-identity thus created is the same on each of the source information systems and the linkability is, therefore, safeguarded. The third PET principle must now, obviously, be applied on each filing system to be merged. The individual files can now be sent to a central location. Additional security (PET) can be achieved in the following manner. After receipt of the files, the original pseudo-identity is hashed once again and it shall replace the original pseudo-identity. Since hashing algorithms shall differ, anonymity is guaranteed.

Encryption

Encryption is a typical example of an existing technology that may aid data protection. This statement may give the impression that encryption technology has no other application than to serve as a PET. On the contrary, encryption is a general security measure to achieve confidentiality. A company shall use encryption to secure its company secrets (financial details as well as product information) even if there are no personal data at stake. In all circumstances where ‘trust’ is an issue, ‘encryption’ might be the solution.

In cryptology, until the mid-1970s a single key was suppose to be used for encryption and for decryption (symmetric ciphers). The venue of public-key cryptography, in which the key for encryption and the key for decryption are different (asymmetric ciphers), gave rise to a range of services to be rendered by a public key infrastructure; PKI.

As well as encryption (and decryption), these services include enabling parties to share secrets without the need of exchanging keys, digital signatures, data integrity and key establishment first (i.e. using a protocol to establish a symmetric key known only to the parties involved).

Authentication, integrity and confidentiality of messages (personal data transferred between senders and receivers) may be considered core services of any PKI, while non-repudiation, the service that ensures that entities cannot falsely deny sending or receiving a message, is a standard service as well.

The power of cryptology may be used for implementing PETs in the following ways. In order to reduce the amount of personal data to be processed, without losing the functionality of an information system, one must make decisions regarding the following at lease: storing, accessing, disclosure and transferring. Using cryptography can strengthen all of these aspects. Examples of the use of cryptography for data protection include:

Data Storage Data must be securely stored. Personal data, in general, and special personal data, in particular, require specific attention. It is recommended to encrypt personal data that are to be stored on computer devices. Integrity of data may be safeguarded by cryptographic means as well;

Authorisation. In order to decide who has access to what, claims about identities must be verifiable;

Data Access and Data Disclosure Applications may be written that manage access by checking authentication and authorisation. By executing encryption and decryption at application level rather than at database level, improper use of personal data may be precluded;

Data Transport Personal data to be disclosed to specific parties via external storage devices such as floppy disks or CD-ROMs must be encrypted. The same holds for data transfers over a network.

Some uses of public key cryptography are valuable for the protection of software as well, e.g., services for authentication and to guarantee integrity.

Encryption may take place at several OSI levels when network-oriented data transfer is involved:

- At the application level (OSI layer 7, logical): data is encrypted by the application before they are sent down to the next OSI layers;
- At the network level (OSI level 1, physical): before the data are actually transferred to the network, they are processed using an encryption box.

This is to ensure that an unauthorised party listening in on a network or an application server cannot interpret the data transferred or processed. In order to achieve the benefits of public key cryptography, public keys must be issued and the connection between parties and their public key must be trustworthy. The latter is achieved by certificates in which a trusted body (Trusted Third Party – TTP) certifies the link between an entity and a public key.

Biometrics

An opportunity and threat at the same time.

Another example of an existing technology that can potentially be used as a PET. PET was recommended to support the implementation of the authorisation scheme in Section 3.1.5

that described identification, authentication and authorisation. The authentication of the user is a very important phase in the logon process. Are you the person you say you are? It is common knowledge that passwords are either too easy to guess or, if the password is well chosen, the user cannot remember it himself and shall, therefore, jot it down somewhere and keep it near his keyboard. Either way, passwords cannot guarantee the identity of the user. Recognition of 'body material' can authenticate the user and can, therefore, establish, without reasonable doubt, the identity of the user. The Dutch Data Protection Authority published a study on Biometrics. We strongly recommend in this study that a central database with related biometric templates is not created. The temples must, instead, be stored locally on a smart card that is carried by the user.

Audit ability

Article 28 of Directive 95/46/EC gives the Supervisory Authorities (the Data Protection Authorities) investigative powers such as powers of access data forming the subject matter of processing operations and powers to collect all the information necessary for the performance of their supervisory duties. A paragraph has been dedicated in Chapter 6 there is to auditing requirements. A description is given of the data that are required by the authority to perform its supervisory duties. Basically, the requirements are twofold:

1. The descriptions of the (automated and non-automated) systems used to process personal data;

2. The monitoring information (audit trails) about the activities of the user who is processing personal data.

The most obvious solution for creating audit trails is by using PETs. A mechanism is defined logging all activities[5] to the database used to store personal data at a database server level.

This principle is, obviously, strongly related to the second PET principle. If the user of the information system is not uniquely identified when information system logon occurs, the logging of the activities linked to this user is of very little use to the supervisory authority using its investigative powers.

3.2 Common Criteria for Information Technology Security Evaluation

Scope[6]

1. This multipart standard, the Common Criteria (CC), is meant to be used as the basis for evaluation of security properties of IT products and systems. By establishing such a common criteria base, the results of an IT security evaluation shall be meaningful to a wider audience.

2. The CC shall permit comparability between the results of independent security evaluations. It does so by providing a common set of requirements for the security functions of IT products and systems and for assurance measures applied to them during

[5] See Chapter 10 for details on the required contents of the audit rails.

[6] This chapter contains a number of paragraphs that are copied from the official CC documents.

a security evaluation. The evaluation process establishes a level of confidence that the security functions of such products and systems and the assurance measures applied to them meet these requirements. The evaluation results may help consumers to determine whether the IT product or system is secure enough for their intended application and whether the security risks implicit in its use are tolerable.

3. The CC is useful as a guide for the development of products or systems with IT security functions and for the procurement of commercial products and systems with such functions. During evaluation, such an IT product or system is known as a Target of Evaluation[7] (TOE). Such TOEs include, for example, operating systems, computer networks, distributed systems, and applications.

4. The CC addresses protection of information from unauthorised disclosure, modification or loss of use. The categories of protection relating to these three types of failure of security are commonly called confidentiality, integrity and availability, respectively. The CC may also be applicable to aspects of IT security outside of these three. The CC concentrates on threats to that information arising from human activities, whether malicious or otherwise, but may be applicable to some non-human threats as well. In addition, the CC may be applied in other areas of IT, but makes no claim of competence outside the strict domain of IT security.

5. The CC are applicable to IT security measures implemented in hardware, firmware or software. Where particular aspects of evaluation are intended only to apply to certain methods of implementation, this shall be indicated within the relevant criteria statements.

6. Certain topics, because they involve specialised techniques or because they are somewhat peripheral to IT security, are considered to be outside the scope of the CC. Some of these are identified below.

   a The CC does not contain security evaluation criteria pertaining to administrative security measures not related directly to the IT security measures. It is, however, recognised that a significant part of the security of a TOE can often be achieved through administrative measures such as organisational, personnel, physical and procedural controls. Administrative security measures in the operating environment of the TOE are treated as secure usage assumptions where these have an impact on the ability of the IT security measures to counter the identified threats.

   (a) The evaluation of technical and physical aspects of IT security, such as electromagnetic emanation control, is not specifically covered, although many of the concepts addressed shall be applicable to that area. In particular, the CC addresses some aspects of physical protection of the TOE.

   (b) The CC addresses neither the evaluation methodology nor the administrative and legal framework under which the criteria may be applied by evaluation authorities. It is, however, expected that the CC shall be used for evaluation purposes in the context of such a framework and methodology.

   (c) The procedures for use of evaluation results in product or system accreditation are outside the scope of the CC. Product or system accreditation is the administrative process whereby authority is granted for the operation of an IT

[7] The CC defines a TOE as ‘an IT product or system and its associated administrator and user guidance documentation that is the subject of an evaluation’. For a system for the ‘processing of personal data’, some parts of the processing may be performed outside the IT domain. A TOE within the context of personal data is, therefore, not limited to the IT domain.

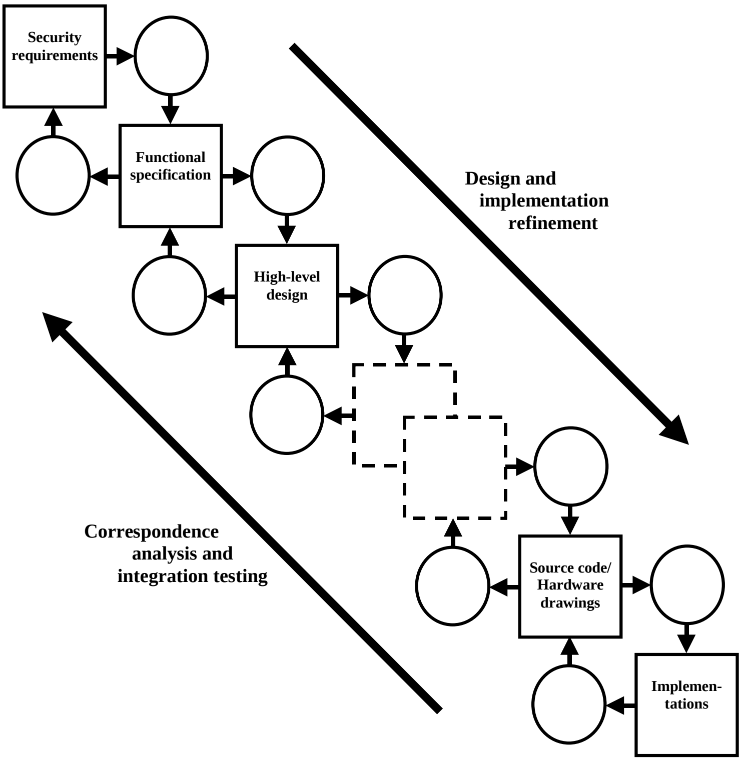

Security requirements
Functional specification
High-level design
Source code/ Hardware drawings
Implemen- tations
Design and implementation refinement
Correspondence analysis and integration testing

Figure 3.2: TOE development model.

product or system in its full operational environment. Evaluation focuses on the IT security parts of the product or system and those parts of the operational environment that may directly affect the secure use of IT elements. The results of the evaluation process are, consequently, a valuable input to the accreditation process. Accreditors must, however, make separate provision for those aspects as other techniques are more appropriate for the assessments of non-IT related product or system security properties and their relationship to the IT security parts.

(d) The subject of criteria for the assessment of the inherent qualities of cryptographic algorithms is not covered in the CC. Should independent assessment of mathematical properties of cryptography embedded in a TOE be required, the evaluation scheme under which the CC is applied must make provisions for such assessments.

Abbreviations Used in the CC In order to fully understand this chapter it is essential to be familiar with some of the abbreviations used in the CC.

- TOE Target of Evaluation
- TSF Target of Evaluation Security Functions
- TSP TOE of Security Policy

The class of the CC that is explained in some detail is the Privacy class. This class uses:

- ANO Anonymity
- PSE Pseudonymity
- UNL Unlinkability
- UNO Unobservability

This paragraph shall show the relation between the seven Privacy-Enhancing Technologies (PET) principles (see Section 3.1.5) and the nine CC classes.

Risk analysis is needed to obtain an effective set of security measures. There are many good methods to perform risk analysis. There are methods that also have an automated tool where the requirements need to be filled in and the tool shall give the necessary 'security measures' as a result. The security measures derived from a risk analysis tool are really more specific requirements for security mechanisms than need to be used. Evaluation criteria such as the Information Technology Security Evaluation Criteria (ITSEC) or the Common Criteria (CC) can be used for implementing these mechanisms. These evaluation criteria, especially the CC, give a very detailed description of the functionality of all possible security mechanisms. These descriptions can be used to build the security for the PISA.

The CC describes the following security functionality classes [CC]:

Class FAU: Security Audit Security auditing involves recognising, recording, storing and analysing information related to security relevant activities (i.e. activities controlled by the TSP). The resulting audit records can be examined to determine which security relevant activities took place and who (which user) is responsible for them.

Class FCO: Communication This class provides two families specifically concerned with guaranteeing the identity of a party participating in a data exchange. These families are related to guaranteeing the identity of the originator of transmitted information (proof of origin) and guaranteeing the identity of the recipient of transmitted information (proof of receipt). These families ensure that an originator cannot deny having sent the message, nor can the recipient deny having received it.

Class FCS: Cryptographic Support The TSF may employ cryptographic functionality to help satisfy several high-level security objectives. These include (but are not limited to) identification and authentication, non-repudiation, trusted path, trusted channel and data separation. This class is used when the TOE implements cryptographic functions, the implementation of which could be in hardware, firmware and/or software.

Class FDP: User Data Protection This class contains families specifying requirements for TOE security functions and TOE security function policies related to protecting user data. FDP is split into four groups of families (listed below) that address user data within a TOE, during import, export and storage as well as security attributes directly related to user data.

Class FIA: Identification and Authentication Families in this class address the requirements for functions to establish and verify a claimed user identity. Identification and authentication is required to ensure that users are associated with the appropriate security attributes (e.g. identity, groups, roles, security or integrity levels). The unambiguous identification of authorised users and the correct association of security attributes with users and subjects are critical to the enforcement of the intended security policies.

Class FMT: Security Management This class is intended to specify the management of several aspects of the TSF such as security attributes, TSF data and functions. The different management roles and their interaction, such as separation of capabilities, can be specified.

Class FPR: Privacy This class contains privacy requirements. These requirements provide user protection against discovery and misuse of identity by other users.

Class FPT: Protection of the TSF This class contains families of functional requirements that relate to the integrity and management of the mechanisms that provide the TSF (independent of TSP specifics) and to the integrity of TSF data (independent of the specific contents of the TSP data). Families in this class may appear to duplicate components in the FDP (User data protection) class at certain levels; they may even be implemented using the same mechanisms. FDP, however, focuses on user data protection while FPT focuses on TSF data protection.

Class FRU: Resource Utilisation This class provides three families that support the availability of required resources such as processing capability and/or storage capacity. The family Fault Tolerance provides protection against unavailability of capabilities caused by failure of the TOE. The family Priority of Service ensures that the resources shall be allocated to the more important or time-critical tasks and cannot be monopolised by lower priority tasks. The family Resource Allocation provides limits on the use of available resources, therefore, preventing users from monopolising the resources.

Class FTA: TOE Access This family specifies functional requirements for controlling the establishment of a user's session.

Class FTP: Trusted Path/Channels Families in this class provide requirements for a trusted communication path between users and the TSF and for a trusted communication channel between the TSF and other trusted IT products. Trusted paths and channels have the following general characteristics:

- The communications path is constructed using internal and external communications channels (as appropriate for the component) that isolate an identified subsets of TSF data and commands from the remainder of the TSF and user data.
- Use of the communications path may be initiated by the user and/or the TSF (as appropriate for the component)
- The communications path is capable of providing the assurance that the user is communicating with the correct TSF and that the TSF is communicating with the correct user (as appropriate for the component)

Table 3.1: PET Principles Versus CC Class.

PET Principles **CC class**	Limitation in the collection	Identification Authentication Authorisation	Standard techniques	Encryption	Pseudo-identity	Biometrics	Audit ability
Security audit **<< Privacy audit >>**							√
Communication				√			
Cryptographic support				√			
User data protection	√	√	√				
Identification and authentication		√				√	
Security management							
Privacy							
Anonymity			√				
Pseudonymity					√		
Unlinkability			√				
Unobservability			√				

Common Criteria CC 2.0 Privacy

Class FPR: Privacy

This class contains privacy requirements. These requirements provide user protection against discovery and misuse of identity by other users.

3.2.1 Anonymity (FPR_ANO)

Family Behaviour

This family ensures that a user may use a resource or service without disclosing the user's identity. The requirements for anonymity provide protection of the user identity. Anonymity is not intended to protect the subject identity.

FPR_ANO.1 Anonymity requires that other users or subjects are unable to determine the identity of a user linked to a subject or operation.

FPR_ANO.2 Anonymity without soliciting information enhances the requirements of FPR_ANO.1 by ensuring that the TSF does not ask for the user identity.

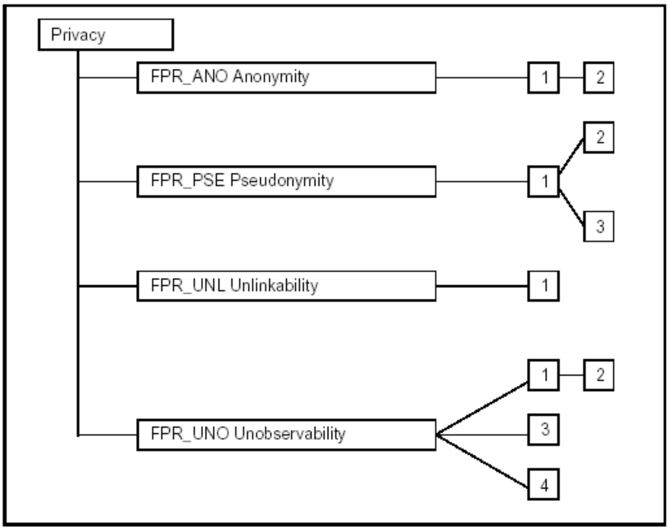


Figure 3.3: Privacy classes and subclasses.

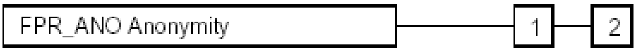


Figure 3.4: Component levelling of anonymity.

3.2.2 Pseudonymity (FPR_PSE)

Family Behaviour
This family ensures that a user may use a resource or service without disclosing its user identity, but can still be accountable for that use.

FPR_PSE.1 Pseudonymity requires that a set of users and/or subjects are unable to determine the identity of a user linked to a subject or operation, but that this user is still accountable for his actions.

FPR_PSE.2 Reversible pseudonymity requires the TSF to provide a capability to determine the original user identity based on a provided alias.

FPR_PSE.3 Alias pseudonymity requires the TSF to follow certain construction rules for the alias to the user identity.

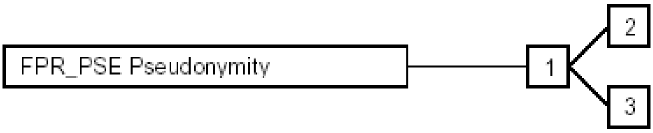


Figure 3.5: Component Levelling Pseudonymity.

FPR_UNL Unlinkability
1

Figure 3.6: Component levelling Unlinkability.

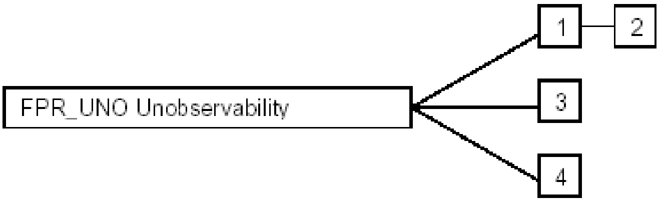


Figure 3.7: Component Levelling Unobservability.

3.2.3 Unlinkability (FPR_UNL)

Family Behaviour
This family ensures that a user may make multiple uses of resources or services without others being able to link these uses together.

FPR_UNL.1 Unlinkability requires that users and/or subjects are unable to determine whether the same user caused certain specific operations in the system.

3.2.4 Unobservability (FPR_UNO)

Family Behaviour
This family ensures that a user may use a resource or service without others, especially third parties, being able to observe that the resource or service is being used.

FPR_UNO.1 Unobservability requires that users and/or subjects cannot determine whether an operation is being performed.

FPR_UNO.2 Allocation of information impacting unobservability requires that the TSF provide specific mechanisms to avoid the concentration of privacy related information within the TOE. Such concentrations might impact unobservability if a security compromise occurs.

FPR_UNO.3 Unobservability without soliciting information requires that the TSF does not try to obtain privacy related information that might be used to compromise unobservability.

FPR_UNO.4 Authorised user observability requires the TSF to provide one or more authorised users with a capability to observe the usage of resources and/or services.

3.3 Privacy by Design
Privacy "act according to the directive" built-in code

Ever since the Privacy-Enhancing Technologies was first thought about, it was assumed that PETs must be used to translate 'soft' legal text into 'hard' system specifications. Once these specifications are translated into system code it would then be impossible, or at least very difficult, to evade these technical measures in contrast to the equivalent organisational ones.

In the Evaluation criteria for compliance auditing (see Section 7.1) nine 'areas of attention' are defined. Each article of the general privacy legislation is allocated to one of these areas. If the method is used, the legislation is grouped into a work plan with nine chapters which can easily be handled by an auditor performing a compliance investigation. The nine areas of attention (Privacy Principles) are[8]:

1. Intention and notification
2. Transparency
3. Finality principle
4. Legitimate grounds of processing
5. Quality
6. Data subject's rights
7. Security
8. Processing by a processor
9. Transfer of personal data outside the EU

It is obvious that some of these areas are easily translated into technical measures (**Security** for one). The largest section of any personal data processing system is performed within the domain of Information and Communication Technology (ICT). Within the domain of ICT, technical security measures are common practice. It is not so obvious, however, that a required set of measures **Data subject's rights** can be implemented through technical measures.

One must start to investigate each of the areas in the firm belief that PETs can be put in place. If they do not already exist for a given problem, a new PET can easily be described and, subsequently, be implemented.

Let us elaborate on the data subject's rights. Let us assume that the controller wishes to give the data subject the possibilities to exercise his rights through an interface on his website. Programmable modules can be designed, which, when executed, shall perform the following. After having authenticated the identity of the data subject, the module shall show the data subject all the personal data items the controller has collected so far. Further functionality of the module could be the implementation of any of the other rights the data subject has been given. From an auditor's point of view, such a PET can be evaluated. If this module is now being made available on the website within a secure environment (e.g. a Virtual Private Network; VPN -) the controller has implemented a legal text into an information system's functionality.

Doing so for all areas of attention is what is now commonly known as 'Privacy by Design'.

[8] Before a privacy audit can be performed all privacy legislation applicable to the environment of the controller and the processing system to be checked must be identified. The contents of these nine areas must be enhanced with the legislation thus identified. It goes without saying that not all areas can be implemented using PETs. Notification to the Supervisory Authority, to name one, cannot be implemented because it is a purely administrative process. The other areas can, at least, partially, be achieved through PETs.

3.4 PET Projects

The goal of the PISA project is not to prove that level 3 personal data can be processed in a secure way and, therefore, the PISA design team has decided that level 3 personal data shall not be processed in the PISA Demonstrator. A proof of adequate protection of level 3 personal data can be found in NBIS[9] offers a presentation for collecting specific research data on burns that is an instrument for best medical practices regarding burn care developed in 2001 for the Nederlandse Brandwonden Stichting.

3.4.1 State-of-the-art PET

Two complementary approaches can be envisioned for the application of PET. The first approach involves the full integration of PET in existing or new information systems. PET functionality is, in principle, enabled in relation to this approach but can be opted out. The second approach involves the development of means that can be used by people when information is communicated. In this case, the PET functionality is switched off by default and concrete action is required (opting in) to enable PET functionality.

Integrated PET

PET can be integrated into several parts of an information system. Integrated implies that PET cannot be separated from the related information system or process. The appropriate manner in which this can be achieved is to consider PET at the design phase of an information system. Integrated PET leaves the user little choice and, therefore, the use of PET is enforced.

ICT Networks

Two applications of PET in ICT Networks can be Recognised: MIX routes [BPS00] and Onion Routers [STRL00]. These two applications enable users to communicate with each other without the communication being noticed by others (unobservability). Both applications are based on network routers. Routers in a network take care of forwarding data packets in the network. The data packets originate from and are destined for computers connected to the network. The routing is based on addresses and, for each address, the router knows in what direction to send the corresponding data packets. Each computer in the network is assigned a unique network address. There is, generally, a fixed relation between incoming and outgoing packets. This allows analysis of the communication flows through a router and, thus, determines between which computers data is communicated. The MIX routes and Onion concepts break the fixed relation between incoming and outgoing data packets and thus prevents, or at least makes it more difficult, to analyse the communication flows and establish a relation between sending and receiving computers.

Information Systems

Currently, only one application is known of a PET that can be applied directly to stored data. The application prevents the improper use of personal data that is stored in a database. Often personal data is stored in a database linked to a unique key such as a social security

[9] NBIS, Maatregelen Gericht op Privacy, presented during the Privacy by Design symposium on 23 May 2002 in The Hague (the Netherlands). The website of The Dutch Data Protection Authority www.cbpweb.nl.

number or a patient number. Persons with access to a database might retrieve all data linked to such a key even if the data is not required for the function of the person. Improper use of personal data can be hampered by implementing a strong access control mechanism based on user groups combined with a mechanism that uncouples groups of personal data from each other and from directly identifiable data. The uncoupling is achieved using so called pseudo-identities. The link between identities and pseudo-identities is stored in an access control list and is coupled to access rights based on the function of the person accessing the database.

PET as Add-on

PET mechanisms can also be used as an add-on to unprotected and privacy protecting information system and ICT networks as well as being integrated in information systems. An important form of an add-on PET is one that is available to the user. The user can then enable specific PET mechanisms to improve privacy.

Anonymisers

An *anonymiser* is an aid to use services anonymously within an ICT network. It is a software program that filters all directly identifiable personal data from the data that are required to establish connections in the network. These are mainly data such as network address, type of computer system, information on the operating system and information on the software that is used. The software program replaces the information with information that does not trace back to the user. The program does not filter personal data from the data that is actually transported over the communication channel.

Profile Managers

Profile Managers enable users to create several profiles that can be considered to be pseudo-identities. A user can use several services while only needing to reveal only those parts of his identity required with these profiles or pseudo-identities. The different profiles can be treated in different ways, thus, improving privacy.

3.4.2 Description of NBIS

The Dutch Burns Information System processes health care data (PD level 3) over the (public) Internet in the Netherlands. End users of the system are physicians. A number of PETs were piled in order to achieve appropriate security. In summary, these PETs are:

- Encryption of all data stored in the database;
- The data in the database is stored in two domains and different pseudo-identities are used in the two domains. Only authorised users shall have access to the second domain after having gained access to the first;
- A Virtual Private Network (VPN) is created to give the end user access to the centralised application and database server;
- Intruder protection through three sequentially placed firewalls of different suppliers;

- Biometrics stored on smart cards are in place to authenticate the end user before access is given to the application server;
- A trusted third party (TTP) is always in place between the end user environment and the central information system.

PET means shall protect Level 1 PD by encapsulating these data within the ISA directly at the time of collection. Only the recipient who is entitled to receive level 1 PD shall have access to the means to unwrap the PET encapsulated level 1 personal data.

3.4.3 LADIS Project

The PISA project uses PET to protect all personal data with traditional PET components such as identity protectors and the creation of anonymity and pseudo-identity domains. When all PD are modified such that the data subject is no longer identifiable, the DPD shall not be applicable any more. An example of this can be found in LADIS. (Landelijk Alcohol en Drugs Informatie Systeem; National Alcohol and Drugs Information System). The PET functionalities were implemented in 1999 into this system[10].

The Dutch Data Protection Authority received a request from a Health Care organisation in 1999 to make a formal statement stating that LADIS was not liable to privacy legislation. The system processes the effect of treatment on drug addicts anonymously. A number of PET measures were applied to personal data before personal data were transferred to this organisation with the result that it was no longer possible to identify the data subjects. The result, therefore, was that personal data was transformed into non-personal data and, consequently, privacy legislation was no longer applicable to the information system. The PETs applied to the personal data are of such a nature that the actions are irreversible: the controller of the non-PD cannot reconstruct the PD in any way. Even if additional resources from outside the organisation are used, reconstruction cannot be achieved. Below a number of examples are given (which also show the PETs simplicity) of the PET data transformations that were applied, after the direct identifiable items, such as name and address, were completely removed to illustrate irreversibility.

- The date of birth was transformed into an age category with a five-year span;
- The full postal code of the data subject was truncated to the first three positions (in the Netherlands the full postal code has 6 positions);
- The job description code was transformed into a code indicating the branch of the job in question (i.e. nurse and physician both get the code for health care employees).

[10] LADIS was presented during the Privacy by Design symposium on 23 May 2002 in The Hague (the Netherlands) in the speech of A.W. Ouwehand, Stichting Informatievoorziening Zorg. The presentation can be found on the website of The Dutch Data Protection Authority: www.cbpweb.nl.

Chapter 4

The Development of Agent Technology

P. Verhaar
verhaar@fel.tno.nl
TNO-FEL, The Netherlands

J.J. Borking
jborking@euronet.nl
CBP, The Netherlands

This chapter is a summary of [BvES99] and provides information regarding agent technology and the link with PET.

4.1 Agent Technology

This section provides information related to agent technology.

4.1.1 Background

Software agents have their roots in work conducted in the fields of software engineering, human interface research and Artificial Intelligence (AI). They can, conceptually, be traced back to the late 1970s when their predecessors, so-called “actors”, were introduced. These actors were self-contained objects, with their own encapsulated internal state and some interactive and concurrent communication capabilities. Software agents developed up to now can be classified under Multiagent Systems (MAS), one of the three branches of distributed AI research, the others being Distributed Problem Solving (DPS) and Parallel Artificial Intelligence (PAI) [NA97]. Technically, they exhibit many of the properties and benefits common to distributed AI systems. These properties include:

Modularity. A modular programming approach that reduces the complexity of developing software systems.

Speed. Parallelism, that is, the concurrent execution of cooperating programs, increases the execution speed of the overall system.

Reliability. Built-in redundancy increases the fault tolerance of an application, thus enhancing its reliability.

Operation at knowledge level. Utilisation of AI techniques allows high-level messaging.

Others. These include maintainability, reusability and platform independence.

4.1.2 Reasons for Software Agents to Exist

Research and development efforts in the area of agent technologies have increased significantly in recent times. This is the result of a combination of ’market pull’ and ’technology push’ factors.

The key factor triggering the ’market pull’ is information overload. The volume of publicly available scientific, corporate and technical information was doubling every five years in 1982. It was doubling every 2.2 years by 1998 and every 1.6 years by 1992. With the rapid expansion of the Internet, one can expect this rate of increase to continue, which means that by 1998 the amount of information shall probably double in less than a year. This dramatic information explosion poses a major problem: how can information be managed so that it becomes available to the people who need it, when they need it? How must one organise network flows in such a way as to prevent massive retrieval of information from remote sources from causing severe degradation of network performance, i.e., how can one ensure that network capacity is used economically? Software agents hold the promise of contributing to providing a solution to this issue. Agent technologies can be used to assist users in gathering information. Agents can gather and select this information locally, thereby avoiding unnecessary network loads. What distinguishes (multi)agent architectures from other architectures is that they provide acceptable solutions to certain issues at an affordable price.

The key factor triggering the ’technology push’ is the rapid development of communication and information technology. At present, communication technology offers communication facilities and solutions with increasing capabilities, both in terms of bandwidth and speed, at decreasing cost. Information technology today offers powerful tools such as object-oriented programming, graphical user interfaces and knowledge engineering techniques, which assist software system developers in keeping the development burden of complex systems manageable.

Interestingly enough, the ’market pull’ factor and ’technology push’ factor reinforce each other. As the communication and information technology gets more advanced, more information can be processed, and, when there is more information to process, the technology to do so needs to be more advanced. This in turn pushes the development of new technology, such as agent technology, designed to solve the issues.

4.1.3 Definition of Agents

There is no general agreement on a definition of the word ’agent’, just as there is no consensus within the artificial intelligence community on a definition of the term ’artificial intelligence’. One can, in general, define an agent as a piece of software and/or hardware capable of acting in order to accomplish a task on behalf of its user.

A definition close to present day reality is that of Ted Selker from the IBM Almaden Research Centre:

> ’An agent is a software thing that knows how to do things that you could probably do yourself if you had the time’.

Agents come in many different flavours. Depending on their intended use, agents are referred to by an enormous variety of names, e.g., knowbot, softbot, taskbot, userbot, robot, personal (digital) assistant, transport agent, mobile agent, cyber agent, search agent, report agent, presentation agent, navigation agent, role agent, management agent, search and retrieval agent, domain-specific agent, packaging agent.

The word 'agent' is an umbrella term that covers a wide range of specific agent types. Most popular names used for different agents are highly non-descriptive. It is, therefore, preferable to describe and classify agents according to the specific properties they exhibit.

An example of an agent is a Personal Digital Assistant (PDA) that is described in the following metaphor [ABN], which describes the cooperative, mobile, and learning processes that are present in a PDA.

> Metaphor:
> 'Bruce awoke instantaneously at 06:00 a.m. sharp, expecting a long day of helping his boss, Hava. He took a look at Hava's daily schedule and then went to the mailbox to see what other meetings and appointments he would have to squeeze in today. There was a request for an urgent meeting from Doris, Seppo's assistant. He contacted Doris, informing her that Hava had half an hour free at 10:00 a.m. or at 5:00 p.m. and that Hava personally preferred morning meetings. Doris confirmed 10:00 a.m. and Bruce posted a note for Hava. Next on his agenda, Bruce went about sorting through the rest of Hava's mail and news bulletins, picking out a select few that he believed would satisfy her reading habits and preferences. At about 9:30 a.m. he caught a message from Hava's best friend that tonight she was free. Knowing that Hava likes going with her friend to movies and that she had not yet seen 'Brave Heart' with Mel Gibson, her favourite actor, Bruce decided to buy them a pair of tickets to the early show and make reservations at Hava's favourite restaurant. He stepped out and zipped over to the mall, to the ticket agency, and discreetly bought the tickets with Hava's VISA number. He returned with a big smile on his face and notified Hava of her evening plans. At about 01:00 p.m. he received an urgent message from Hava telling him that she was happy about tonight's arrangements, but did not want to see 'Brave Heart' because it was too violent for her. Bruce noted Hava's aversion to violent films for future reference and hurried back to the mall to try and sell the tickets to someone else and then buy tickets to 'Sense and Sensibility' (Hava just loves Emma Thompson). At 7:00 p.m., before leaving for the movie, Hava notified Bruce that he had done well today and then she turned off the computer (and Bruce of course) for the night.'

Bruce is not a human secretary, but a personal digital assistant. This assistant is trusted by its (controlling) user Hava on many matters: deciding about meeting schedules, money and personal matters, such as entertainment and dining. Moreover, the personal assistant Bruce has to ensure discretion by not revealing any privacy sensitive information about Hava, unless instructed to do so by Hava.

The information Bruce possesses about Hava is recorded in the so-called user profile. A user profile contains all personal data an agent possesses about its user. In the metaphor, the user profile Bruce has of Hava contains at least the following information:

- The name of its user: Hava;
- Hava's daily schedule;

- Hava's mailbox;
- The name of one of Hava's acquaintances and the agent that works for this acquaintance: Seppo and Doris;
- Hava's reading habits and preferences;
- Hava's best friend, and their mutual hobby;
- Hava's favourite actor: Mel Gibson;
- Hava's favourite actress: Emma Thompson;
- Hava's favourite restaurant;
- Hava's aversion to violence.

This is only a fragment of the personal data that could be present in Hava's user profile. Bruce could have collected far more information, such as:

- Hava's address, telephone numbers and electronic mail addresses;
- Hava's relatives;
- Other acquaintances Hava may have;
- Not just Hava's reading habits and preferences but also all other habits and preferences.

An infinite amount of personal data could, therefore, be recorded in a user profile. The only restrictions are the technical restrictions of the agent's memory (its capacity) and the agent's processing capacity.

4.1.4 Agent Ownership

Agents could be owned by individuals or organisations. These agent owners can use their agents to:

- Carry out tasks to fulfil their owner's purposes;
- Offer agent services to individuals or organisations that are not in a position to own an agent.

The agent Bruce could be owned by the boss Hava in the metaphor provided above, but Hava could also have hired Bruce from a company or organisation that provides agents. There are a number of reasons why Hava would not be in a position to own her own agent. One of the reasons relates to the cost of purchasing an agent or the hardware needed for the proper operation of the agent. Another reason could be the number of tasks that Hava wants to delegate to the agent. If the number of tasks is very small, for example, fewer than three tasks a year, it is better to hire an agent than to use her own agent.

Service providers, such as Internet service providers, could provide a network infrastructure with strong network servers and local workstations with only the necessary hardware and software to connect to the network servers. This structure could also be provided by cable TV companies, which already have the cable infrastructure and want to provide more services to their subscribers. Such a network infrastructure shall reduce the costs of the workstations and, therefore, increase the possibilities for financially less well-endowed individuals to use the Internet. These workstations leave practically no room for the installation of additional (local) software, including user owned agents. People who use these services shall end up using agent services that are provided by the network provider.

When using an agent provided by an agent provider, the personal data that is provided to the agent in order to create a user profile can be passed on to, and recorded by, this agent provider. This could be an undesirable situation for an individual, especially for individuals who are concerned about their privacy. This might be an argument for only using an agent that is owned by the individual. It could also be a good reason to draw up an agreement between the individual and the agent provider which contains, for example, a privacy intrusion liability clause.

4.1.5 Interaction between Users and Agents

When a user activates an agent, said user not only delegates tasks to it but also delegates responsibility and competence. The interaction between a user and the agent might be compared to the interaction between a boss and a secretary or a master and a servant. The user loses control over a considerable amount of the agent's activities when the user delegates tasks, responsibilities and competence. It is, therefore, important that the user can trust the agent that said user uses, just as the boss trusts his or her secretary and the master trusts his or her servant.

A lack of trust could be the result of a difference between the working methods of the user and the agent [Nor94]. If the user does not know what his agent is doing, or is not content with the way the agent works, he might consider never using this agent again. A certain type of agreement must be in place between the agent and the user, as there is between secretary and boss where the agreement is often based on mutual engagements. The agreement is tried out for a probation period. During this period, both parties can decide whether they accept the agreement. A user must have a description of the working method of the agent in order to learn more about it before using the agent. The user shall then know what to expect from the agent and can decide the extent to which he can trust the agent.

A lack of trust could also be avoided by increasing the discretion of the agent. The longer an agent works for its user the more it shall know about him or her. This is the case in the relation between master and servant, where the servant knows practically everything about his or her master making it very important that the information is dealt with using the highest degree of discretion. The servant shall be engaged on account of this quality. It is important that agents have the means to protect the privacy of their users. These means are called Privacy-Enhancing Technologies (PET), which shall be discussed in Chapter 3.

Another way to increase trust is to provide assurances about the level of control individuals have over their agents. The following items have to be taken into account to give users the feeling that they are in control of their agents [Nor94]:

- A description of the way the user and the agent interact;
- Safeguards to prevent unwanted situations;
- The setting of accurate expectations to minimise false hopes.

4.1.6 Classification of Agents

Agents can be classified according to the specific properties, or attributes, they exhibit [NA97]. These include the following:

Mobility. This refers to the extent to which an agent can move around a network. This leads to a distinction being made between static and mobile agents. Sometimes this includes cloning to distribute subtasks in a remote environment.

Deliberative behaviour. Deliberative agents possess an internal reasoning model and exhibit planning and negotiation skills when engaged with other agents in order to achieve their goals. In contrast with deliberative agents, reactive agents lack an internal reasoning model, but rather act upon the environment using a stimulus/response type of behaviour.

Primary attributes. The most important attributes of an agent are referred to as primary attributes; less important, or secondary attributes, are listed below. The primary attributes include, at least, the following three:

- Autonomy: reflects the ability of agents to operate on their own, without immediate human guidance, although the latter is sometimes invaluable.
- Cooperation: refers to the ability to exchange high-level information with other agents: an attribute which is inherent in multiagent systems (MAS).
- Learning: refers to the ability of agents to increase performance over time when interacting with the environment in which they are embedded. Agents combining several of the primary attributes are referred to by different names again in [ABN]: autonomous agents that cooperate are called collaborative agents, those that learn are referred to as interface agents, and those that do both are termed smart agents.

Secondary attributes. Agents can be classified according to a number of other attributes, which could be regarded as being secondary to the ones described above. Rather than a comprehensive list, some examples of secondary attributes that agents may exhibit shall be given. Agents may be classified, for example, by their pro-active versatility, that is, the degree to which they pursue a single goal or engage in a variety of tasks. One might, furthermore, attribute social abilities to agents, such as truthfulness, benevolence and emotions (anger, fear), although this last ability is certainly controversial. One may also consider mental attitudes of agents, such as beliefs, desires, and intentions (BDIs in short).

Hybrid agents and heterogeneous agents can be constructed by combining these properties and attributes, [CH97]. With hybrid agents, two or more properties and/or attributes are combined in the design of a single agent. This results in the combination of the strengths of different agent design philosophies in a single agent, while at the same time avoiding their individual weaknesses. It is not possible to separate such an agent into two other agents. Heterogeneous agents combine two or more different categories of agents in such a way that they interact via a particular communication language.

4.1.7 Intelligence and Agency

An agent's intelligence can range from more to less intelligent by varying the extent of the learning attribute. By varying the extent of the autonomy and cooperation attributes,

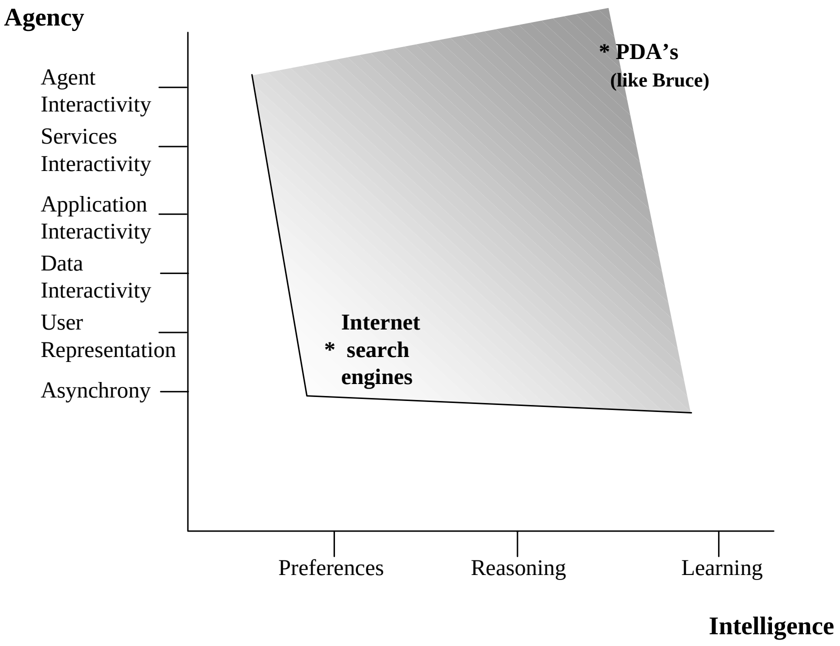


Figure 4.1: The position of intelligent agents in relation to intelligence and agency.

an agent's agency can vary from no interactivity with the environment to total interactivity with the environment.

Intelligence relates, in this case, to the way an agent interprets the information or knowledge to which it has access or which is presented to the agent [CH97]. The most limited form of intelligence is restricted to the specification of preferences. Preferences are statements of desired behaviour that describe a style or policy the agent needs to follow. The next higher form of intelligence is described as reasoning capability. Preferences are combined with external events and external data in a decision-making process with reasoning. The highest form of intelligence is called learning. Learning can be described as the modification of behaviour as a result of experience.

Agency relates to the way an agent can perceive its environment and act on it [CH97]. Agency begins with asynchrony, where the agent can be given a task which it performs asynchronously with respect to the user's requests. The next phase of an agency is user representation, where an agent has a model of the user's goals or agenda. The agent is able to perceive, access, act on, communicate and interact with data, applications, services and other agents in subsequent phases. These phases are called: data interactivity, application interactivity, service interactivity, and agent interactivity.

When intelligence and agency are combined, it becomes possible to indicate where 'intelligent' agents are positioned. Figure 4.1 illustrates this. Agents that are positioned in the shaded area are more or less 'intelligent' agents.

4.1.8 Examples of Agents

Agents can be classified according to the properties they exhibit. This section shall provide some examples of actual implementations of software agents:

Collaborative agents. Collaborative agents interconnect existing legacy software, such as expert systems and decision support systems, to produce synergy and provide distributed solutions to problems that have an inherent distributed structure.

The Pleiades System, a visitor hosting system of Carnegie Mellon University, is an example. This system uses two specific types of agents, known as task agents and information agents. The former are used to arrange appointments and meetings with users and other task agents and the latter are used to provide task agents with information (user preferences, agendas, etc.), which they, in turn, retrieve from databases. Other examples of collaborative agents include Personal Digital Assistants (PDAs) and systems for financial portfolio management, for emergency medical care and for workflow management.

Interface agents. Interface agents provide personalised user interfaces for sharing information learned from peer observation and for alleviating the tasks of application developers. Interface agents adapt to user preferences by imitating the user, by following immediate instructions of the user or through the Pavlov effect (learning from positive and negative responses of users). One has to realise that interface agents can only be effective if the tasks they perform are inherently repetitive (otherwise, agents shall not be able to learn) and if the behaviour is potentially different for different users (otherwise, a knowledge base would be a better option).

Well-known examples of interface agents include news filtering agents (e.g., PointCast), calendar agents, web browsers and World Wide Web (WWW) cookies. The task wizard under MS Windows ’95 or Office ’97 might also be considered a (primitive) interface agent. Other examples include Yenta, a match-making agent that brings together people with shared interests, Kasbah, a classified advertisement service on the WWW that filters information, and Ringo and Firefly, recommendation systems for music based on social filtering, that is, a technique similar to word-of-mouth recommendations.

Mobile agents. Mobile agents reduce communication costs and overcome limitations of local resources. Decentralisation of the selection process prevents unwanted information being sent over networks, thus economising on network utilisation. The user may, for example, have to download many images from a remote location just to pick out one. Mobile agents could ’go’ to that location and only transfer the selected compressed image across the network. General Magic’s Telescript Development Environment is an example of this situation. The Java programming language from Sun Microsystems also supports mobile agent system development. Other examples of mobile agents include communication super services such as speech-to-text applications.

Information agents. Information agents circumvent ’drowning in data, but starving for information’. This corresponds to solving the problem of information overload mentioned earlier in the Introduction. The best-known example is Netscape’s web browser. Other examples are search engines, like Alta Vista and Yahoo!. The history of Netscape, Inc., makes it clear that the financial incentives to develop information agents can be awesome.

Reactive agents. Reactive agents have as primary advantages that they are robust and fault tolerant yet, in spite of their simple stimulus response communication behaviour, allow for complex communication behaviours when combined. Examples include sensors and robotics.

Role model agents. These are agents that are classified according to the role they play, e.g., World Wide Web (WWW) information gathering agents.

Hybrid agents. Hybrid agents combine the strengths of different agent design philosophies into a single agent, while at the same time avoiding their individual weaknesses. Most examples involve hybrid agents that combine deliberative agents with reactive agents. The reactive agent is used for tasks that are behaviour based and that involve relatively low-level messaging; the deliberative agent is used for tasks that involve local planning or coordinating planning activities with other agents or the user. Specific examples include FORKS, an automated loading dock with forklift robots , computer games and entertainment software.

Heterogeneous agents. Heterogeneous agents combine two or more different categories of agents in a single application, which can interact via a particular communication language. These agents provide for interoperability of existing software products in order to produce synergetic effects. The key issue is to develop an Agent Communication Language (ACL) that forms the basis for interoperability. Implementation of ACLs involves one of the following: a (costly) rewrite of the existing software, a transducer which acts as an interpreter of the original software's communication protocol and converts it to the ACL, and a wrapper which injects modified communication software into the existing software.

A number of the examples given above refer to prototype systems. The introduction and further development of agent systems, moreover, usually involve having to overcome technical as well as social, legal and ethical hurdles.

4.1.9 A General Agent Model

The previous section provided a definition of agents as pieces of software and/or hardware capable of acting in order to accomplish a task on behalf of its user. There were, moreover, a number of properties and attributes listed that agents can exhibit. Most of these properties and attributes were illustrated by giving examples of existing implementations of agents. Although these properties, attributes and examples give a flavour of the scope of agent research and the potential practical uses of agents, they hardly describe how agents actually work. The way in which agents work shall be addressed in this section.

An agent acts in order to accomplish a task on behalf of its user. Conceptually, several steps can be discerned. First, an agent establishes a profile of its user. Once this profile has been interpreted, an agent derives tasks from it, taking environmental conditions and internal reasoning models into account. These tasks are accomplished by performing a series of internal and external actions. The internal actions reflect the execution of algorithms[1], while the external actions reflect actions (e.g. communication) with the environment and possibly with other agents. After completion of the tasks, the results are mediated back to the user.

It is clear, from the conceptual description given above, that a general understanding of the working of agents requires an understanding of its internal workings as well as an

[1] Algorithm: a prescribed set of well defined rules or processes for the solution of a problem in a finite number of steps.

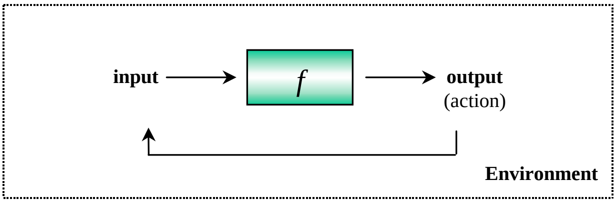


Figure 4.2: Black box model of an agent.

understanding of the mechanisms that underpin the communications behaviour amongst agents. It must be noted here that in real world applications agents have limited resources and act in a time-constrained environment.

Agents can be described using the black box model of Figure 4.2. This figure describes the processing of received messages (input) via some function f to performed actions and transmitted messages (output).

An external authority does not directly control the f mapping: the agent is autonomous. The distinctions between agent models stem from differences in the nature of the f mapping that determines the behaviour of an agent.

The main flaw of this black box model is that it is too general: any information system can be described using the model of Figure 4.2. There is, therefore, a gap between the legitimacy of this model and its usefulness. It is useful first to discuss the parent disciplines of agent design in more detail in order to derive a more appropriate model, that is, one that captures the essential characteristics of agents. These disciplines are control theory, cognitive psychology and classical AI planning theory.

4.1.10 Control Theory

Classical control theory provides a mathematical framework describing the interaction between controller and environment (both viewed as deterministic finite state machines acting in a dynamic system). Determining a sequence of actions suitable for achieving a certain goal is called the control problem. Since there is usually no exact model of the environmental status, it must be estimated. A distinction is made between feed forward and feedback control. Actions are based on the monitoring of the behaviour of the environment and changes therein with feedback control while the reaction of the process to be controlled can be predicted with feed forward control. The analogy with agents is that agents recognise situations, derive goals and engage in planning and scheduling in order to act according to the goals set. The interpretation of recognised situations employs (symbolic) reasoning using a specific perception of the actual model of the world, and, therefore, may be incomplete and erroneous. Feed forward planning uses optimal planning according to some hypothetical model of the world. Feedback planning involves actions in response to preconditions triggered by actual situated rules.

The contrast with agents is that control theory usually copes badly with complex environments which can, at best, be only partially modelled. Agents use explicit representations of knowledge in reasoning processes, which allows for reasoning with incomplete or inconsistent data. Agents, however, usually require insight into the intentions of the environment from which it derives information (truthfulness, benevolence, etc.).

4.1.11 Cognitive Psychology

Cognitive psychology, in particular, motivational theory, investigates how goals and intentions of human agents emerge and finally lead to the execution of actions that change the state of the world. One can distinguish two main subprocesses:

- Formulation of intentions: starting from a set of (possibly inconsistent) motivations, the resulting motivational tendency, which forms the basis for the formation of (an internally consistent set of) intentions to act, is derived;
- Activating processes: the process of putting intentions into practice, i.e., the process of deciding how and when actions are to be initiated in compliance with these intentions.

Motivational theory comprises two schools of thought, a person-centred (Descartes) and a situation-centred (Darwin) approach, which derive from a debate about whether human intelligence is the result of an evolutionary process (mutation, selection) or a fundamental quality inherent to human beings (rationale versus instinct). The Dynamic Theory of Action (DTA) is a formal method for deciding which goals to pursue as a function of the current situation and the mental state of a person. The agent needs to take into account instigating forces, consummatory forces, inhibitory forces and forces of resistance when making decisions. Implementing the DTA requires solving bottlenecks concerning how to represent and recognise situations and motives, and how to model the (inter)dependencies of the four forces at play to ensure smooth behaviour patterns result.

4.1.12 Classical Artificial Intelligence Planning Systems

The problem solving behaviour of agents is viewed as a sense-plan-act (input, function(f), output: from Figure 4.2) cycle. Given a problem description in terms of an initial world state, a goal state, and a set of operators, one may view planning as selecting a set of actions (operator executions) that transforms the initial state into a goal state. Planning can thus be viewed as searching for a state space in order to realise a goal of the agent in question. Most classical AI planning systems require complete and up-to-date information, changes to the world state to be a function of the actions of the agent alone, actions to be deterministic, with correct specifications, and correct implementations (e.g., without system failure). Results have been achieved where a number of the constraints were relaxed, e.g. planning under resource limitations, interleaving of planning and execution, etc. Usually, however, AI approaches are strongly influenced by the classical AI approach, where planning involves symbolic representation of knowledge, skills and goals, and the process of planning and plan execution is viewed as achieving transitions in a discrete state space.

4.1.13 The Agent Model

A detailed model of the internal workings of an agent can be given with this insight into the parent disciplines of agent design. First a layered model of an agent that is deliberative, reactive and cooperative shall be given. Such an agent consists of three layers. These layers are: the behaviour-based layer, the local planning layer and the cooperative planning layer. The agent also uses three models that represent different parts of reality: the world model, the mental model and the social model. The world model contains a description of the agent's environment and is linked to the behaviour-based layer. The mental model

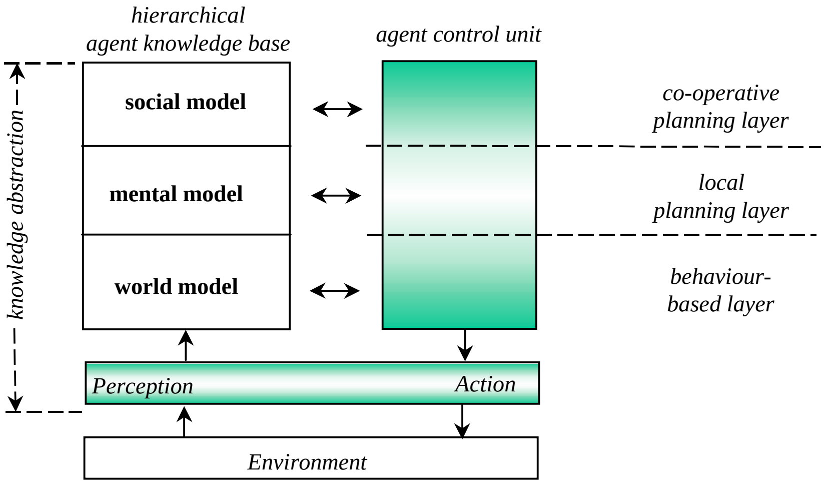


Figure 4.3: The layered design of an agent that is deliberative, reactive and cooperating (cf. [Mül96]).

describes the inner state of the agent itself and is linked to the local planning layer. The social model describes the inner states of other agents in the environment and is linked to the cooperative planning layer.

The models are hierarchical, see Figure 4.3. The agent shall use the models to interpret the input from the environment and to plan possible internal and external actions. The world model helps the agent to make decisions about its environment. The mental model helps the agent to plan possible actions the agent can perform to fulfil its task. The social model helps the agent to control actions taken by other agents in order to cooperate with these other agents and to avoid conflicts with them.

An example of an automated loading dock could illustrate this model. Autonomous agents are carrying out tasks, such as moving goods from the storage facilities to the transportation area and vice versa, in this automated loading dock. The agents need to know, at first, what their tasks are and what these tasks mean. The agents shall use the world model to interpret their tasks and fix their behaviour. The mental model shall help the agents to plan all actions they need to execute to move goods from the storage facilities to the transportation area, or vice versa. The agents need to know what the other agents are doing to avoid running into each other and getting blocked by the other agents. Agents could each decide to make a random move to get out of this situation with the social model. Another possibility for resolving the ’traffic jam’ could be that the agents exchange their goals, which shall lead to mutually agreed actions.

Another way to describe an agent that combines deliberative, reactive and cooperative properties and attributes, is given in Figure 4.4 [Mül96].

Figure 4.4 depicts the agent control unit (see Figure 4.3) of the hybrid agent, its conceptual building blocks and their relationships. Conceptually, one can distinguish the following components:

1. Perception. This refers to the symbolic representation of the information communicated to the agent.

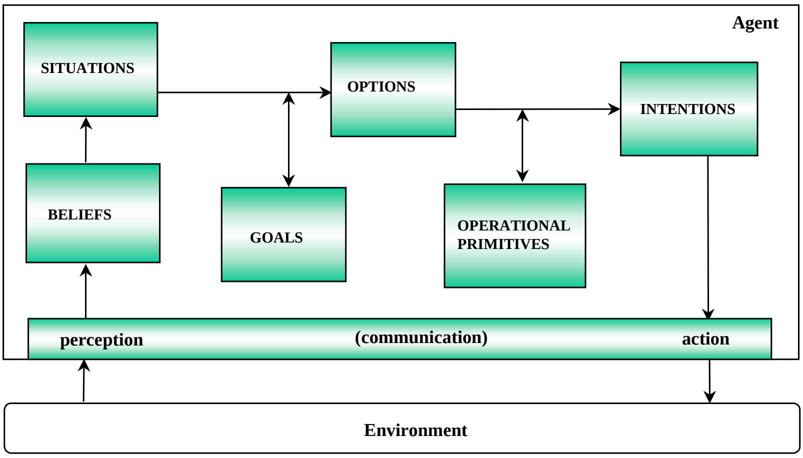


Figure 4.4: The agent model - a conceptual model of an agent that combines deliberative, reactive, and cooperating properties and attributes (cf. [Mül96]).

2. Beliefs. These express the expectations an agent has about the current state of the world and about the likelihood that a specific action produces certain effects.

3. Situations. These enable the agent to identify the need for activity. There are three classes of situation according to the three models. Firstly, there are the behavioural situations, which are a subset of the agent's world model. Secondly, there are the situations describing local planning. These situations are based both on the world model and on the mental model. Lastly, there are the situations that describe cooperative planning. These situations are based on the social model.

4. Goals. It is possible that an agent has a set of goals. These goals are context independent. Goals can be classified into reaction goals, local goals and cooperative goals. Reaction goals are goals that are triggered by external events. These goals require a fast reaction and are short-term. Local goals refer to the goals of the agent itself. Cooperative goals are goals that are shared among a group of different agents.

5. Options. An agent can also contain a set of options. The options represent the agent's motivational state. Based on the current situation, a set of context dependent options is selected. These options are related to the agent's goals. Given the selected option, operational primitives are selected to achieve the current goal(s).

6. Operational primitives. These primitives enable an agent to achieve certain goals. Once selected, these operational primitives are merged into an execution schedule.

7. Intentions. An agent also has intentions. These intentions define the action an agent is going to take (the deliberative state of the agent). The intentions lead to the execution of the operational primitives from the execution schedule.

4.1.14 The Communication Model

All agents need to be active in the same network infrastructure to make communication possible. This network architecture needs to contain one or more of the following facilities:

- Facilities to run an agent (program);
- Facilities to support communication between agents of different types;
- Facilities to allow movement of agents from one system to another;
- Facilities to allow cloning of a mobile agent in a local environment;
- Facilities to encapsulate agent information;
- Facilities to identify and ’fingerprint’ (authenticate) agents.

There is no need for agents to stay in the same place. If it is more efficient to accomplish a task or achieve an objective at a different location, the agent might move to that location. Figure 4.2 presents a black box model of the way agents communicate with their environment independently of their exact location in the environment.

4.1.15 The Future of Software Agents

Some say ’intelligent’ agents are Science Fiction, but is this really so? No, the future is close at hand. Many current developments in R&D laboratories deal with the problems of intelligence, adaptive reasoning and mobility. People have, nevertheless, exaggerated expectations about agents due to the natural enthusiasm of researchers. Researchers see far into the future and imagine a world of perfect and complete agents. Most agents available today are, in practice, used to gather information from public networks, such as the Internet. Many user-initiated actions are still needed for these agents to accomplish their tasks. This means that most agents are still reactive, and have not yet developed as far as most researchers would like. Today’s agents are, therefore, simple in comparison to those that are being planned’ [Nor94, MR94].

4.2 Software Agents and PET

As stated in the previous section, various governments and international governmental organisations have drawn up privacy regulations and privacy guidelines. Tough measures are needed to enforce these regulations. These have, up to now, taken the form of inspections or audits to verify whether all organisations that collect personal data are complying with privacy regulations. These inspections are time-consuming and, therefore, expensive. There is, obviously, a need for technologies capable of replacing inspections for enforcing privacy regulations.

This section describes the potential and implications of using technologies to manage the threats described in the previous chapter and improve the privacy of individuals in an agent-based environment. These threats can be managed by using an Identity Protector (IP) described in [HB98]. [HB98] also describes the technologies to implement an IP. These technologies are defined as Privacy-Enhancing Technologies (PETs). An IP controls the exchange of the user’s identity within an information system. An IP can be used in two ways in an agent-based environment:

1. Between the user and the agent, see Figure 4.5(a);
2. Between the agent and the external environment, see Figure 4.5(b).

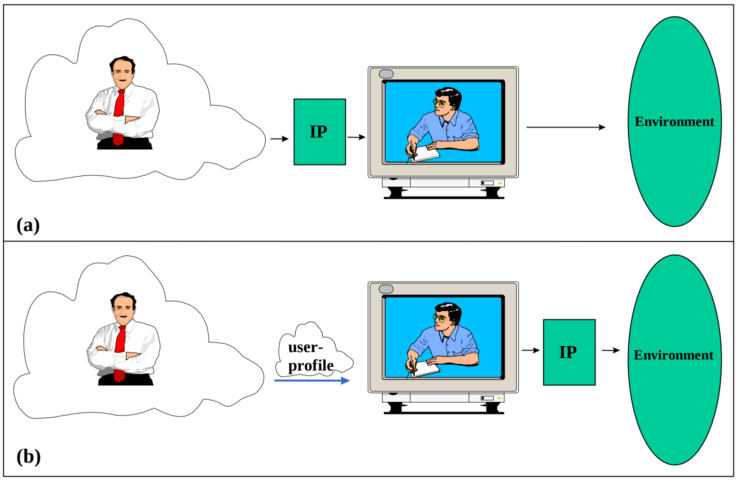


Figure 4.5: The Identity Protector (IP) placed in an agent-based environment: (a) the IP placed between the user and the agent; (b) the IP placed between the agent and the external environment.

There shall be no exchange of personal data from the user to the agent without the approval of the IP and the user when the IP is placed between the user and the agent. The user can, thus, control the amount of personal data that is recorded by the agent. This option could be used to protect the user against threats to privacy caused by agent providers.

Placing the IP between the agent and the external environment gives the agent comprehensive powers to obtain and record personal data from its user. The IP shall help the agent to protect the personal data of its user against unwanted dispersion.

The PET described in [HB98] to implement an IP are only capable of managing a few of the threats. Existing security technologies that are not yet defined as PET need to be applied in such a way that they can improve the privacy of individuals to ensure that the remaining threats can be managed. These technologies shall, eventually, also be called PET.

Agent threats can be divided into two groups: Threats caused by agents acting on behalf of a user and threats caused by foreign agents that act on behalf of others. The potential and implications of using PETs to counter threats shall be studied for each group. The result of this shall ensure that PETs are given solutions for each group of threats. An overall solution for an agent that protects both the privacy of its user and the privacy of the individuals in its environment is given by combining the PET's solutions for both groups.

Irrespective of the fact that privacy is not a commodity but a fundamental human right, it has to be said that the protection of an individual's privacy is still the individual's own responsibility and choice. It is, therefore, up to each individual whether privacy is protected or not. This leaves the individual with the consideration of whether or not to use PETs to secure his or her agent. If individuals choose to protect their privacy, they still needs to make a choice about the extent of the protection offered by the PETs. The extent of protection could be defined by the relationship the individuals have with their environment. This relationship can consist of political, social, public, or other types of interactions. If the individual decides to take an agent with PETs with a high degree of privacy protection, this shall have consequences for the performance of the agent.

4.2.1 PET that Manage the Identified Threats

The following threats have been identified:

- Loss of control;
- Agent providers;
- The agent exchanges personal data with its environment:
- When communicating with service providers;
- By forwarding tasks to other agents or clones;
- The agent runs into an agent that is in disguise;
- The agent runs into an agent that is more powerful (e.g. has more processing power or capacity) than itself;
- Agents can perform traffic flow analysis;
- Agents can enter the privacy domain of a user and collect whatever they want;
- Agents can enter databases and collect personal data;
- Agents can filch personal data from the users agent.

Loss of Control

Increasing the users' trust towards their agent can prevent loss of control. This can be achieved by the certification of the agent's working method and the logging and auditing of all the agent's internal and external actions.

The evaluation of the agent's method of operation by an independent organisation is the first step towards increasing users' trust in their agent. A certificate of this evaluation, in combination with the 'digital signature' of the agents themselves, shall provide users with a guarantee that the agents can be granted a certain level of trust and discretion. This 'digital signature' is the equivalent of a digital signature placed over an electronic document, where the electronic document is replaced by the agent's source code. A detailed description of digital signatures is given in [HB98].

A user must have the possibility to verify the correctness of the agent's working method, besides certification, and thereby consolidate their relationship. All actions, therefore, taken by the agent need to be logged besides certification of the working method. The user of the agent must be able to audit all logged actions. The threat of loss of control can be kept under control with the certification of the working method and logging and auditing of the agent's actions. The level of control depends on the level of assurance provided by the evaluation and the implementation of the logging and auditing mechanism. The logging and auditing mechanism covers part of the privacy principle of transparency, since this shall show the user when, where, and what personal data was exchanged with the external environment. A well-implemented logging shall also help the user to backtrack a decision (automated decision) made by the agent.

Agent Providers

The following measures can be used to decrease the impact of the threats caused by agent providers depending on the confidence that can be assigned to an agent provider:

- Certifying the agent's method of operation;
- Concluding a contract or agreement between the user and the agent provider;
- Using the IP if there is no confidence in the agent provider at all.

Certification of the agent's working method has already been described in the previous paragraph. The agent's working method must be checked for activities directed towards the agent provider. If there are no such activities, the working method can be certified. If there are such activities, the intentions of these activities need to be clear and acceptable to the agent user. The agent can still be certified but there also needs to be a logging and auditing mechanism to help the user to control these activities. An agreement, or contract, between the user and the agent provider can help to increase the user's trust in the agent provider. The recording behaviour of the agent provider can be controlled by inserting statements such as 'the agent provider shall not record any information about the agent user' or 'the agent provider is not allowed to record any information about the agent user except for the information needed to supply the agent'. A user can place an IP between the user and the agent if there is no confidence in the agent provider at all as illustrated in Figure 4.5(a). This IP can be implemented with the PETs described in [HB98].

Exchanging Personal Data

Personal data only needs to be exchanged if the agent or the agent's user has to account for a specific action that has been executed (see the 'buying flowers' example in [BvES99], and [HB98]. The agent must remain anonymous for all other actions. The technologies that can be used to secure 'anonymity' can be found in [HB98].

Other technologies can also be used to protect the privacy of a user of agent technologies. The personal data must be accurate if an agent needs to exchange personal data with its environment for the purposes of a specific action. It is also necessary that the (personal) data are kept accurate when they are received by service providers, other agents, or clones of the original agent. This means that the integrity of the (personal) data needs to be guaranteed. The integrity of the (personal) data can be safeguarded by various means, including parity bits, checksums or digital signatures. A digital signature is similar in nature to a hand-written signature. The integrity of (personal) data with a signature can be checked for authenticity.

Each time that (personal) data needs to be exchanged, a (new) unique digital signature shall be calculated. This signature shall accompany the (personal) data when it is exchanged. The verifying party needs to execute the same calculation to verify the integrity of the (personal) data. This shall result in another digital signature. Data shall be deemed authentic if this signature is identical to the signature received with the data. (Personal) data shall have been modified if they are not identical. Authenticity can be verified as long as data accompany the signature. Data that are not accompanied by a signature cannot be verified, and, therefore, need to be treated as dubious.

Service providers, other agents or clones can easily change the personal data. These parties can also easily change the parity bits or checksums. The parties involved, therefore, need to be trusted when using parity bits or checksums to guarantee the integrity of personal data. It is still easy to change personal data with a digital signature, but the signature cannot be changed as easily and, therefore, it is easy to verify the authenticity of the data.

Service providers, other agents or clones that receive personal data need to protect that data to prevent it from being unlawfully acquired by others. This topic shall be addressed when describing the measures for tackling threats where 'the agent runs into an agent that is in disguise' and 'the agent runs into an agent that is more powerful (e.g. has more processing power or capacity) than itself' in the next section.

The impact of the threat 'the agent exchanges personal data with its environment' can be reduced by using the technologies described in [HB98], integrity mechanisms or logging and auditing mechanisms. The strength of the technologies from [HB98], the integrity mechanisms and the logging and auditing mechanisms shall depend on the level of privacy protection that is required.

Malicious Agents

Agreements have to be made between friendly and trustworthy agents to avoid interaction with unwanted and malicious agents. One of these agreements needs to describe the way that agents must identify themselves to other agents. The agreed identities need to be kept secret since, otherwise, a malicious agent can filch the identity of a friendly agent and present itself as this friendly agent. If the identities cannot be kept secret, it is necessary that agents present supplementary information to authenticate themselves. The identity of the agents, in this case, no longer needs to be a secret but the information necessary for authentication still needs to be kept secret. Examples of information used for authentication are PIN codes, passwords and biometrics information. There are many options for keeping the information necessary for authentication safe from malicious agents. One-time password generators and challenge response mechanisms are examples of practically safe ways to authenticate. The threat posed by 'the agent running into an agent that is in disguise' can be reduced with identification and authentication. The impact of this threat can be reduced by logging and auditing all actions the malicious agent executes on the personal data kept in the user's agent.

More Powerful Agents

The use of identification and authentication alone is not sufficient to reduce the occurrence and the impact of the threat posed by 'the agent running into an agent that is more powerful (e.g. has more processing power or capacity) than itself'. There also needs to be a way to control the actions that other agents execute on the personal data that is kept in the user's agent. The user agent can control all activities inside itself by granting rights to other agents. This can be done with an access control mechanism.

Traffic Flow Analysis

The impact and occurrence of the threat 'agents can perform traffic flow analysis' can be minimised by using conventional security measures, such as transmitting a permanent random bit stream. This measure makes it impossible for others to analyse the traffic flow that is generated by a specific user. This measure, unfortunately, neutralises the advantage of the agent's property of mobility, namely preventing network overload. The best way for agents to prevent others from performing traffic flow analysis is to use a different pseudonym each time an external action is executed. How to use these pseudonyms is described in [HB98].

An individual needs to be protected against possible automated decisions. Every individual has the right to know on what grounds a decision has been made. The reasons for the automated decisions can be traced when necessary by logging all internal and external actions.

Protection of the Privacy Domain

Users that are connected to an internal or external network where agents are active need to be aware that these agents can approach the personal data stored in the computer system by which they are connected to the network. These users can choose not to store any personal data on their computer systems, but sometimes these data have to be stored for a smooth performance of daily duties. If so, the users need to secure their computer systems in such a way that unwanted agents cannot approach the personal data. This can be done by applying an identification and authentication mechanism, an access control mechanism and a logging and auditing mechanism. The access control mechanism helps the user to give rights to others. The logging and auditing mechanism helps the user to verify all actions performed by others that are addressed to the user's personal data. These three measures shall decrease the risks that are generated by the threat 'agents can enter the privacy domain of a user and collect whatever they want'.

Protection of Databases

Controllers also need to secure the databases that contain personal data. The risks of the threat 'agents can enter databases and collect personal data' can be decreased in the same way as the risks of the previous threat.

The measures that can be used to decrease the risks of the threats 'agents can enter the privacy domain of a user and collect whatever they want' and 'agents can enter databases and collect personal data' need to be applied to the computer systems of the parties involved and not to the collecting agents. It is very difficult to enforce measures on collecting agents because there is not always a direct relationship between a user, or a controller, and the owners of the collecting agents. The owner of an agent can have questionable intentions. All parties can draw up an agreement about how to handle personal data if there are no questionable intentions. Agents need to have a logging and auditing mechanism, as well as the agreement, to preserve personal data transparency. Transparency means that the agents need to inform the persons concerned about the type of personal data they collected and when and where they collected it.

Protection of the User's Agent

The same technologies can be used to protect the user's agent from dubious agents as the technologies that can be used to reduce the risks of the threat posed by 'the agent running into an agent that is more powerful (e.g. has more processing power or capacity) than itself'.

4.2.2 PET Placed in the Generic Agent Model

The evaluation and certification of the working method needs to be executed in compliance with an internationally agreed evaluation and certification scheme to receive an internationally accepted certification. The working methods can be certified by independent organisations, such as the College bescherming persoonsgegevens.

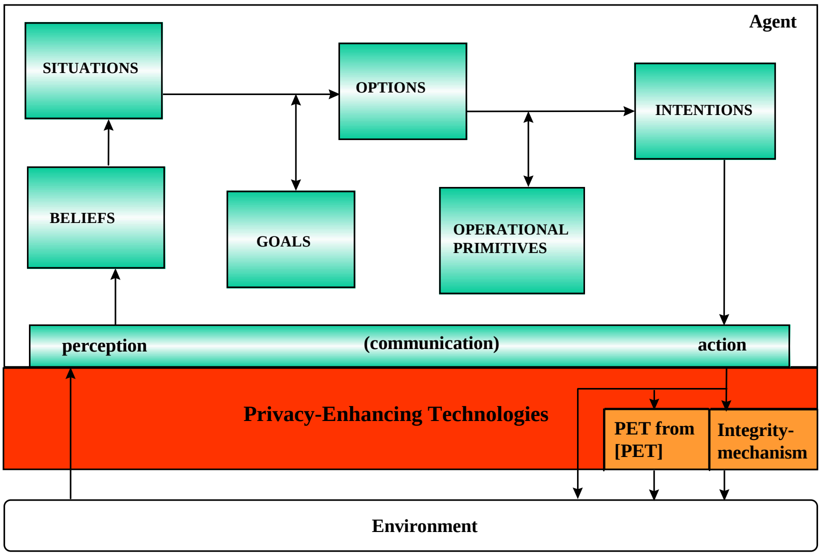


Figure 4.6: PET wrapped around the agent.

The security measures that are used to enhance the privacy of agent users can be applied to the agent in many ways, although they are subject to two limitations. The design extremes are either wrapping the Privacy-Enhancing Technologies around the agent or the total integration of these technologies into the agent. Every combination of integrating PETs and layering PETs is possible between these two extremes. The two extremes are represented in Figures 4.6 and 4.7.

The wrapping of PETs around the agent can be compared with placing the IP between the agent and the environment, which has been illustrated in Figure 4.5(b). The integration of PETs in an agent can be seen as the integration of an IP in an agent. The wrapping of PETs around the agent can have certain advantages. One of them is that a user can separately buy a relatively cheap agent and PET modules (PET tools) containing only the specific protection functionality that is required by the user. This is in contrast to the PET integrated variant, where the user has to deal with the protection functionality that is put into the agent by the manufacturer. This could mean that the privacy protection functionality of the agent differs from the functionality requirements the user has. The PET integrated agent shall also be relatively expensive.

A disadvantage of wrapping is that only external activities of the agent can be logged and audited. A combination of wrapping and integration of PETs using the correct ratio could provide an agent with the advantages of both wrapping and integration.

4.2.3 Alternative Use of Privacy-Enhancing Technologies

There could still be individuals who cannot afford to wrap or integrate the abovementioned PETs around or into their agents. These individuals must have an alternative way of protecting themselves. This can be done by using PETs to create an infrastructure of components that can be trusted within the environment. This infrastructure of trusted components must handle privacy sensitive data (personal data) in a manner that respects privacy.

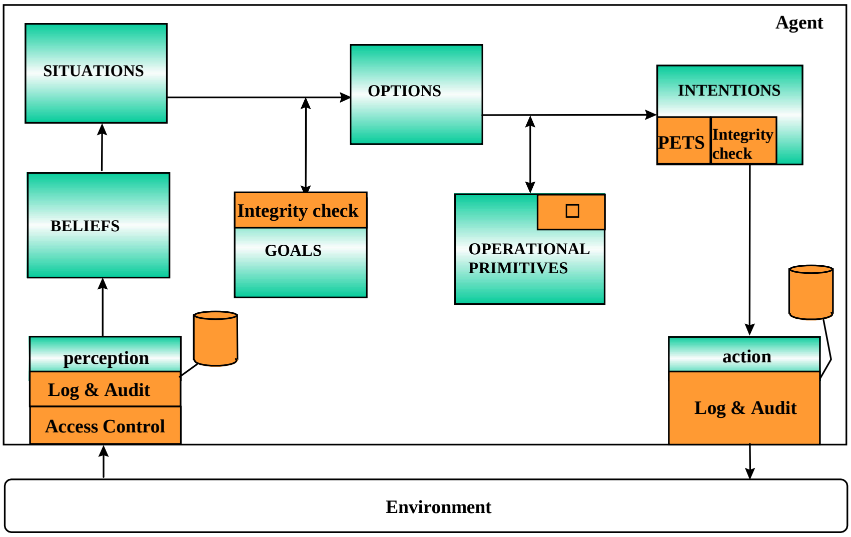


Figure 4.7: PET integrated in the agent.

The trusted components are computer systems that are protected by security products, and PETs, consisting of, for instance, identification, authentication, integrity, logging and auditing mechanisms. The trusted components need to be evaluated. These evaluations shall lead to the certification of these components with an indication of the guaranteed level of trust. Individuals who cannot protect their agents with PETs can still interact with other participants without giving up their privacy with this alternative.

How does this alternative work? An agent can directly interact with other agents which may be unknown, not to be trusted or may have the risk whereby it reveals its owner's personal data. An agent can also indirectly interact with other agents by using a trusted component and ,thus, protect the personal data of its owner. The agent shall move to the nearest trusted component, identify and authenticate itself to the trusted component and authenticate the trusted component in order to do so. The movement of the agent is visualised in Figure 4.8 as the agent putting itself in an envelope and going to the trusted component.

After the mutual authentication, the agent shall ask the trusted component to execute the intended activities with the other agents (which could be unknown or not trusted). The trusted component shall execute the desired activities in a privacy secure manner and shall come up with the results.

The agent that wants to use the services of the trusted components needs to subscribe to a participant list, which may involve payment of a subscription fee.

The functionality of the trusted component infrastructure is not sufficient to secure the privacy of its users. The functionality needs to be implemented in a safe manner. This is also applicable for the use of PETs in general.

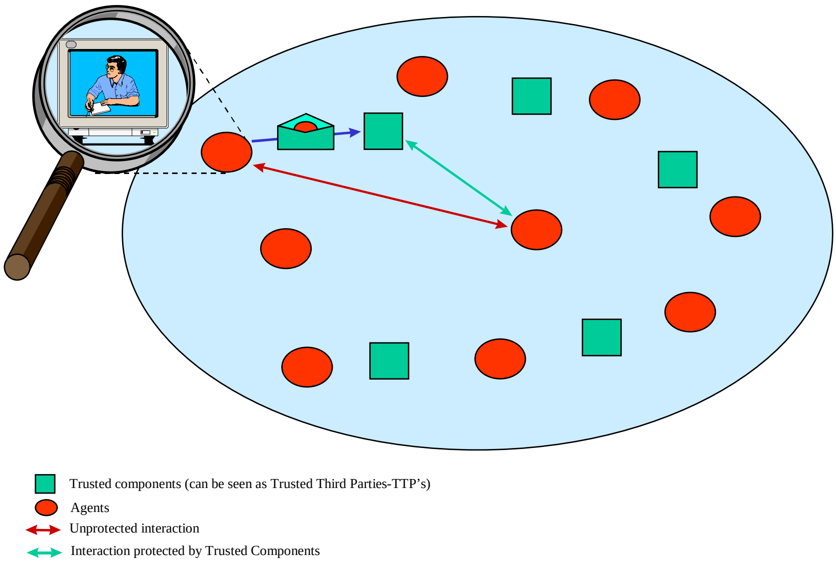

🟩 Trusted components (can be seen as Trusted Third Parties-TTP's)
🟥 Agents
↔ Unprotected interaction
↔ Interaction protected by Trusted Components

Figure 4.8: PET used to create trusted components.

4.2.4 Consequences of Using PET

There is no need for measures to guarantee privacy when there is no need for privacy. When there is, however, a need for privacy, at least all the agents that a user agent wants to communicate with need to have a unique identity. They must also authenticate themselves. The agents, therefore, must be registered. There are different ways to register the agents. One of them is to let the agents handle an authentication table themselves. When an agent wants to communicate with another agent, it identifies itself, and both agents shall add the new identity to an authorisation table. Another way to register could be that a new agent needs to apply to a mutually recognised party that shall add the new agent to a participant list. This list, in turn, is distributed to all attending participants. The next step that can be taken when this is not enough because more privacy is required, is to make every participant known in the environment. This means that the environment shall be limited to the group of participants listed in the participant list.

The measures taken need to be very strong is users set high requirements for the protection of their privacy. It also means that the measures need to be enforced by the agent software. It must, therefore, be impossible for the measures to be bypassed. It is uncertain whether the measures can still be enforced when the agent is sent to different places or if the agent is cloned. The following would apply when users demand strong protection of their privacy:

- Mobility of the agent is not allowed;
- Cloning of the agent is not allowed;
- The use of agents that are provided by an agent provider is not allowed;
- The performance of the agent shall be reduced, because properties, such as mobility and cloning, are not allowed;

- The costs of an agent that satisfy the requirement for strong protection of privacy shall be high.

Privacy consciousness shall also limit the advantages of the agent's properties. A user who wants strong privacy protection can only make use of an agent that is expensive and dedicated to a specific computer system.

4.2.5 The Supply of PETs for the Consumer Market

There are three options for supplying PETs to the consumer market. These are to supply:

- Agents in which PETs are integrated;
- PET tools to wrap around an unprotected agent;
- An infrastructure of trusted components.

The supply of PET integrated agents could provide a complete solution for the future user, especially if the agent is designed according to the user's requirements or if the design corresponds with the user's requirements. These agents shall be relatively expensive.

PET tools or trusted components could offer a better solution for individuals who use agents with no privacy protection. Organisations that develop security tools or agents must actively start developing PET tools. These tools can be used as an IP to help individuals protect their privacy. Everybody shall be able to put together their own privacy protection agents according to their own specific privacy requirements when these tools are available.

The development of trusted components shall also help individuals to protect their privacy, but this is slightly more difficult than the development of PET tools. The development of a trusted component infrastructure calls for a network-wide approach. This could lead to a nationwide, or even a worldwide, approach. If such an approach is needed, a lot of time and money shall have to be spent on setting up mutual agreements between all participating parties. Although this is a time-consuming and costly process, it shall be a good alternative for individuals who want to protect their privacy in an agent-based environment. A subscription to a trusted component infrastructure shall be less expensive than the procurement and maintenance of a PET integrated agent or PET tools.

4.2.6 PET Design Criteria for Agents

As mentioned before, the four ways of using PETs to protect an individual in an agent-based environment are:

- Wrapping PETs around the individual's agent;
- Integration of PETs in the individual's agent;
- Combining wrapping and integration of PETs;
- Using PETs to create an infrastructure of trusted components.

User, designers, developers, suppliers or providers of agents can ask themselves how the privacy of users and all other individuals involved can be protected. A checklist of considerations during the different phases of the design process is available to provide assistance with this issue.

It must become clear, during the analysis phase of the development of an agent, whether the agent shall collect and handle personal data of both the future user and other individuals. The personal data of the future user (user profile) must be protected with proper PETs. In addition, collection of personal data of other individuals, particularly identifying data, must be minimised in accordance with the privacy regulations described in Section 4.1.

During the design phase, the way PET shall be used needs to be defined. Decisions need to be made about whether to integrate or wrap PETs around the agent. The use of trusted components also needs to be considered.

Deciding which specific techniques can be used shall take place during the implementation phase. The main issue is that the agent must not allow personal data to leak from the user profile or other internal resources into the environment without having given permission.

Figure 4.9 indicates how the designer can take the privacy of everyone involved into account during the different phases of the design process.

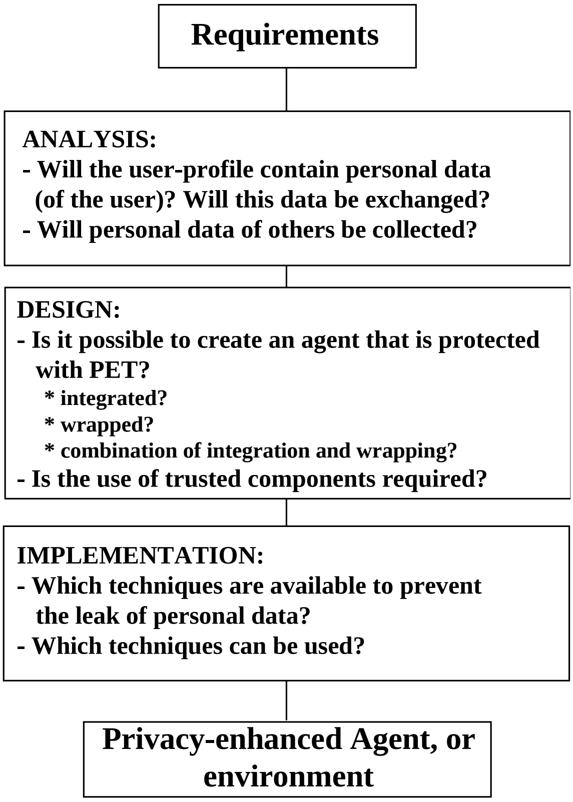

Requirements
ANALYSIS:
- Will the user-profile contain personal data (of the user)? Will this data be exchanged?
- Will personal data of others be collected?
DESIGN:
- Is it possible to create an agent that is protected with PET?
* integrated?
* wrapped?
* combination of integration and wrapping?
- Is the use of trusted components required?
IMPLEMENTATION:
- Which techniques are available to prevent the leak of personal data?
- Which techniques can be used?
Privacy-enhanced Agent, or environment

Figure 4.9: Aspects to take into account during the different phases of the design process of a privacy protecting agent.

Chapter 5

Providing privacy to agents in an untrustworthy environment

K. Cartrysse J.C.A. van der Lubbe
K.Cartrysse@ewi.tudelft.nl J.C.A.vanderlubbe@ewi.tudelft.nl
Delft University of Technology, The Netherlands

Agent technology gives many promises for future IT-systems, but providing security and privacy in particular is a difficult challenge. This chapter addresses the privacy problems in agent technology and provides several solutions to overcome some of these problems. It is shown that if an encryption algorithm is used in a different form it can also provide privacy in communication for agents. Second, an agent digital signature is described that takes into consideration that the agent's private key must be seen as the most privacy sensitive information of an agent. Finally, using information theory, a look is taken at the theoretic aspects of agent technology to find boundaries to what level of privacy and security a mobile software agent can be protected.

5.1 Introduction

Agent technology is nowadays seen as one of the technologies that will play a key role for future IT- applications. As this technology evolves, awareness of security for this kind of technology is increasing. However, solving the security problems is a difficult challenge and adding privacy makes it even more difficult. To get an impression of the challenge, imagine an agent, owned by a user, that is sent out to be executed at some foreign host. Nothing may be known about the foreign host, therefore the agent must be well protected to make the user feel comfortable in sending his agent off to such a possibly dangerous location. Compare it to a situation where a person takes its most private possessions to a location with a high chance of being attacked. The person must be very well protected and he/she must be aware of that protection to survive in such an environment. In the digital world the fear may even be bigger as one is not physically present where the agent (including the user's personal data) is processed.

From the above example, it is clear that not only an agent's security should be guaranteed, but also its privacy such that the user's personal data is protected. Especially the privacy aspect of agent technology is an area where not much research has been done and even the security problems are not completely solved yet. This chapter describes these privacy problems and gives several solutions that contribute to provide privacy in agent technology. All of this is done in the context of cryptography. Shortly it can be stated that the main difference in providing privacy to a software agent and conventional it-systems is the fact that the execution environment does not belong to the user.

5.1.1 Related work

Over the years much research has been done in the area of privacy in conventional it-systems and many good solutions have been presented. The term PET (Privacy-Enhancing Technologies) is used to describe all types of technologies that provide privacy to a user [HB98]. Typical cryptographic techniques that can be called PET are blind signatures [Cha82, Cha92, CPS94], partial blind signatures [AF96] and pseudonym systems [LRSW99]. Each of these techniques have their own applications but they are all based on the assumption that the computers where the computations are performed can be completely trusted, which is not the case in a mobile agent system. PET are mainly used in applications where the privacy aspects determine the success of the product. An application where privacy is of great importance is electronic voting. Most electronic voting schemes make use of blind signatures. A blind signature allows one to sign a message without being able to read the content of the message. In the electronic voting application this means that the voting committee signs the vote to declare it a legitimate vote, but it is not able to view who the voter voted for. Hence, by using blind signatures anonymity can be provided. In electronic cash application [Cha92, Bra94, CFN90] a partial blind signature [AF96] can be used to provide anonymity, in the same sense as in electronic voting, but here the amount of money should not be blinded, only the user's name should be. A partial blind signature has this property where the amount of money can be public, but the name is kept anonymous. A fair blind signature [SPC95] gives the possibility in case of dispute to reveal the connection between the message and signature.

Then, there are other cryptographic techniques that may be of importance when privacy must be provided in a system. Zero-knowledge techniques [vdL98, TW02] allows one to proof the knowledge of something without actually providing the secret. Using zero-knowledge techniques it is possible to proof knowledge about a password without providing it. A second useful concept to provide PET may be secret sharing schemes [TW02]. A (t, w)-threshold scheme is a method of sharing a message M among a set of w participants such that any subset consisting of t participants can reconstruct the message M, but no subset of smaller size can reconstruct M.

Over the last few years when PET were evolving rapidly, mainly new techniques were invented in the area of network privacy. Examples are the Mix network [Cha81b], onion routing [GRS96, RSG98] and the crowds system [RR99, RR98a].

Many solutions have been proposed to protect the user's privacy. However, there are several drawbacks. First, the solutions described above, all have the assumption that the computations take place on a completely trusted computer. This is an assumption that cannot be made in agent systems if full benefit is to be taken from agent's characteristics. A second drawback is that all these techniques provide privacy to the user's identity (blind signatures) or to privacy sensitive data (zero-knowledge techniques, secret sharing schemes), but they do not provide privacy to one's actions, which is necessary in case of agent systems. Nevertheless these privacy techniques will proof to be useful in the context of agent technology.

Next to PET for conventional it-systems, many security techniques have been invented for mobile software agents to protect them from malicious hosts [Ng00]. Several schemes have been described to provide integrity of partial results [Sta03, KAG98]. In [Sta03] a solution based on PRAC (Partial Result Authentication Code) is given. A PRAC provides forward integrity of the agent's partial results. Forward integrity means the results obtained at the previous hosts cannot be modified. An improvement of PRAC is given in [GRS96]. An second method to provide computation integrity is by identifying a trusted host. If a host is trusted, computation integrity is ensured, in other cases not [Rot98, Sch97]. A different approach to provide integrity checks of code comes from the field of watermarking [CT98, CT99]. A special data structure is embedded in the program such that even after the execution of the program the watermark can be detected by the user and it makes it possible to detect any malicious modification of the program. Many schemes provide accountability in the sense that afterwards the computations can be done. In [Vig98] and [BMW98] execution traces are computed by the host and they can be verified later to detect whether suspicious computations have taken place. Farmer et al [FGS96] described a method to check the state of the mobile agent for inconsistencies.

Many articles have been written about confidentiality, and most of them define confidentiality for agent as protecting their code such that it is impossible to determine the agent's strategy. Hohl [Hoh98] described a mechanism called "time limited black box security", where the idea is to obfuscate the source code such that it takes more time to understand the code than the programmed time limit. A more cryptographic method is presented in [ST98] where Sander and Tschudin encrypt functions that can be executed in its encrypted form. This method works for polynomials and rational functions [FSW]. Young et al. [SYY99] extended the results to all functions computable by circuits of logarithmic depth and further generalized to arbitrary functions, provided they can be represented by a polynomial-size circuit. As far back as 1990, Abadi and Feigenbaum [AF90] described a method that provides confidentiality for circuit evaluation. A disadvantage of this method is that many interactions are required to provide confidentiality. Many other solutions have been published to provide secure circuit evaluation, but none of them is very practical and efficient. Loureiro et all. described how functions can be hided using coding theory [LM99]. Several more practical methods have been proposed, but they are all based on either trusted hardware located at the host [Yee99] or on the presence of a trusted third party [CKAC01]. Next to protecting the agent's code, also some data the agent receives must be protected against eavesdropping. A method is to use sliding encryption [YY97]. It provides encryption for small amounts of plain text resulting in small amounts of cipher text without loss of security.

5.1.2 Chapter outline

An agent interacts with its environment and during communication it must be possible to provide confidentiality. Most multi-agent systems use a protocol similar to SSL [Sta03], but that means that the host is capable of eavesdropping the communications. As far as we know, no method has been proposed yet to have secure communication on an agent platform without the host being capable of eavesdropping. In this chapter a solution will be given that under certain conditions provides this confidentiality. The details are described in [CvdL02b].

A second mechanism that is not provided yet for agent technology is a privacy protected signature scheme, which is an agent signature that cannot be forged by the host. This is not trivial as an agent must use a private key in order to sign a message. This process is open to the host and can therefore copy the agent's private key. A private key is the agent's most personal information as it represents its identity.

Summarizing, it can be said that many PET have been developed, but for conventional IT-systems and these cannot directly be used in agent technology. On the other hand for mobile code and mobile agents, many security techniques are present, but most of these techniques hardly consider privacy. This paper addresses both concepts and gives an overview of the various cryptographic mechanisms as they are developed during the PISA-project [CvdL02a]. It covers several open problems of security and privacy in agent technology such as confidential communication, an agent digital signature and a theoretic approach how much protection an agent can be given.

The remaining of this chapter is organized as follows. In Section 5.2 the general agent model is presented as it is used during the PISA project. This model differs in assumptions made from the more practical approaches, hence it must be defined here. It also gives an overview of the various threats an agent must overcome. Section 5.3 provides solutions to the problem of secure communication and providing agent digital signatures. In Section 5.4, one step further is taken. A study is described that provides some theoretic boundaries on the protection of an agent. Finally, Section 5.5 gives conclusions and recommendations for future work.

5.2 System model

Several assumptions must be made about the agent system before the solutions can be described. This is followed by a list of possible threats to an agent and its user.

5.2.1 System model and assumptions

Figure 5.1 gives a general overview of the system elements. The colours indicate the assumption of the level of trust for each element. A user owns an agent platform on which he can launch his own agents. The user has complete control over these elements and therefore they can be trusted completely (green in Figure 5.1).

Once the agent is initialized it can travel over the network to different platforms on foreign hosts. Beforehand it is not known whether these hosts are trustworthy or not. Hence, these are considered untrustworthy. However, for these foreign platforms untrustworthy means that they are interested in the actions and data of the agent, but they do not actively attack the agent by altering its code. These hosts execute the agent correctly and can do that as many times as is required for them to obtain information about the agent's goals or data. These hosts do not change the agent's code, therefore these attackers are called passive attackers (blue in Figure 5.1).

All the other elements, users and other agents, are considered to be active attackers. They do actively try to attack the agent using all possibilities they have (red in Figure 5.1). These attacks can occur anywhere in the system, during transmission of an agent or when it is located at a foreign host.

5.2.2 Threats

Agents can either be owned by the user or by an agent provider. If an agent is owned by an agent provider some additional treats will exist. General threats that an agent must overcome are:

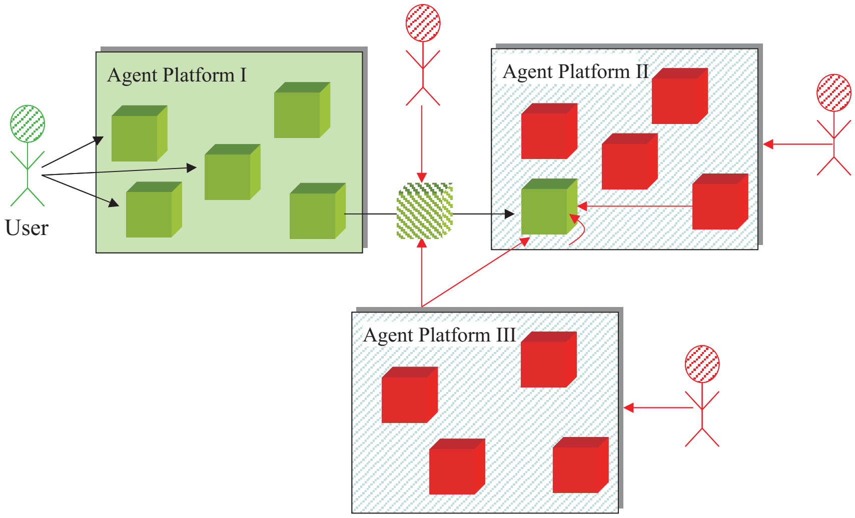


Figure 5.1: System model.

- **Altering the agent's functionality.** If the agent has a lack of security it might be possible for other entities to change one or multiple functions of the agent. A consequence can be that the user looses control over his agent, without knowing it. Two things should be considered: first it should not be possible for other entities to change the agent's functionality. Second, if the first condition cannot be guaranteed, somehow the user should be notified if his agent's functionality has changed.

- **Duplication of the agent.** It might be possible to clone the agent without having permission for it. The cause of this threat could come from all other entities of the system. This threat might have impact on the network performance, but may also give serious problems to the user. If an agent is cloned, each new agent might perform the same tasks the agent was supposed to perform and therefore the task will be performed multiple times, which could damage the user.

- **Damaging the agent.** Malicious entities might be able to damage the agent, such that the agent is not able to function as it is supposed to.

- **Masquerading.** Other agents might pretend to be an agent they are not. By pretending this they could learn more about the original agent. This can be done by talking to agent or sites the original agent was supposed to talk to. By trying to track the agent's history by communicating with the agent's identity it might also learn more information about the user.

- **Fraud with user's identity.** In this threat a user might claim to be an agent's owner but is not. A consequence is that the agent would trust that user and could give all his information to that user or the agent could respond to orders given by the malicious user. In this case it would be possible that both user and agent loose privacy.

- **Unsecured data storage.** In all computer systems this is a serious threat, but when talking about intelligent agents it is even more serious. If data is stored unsecured in the agent while the agent is traveling around the Internet, it might be possible for other entities to have access to the agent's data. This data could be data just obtained at his search, but it could also be data about the user itself.

- **Unclear distinction between public and private data.** If the agent does not exactly know how to distinct private data from public data, it might give information to other entities, which it is not supposed to give. The agent would not even know that it is not performing well and the user would not find out his information is given to other entities.

- **Duplicating agent's actions.** If an agent is executed at an agent platform the platform might get control over the agent and could send the agent to another platform where it might perform the same function. In this case the agent would perform the same function twice, which could have negative consequences for the user.

- **Unsecured agent platform.** If the platform where an agent is executed is not secure, this might damage the agent. It might be able to change functionality or adds a virus to the agent or change the agent's data.

  If an agent provider owns the agent, an additional threat is the agent provider itself. It might not be wise to trust the agent provider at all times. Laws should be available to protect the user from malicious agent providers.

  The agent must not only protect itself against external dangers, but as it is the interface between the agent world and its user, an attack against an agent may also be an attack directly against the user.

- **Access to the user through the agent.** If the agent is not protected well enough, it might be possible for other entities to have access to the user via the agent. This can happen either by accessing user data in the agent or by setting up a communication link between agent and user.

- **Agent is set up against user.** If a malicious entity is able to change the agent in such a way that it is set up against the user, it might give information about the user to this malicious entity. But it may also order the agent to perform in another way, as it should have done if the original user were in control. The damage could be enormous, because the agent could start doing illegal things without the user knowing it.

- **Treat coming from agent provider.** If an agent provider owns the agent, the user should be aware of a privacy threat caused by the provider.

Summarizing, many threats exist against agents and not all of them are addressed in this chapter. Based on the assumptions made not all possible threats occur in the system (e.g. damaging the agent by an agent platform) and many of these threats cannot be solved using cryptographic techniques.

5.3 Proposed solutions

Many threats exist, but with respect to privacy the two main points are that the agent should have the ability to provide confidentiality (during communication and data storage). The second item is that when an agent must sign a document it should be impossible for a host to forge the agent's signature. To these two threats solutions are proposed.

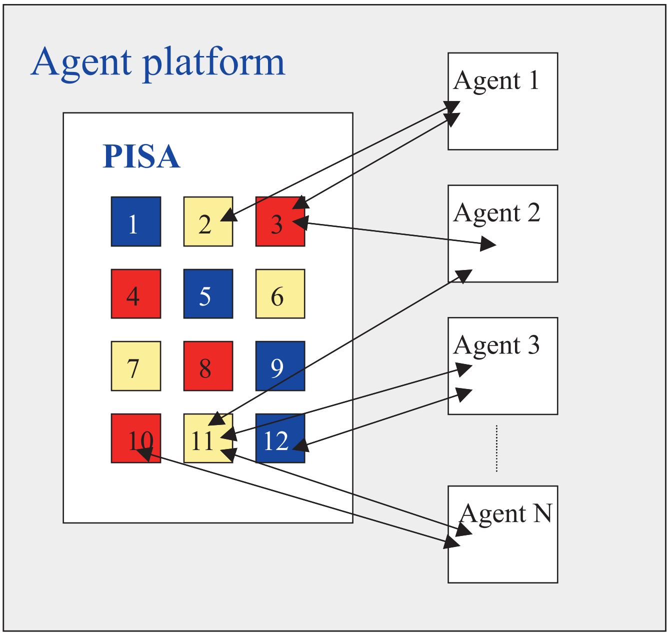


Figure 5.2: Problem with confidentiality.

5.3.1 Confidentiality

The problem with data storage that requires the property confidentiality is the fact that other parties must be able to have access to this data according to their rights, although beforehand it is unknown who these parties are and in what environment (trusted/ non-trusted) the agent operates. This is shown in Figure 5.2. The agent consists of many data blocks (the colored squares). Other agents (but also users) will at a certain moment have access to several data blocks given by the agent, but the agent should give as little data as is necessary. For example agent 1 may have access to data block 2 and 3, while agent 2 has only access to 3. Agent 2 should not be able to access block 2.

Summarizing, about the problem of confidentiality in data storage it can be said that it differs from classical confidential storage in the following ways:

- It is not known beforehand who will have access to which data.
- It is not known whether the environment, the agent is operating in, is a trusted environment.
- The data that is stored in the agent beforehand can also be stored at the users computer as a backup.

The third point makes the problem a little less complex, because in conventional systems where confidentiality of stored data is required, the data can nowhere be stored in the clear, hence if something goes wrong with the key the data is lost. This is not the case in an intelligent software agent system, where it is also possible to make a backup of the data at the user's computer.

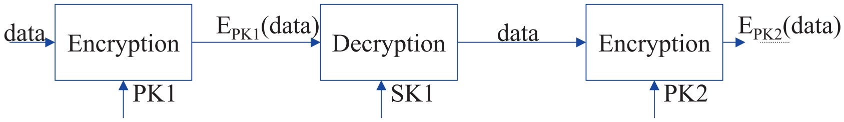


Figure 5.3: Convential mechanism to provide confidentiality for data storage and access.

It must be stated here, that the problem is not the encryption process itself, because that can occur at a trusted location (e.g. the user's computer), but the access to the data occurs at an unknown location. In order to provide a solution to this last problem, it is important to look at the complete picture of encrypting data and later accessing this data.

In this solution the data in the agent is encrypted using a key that is specific to the agent. Hence, if public key cryptography is used, the public key of the agent is used to encrypt the information and the agent's private key is used to decrypt it. In the case where symmetric cryptography is used, the key for encryption and decryption is the same and must be kept secret. In Figure 5.3 the procedure is shown for encrypting information for storage and how it can make the data ready to be sent to the communicating partner.

The first step is to encrypt the data using the agent's encryption key $PK1$ (in case of public key cryptography, this would be the public key of the agent): $E_{PK1}(data)$. This operation can be performed at the user's computer and the result can then be stored in the agent. At the moment the agent needs to send this confidential data to another party, the agent decrypts the data using its personal decryption key $SK1$ (private key in the case of public key cryptography). The result of this operation is plaintext data. Then the data is again encrypted, but this time the communicating partner's encryption key $PK2$ is used: $E_{PK2}(data)$. Now the data is ready to be sent to the communicating partner and this entity can decrypt the data, because he has access to his decryption key (this last step is not shown in Figure 5.3). A simpler solution would be that the user encrypts the data directly with the communicating partner's key, such that the data should not be first decrypted and then encrypted again before it can be sent to the communicating partner (last two blocks in Figure 5.3). However, this is not possible, because the agent and the user do not know beforehand who they will be communicating with, so they do not know which key to use to encrypt the data. In case it is known to which parties the agent will be communicating, it would still be inefficient to use the communicating partner's key immediately for encryption, because sometimes the same data must be sent to different entities. This would mean, in case public key cryptography is used, that in case n entities need the data, it must be stored n times using n different keys.

The advantage of the solution, as shown in Figure 5.3, is that it is simple and efficient. All the data to be stored is encrypted with the same key and only when needed it is transformed to an encryption with the appropriate key. Also it is an advantage that beforehand it must not be known whom the agent will talk to, because the encryption for the communicating partner occurs at the time of communication.

This solution would be sufficient and adequate in a scenario where the agent is in a trusted environment and where privacy is not a priority. During the transformation from encryption with the agent's key to encryption with the communicating partner's key, the data is in the open for a moment in time. If the host can be trusted this should not be a problem, but if the host cannot be trusted it can have access to the data in the agent. Of course this situation should not occur. A second problem is that not only the data is at a certain

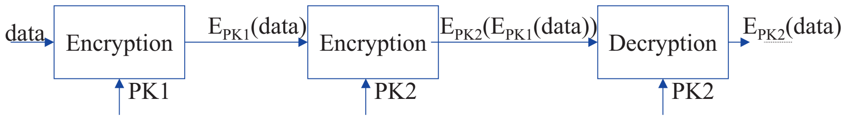


Figure 5.4: Proposal of new confidentiality mechanism for storage and data access.

moment readable to the host and maybe to other parties, but also during the decryption process the host has access to the decryption key of the agent. In some cases this might be a serious threat. For instance, when public key cryptography is used. Then during the decryption phase, the host will have access to the agent's private key. Concluding, it can be said that this is an adequate solution to the stated problem, however, only in a trusted environment.

A second possibility is presented in Figure 5.4. In this solution the data is first encrypted using the encryption key of the agent. At the moment data must be exchanged to another party, the data is again encrypted, but this time with the encryption key of the communicating partner. A decryption process follows this where the decryption key of the agent is used, such that the overall result is encrypted data, which can only be deciphered by the communicating party. To this solution will further be referred as E-E-D.

An encryption algorithm should fullfil the following expression (Equation 5.1) in order to make E-E-D possible.

$$D_{SK1}(E_{PK2}(E_{PK1}(m))) = E_{PK2}(m) \tag{5.1}$$

where $PK1$ and $PK2$ are the public keys of the agent and communicating party respectively. $SK1$ and $SK2$ are their corresponding private keys. What is important here is that the order of the functions cannot be changed, because otherwise the data (and possibly the key) might be available to anyone. Two options are possible, see Figure 5.5:

1. Second encryption and first decryption occur at agent platform,
2. Second encryption takes place at agent platform, decryption takes place at user or at some known trusted platform.

The first option is of course the most satisfying one, but here it is important that the order of operations on the agent platform is not interchanged which is as such according to the assumption in paragraph 5.2.1. In the case where it is possible for the agent platform or some other entity to change the order of operation such that the second encryption occurs after the first decryption, the situation is equal to the one in Figure 5.3, hence at some moment in time the data will be available to the platform. The second option is less efficient communication-wise, it is not preferred that the agent must contact its user (or maybe a TTP) each time it needs to exchange confidential information.

It is possible to use a standard encryption algorithm that fulfills Equation 5.1 and because the decryption key of the agent must not be stored in the agent, the order of the operations is (in the second case) guaranteed. Many public key encryption algorithms exist, but not all of them satisfy Equation 5.1. It is not possible to use algorithms based on the factorization of integers (e.g. RSA), as the modulus is the public key and these are different for each party. It is not possible to interchange the operations of encryption and decryption as the

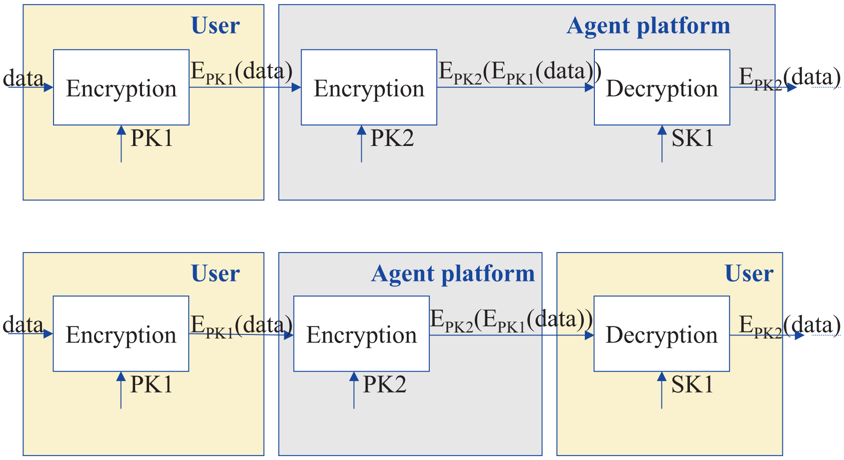


Figure 5.5: Two possibilities for the E-E-D mechanism.

modulus is different for the operations D_{SK1} and E_{PK2}, then Equation 5.1 will not hold. Algorithms based on elliptic curves and algorithms in Z_p do not encounter this problem as the modulus can be chosen equal for both parties. The encryption algorithm defined by ElGamal fulfills Equation 5.1 [vdL98]. Full details can be found in [CvdL02b].

5.3.2 Integrity

One of the security mechanisms an agent should have access to is the digital signature. During its life cycle it will need to be able to sign documents (to provide integrity) or authenticate itself by using digital signatures. The problem is that signing a document involves the agent's (or user's) private key and as said before this is the most privacy-critical information of the agent, hence in case the agent platform cannot be trusted how can an agent sign a document without the platform being able to use its private key for other purposes. In this report several solutions are proposed to this specific problem of how an agent can sign some document without anybody (except the user) can have access to the agent's private key. It can be seen as a special case of computing with secret data.

The approach taken here [CvdL02a] is different from other proposed mechanisms [LRSW99, Cha92, Cha82], in a way that not the entire function is hidden, but only the data (private key), while maintaining the property that a signature can only be set once. The advantage is that the signature and verification formulas can be used with a so-called hidden key, but also like a conventional signature using a normal private key. The idea to provide a solution is that a transformation on the private key is needed, which results in a hidden private key. The original private key is stored at the user's trusted computer and the agent only has access to the hidden private key. It is clear, that it must not be possible to calculate the private key directly from the hidden private key.

In several applications, such as electronic voting and anonymous electronic cash [Cha82, Bra94], the idea of computing with secret data is already well established, but is not always seen as such. These applications are based on blind signatures, first introduced by Chaum [Cha82].

A blind signature allows a person to let somebody else digitally sign a message, without this signer knowing the content. This process is shown in Figure 5.6. The user (Bob)

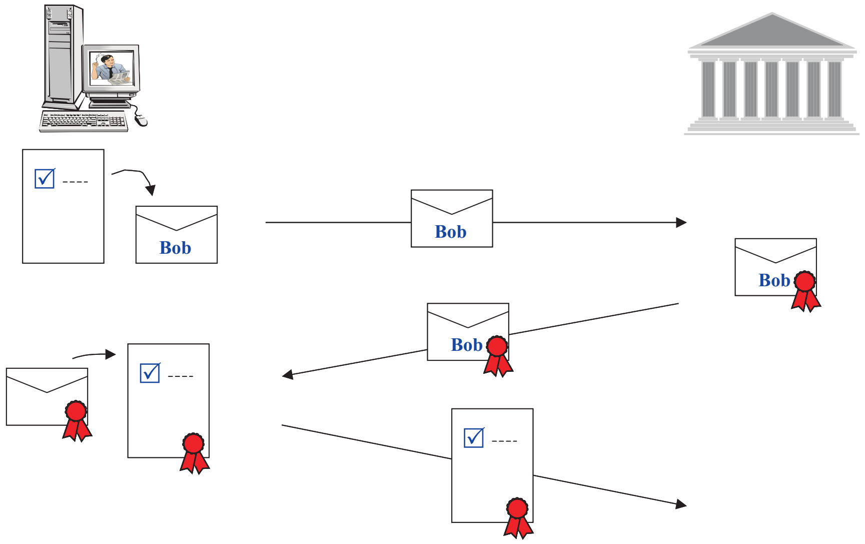


Figure 5.6: Digital blind signature.

generates a message and puts it in a digital envelope. This envelope can be seen as an encryption of which Bob is the only person to be able to extract the original content. He sends this envelope to the signing authority who signs the envelope without being able to read Bob's original message. The signed envelope is sent back to Bob who decrypts the envelope, such that the result is the original message signed by the signing authority. Everyone can verify the signature. This procedure is only possible if the following equation holds:

$$D_{KB1}(S_{SKA}(E_{KB2}(m))) = S_{SKA}(m) \tag{5.2}$$

The secret in this signature function is the original message. In agent technology this idea of blinding data can be used. Instead of blinding the message, the private key should be blinded. This signature will then exist out of the following steps:

1. Key generation
2. Blinding operation on private key
3. Signature operation
4. Activation of signature
5. Verification

Steps 1, 3 and 5 are necessary in any conventional digital signature algorithm. Step 2 is the transformation from private key into a blinded version and in step 4, the signature is activated. This step is necessary because in step 3 a signature is set using a blinded private key. This signature cannot yet be verified by using the agent's public key, because the blinded private key and the agent's public key are not related as such. Hence an activation procedure must be added. Steps 1 and 2 must be performed in a trusted environment, e.g. the user's computer. Step 3 and 4 are done at the foreign host. To activate the signature, the private key is not necessary, only one part of the private key must be used, but for the host it is infeasible to calculate the complete private key. Finally the verification can be done anywhere in the system.

Using the agent signature gives the agent privacy in the sense that its private key will not be public to a host. however, one major drawback exists. It is possible for the host to repeat actions 3, 4 and 5 but then on a different message. Now, the signature looks like an original agent signature, while the agent is not aware of it. This is a fundamental problem, if the activation is done at the host, it will always be possible for the host to repeat the signing procedure. To overcome this problem two solutions are possible. The first is that the activation should take place at a trusted location (a TTP). This limits the behaviour of an agent somewhat, but it provides more security. It may even be possible to activate the signature in a distributed way. A second option is to include the host's signature in the agent signature, such that the host's identity is revealed during verification. The advantage is that no extra communication or data transmission is necessary, however, this solution provides an extra threshold against forgery, but it is still possible for the host to repeat the signature procedure, but his identity will be revealed.

Summarizing, an agent digital signature is presented that includes providing privacy in the sense that the private key is not exposed. The mathematical details can be found in [CvdL02a]. The next step in these digital signatures would be to provide non-repudiation not only for the host but also for the agent. Non- repudiation for the agent is not provided yet, because in case a malicious host has the purpose to sell products, it can double sign an order and the agent is not capable of proving whether it gave permission for this order, or it cannot be proven that the agent gave its permission, because in this case it's in the host's advantage to have its name on the signature. A solution must be found to this problem in order to provide full functionality of a digital signature.

5.4 One step further: theoretical boundaries

Several solutions have been provided to solve parts of the security and privacy problems, but no such thing as the one solution to attack these problems does exist. In this context it is of interest to determine how much security and privacy can be provided in theory, such that boundaries can be defined and one is not searching for something that is in theory not possible. Back in 1949, Shannon described a model for secrecy systems and gave a definition for absolute security of such a system [Sha49]. Analogous to this model it is possible to develop a new model that is suitable for mobile software agent applications.

5.4.1 Shannon's secrecy model

Shannon described a secrecy model shown in Figure 5.7. A message M is generated, followed by an encryption using a key K (generated by a key source). The encrypted message (C) is sent to the communicating partner over a possibly insecure channel. The recipient decrypts the cipher text by using key K and obtains the plain text M. The exchange of the key must be over a secure channel. Important is that the locations where computations are performed can be completely trusted. An attack can only occur when the encrypted data is in transmission. The objective of the attacker is to obtain the plain text and/or the key given the cipher text.

Shannon defined the term perfect secrecy as the level of security where the cipher text does not reveal anything substantial about the plain text and/or key. Even with unlimited amount of computation power it is not possible to break a perfect secure system. A consequence of the definition of perfect secrecy is that a sufficient condition to obtain perfect secrecy

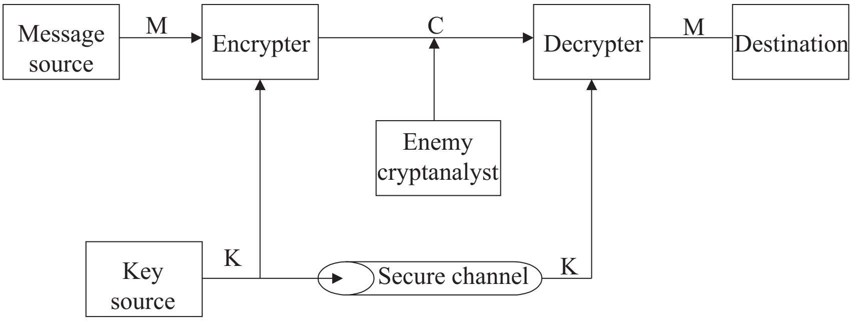


Figure 5.7: Shannon's secrecy model.

is that the key length should be as large or greater than the plain text length, under the assumption that the key is a random sequence of symbols.

It is possible to represent the different parameters in terms of uncertainty (entropy). Let $M_I, i = 1, \ldots, n$ be the possible plain text and $p(M_i)$ is the probability of occurrence of the possible plain text, then the entropy of $M, H(M)$ is defined as [Sha48]:

$$H(M) = -\sum_{i=1}^{n} p(m_i) \log p(M_i) \tag{5.3}$$

Equivalent expressions can be given for the entropy of the cipher text and key. The following equation holds:

$$H(K/M, C) = H(K/C) - H(M/C) \tag{5.4}$$

The entropy $H(K/C)$ gives the uncertainty about the key K when the cipher text C is given. Equation 5.4 can be seen as a cryptographic dilemma. From a designer point of view it is the objective to maximize $H(K/M, C)$, $H(K/C)$ and $H(M/C)$ because then it is infeasible to obtain the key (whether the C is known or M and C) or the plain text (when the cipher text is known). However, due to the correctness of Equation 5.4 $H(M/C)$ has a negative influence on the value of $H(K/M, C)$. A larger value for $H(M/C)$ will result in a smaller value for $H(K/M, C)$, hence a dilemma exists. A practical consequence is when $H(K/C)$ and $H(M/C)$ are optimized it should be practically impossible to obtain a plain text-cipher text pair, because $H(K/M, C)$ will be small.

5.4.2 Mobile agent secrecy and privacy model

Similar to Shannon's model, a new model can be developed that represents the situation for mobile agents or mobile code in general. The model is shown in Figure 5.8 [CvdLL02]. A mobile agent must protect a task T. This can be achieved by encrypting the task using key PK. The result is an encrypted task $\widetilde{T}$, which is stored in the agent. The agent is then sent over a public channel to an untrustworthy environment where it is executed by a host. For the execution some input, X, is necessary from the host. The host must first encrypt its input before the encrypted task can be executed. It will be explained later why the input must encrypted too.

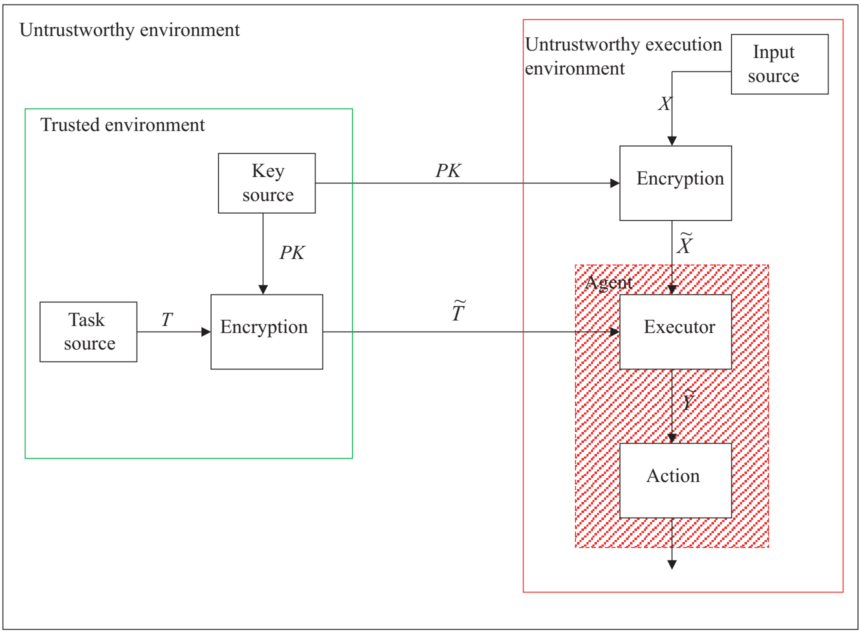


Figure 5.8: Model for agent environment.

The function $\widetilde{T}(\widetilde{X})$ is run and the outcome $\widetilde{Y}$ is some encrypted value that determines the agent's next action. An important aspect is that no decryption takes place. Imagine a system where $\widetilde{Y}$ must first be decrypted before the agent can determine its next action. The decryption can take place in the untrustworthy environment, but then it will be possible for the host to decrypt the task itself, because the decryption key will be available at the host. Therefore, the agent must leave the untrustworthy environment and move to a trusted one where it can perform a decryption (e.g. the user's one). This results in a limitation of the agent and more interactions on the network. Considering the drawbacks of these two options, the best solution is to have no decryption at all, but then it must be assumed that it is possible to determine the agent's next action based on an encrypted value without using a key that makes it possible to decrypt the complete task. Research is necessary to see whether this is possible in an efficient way.

It must be stated that the key used for encryption is not equal to the one for decryption. Public key cryptography [vdL98] makes this possible. Parameter PK represents the public key and K is the parameter that makes correct decryptions possible. K is the private key such that the following holds:

$$D_k(\widetilde{X}) = X, \tag{5.5}$$

$$D_k(\widetilde{Y}) = Y, \tag{5.6}$$

$$D_k(\widetilde{T}) = T, \tag{5.7}$$

where $D_k(c)$ is a decryption using private key K. In this model one must see a task as a sequence of mathematical functions that can be executed.

Many similarities exist between the agent model and Shannon's original model, however, the differences result in a more complex system. First, a large difference is that the host (e.g. the execution environment) cannot be trusted. This means that the host can execute the encrypted task and observe the input-output relations.

A second difference, connected closely to the first, is the number of attackers. Here, attackers can be present at two locations in the system. The first is the conventional attacker that eavesdrops the communication channel. He can observe $\widetilde{T}$ and his objective will be to determine T and/or key K. The second attacker might be an agent on the platform that is able to observe the agent's behaviour and by doing that it might be possible to determine the details of the task. Note, that when this second attacker obtains the knowledge that the agent must buy a flight ticket, this may not be seen as a successful attack. The attack is successful if the attacker obtains the criteria on which an agent decides to purchase the ticket. A third attacker is the execution environment itself. This is the most powerful attacker. It may observe the parameters $\widetilde{X}, \widetilde{T}$ and $\widetilde{Y}$, but it is also able to execute the agent as many times as it desires. As is stated in Section 5.2, the assumption is made that the host is a passive attacker, where it does not alter the agent's code. Obviously, if protection can be guaranteed to this last attacker, protection is guaranteed to the other two attackers too.

Finally, a third difference between the new model and Shannon's model is the objective of the attacker. In Shannon's model the objective is to obtain knowledge about the key and/or the message. In the new model the attacker may have a third objective besides obtaining the key and/or the plain task. This additional objective might be the determination of an optimal value for X. It may be of interest to the attacker inside the execution environment to determine an X-value given a desired $\widetilde{Y}$-value. For example, an agent purchases a flight ticket if the price is less than EUR 500,-. If the input X (e.g. price) is then EUR 499,99 this can be considered as the optimal X-value as this is the highest price an agent is prepared to pay.

The fact that three types of attackers are present and that they can have three objectives, makes the system more complex than Shannon's model. A consequence is that the definition for perfect secrecy must be redefined and will be more complex.

Perfect secrecy in mobile code/ agents is defined as follows. The following equations must hold to obtain perfect secrecy in tasks of mobile agents:

$$H(T/\widetilde{X}, \widetilde{Y}, \widetilde{T}) = H(T) \tag{5.8}$$

$$H(X/\widetilde{Y}, \widetilde{T}) = H(X) \tag{5.9}$$

The first condition is about the secrecy of the task and states that the uncertainty about the task should not change whether $\widetilde{X}, \widetilde{T}$ or $\widetilde{Y}$ is known. In other words, T should be independent from $\widetilde{X}, \widetilde{T}$ and $\widetilde{Y}$. The second condition deals with the repetition attack, e.g. the attack performed by the host where it keeps executing $\widetilde{T}$ using each time different inputs X to obtain the optimal value. According to Equation 5.9 it should be just as hard to find the optimal X-value when $\widetilde{Y}$ and $\widetilde{T}$ are known as when these parameters are not known. Equation 5.9 also implies that X should be independent from $\widetilde{X}$. This is important, because this independence makes it impossible for the attacker to reverse the process. For a given $\widetilde{Y}$ it cannot compute X, due to the independence between X and $\widetilde{X}$. This is also the reason why an encryption of X is absolutely necessary. In case the encryption of X is perfect secure, it will still be possible for the host to compute $\widetilde{X}$ from a given $\widetilde{Y}$, but due to the encryption the host cannot compute the corresponding X.

Similar to the perfect secrecy definition for agents, the cryptographic dilemmas for agents are more complex than the original one in Shannon's model. The following cryptographic

Table 5.1: Comparison between Shannon's model and the mobile agent's model.

Shannon's model	Mobile agent's model
One type of attacker: during transmission	Two types of attacker: during transmission and execution
Two goals for attacker: determination of M and/or K	Three goals for attacker: determination of T, K and/or X
Model holds for both symmetric and asymmetric encryptions	Asymmetric keys must be used

dilemmas exist in the mobile agents model [CvdLL02]:

$$H(K/X, T, \widetilde{T}, PK) = H(K/X, \widetilde{T}, PK) - H(T/X, \widetilde{T}, PK) \tag{5.10}$$

$$H(T/X, \widetilde{T}) = H(T/\widetilde{T}) - H(X/\widetilde{T}) \tag{5.11}$$

$$H(K/X, \widetilde{T}, PK) = H(K/\widetilde{T}, PK) - H(X/\widetilde{T}, PK) \tag{5.12}$$

The proofs are given in [CvdL04]. Equation 5.10 is similar to the dilemma in Shannon's conventional model. Increasing the uncertainty about the task given the encrypted task and X will result in a decrease of the uncertainty of the private key K given X, T and $\widetilde{T}$. However, in the ideal case it is necessary to optimize all three of the entropies in Equation 5.10. The second dilemma is the repetition attack versus the confidentiality of the task. To obtain a maximum for $H(T/X, \widetilde{T})$, it will not be possible to have also a maximum value for $H(X/\widetilde{T})$, e.g. a trade-off must be made between securing the task against attacks where X and $\widetilde{T}$ are known, and securing X when $\widetilde{T}$ is known. Finally, the third dilemma represents a trade-off between the protection of the private key K and the optimal value X. Increasing the uncertainty about X when the encrypted task is given will result in a decrease of the uncertainty of key K when X and $\widetilde{T}$ are known.

5.4.3 Evaluation

As was described in the previous paragraph several differences exist between Shannon's conventional model and the new mobile agent model. Table 5.1 shows a summary of these differences.

The first difference is that in the new model more attackers are present, but these attackers can also have more objectives. The fact that in Shannon's model the attacker can only have access to the cipher text when it is in transmission gives much less security risks than the fact that the encrypted task can be intercepted during transmission, but also attacks can take place during execution. The attacker has more methods to attack the mobile code as it has full control over its own execution environment.

The second difference is the objective of the attacker. If the attacker only has the objective to determine the plain task and key (equivalent to the objective in Shannon's model), the security only depends on the strength of the used encryption scheme and its correct usage. However, in this case, the attacker is also interested in finding an optimal input X. In this model the optimal input is not protected by a security function and therefore it may be necessary to put strong requirements on the space of X.

Finally, it can be shown that it is not possible to achieve secrecy of tasks using symmetric cryptography. It is required that the executor who delivers input X encrypts the input before $\widetilde{T}$ is executed. This means that the executor should have access to some kind of encryption key, but this key should not allow him to decrypt $\widetilde{T}$. To show that asymmetric keys are not possible, assume that the executor does not encrypt his input X. The originator (agent's owner) has encrypted its task using key K:

$$E_k(T) = \widetilde{T} \tag{5.13}$$

The execution of an encrypted task would then result in:

$$Y = \widetilde{T}(X) \tag{5.14}$$

and the adversary can calculate any X for any given Y by inverting $\widetilde{T}$:

$$X = \widetilde{T}^{-1}(Y) \tag{5.15}$$

obviously this should be prevented. If Equation 5.14 should not be possible, that means that $\widetilde{T}$ must be a one-way function. This is a strong assumption as this would limit the possible tasks to be encrypted and it may not even be possible to encrypt a task, such that the result is a one-way function. Therefore, the solution to prevent Equation 5.15 is to encrypt X. Let an encryption of an input X using key K be written as follows:

$$E_k(X) = \widetilde{X} \tag{5.16}$$

Executing the encrypted task gives:

$$\widetilde{Y} = \widetilde{T}(X) \tag{5.17}$$

Calculating X for a given $\widetilde{Y}$ is now only possible if the key K is known to the adversary. However, K is known by the executor to encrypt X. Hence, the key for encryption and decryption must be different and knowing the encryption key it must be impossible to calculate the decryption key. It may also be possible that there is no decryption key as a decryption is not done in this model.

5.5 Conclusions and future work

Protecting mobile software agents is difficult and by far not all problems have been solved. As was shown the main difference with conventional software systems is that execution does not occur in a trustworthy environment. In the PISA-project, some solutions are proposed to the open problems and the main ones have been described in this paper. Many threats have been identified and based on the system model and assumptions several of these threats have been chosen to provide a solution for. The system model and assumptions made in this paper were different from the model used by the other partners in PISA, because we approached the problem of providing privacy and security at the long term. In the long term it is not feasible to assume that each agent platform can be trusted completely.

The first problem to solve is to provide confidentiality in an untrustworthy agent environment such that the sensitive data is never in the clear at an untrustworthy location. To provide this, a solution is proposed where the order of encryption and decryption is reversed. An assumption is that it must be possible to compute with encrypted data if the sensitive received data must be further processed.

A second problem to solve is the privacy problem when an agent needs to sign a document. At the moment of signing a private key is used and in an agent environment this means that the private key must be revealed to the host. A solution is proposed such that this is not necessary anymore. A drawback is the possibility that the host is able to repeat the procedure but with a different message and the outcome will be a valid signature. Several methods have been provided to decrease the likelihood of this occurring.

Instead of only looking at small problems and solving each of them individually, a look was taken at the broader picture where the idea is that an agent owns a task and that must be protected. Description of a new model on agents and protecting tasks lead to a new definition of perfect secrecy and several dilemmas were derived such that the theoretical boundaries become clear.

As by far not all problems have been solved, much research must still be performed in the area. Especially in the area of computing with encrypted data in an agent environment and making decisions on encrypted data much research is necessary, because the solutions shown here depend on possibility that these solutions will become available.

Chapter 6

Public Key Infrastructure

A. Youssof
abdel.youssof@ubizen.com
GlobalSign, Belgium

R. Lachman
lachman@fel.tno.nl
TNO-FEL, The Netherlands

This chapter discusses the public key infrastructure (PKI) to be used by a PET solution for agent technology. Please note that PKI is not to be associated with encryption and decryption algorithms, PKI is simply the infrastructure for public key distribution.

6.1 Overview of architecture

The identification of threats related to the agent platform highlighted a number of important issues, which need to be addressed in the completion of this task, overall requirements imposed by the agent and PKI platforms. Issues specifically relating to the software to be developed in agent platform, include the security infrastructure to be in place, algorithms, key lengths, standards, key recovery equipment, scalability.

A CA system implements a Public Key Infrastructure (PKI) and offers authentication, integrity and confidentiality services to concerned application. These application programs aim to secure communication between agents inside an agent platform. The communication between agents may be of critical importance and may deal with confidential or very sensitive personal information, this is why securing communication is an ultimate objective of introducing a PKI platform.

The nature of the privacy incorporated software agent PKI platform dictates that a number of important issues are in need of careful consideration when the actual design phase of a PKI architecture is being undertaken. The diagram pictured in Figure 6.1 is just an architectural overview, which shows the overall layout of an agent PKI platform.

At the centre of the PKI model are the core services that will be provided. These include certification authority, registration authority, directory service, database and signing engine. These are broken down as follows:

- PKI Certification Authority includes Certificate issuance, key generation, entity authentication, certificate revocation, and signing the OCSP if implemented.

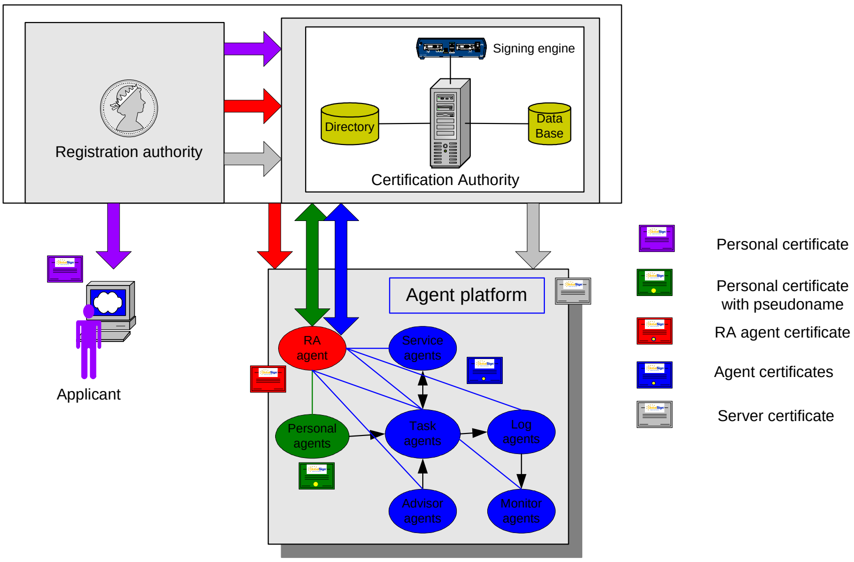


Figure 6.1: Interaction PKI-Agent platform.

- The registration authority relative to PKI is subdivided into two parts:
  - The classical registration authority: is the entity that approves the issuance of certificates after checking credentials of the requestors. The concerned PKI certificates are: end user (applicant) certificate, server certificate, and the RA agent certificate. Those are classified as offline certificates because they are issued only ones.
  - RA agent: is a set of API running on the agent platform, dedicated to manage agent related certificates.
- Signing engine allows signing certificate and CRLs. The Hardware Secure Module is FIPS level 3.
- Directory Services providing for the publication of certificates and certificate revocation Lists (CRL's).
- Database storing information about users (their passwords), certificates requests and certificates, payment if there is any, etc.

6.1.1 Architectural features required for PKI

Given the fact that the PKI's aim is to generate appropriate certificates to different agents, and distribute them according to the requirement listed before, it is imperative that some features be implemented in the design phase, prior to the actual set up of the different component. The approach that will be used here is to use a scalable structure, which will allow the system to be broken down into a series of logical and discrete components. This will allow the load to be distributed, with consequent benefits which are described below;

- Resource allocation; this feature allows to distribute the roles inside PKI platform: who is responsible for what.

- Security; the PKI platform must be located in a secure environment and accessible only by authorised persons.

- Scalability; the system must take into account the fact that the number of certificates will be very high, hence the need to facilitate an easier migration of some of it components. This approach enhances system modularity, flexibility and security.

- Robustness; this is another benefit that comes from distributing the system across a number of discrete and independent components.

- Interoperability; may be one of the long term objectives of the PKI system, and answer the following two requirements:
  - What happens in case the system is using multiple agent platforms and multiple PKI-platforms?
  - Is the system capable of using multiple agent providers and PKI providers at the same time?

Hence at the design stage, one has to keep in mind how to facilitate interoperability with other certifying authorities - this does pose a number of clear technical problems, at a number of levels, in particular when dealing with the application level, where agent end-entities may have difficulties in dealing with certificates issued by other CAs.

6.1.2 Overall Trust hierarchy

Taking into account the trust requirements, the PKI architecture will take hierarchical structure as shown in Figure 6.2.

A Hierarchical structure uses a top down approach, consisting of a Top Level CA acting as the Root CA, which in turn certifies Sub CA beneath it, called operational CA. There are many advantages associated with this form of architecture. Scalability is the key feature of a Hierarchical structure. It offers the main benefit of ease of scalability when a system needs to grow and to integrate other PKI CAs.

However, Root CA is not embedded in any browser and/or any client application. Hence the Root CA will not be recognised, and per consequent neither will any agent certificate. To overcome this problem, GlobalSign proposes a simple solution for Root CA to chain itself under GlobalSign's embedded Root Certificate. This program is called RootSign.

The diagram pictured in Figure 6.3 shows a possible global PKI hierarchical architecture. Which leads to trust and confidence for all agent certificates.

6.1.3 Description of the PKI

The PKI will be built and designed according to the architecture in Figure 6.4.

6.1.4 Design Principles

Scalability

One of the key issues of the implementation of PKI system is scalability. The key point here is the fact that significant growth of certificates must be forecasted in the future.

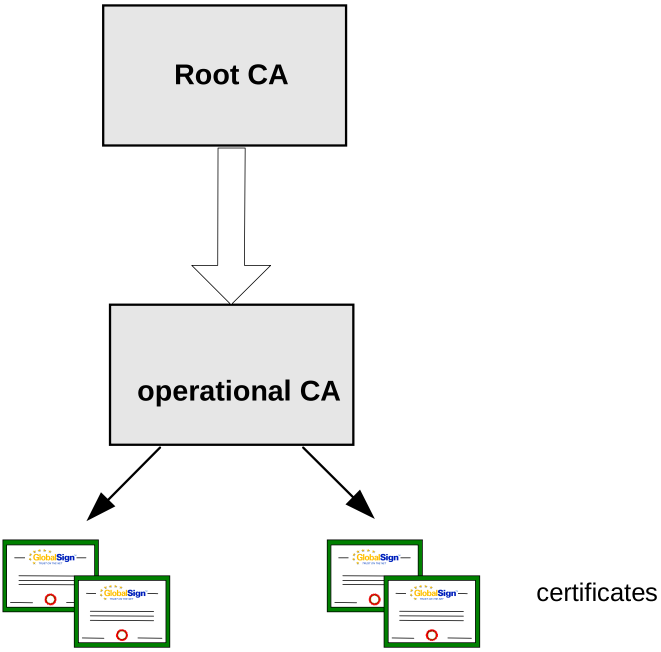


Figure 6.2: PKI hierarchical architecture.

Scalability is a feature that will mark strongly in PKI's future especially as the project will start as a small-scale pilot and eventually be built up to full-scale agent platforms.

Given that, an architecture is required which can be easily designed to scale from a small configuration, running on a single PC, hosting the CA, RA and database right up to an installation connecting to a real TTP in the future (for instance GlobalSign), and may be in the future inter-operating with many PKI and agent platforms.

Flexibility

PKI is designed to be easily extendible to accommodate agent platform specific requirements. This will be an important feature when it comes to adapting the RA system in accordance with its needs. The PKI Architecture enables, according to some requirement, to issue agent certificates. PKI offers multiple delivery mechanisms for certificates (off line and on line) and supports all popular X.509 certificate standards. PKI is designed to cope with a wide variety of PKI requirements, including: different certificate types (personal certificate, certificate with pseudoname, code signing certificate, server certificate, registration authority certificates).

Versatility

Initially PKI may use a CA system for a dedicated application (secure communication between agents). However, the CA system should be able to easily handle new applications or custom systems. PKI is designed to be modular, so that new components can be individually added, modified, upgraded or removed as those needs evolve.

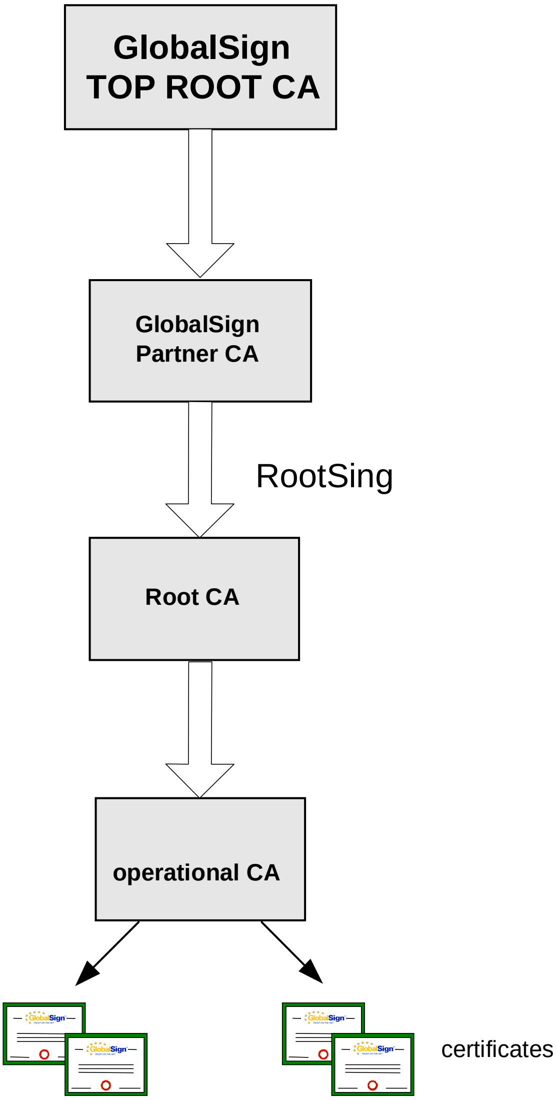


Figure 6.3: Global agent Trust hierarchy with GlobalSign's RootSign program.

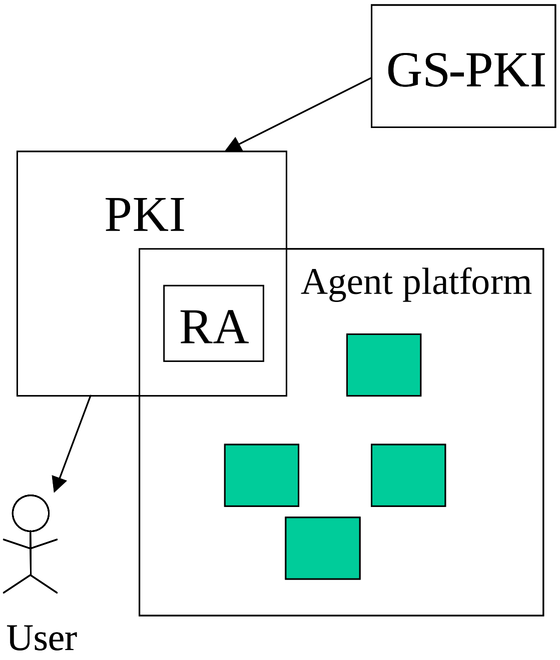


Figure 6.4: PKI Architecture.

PISA-PKI interfaces

PKI must offer configuration options for deployment of CA and RA systems. The PKI platform will provide an easy to use graphical user interface system which allows a CA operator to create/modify and configure all the needed functionalities of the PKI, and RA to create, suspend, activate, revoke certificates.

Tools

PKI must be designed to have a low-cost administration, such as for instance to enable 3rd party applications to generate keys, issue standards compliant certificate requests (PKCS#10) and to import certificates to the client application.

PISA PKI must supports the following standards and algorithms:

- X.509 version 3 certificate.
- PKCS#7: Cryptographic Message Syntax Standard
- PKCS#10: Certification Request.
- PKCS#12: Personal Information Exchange Syntax Standard. (When agent can travel)
- Publication of certificates and Certificate Revocation Lists (CRLs) to LDAP directory.
- RSA and DSA: Digital Signature algorithm.

- Elliptic Curve Digital Signature Algorithm.
- SQL for Database Management Operations.

Key length The treatment of key generation is of paramount importance to the integrity of a PKI. PKI supports the following key length: 512, 1024 and 2048 bit RSA keys.

Generation of the Root CA key pair: The CA's key pair is generated in the Hardware Security Module (HSM), in an off line way, its length is 2048 bits.

Generation of PISA operational CA key pair: The CA's key pair is generated in the Hardware Security Module (HSM), in an off line way as well, it length is 1024 bits.

Generation of RA agent key pair: It is generated in keystore of JAVA platform on which is running the agent software, its length is 1024.

Generation of other PISA entity key pairs: they are generated in the application program where the certificate request is formulated, their length is 512 bits.

Algorithm Support The following cryptographic algorithms can be implemented in PISA-PKI platform:

- Encryption: DES, Triple DES, RSA and IDEA
- Digital Signing: RSA, DSA
- Hashing: SHA-1, MD2, MD5
- Key Agreement: Diffie-Hellman

6.2 Functional descriptions of PKI

6.2.1 Certificate request

Interactions between PKI and its entities will be managed through registration authorities. Typically, two dedicated registration authorities are designed for PKI platform. The first one is a classical registration authority, and the second one is an automated registration authority named “RA agent”. The first registration authority handles the certificates that should be installed only once, i.e. server certificate for agent platform, personal certificate for applicant and admin certificate for the registration agent, while the RA agent deals only with agent certificates.

The certificate request is initiated by the entity who will own the certificate if issued. Depending on the enrolment procedures, the entity provides the RA with all the necessary identity information. Some certificate issuance involves the verification of entity's identifying document. (May be personal physical presence if the entity is a person.)

The RA initiates a Certificate Signing Request to the CA, after the receipt and verification of the user identifying data. This function takes as input the valid registration request and produces as output a binary value indicating whether the requester is really the one that he claims to be. As a consequence, a one-to-one relation is formulated between the requester and the corresponding public key. The party generating the key greatly affects the way the certificate requests are handled and on the way the certificates and the private keys are handed out to the end entities.

For the different agents, the certificate requests will be created and managed by the RA agent. The registration agent sends a certificate signing request to CA server, an identifier called cert-order-id is sent back to RA agent, which should be used in order to (download) obtain the the matching certificate.

For other agent entities, the certificate requests are handled by the classical registration authority, their private keys are generated at their side (client side), and follows the classical procedure depending of course on enrolment procedure.

All PKI certificate requests will be formulated in PKCS10 format. Bulk certificate request must also be investigated.

6.2.2 Certificate issuance

At the end of the registration procedure, the PKI platform generates a certificate and delivers it to the requesting entity. It is through certificate generation that the binding between the requester's identity and its public key is made which, in turn, is based on the appropriate user authentication and identification policies and procedures. This binding is the key to provide authentication services, during the communication of PKI entities, through the use of selective techniques (e.g. digital signatures). Certificates are signed by the operational CA, using its private key (which should be hold by signing device. This signature shows that the PKI platform vouches for the authenticity of the information contained within the certificate. A copy of issued certificate is submitted to a certificate repository (directory).

All PKI certificates will be formatted according to X509 v3 recommendations. Bulk certificate generation must also be investigated.

6.2.3 Certificate storage

PKI platform should perform management functions on the certificates they generate. These functions would effectively be performed only if PKI platform allows the storage and retrieval of all generated certificates. In addition, PKI platform should make publicly available all the valid certificates. This can be achieved by sending a copy of the certificate to a directory server.

Furthermore, the PKI platform may also want to maintain a back-up file of certificates, in case the certificates are needed for recovery. It will not be efficient for the certificate management service to undertake the publication of certificate related information. For this reason, the certificate authority has to forward the certificate list to a publication service, or directory. Entities may communicate with the publication service and get the information they want. The certificates will be stored in a directory, implementing the LDAP protocol.

6.2.4 Certificate revocation

In general, certificates may become invalid due to many reasons, among of them: validity period termination, key compromise or loss of key. Entities should always be kept informed for an invalid certificates.

The PKI platform should maintain a central repository listing the invalid agent certificates. This central repository is usually referred to as the Certificate Revocation List (CRL).

The PKI platform will periodically generate signed CRLs for the certificates that it has generated. The PKI platform will need to ensure that the information held by the CRL is as current as possible. Therefore the PKI platform should periodically update its CRLs in order to incorporate new invalid certificates.

The agent operational CA will sign the CRL and distribute it to directory for publication purpose. Functions performed during this procedure may vary depending on the CRL format (e.g. delta CRLs), the time of service with respect to the bilateral transaction between agents.

As in the case of certificates, the PKI platform will need to store CRLs it has created and retrieve them to update, to restore or to distribute them. Similar to certificate, the CRL storage and retrieval may be accomplished through the use of an LDAP directory. The PKI CRL format will be formatted according to X.509 v 2 recommendations.

Certificate revocation will be done upon request coming from RA or from an agent entity. When and how an agent can request revocation of it certificate is subject to more investigation in the future.

The certificate revocation request format must be fully compliant with the user interface used between the end entity and the TTP.

6.2.5 Certificate renewal

The CA initiates the certificate renewal. The end user will be informed that the validity of the certificate expires and he/she should request the Registration Authority for renewal of his/her certificate, or issuing of a new certificate.

In order to avoid misconceptions, the TTP must ensure that prior to a certificate's expiry date, the corresponding requester is informed about the forthcoming certificate expiration.

6.2.6 Verification

The digital signature, certificate and the clear data will be read from the message. The new hash code will be calculated again from the data. The received certificate must be verified at the CA. Form the certificate will be read the public key and with the public key, the encryption layer will be removed from the digital signature to recover the original hash code. When the original hash code and the new hash code are not equal, then the document has been modified.

6.2.7 Authentication

In an agent environment, before fulfilling any transaction, the most important basic need is assurance regarding the identity of the agent entities (applicant, agent platform, RA agent, agents). In such environments this security service is referred to as entity authentication.

Many protocols can be implemented to achieve such propriety. As we are in the context of PKI, one uses the digital signature where a sender agent must authenticate the receiver agent.

First, the sender sends a random string to the receiver. The receiver proves possession of a particular private key by signing this random message and sending it back to the sender.

The sender checks the signature by decrypting the signed message with the corresponding public key. If the signature matches, then the transaction can follow.

When the sender uses his private key to encrypt a message, the receiver can only decrypt the message using the sender's public key and is as such provided with the authentication service. This mechanism (encrypting with the originator's private key) is being used in the signing mechanism to encrypt message digests and to provide the service of integrity authentication.

6.2.8 Confidentiality

The sender of a message uses the receiver's public key to encrypt the original text into ciphertext. This public key is typically published in a certificate which itself is made tamper-proof by the CA who issued the certificate by signing this certificate (see further). This guarantees that the published public key of the sender is really the public key, which belongs to the sender (integrity of the public key + authenticity). The receiver can decrypt the ciphertext by using his private key. Using this set-up the sender and receiver have achieved the confidentiality service.

6.2.9 Non-repudiation

Non-repudiation services ensure that a certain action is undeniable (repudiation is the denial of agreements and procedures, and of techniques).

Objective

The goal of the non-repudiation service is to generate, maintain, make available and validate evidence concerning a claimed event or action in order to resolve disputes about the occurrence or non-occurrence of the event or action. Non-repudiation is different from the other security mechanisms. It does not prevent transacting parties from denying that an action or event has taken place, but it protects the parties by generating evidence to enable the settlement of (possible) disputes. Another aspect that makes non-repudiation different is the fact that it collects evidence between known legal parties instead of the protection against the unknown. Also it does not prevent attacks (by others then the legal parties) but may assist in the tracing of the attacks and indicating responsibility for protection against the attack.

Context

Non-repudiation may be implemented as application-to-application security. Non-repudiation services establish evidence. Because of this evidence, partners involved in a transaction can no longer deny having received or submitted an order or message. There are different non-repudiation services such as:

- Non-repudiation of Origin (NRO); Parties cannot falsely deny having send a message.
- Non-repudiation of Delivery (NRD); Parties cannot falsely deny having received a message.

- Non-repudiation of Submission (NRS); The proof that a delivery authority[1] has accepted the message for transmission.

- Non-repudiation of Transport (NRT); The proof that a delivery authority has delivered the message to the intended recipient.

For the agent platform is the non-repudiation of Origin and the non-repudiation of Delivery must consider.

Non-repudiation of Origin evidence will be generated by the customer agent, after the customer agent has confirmed the order; accepted all terms and signed the order form digitally. The non-repudiation of Origin service saves the 'order form' with the digital signature on the client side and sends a copy to the shop agent.

Non-repudiation of Delivery evidence will be generated after the delivery of the actual products to the customer. The customer is prompted a directory selection box to select the destination directory in which the product has to be downloaded. By selecting a directory and agreeing with signing, a digital signature of the customer will be generated by the customer. After the transaction has ended successfully, the customer's digital signature together with the checksum results evidence will be send to the shop agent.

Evidence will be generated using cryptographic techniques. There are two types of evidence depending on the cryptographic technique employed:

- Secure Envelope, which is generated using symmetric cryptographic techniques.

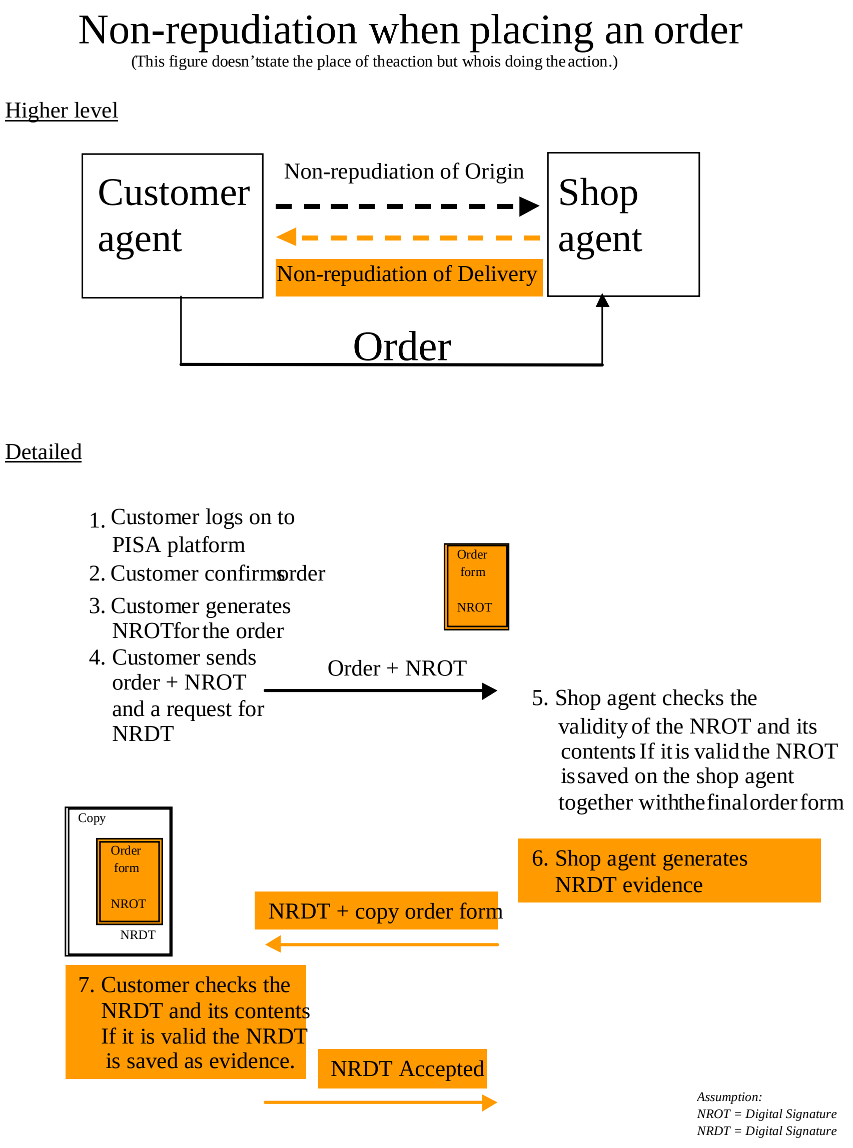
- Digital signatures, which is generated using asymmetric cryptographic techniques.

For the agent platform, asymmetric cryptographic techniques with digital signature will be implemented with a Certificate Authority (CA) being responsible for the generation of the digital certificates.

There are two different types of evidence generation: one type is generated by an evidence generator (internal) and the other by a Trusted Third Party, TTP (external). The evidence generated by a TTP is called a non-repudiation token. Every non-repudiation service has its own token. These tokens are called non-repudiation of Origin token (NROT) and non-repudiation of Delivery token (NRDT). A token consists of a digital signature and additional information. An important token, besides these two tokens, is the Time Stamping token (TST). The time stamping token offers a trusted time which is generated by a Time Stamping Authority (TSA). Time stamping tokens are additional information to the digital signature evidence. By combining the signature generation and a time stamping, strong non-repudiation evidence will be created and the overall process will be simplified.

The choice of which tokens to generate, TTP's and additional information depends on the Non-repudiation Policy in operation on the PISA platform. Figure 6.5 illustrates the Non-repudiation of Origin and the Non-repudiation of Delivery data flow when placing an order. This is an example how both transacting parties can be protected. The Non-repudiation of Origin (NROT) protects the PISA platform and the Non-repudiation of Delivery (NRDT) protects the customer.

[1] A delivery authority is an authority trsuted by the sender to deliver the data and to provide the sender with evidence on the submission and transport of data.



Non-repudiation when placing an order

(This figure doesn’tstate the place of theaction but whois doing the action.)

Higher level

Detailed

1. Customer logs on to PISA platform
2. Customer confirmsorder
3. Customer generates NROTfor the order
4. Customer sends order + NROT and a request for NRDT

5. Shop agent checks the validity of the NROT and its contents If it is valid the NROT is saved on the shop agent together withthe finalorder form

6. Shop agent generates NRDT evidence

7. Customer checks the NRDT and its contents If it is valid the NRDT is saved as evidence.

Assumption:
NROT = Digital Signature
NRDT = Digital Signature

Figure 6.5: Non repudation.

6.2.10 RootSign function

The agent platform has decided to run its own Certificate Service Provider e.g. PKI platform. However, major browsers and client applications do not recognise the agent Certification Authority, and per consequent all agent certificates. The result of this is a lack of trust and confidence.

To solve this problem, GlobalSign proposes a simple solution for CAs to chain themselves under GlobalSign's embedded Root Certificate. This program is called RootSign. RootSign is a uniquely developed service for Corporate Certification Authorities who do not have their Root Certificate embedded in popular browsers. Globalsign's Root Certificate is embedded & configured in all popular browsers and operating systems.

Chapter 7

Evaluation and Auditing

G.W. van Blarkom J.J. Borking
gbl@cbpweb.nl jborking@euronet.nl
CBP, The Netherlands

7.1 Privacy Audit Framework

The protection of personal data affects everyone. The extent of the measures to protect personal data against misuse or improper use depends on the data's contents, the amount of data, the purpose of the processing, the processing method and the circumstances surrounding the processing operations. Additional factors such as technological developments and social and personal vision also play a role. In short, a complex range of factors affects the method of implementation of the Wet Bescherming Persoonsgegevens (Dutch Data Protection Act – WBP)[1] in organisations and especially in ICT facilities.

The complexity of many aspects of the WBP requires interpretation and practical day-to-day application. This practical application is also needed in order to supervise the way in which the processors (those processing personal data) deal with and use the personal data. To put this application into practice, the College bescherming persoonsgegevens (the Dutch Data Protection Authority, CBP) has set up a co-operation group. This co-operation group has developed a set of assurance products that allow organisations, with different levels of depth, to check primarily by themselves how their own situation relates to the WBP. The contents and meaning of these assurance products have been further elaborated in Section 7.1.2.

[1] The Dutch Data Protection Act (Wet bescherming persoonsgegevens – WBP) is the Dutch implementation of the Data Protection Directive (95/46/EC). The assurance products developed in The Netherlands are legally based on the WBP. The Privacy Audit Framework, extensively described in this chapter, refers to articles in the WBP. A table of the correlation of the article numbers of the WBP and the ones of the DPD and vice versa has been published. This and the fact that the WBP does not to a great extent deviate from the DPD, makes the Privacy Audit Framework applicable when performing a compliance audit against the DPD. The evaluation of the PISA Demonstrator shall, therefore, use this framework as the basis work plan for the investigation to be carried out.

7.1.1 Introduction

The Privacy Audit Framework was set up to carry out Privacy Audits in organisations where personal data are processed. Privacy Audits must be carried out in careful consideration: not every organisation is initially ready to undergo a Privacy Audit. A thorough analysis to assess whether a Privacy Audit has added value for an organisation must take place before a Privacy Audit is carried out. This is to prevent that the results of the Privacy Audit are disappointing for the client. If the aforementioned analysis shows that a Privacy Audit does not provide sufficient added value for the organisation at that time, then the organisation must take proper measures first. The WBP Self-assessment can be used for this purpose if so desired. The auditor can help an organisation by providing advice during the improvement process.

Organisations can have different motives for carrying out a Privacy Audit. The auditor must be informed of the client's motives. Broadly speaking, two important motives can be recognised:

1. An economic motive (primarily aimed internally);
2. A social motive (primarily aimed externally).

Economic Motive
Organisations are obliged to comply with the legal requirements for the protection of personal data. Organisations must, therefore, convert the WBP conditions into a satisfactory system of measures and procedures within their own organisation, within the set transitional period. It is also in the organisations' interest to prevent sanctions and negative reports that can result from non-compliance with the act. The management of an organisation can, for these reasons, instruct an auditor to carry out a Privacy Audit which shall give the management certainty on the implementation and compliance with the Act.

The CBP encourages self-regulation by organisations and via branch and umbrella organisations. An active approach of the management in relation to the adequate implementation of the legal conditions within their own organisation fits within this framework.

Social Motive
Proper compliance with the WBP requirements can give organisations a publicity advantage with regard to the competition. The careful processing of personal data can have a positive effect on an organisation's image and, therefore, has commercial value. In this way organisations can present their positive image to clients, suppliers, employees, the public and so on with regard to privacy. Reports on the Privacy Audit can also be used in a social sense.

7.1.2 Target Group and Application

The Framework was written for auditors who are responsible for the implementation of the Privacy Audit. Correct use of this framework requires sufficient knowledge and skill of auditing in general and IT audits in particular. The auditor must also have sufficient knowledge of the WBP. If the auditor lacks the correct legal knowledge, then the audit must be jointly set up and carried out in collaboration with a specialised legal adviser.

The Framework offers guidance for the set up of an audit plan. Use of an audit plan that is geared to the organisation's specific situation is essential for an effective and efficient implementation of the Privacy Audit. The application of this Framework requires the thorough and well-considered judgment of the auditor at various moments. The Framework does not offer a tailor-made solution. Specific work programmes have to be developed on the basis of the Framework.

Standards and Scope
The Framework does not give standards for the criteria that the law demands of organisations with regard to the protection of personal data. The law gives organisations the room to provide detail to certain legal requirements. This can be deduced from Article 13 of the WBP, which states that, ”appropriate technical and organisational measures must be taken to protect personal data against loss or any form of unlawful processing.” It is not possible to state beforehand what is considered appropriate in a specific situation.

Criteria, which must be considered when deciding whether the measures are appropriate, include:

- State-of-the-art technology;
- The costs of implementation;
- The risks, regarding processing, nature and volume of data.

The auditor defines the specific test standards according to the legal conditions, taking into account the desired scope and depth of the audit, the technical ICT infrastructure and the meaning of the privacy protection measures. Next, the consultation with the organisation that is being audited takes place, after which decision-making with regard to test standards takes place. Consultation with colleague auditors and WBP legal advisers can be very useful at this stage.

7.1.3 Privacy Audit Framework Introduction

Introduction

The entry into force of the WBP affects all organisations that process personal data. The Act concerns both automated and non-automated processing of personal data. The management must see to it that the WBP is implemented satisfactorily within the organisation. This requires purposeful implementation of the measures that must be taken in the framework of this law. The system of measures and procedures already in place for the processing management and security must explicitly be tested with respect to the WBP objectives and reconsidered if necessary.

This chapter further explains that implementation of the WBP is primarily an organisational issue and is, therefore, also the primary responsibility of the management of an organisation. The auditor must be aware of this when communicating with the client.

The process through which proper privacy protection is achieved is not realised overnight. Organisations need time to bring about the awareness process. The process also requires supervision. The auditor can help with this within his or her natural advisory function. It is advisable for the management who is responsible for the implementation of the WBP to appoint one or more contact persons who are responsible for the specific coordination of the measures to be undertaken and the evaluation of the measures which have already been taken. The personal data protection official or the security officer could be the responsible person.

Positioning Privacy Audit

The WBP establishes requirements for the processing of personal data and has consequences for the procedures and measures that an organisation has taken to properly protect and manage its data processing operations. The quality range for the protection of personal data (please refer to Section 7.1.3) is less wide than the quality range for data processing in a broad sense (reliable, efficient, effective, exclusive, integer, continuous and auditable).

The Co-operation Group Audit Strategy has developed three products to help organisations analyse the actual situation regarding to the protection of personal data and implement the desired situation. These products are called Quickscan, WBP Self-assessment (possibly with review) and Privacy Audit Framework.

The Quickscan allows officials within an organisation to gain quick insight into the extent of the awareness regarding the protection of personal data. The Quickscan's scope does not go further than creating awareness within the organisation and can be regarded as a global checklist. No statements are made about the extent to which the WBP conditions are met or not.

The WBP Self-assessment is a more extensive product that must be carried out by officials involved in privacy protection. The WBP Self-assessment is a systematic method for the independent assessment of the quality of privacy protection within an organisation. The results of the WBP Self-assessment give a clear picture of the current situation and the necessary points of improvement. Organisations can have the internal WBP Self-assessment reviewed by an internal or external auditor, for example, an accountant or IT auditor, if so desired.

The Privacy Audit Framework forms the tailpiece of this set of products. An expert auditor/lawyer or team of experts must carry out the Privacy Audit. This is a full scope audit into the method and the extent to which the organisation meets the requirements set down by the law for the protection of personal data.

The Registratiekamer, (”Dutch Data Protection Authority”) the legal predecessor of the CBP, issued the ”Background study & Investigation”, no. 23 ”Security of Personal Data” [vBB01]. This study describes the necessary security measures that must be taken for the processing of personal data in different situations. This study can be ordered from the CBP[2].

The mutual relation between the three developed products and the study ”Security of Personal Data” is shown in the diagram shown in Figure 7.1.

This document contains the Privacy Audit Framework. The diagram shows that, of the products mentioned, a Privacy Audit has the most depth and is, therefore, the most extensive investigation.

WBP Outline

The WBP has entered into force in September 2001. This means that The Netherlands meets the EU data protection Directive's requirement to bring the national legislation in line with this Directive. The WBP replaces the 1989 Wet persoonsregistraties (”Data Protection Act” , abbreviated as WPR in Dutch) and comprises general legal rules to protect the privacy of citizens.

An important difference with the WPR consists of the expansion of the scope of application. The WPR mainly regulated the requirements with regard to the so-called ”personal

[2] A summary of this study can be found in Section 2.6.2.

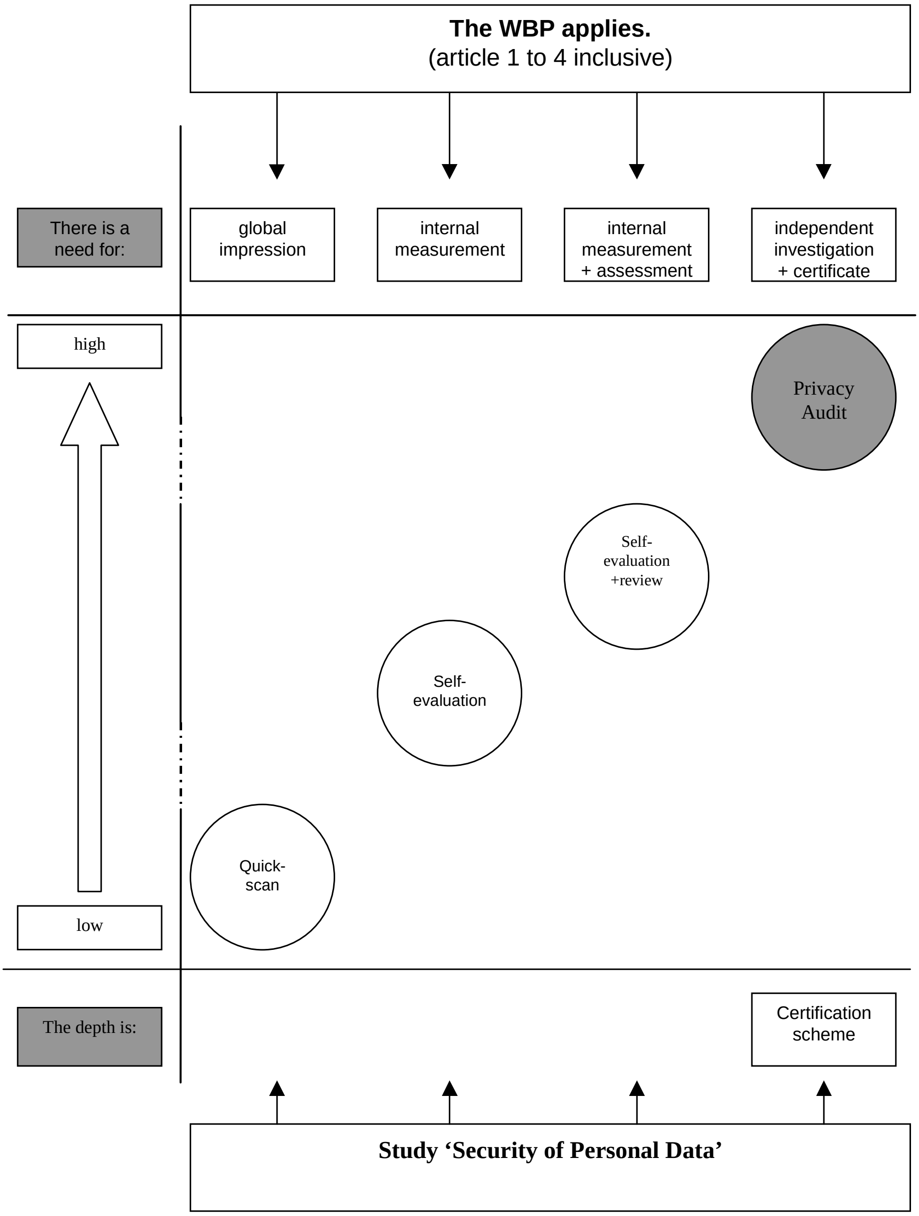


Figure 7.1: Mutual relation between the three developed products and the ”Security of Personal Data” study [vBB01].

data registrations”, whereas the WBP sets requirements on the entire processing chain, which includes the collection, recording, storage, alteration, linking and consultation of personal data as well as the disclosure of personal data to third parties and the deletion or destruction of personal data.

The law offers citizens safeguards for careful and purpose-limited processing of personal data, and gives them opportunities to correct the processed personal data. Data subjects (those persons whose personal data are processed) can also object to the processing of their personal data. This does not mean that certain forms of processing are prohibited but the law does attach clear conditions to this.

The WBP can be summarised from two points of view:

1. Legal:
   The collection of personal data takes place for specified, explicit and legitimate purposes, for example, with the consent of the data subject or based on a legal obligation. Further processing of personal data must be compatible with the original purpose; the personal data must be relevant, not excessive, adequate and accurate.

2. General:
   The processing of personal data offers safeguards that the correct personal data are available for the right purpose, on the right grounds, for the right people at the right time.

The law distinguishes categories of personal data to which strict conditions for use apply. This applies to the so-called ”special categories of personal data”, for example, data on race, political opinions, health and sex life. These personal data may only be processed by legally authorised bodies or in situations described in the law or with explicit consent from the data subject involved. The WBP does not distinguish between processing personal data by public authorities or by private companies.

The WBP institutes the CBP to supervise compliance with this privacy act. The CBP is competent (in accordance with Article 60 of the WBP) to investigate the way in which the current privacy law is implemented within a specific personal data process. When carrying out investigations the CBP shall use the Privacy Audit Framework as point of departure.

Privacy Protection as Part of the Management Cycle

The requirements formulated in the WBP must be implemented within the organisation in an effective way in order to guarantee the rights of citizens in an adequate way. This requires proper general processing measures and procedures system that must take into account the specific protective measures for the processing of personal data. Privacy protection shall generally lead to an additional system of measures and procedures on top of the usually required processing and security measures. This policy must occupy an important place in the management cycle in order to achieve a well-balanced processing policy for personal data and to implement and maintain this properly. Pursuing a policy that aims at protecting privacy is also in line with the management’s objective of total quality and socially responsible business.

Accomplishing business objectives, through the management cycle, generally occurs via the following three phases: the organisation of the processes (including policymaking), the processes themselves and the assessment and adjustment of the processes. These phases are further detailed in Section 7.1.3. The work approaches and methods on which the development of this Privacy Audit Framework are based are in line with this.

Defining, Implementing and Assessing Privacy Policy

The previous paragraph explained the importance of the privacy policy being the responsibility of the management and part of the management cycle. A number of phases must be systematically worked through in order to define a processing policy and then implementing and maintaining it. This process is, in practice, not very formalised. This must, however, not be an argument for the management not to have a structured approach to the process.

It is not the intention to discuss the best method to define a processing policy and the implementation and maintenance of measures and procedures linked to this here. The following explanation is aimed as guidance; it focuses on the implementation of the WBP within an organisation.

Phase 1: Policy and Organisation
The starting point is to develop a privacy policy based on the organisation's objectives. A policy for processing personal data can be formulated based on this. Additionally, the existing privacy policy shall be assessed and differences with regard to implementation of the WBP conditions shall be mapped.

Phase 2: Processing (in Terms of the WBP)
The formulated policy must give tangible form to specific measures and procedures for the processing cycle of personal data. Defining tangible measures and procedures occurs after thorough risk analysis, i.e. making a list of the threats which the processing of personal data is exposed to. Within this context, the strong and the weak points of data processing are laid down. The risks together with the strong and the weak points of the processing organisation and a cost-benefit analysis which is based on the defined privacy policy, result in a carefully considered choice for the organisational and technical measures to be undertaken. The management is responsible for the implementation of the chosen provisions at a satisfactory level.

Phase 3: Control and Assessment (of Phases 1 and 2)
Management must, with the help of a monitoring system, examine to what extent the taken measures fulfil the objectives of the formulated privacy policy. Management must indicate the form and quantity of monitoring data it wishes to receive. The required corrective actions, adjustment of measures and procedures taken or adjustment of the formulated policy are based on the results of the performed monitoring.

The three phases of the management cycle are displayed in the diagram shown in Figure 7.2. The specific detailing of the various elements of the diagram has been incorporated in the Framework as follows:

- The requirements for the processing of personal data (V.I, V.1 to V.9 inclusive) have been described in Section 7.1.5;

Design of Privacy Audit Framework

The essence of the Privacy Audit Framework is that by conducting the Privacy Audit, the respected level and location of the implementation of the requirements of the WBP within the operational organisation are investigated, and which additional measures must still be undertaken to ensure the satisfactory protection of personal data.

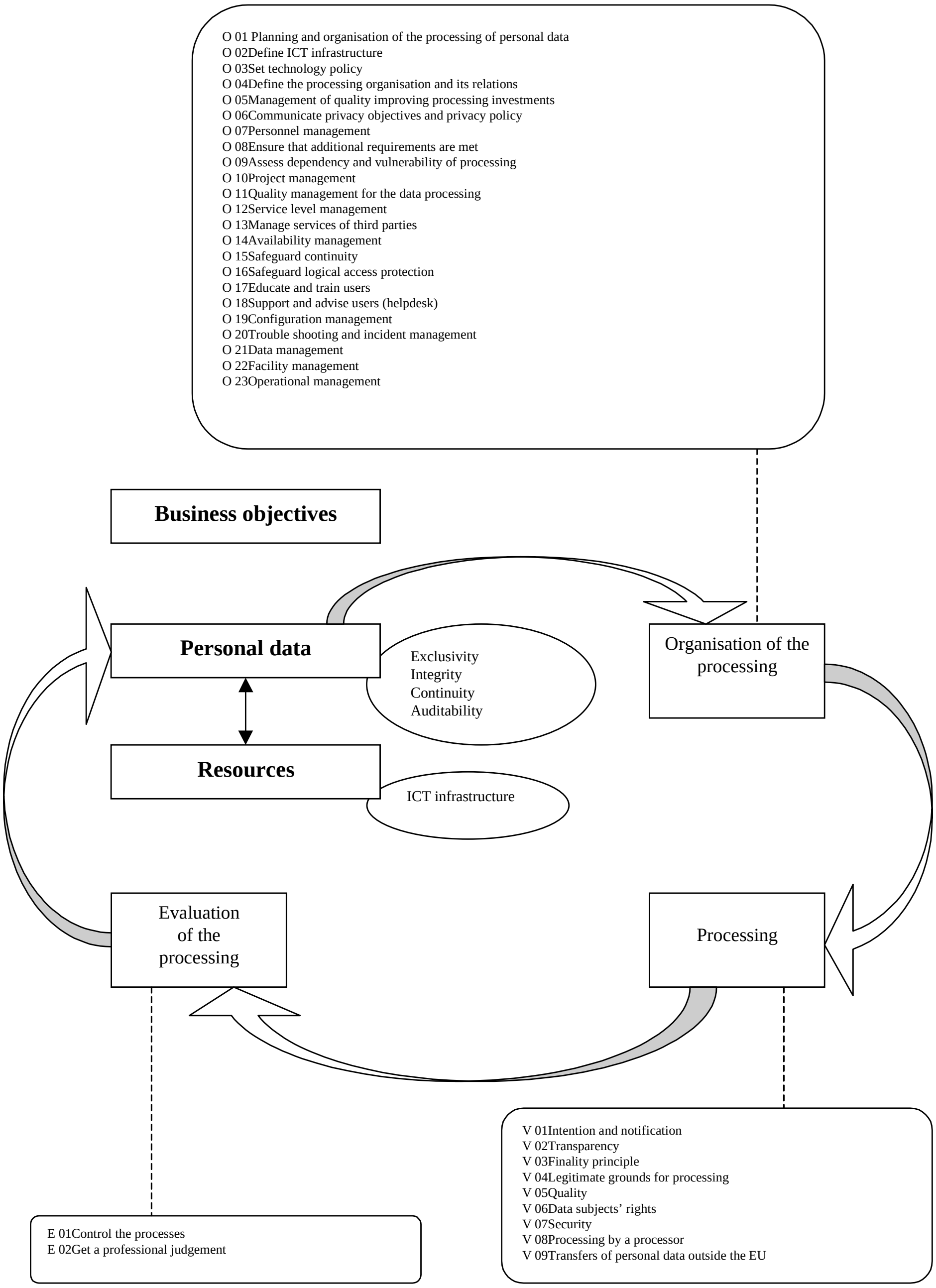

O 01 Planning and organisation of the processing of personal data
O 02Define ICT infrastructure
O 03Set technology policy
O 04Define the processing organisation and its relations
O 05Management of quality improving processing investments
O 06Communicate privacy objectives and privacy policy
O 07Personnel management
O 08Ensure that additional requirements are met
O 09Assess dependency and vulnerability of processing
O 10Project management
O 11Quality management for the data processing
O 12Service level management
O 13Manage services of third parties
O 14Availability management
O 15Safeguard continuity
O 16Safeguard logical access protection
O 17Educate and train users
O 18Support and advise users (helpdesk)
O 19Configuration management
O 20Trouble shooting and incident management
O 21Data management
O 22Facility management
O 23Operational management
Business objectives
Personal data
Exclusivity
Integrity
Continuity
Auditability
Organisation of the processing
Resources
ICT infrastructure
Evaluation of the processing
Processing
V 01Intention and notification
V 02Transparency
V 03Finality principle
V 04Legitimate grounds for processing
V 05Quality
V 06Data subjects' rights
V 07Security
V 08Processing by a processor
V 09Transfers of personal data outside the EU
E 01Control the processes
E 02Get a professional judgement

Figure 7.2: The Three Phases of the Management Cycle.

Privacy Audit Design

The requirements arising from the legislation have been classified into nine " areas of attention" in Section 7.1.5. The implications of the legal provisions for processing personal data are discussed and laid out in that chapter. Every area of attention in Section 7.1.5 has implications for the detailing of the administrative organisation and measures of internal control and protection. Assessing the question of whether an area of attention is relevant and if so, to what extent attention must be paid to it in the Privacy Audit, depends on the concrete situation. Typology, nature and size of the organisation concerned as well as the nature and volume of processing of personal data within that organisation play an important role in determining this.

The legal provisions have been formulated in such a way that they are applicable to all organisations and for all types of processing. No specific details have, therefore, been provided regarding to the typology and size of the organisation or the nature and volume of the personal data processing operations. This means that organisations shall need to further define the requirements arising from the legislation specifically for their own situation. In addition to that, appropriate technical and organisational measures must be taken to provide protection against loss or unlawful processing of personal data (Article 13 of the WBP) and measures and procedures to safeguard compliance with the other legal provisions.

The auditor first investigates whether the organisation has been set up in a way that makes it possible to comply sufficiently with legal conditions (the design) when carrying out the Privacy Audit. Next, the auditor assesses the existence of measures and procedures taken by the organisation in order to assure compliance with the legal requirements. Lastly, the auditor shall concentrate on testing the operation of the measures concerned over a predetermined period.

An organisation's management can determine the way in which technical and organisational measures are taken in order to safeguard the protection of personal data. It shall try to adapt this to the existing organisation and further detailing of administrative organisational and technical measures and procedures to safeguard (automated) data processing. Based on the existing set of control instruments, management can further implement the WBP requirements in an effective and efficient way. The law currently does not impose any compulsory set up on organisations with regard to these technical and organisational measures.

Framework

It has been previously mentioned that the nature and volume of personal data, the purposes of the processing and the method by which it occurs differs per organisation. This product has, therefore, been given the title "Framework" as a feature. This indicates that the elaboration, as given in this document, is based on the most accepted common starting points of organisational theory. Based on these starting points, it must be checked, per processing operation, to determine to what extent the application of this Framework demands additional work or more specific detailing in relation to the audit object.

Conducting a Privacy Audit

Phases

The Privacy Audit must, as is the case for any audit, be set up and carried out in a structured manner to guarantee effective and efficient realisation of the audit.

Privacy Audits generally consist of the following steps:

- Determine the audit assignment;

- Prepare the audit;
- Carry out the audit;
- Evaluate and report results.

Determine the Audit Assignment

The auditor and the client must reach an agreement on the assignment's goal and scope. The minimum scope of the audit is embedded in law. Agreements must also be made as to both parties' responsibilities, realisation and the method of reporting. It is recommended that the audit assignment be laid down in writing in an assignment confirmation before commencing the audit.

The following quality aspects are relevant for compliance with the requirements as defined by the WBP and compliance monitoring in the framework of the Privacy Audit:

1. **Exclusivity/Confidentiality**
   Only authorised people have access to and can make use of personal data.
2. **Integrity**
   Personal data must be based on the projected part of reality and nothing may be wrongfully held back or made to disappear.
3. **Continuity**
   Personal data and the information derived from these must be available without restrictions in accordance with the agreements made to that respect and the existing legal regulations. Continuity is defined as ”undisturbed progress of data processing”.
4. **Audit ability**
   Audit ability is the extent to which it is possible to gain insight into the structure (documentation) and working of an object. The quality aspect of audit ability also encompasses the extent to which it is possible to determine that processing personal data has been carried out in accordance with the requirements with regard to the aforementioned quality aspects.

The extent to which these aspects must be used in a concrete situation partly depends on risk analysis performed by the auditor. The choice for quality requirements per audit object must be explained in the audit plan. The extent to which the quality aspects mentioned are relevant for obtaining a certificate shall be worked out in the certification scheme.

It is possible to use a wider scope than indicated in this Framework in an assignment confirmation for a Privacy Audit. A client may want the auditor to assess the efficiency of the measures and procedures taken. Such an extension of the scope does not affect the audit's base. Limiting the audit's scope is not permitted.

Preparation of the Audit

Investigation of the Organisation and Control Environment

• The auditor needs insight into the organisation and control environment in order to carry out the audit effectively and efficiently. This activity forms the foundation for the audit plan to be developed in this phase. This examination comprises, in all cases, the following elements:

- Organisation (knowledge of business activities, group structure, organisation diagram, information policy, privacy policy, presence of a personal data protection official);

Table 7.1: Risk Classification of Personal Data.

Nature of the Data:		Standard Personal Data	Special Categories of Personal Data Article 16 of the WBP
Amount of Personal Data (nature and volume)	*Nature of processing*		
Few personal data	Low processing complexity	**Risk Class** **0**	**Risk Class** **II**
Many personal data	High processing complexity	**Risk Class** **I**	**Risk Class** **III**
Financial / Economic Personal Data		**Risk Class** **II**	

- Organisation management (integrity, attitude with regard to privacy protection and data processing in general, use of control instruments);
- Other relevant legal requirements (specific sector-related or horizontal legislation, codes of conduct);
- Nature and volume of personal data (types of personal data, processing special categories of personal data, impact of social damage for data subjects in the event of unlawful processing);
- Organisation of the processing environment (data flows, organisation ICT infrastructure, organisation physical processing environment, administrative organisational procedures and measures).

Prior consultation with the officials involved in the protection of personal data within the organisation is advisable. The highest officials, the officials responsible for processing, the processor and the personal data protection official (as defined in Article 62 of the WBP) must be the persons that can be contacted.

Risk Analysis
Thorough risk analysis is essential in order to set up an effective and efficient Privacy Audit. The result of this analysis determines, to an important extent, the type and number of control activities with respect to the technical and organisational measures in place to protect the processing of personal data. The organisation's management is expected to gear these measures to face the potential threats regarding the nature of the personal data, the volume of the processing operations and the influence of the processing on the social position of the data subjects in the event of unlawful processing. The CBP advises organisations to define the system of technical and organisational measures taking into account the risk categories defined in Study No. 23, ”Security of Personal Data”. The diagram to determine risk categories is shown in Table 7.1. See Study No. 23 for an explanation of the diagram.

Risk analysis as part of the Privacy Audit also comprises, in addition to the abovementioned analysis, an assessment of the other elements mentioned above in the ”Investigation of the organisation and control environment” paragraph.

Use of Results of (Different) Audits
The Privacy Audit can be characterised as a frequent form of auditing with its own specific scope of examination. The Privacy Audit can be fitted within the existing framework that

applies to auditing financial accounts and IT audits. The auditor can use the results of earlier audits where possible when carrying out the audit. External auditors can also use results of investigations carried out by internal auditors such as, for example, form previous Privacy Audits for their own Privacy Audit.

It is worthwhile to look at IT audits of computing centres or other processing environments, when using results of previous audits, within which the examined processing of personal data takes place as well as system audits of applications with which personal data are processed. The auditor can decide, based on the audit plan and the results of these audits, which audit objects and quality aspects must be used and to what extent.

The audit results must be prepared in an audit plan that concentrates on the organisation to be examined, the processing of personal data and the consecutive work programme.

Carry out the Audit

The Privacy Audit's key activity is to examine whether the processing of personal data in an organisation complies with the WBP. The auditor must determine whether all areas of attention, relevant to the organisation, have been sufficiently implemented. The irrelevance of one or more areas of attention must be laid down explicitly and must be substantiated by the organisation or auditor in order to allow the correctness and completeness of this explanation to be assessed afterwards if necessary. Next, the auditor must evaluate the adequacy of the measures taken by the organisation. This requires a carefully considered professional judgment by the auditor and must, where necessary, include consulting specific legal support.

Below follows an inventory of the aspects to which the auditor must pay attention in the framework of the assessment of the technical and organisational measures taken to guarantee a lawful processing in accordance with the requirements of the WBP.

Firstly, the auditor must make an inventory of how many processing operations take place within the organisation. Chapter 2, Paragraphs 1 and 2, of the WBP regulates the processing of personal data. The organisation must define which types of processing there are. The examination must include the following points:

- From whom the data are obtained;
- To whom are disclosed which type of (special categories of) personal data;
- What are the data used for;
- The organisation's structure and internal and external relations;
- Which information systems and types of users are distinguished;
- Within which ICT infrastructure processing takes place;
- Which bodies and information systems supply and receive data and the method by which this is done.

Next, the auditor must make an analysis per processing operation, paying attention to the following aspects:

- Whether there is an obligation to notify;
- Whether there are further requirements to process special categories of personal data;
- Who is the controller and who is (are) the processor(s);

- Who is (are)the internal manager(s);
- Which categories of data subjects are involved;
- Which type of data is or shall be processed;
- What is the source of the data;
- What are the grounds for processing;
- Which categories of third parties there are
- Which persons and bodies are obliged to notify the processing of personal data;
- Determining the purpose of collecting the data;
- Who is responsible for the audit (internal or independent auditor);
- Who are the recipients of the personal data.

Measures and procedures must be specified and developed for processing personal data. Articles 6 and 13 of the WBP help define these instruments. The auditor must assess the adequacy of these measures and procedures, including:

- Whether persons who process personal data are sufficiently aware of the problems involved;
- The obligation to notify the CBP or to the personal data protection official, Article 62 of the WBP;
- The lawfulness of the processing;
- The fairness and quality of the processing of personal data;
- The transparency of the processing of personal data;
- Right of access, correction and objection of the data subjects;
- Ensuring that measures are kept up to date;
- Ensuring the enforcement and compliance with the WBP (right of responsibility and right of inspection/audit obligation).

Evaluation and Report Results

The auditor must obtain a thorough basis to give a judgment on the extent to which the organisation complies with the legal provisions for the protection of personal data. The auditor must collect, for this purpose, sufficient audit evidence and lay this down in an audit file. The considerations that led to the judgment must also be laid down therein. The auditor's judgment must always be clearly formulated.

During the Privacy Audit it could emerge that specific measures, already present in the organisation, must be adjusted or that new measures must be taken to ensure that the WBP requirements are met in the given circumstances. This can be implemented through advice given to the responsible management body.

7.1.4 Legal Framework for Privacy Protection

Constitution

Respecting individual privacy is one of the fundamental rights of the Dutch legal system. The right of respect to private life has been laid down in Article 10 of the Dutch Constitution:

1. With the exception of or in accordance with legal restrictions, everyone has the right to respect of their private life.
2. The law prescribes rules to protect individuals' privacy with respect to the processing and disclosing of personal data.
3. The law prescribes rules concerning the right of people to access data that have been processed on them and to be informed about the way in which these data are used, as well as the right to correction of such data.

How are these principles applied in practice? Since 1989 they have been put into effect through the WPR (Data Protection Act), which contains rules for the lawful and careful handling of personal data. The WPR has been replaced by the new Data Protection Act (WBP) in 2001. This new act differs on a number of important points from the WPR. The adjustments reflect the strongly growing and still increasing possibilities of Information and Communication Technology (ICT). The main lines of the WBP are the same as those of the European Directive 95/46/EC, which was adopted on 25 October 1995. This Directive regulates how Member States must deal with the processing of personal data.

7.1.5 Legal Requirements for the Processing of Personal Data

(Article 1, 2, 3, 4) The scope of the law
Article 2, first Paragraph of the WBP reads:

> ”This act applies to the fully or partially automated processing of personal data, as well as to the non-automated processing of personal data entered in a file or intended to be entered therein.”

This article implies that processing personal data does not ,by definition, have to fully take place within the ICT domain. Processing personal data (or parts of the data) that are recorded on other media such as paper, audio or video is also included within the scope of the WBP and is, therefore, subject to audit during a Privacy Audit.

Framework Point of Departing
The point of departure for the Framework is that it has been determined that personal data are being processed and that this processing falls under the scope of application of the Act (Articles 1 to 4 inclusive of the WBP).

I Introduction

(Article 25, 26, 29)

The WBP provides the normative framework from which the specific technical and organisational measures for an organisation must be derived. In addition to this act regulating privacy protection in general, it can certainly be the case that other legislation and regulations are applicable, from which standards must also be derived for the organisation. The Privacy Audit assesses compliance with all relevant legislation.

- Decree on exemption from notification
  If a processing operation of personal data is exempted from notification, the Decree on exemption from notification contains additional provisions, which ensue for organisations from the WBP.

- Sector-related legislation
  These are pieces of legislation that have been developed for a specific sector and in which privacy is one of the issues regulated. Examples of sector-related legislation concern municipal registers (Wgba), medical treatment agreements (Wgbo), police registers (Wpolr), medical examinations (Wmk) and the organisation of social security (Osv).

- Other legislation and regulations
  This refers to regulations which have effect across different sectors. The Telecommunications Act is an example of this.

- Codes of conduct
  Certain business sectors have developed a code of conduct (Article 25 of the WBP). These codes of conduct, which must be approved by the CBP, contain rules that must be used as standard in the sector for a Privacy Audit.

- General administrative orders
  Further rules may be issued by General Administrative Order concerning the application of Articles 6 to 11 inclusive and 13 (Article 26 WBP). Such measures also include rules that must be used as standard for a Privacy Audit.

V.1 Intention and Notification

(Articles 24, 27, 28, 29, 30, 31, 32, 43)

> Personal data are processed within an organisation to which the WBP applies. The CBP or the personal data protection official must be notified of this.
> During the Privacy Audit a judgment must be given on one of the following two points:
>
> 1. If an appeal has been made to an exemption for notification, it must be decided whether this has been done on correct grounds.
>
> 2. If the CBP or the personal data protection official has been notified of the processing of personal data, it must be determined whether the notified information corresponds with the actual situation in the organisation.

1 Determine the nature of processing and the obligation to notify
The first step is to characterise the processing operation. This step is essential to assess whether the processing operation is exempted from the obligation to notify to the CBP or the personal data protection official (Article 62 and, subsequently, Article 27 of the WBP).

Determine:

- The nature of processing;

- Whether the given data processing is mentioned in the decree on exemption from notification;

- Whether the processing operation corresponds with the description in the appropriate article of the decree on exemption from notification;
- The information materials of the, for example, notification disk or website gives, after having been correctly filled in, an assessment as to whether notification must take place or not. The means made available by the CBP for notifications contain an overview of the processing operations exempted from notification;
- That the processing operation is exempted from the obligation to notify the CBP or the data protection official;
- That the processing operation has been notified to the CBP or personal data protection official;
- That action is taken to carry out the notification to the CBP or personal data protection official.

2 Notification

A number of matters must be known before notification can take place in the cases in which the processing operation that is not exempted from the obligation to notify. Exemption to notify does not relieve the organisation from compliance with all other provisions of the WBP.

The auditor must determine that all legally required information has been included in the notification, namely:

- Name and address of the controller;
- The purpose or purposes of processing;
- A description of the categories of data subjects and of the data or categories of data which relate to them;
- A description of the recipients or categories of recipients to whom data may be disclosed;
- Notification of intended transfer(s) of data to countries outside the European Union;
- A general description of the technical and organisational measures which are taken to ensure the security of personal data.
- A description of the purpose or purposes for which the data or categories of data have been collected.

3 Prior Checking by the CBP

A CBP inquiry can precede the notification of the processing operation in certain cases. This is the case if:

- There is an intention to process a number to identify persons for a purpose other than for which the number is specifically intended or expressly permitted by law or general administrative order in order to link data with other data which are processed by another controller. The following exception applies in relation to this: if the number is used for the purposes of the law which prescribe use of the number;
- There is an intention to record data based on the controller's own observation without informing the data subject;

- There is an intention to process criminal data or data on unlawful or obstructive behaviour on behalf of third parties other than in accordance with a permit on the grounds of the Private Security Organisations and Investigation Bureaus Act.

It is determined which of the abovementioned cases is applicable. The following must be recorded in relation to this:

- Notification to the CBP;
- Suspension of processing;
- The time when the result of the CBP inquiry is determined.

4 Central Record
An overview of the notifications of processing personal data is kept at a central point in the organisation, for example, with the personal data protection official. The overview shall, in any case, include the prescribed information for notifications and, where relevant, also the confirmations of receipt of the notifications to the CBP. Everyone can consult this overview free of charge.

5 Periodic Assessment Regarding the (Exemption from) Notifications
Assessments as to whether a processing operation still meets the condition for exemption and that a notification is correct shall take place in the event of adjustments or periodically. The following shall be determined:

- A processing operation which deviates from the notification. This shall be stored for at least three years;
- Whether the processing operation is more than just incidental and whether the notification to the CBP or data protection official must be supplemented and must be notified.

6 Providing Information on Processing
On request of any person, information shall be supplied on the processing operations as it has been notified to the CBP or data protection official or on the processing operations that are exempted. The following shall be recorded:

- The way in which information is given relating to the processing of personal data.

If these requests are not met, it shall be recorded whether this is necessary in the interest of:

- State security;
- Prevention, detection and prosecution of criminal offences;
- Important economic and financial interests of the State and other public bodies;
- Supervising compliance with legal regulations that are in place for one of the abovementioned interests;
- Protecting the data subject or the rights and freedoms of other persons.

V.2 Transparency

(Articles 33, 34, 41, 43, 44)

Everyone must be informed about what is done with their personal data. The data subject must also be informed of this.

1 Providing information to the data subject

Data subjects must be informed about the data processing. Said data subjects have the right to be kept informed of the processing of their personal data. The following two situations can be distinguished:

The personal data are collected from the data subject. Determine:

- The data subject shall be informed about the following before the data are collected:
  - The identity of the controller;
  - The purposes of the processing for which the personal data are intended.
- Whether further information must be provided to guarantee a lawful and fair processing. Attention must be paid to the following points:
  - Nature of the personal data;
  - The circumstances in which they were collected;
  - How the data are used.

2 The personal data are collected in another manner.

Determine that the following actions have been taken at the latest at the moment when the personal data are recorded or when they are disclosed to a third party at the time of the first processing:

- The data subject shall be informed of:
  - The identity of the controller;
  - The purposes of the processing for which the personal data are intended.
- Further information must be provided to guarantee a lawful and fair processing. Attention must be paid to the following points:
  - The nature of the personal data;
  - The circumstances in which they were collected;
  - How the data are used.

If the data subject is not informed, determine if:

- The data subject has already been informed of the processing;
- It is impossible or shall involve a disproportionate amount of effort to inform the data subject;
  - In this case, the source of the data must be recorded.
- Recording or distribution occurs in accordance with the law;

  - In this case, information on the legal regulation concerned must be recorded.

- This is necessary in the interest of:

  - State security;
  - prevention, detection and prosecution of criminal offences;
  - important economic and financial interests of the State and other public bodies;
  - supervising compliance with legal regulations that are in place for one of the abovementioned interests;
  - protecting the data subject or the rights and freedoms of other persons.

3 Specifications concerning the information provision to the data subject
The following must be determined:

- That if the organisation is an institute or service for scientific research or statistics, measures have been taken to ensure that the personal data can be used solely for scientific and statistical purposes. In that case, the data subject does not have to be informed;

- Whether these personal data are part of the archive records that have been transferred to a storage place in accordance with Article 12 or 13 of the Archives Act (Archiefwet 1995). In that case, the data subject does not have to be informed neither.

4 Information in the case of recruitment for commercial or charitable purposes
Determine whether personal data are processed in connection with the creation or maintenance of a direct relationship between the controller or a third party and the data subject in view of recruitment for commercial and charitable purposes. The following must be arranged:

- A direct message is sent to the data subjects and, in each case, it shall be pointed out to them that they may object to the processing;

- The data subject must be informed of the possibilities to object to the processing if there is the intention to disclose the personal data to third parties or to use them on behalf of third parties. The announcement shall be made via one or more newspapers, free local papers or in another suitable way;

- Such an announcement shall be made at least once a year if personal data are frequently disclosed to third parties or are used on their behalf.

V.3 Finality Principle

(Articles 7, 9, 10)

> Personal data are only collected for a specified predetermined purpose. The data can be processed for that purpose and under certain conditions for other purposes.

1 Finality Principle
Collection of personal data takes place for specified, explicit and legitimate purposes. Determine for which purpose the data of the examined processing operations have been collected. The information requested here can be found in V.1. Determine whether the purpose of collection has been described in sufficiently concrete terms.

2 Compatibility of the Data Processing
Personal data are processed in a manner that is compatible with the purpose for which the data have been collected. Determine if in the examined processing operations, the purpose of the processing is compatible with the purpose of the collection. The following points must be taken into account for this assessment:

- The relation between the purpose of the intended processing and the purpose for which the data have been collected;
- The nature of the data concerned;
- The consequences of the intended processing for the data subject;
- The way in which the data have been collected and the extent to which adequate safeguards have been put in place regarding the data subject.

If processing is incompatible with the original purpose, it is necessary to substantiate whether this occurs on the grounds of one or more of the following exceptions:

- State security;
- Prevention, detection and prosecution of criminal offences;
- Important economic and financial interests of the State and other public bodies;
- Supervising compliance with legal regulations that are in place for one of the abovementioned interests;
- Protecting the data subject or the rights and freedoms of other persons;
- Processing of data takes place for historical, statistical or scientific purposes. In this case, the measures taken to ensure that further processing only takes place for these specific purposes must be mentioned.

3 Personal Data Storage
Personal data shall not be kept in a form which permits identification of the data subject for a period longer than necessary for the purposes for which they have been collected or, subsequently, processed. Personal data may be kept longer than provided in as far as this is for historical, statistical or scientific purposes and the necessary measures have been taken to ensure the data are only used for these specific purposes. The storage term can also be determined by legal rules in certain cases, for example, the Act on state taxes(Algemene Wet inzake Rijksbelastingen) or the Act on medical treatment agreements(Wet Geneeskundige Behandelovereenkomst).

- Determine whether a (legal) storage term has been set;
- Determine whether the personal data are kept in accordance with the storage term set or, in the cases where no storage period has been fixed, whether the storage term used in practice are acceptable in view of the mentioned purposes.

4 Obligation of Professional Secrecy
Processing of personal data shall not occur where this is precluded by an obligation of professional secrecy by virtue of office, profession or legal regulation. Determine for the examined processing whether there is an obligation of professional secrecy and ensure that no processing takes place (outside of the professional secrecy regulations).

V.4 Legitimate Grounds for Processing

(Articles 6, 8, 16, 17, 18, 19, 20, 21, 22, 23)

> Personal data may only be collected and processed if grounds for it can be found in the WBP. Specific rules apply for special categories of personal data.

1 Grounds for Processing Personal Data
Personal data are solely processed if one or more of the following grounds apply:

- Data subjects have unambiguously given their consent for processing;
- Data processing is necessary for the performance of a contract to which the data subject is party or in order to take pre-contractual steps which are necessary to conclude a contract at the data subject's request;
- Data processing is necessary to comply with a legal obligation to which the controller is subject;
- Data processing is necessary to protect the vital interest of the data subject;
- Data processing is necessary for the proper performance of a task carried out in the public interest by the administrative body concerned or by the administrative body to which the data are disclosed;
- Data processing is necessary for the legitimate interests pursued by the controller or by a third party to whom the data are disclosed, except where the interests or the fundamental rights and freedoms of the data subject, in particular the right to protection of the private life, prevail.

Determine in which of the aforementioned cases the recorded personal data are processed. The information obtained under part V.1 serves as guideline.

2 Processing of Special Categories of Personal Data
Special categories of personal data are not processed unless the law offers grounds for this. Special categories of personal data are personal data on a person's religion or philosophical beliefs, race, political opinions, health, sex life, trade union membership, personal data concerning a person's criminal convictions and personal data on unlawful or objectionable behaviour in connection with a ban imposed with regard to such behaviour (Article 16 of the WBP).

Determine whether one of the abovementioned types of data is being processed. If this is the case, examine whether this is in accordance with the situations referred to in Articles 17 to 23 inclusive of the WBP.

V.5 Quality

(Articles 6, 10, 11)

> Processing of personal data must comply with quality requirements. Quality means that the personal data are adequate, relevant, not excessive, correct and accurate in relation to the purposes for which they are collected and, subsequently, processed.

1 Data Processing Quality
Determine that the following is properly taken into account in the data processing:

- Storage terms (see also V.3);
- Measures for processing special (diacritical) signs;
- Periodical clearing;
- Information on the disclosure of corrected data to third parties to whom these data have been previously disclosed;
- Final inspection for automated decisions;
- Accuracy, completeness and authorisation inspection for data input (among others if inputted data are validated and processed at a point as close as possible to the source).

2 Errors in Data Processing
Errors are made when working with data. Determine the following in order to limit errors in data processing:

- Measures have been taken to minimise errors or to prevent omitting data input (among others integrity of data processing);
- There is a procedure for the handling of established errors (accurate, complete, in time) and for the checking of irregularities while drawing up basic documents, including notification;
- Measures have been taken to detect and notify errors (correct and complete) during data input;
- There is a correction procedure to correct incorrect data input;
- care is taken of errors pointed out by the data subject.

V.6 Data subject's rights

(Articles 5, 35, 36, 37, 38, 39, 40, 41, 42)

> Persons whose data are collected have a number of rights including the right of access, rectification, deletion, blocking and objection.

1 Implementing Right of Access
Data subjects have the right to access their personal data. The following has to be arranged:

- How and where the request must be submitted;
- Data subjects receive within four weeks confirmation in writing as to whether their personal data are processed;
- That, if this is the case, a full overview is provided at the same time, in an intelligent form, of the relevant personal data, a description of the purpose or purposes of the processing, the categories of data to which the processing relates, the recipients or categories of recipients and the available information as to the source of the data;
- That if third parties are expected to have objections, said third parties are given the opportunity to put forward their view, unless an explanation is given why this is impossible or requires a disproportionate amount of effort;
- That at the data subject's request, information is given on the logic which underlies the automated processing of the data concerned;
- That, if no information is given as to the logic, it is substantiated that an appeal has been made on the grounds of one or more of the following cases in the necessary interest of:
    - State security;
    - Prevention, detection and prosecution of criminal offences;
    - Important economic and financial interests of the State and other public bodies;
    - Supervising compliance with legal regulations that are in place for one of the abovementioned interests;
    - Protecting the data subject or the rights and freedoms of other persons.

Special circumstances can occur with requests for access. Measures have been taken so that:

- If an important interest of the requestor so demands it, the request shall be complied with in a form other than in writing, taking due account of that interest;
- Their legal representatives shall make the request with regard to minors, who have not yet reached the age of 16 and persons placed under legal restraint. The communication concerned shall also be made to the legal representatives;
- The identity of the requestor is properly established.

2 Rectifying, Supplementing, Deleting or Blocking

An access request can be followed by a request to rectify, supplement, delete or block personal data if they are factually inaccurate, incomplete or irrelevant for the purpose or the purposes of processing or, otherwise, if they are processed in conflict with a legal regulation. In that case, attention must be paid to the following:

- Whether the request contains the amendments to be made;
- Whether the requestor is notified in writing within four weeks after receipt or to what extent the request is met;
- Whether a refusal to a request is motivated;
- Whether a decision to rectify, supplement, delete or block is carried out as quickly as possible;

- Whether, where personal data have been recorded on a data carrier to which no adjustments can be made, the data user is informed about the impossibility to rectify, supplement, delete or block the data;
- Whether the request with regard to minors who have not yet reached the age of 16 and persons placed under legal restraint is made by their legal representatives. The communication concerned shall also be made to the legal representatives;
- Whether, where personal data have been rectified, supplemented, deleted or blocked, the third parties to whom the data have been previously disclosed, are informed as quickly as possible of the rectification, supplement, deletion or blocking, unless it has been substantiated that this is impossible or shall require a disproportionate amount of effort;
- Whether, when the data subject requests this, a statement is given of those parties who have been informed of the abovementioned actions;
- Whether the payment is refunded when the data are rectified, supplemented, deleted or blocked on request, on the recommendation of the CBP or by order of a judge.

Exception: there are public registers set up by law that include a special procedure to rectify, supplement, delete or block data.

3 Objection
Objection means registering an objection against the processing of personal data.

3.a Relative Objection
The data subject can at any time register an objection to the processing with the controller in connection to said data subject's particular situation. The controller shall not comply in the following cases:

- Data processing is necessary for the proper performance of a task carried out in the public interest by the administrative body concerned or by the administrative body to which the data are disclosed;
- Data processing is necessary for the legitimate purpose of the controller or of a third party to whom data are disclosed, except where the interests or the fundamental rights and freedoms of the data subject, in particular, the right to respect the private life, prevail.

The following must be arranged with regard to the objection:

- Assessment as to whether the objection is justified occurs within four weeks after the request's receipt;
- If the objection is justified, processing shall be terminated immediately;
- Payment of expenses in order to deal with the objection does not exceed an amount which is determined by general administrative order;
- Payment is refunded if the objection is legitimate.

The aforementioned does not apply to public registers which have been set by law.

3.b Absolute Objection (objection Against Processing for Commercial or Charitable Purposes)
If personal data are processed in connection with creating or maintaining a direct relationship between the controller or a third party and the data subject in view of recruitment for commercial and charitable purposes, the data subject has the right to object to this processing. Such a request must always be complied with. Determine whether the following has been arranged:

- The data subject can at any time register an objection, free of charge;
- When a direct message is sent to data subjects, they are informed they have the option of objecting in each case;
- Measures are taken to terminate processing immediately;
- If the intention is to disclose personal data to third parties or to use them on behalf of third parties, appropriate measures must be taken to inform the data subject of the possibilities to object;
- The announcement shall be made via one or more newspapers, free local papers or in another suitable way;
- If data are frequently disclosed to third parties or used on behalf of third parties, this announcement shall take place on an annual basis.

4 Automated Decisions on a Person's Personality
Nobody may be subject to a decision which produces legal effects concerning their personality or significantly affects their personality, if that decision is based solely on the grounds of automated processing of personal data intended to gain insight into certain aspects of a person's personality. If this is the case, it must be determined why this is done. Determine whether:

- The decision has been taken in the course of entering into or the performance of a contract and
  - The data subject's request has been satisfied;
  - The data subject has been given the opportunity to put forward his or her view.
- The grounds of processing are based on an Act in which measures to safeguard the data subject's legitimate interests have been laid down;
- The data subject is informed of the logic that underlies the automated processing of his or her data.

V.7 Security

(Articles 6, 12 and 13)

> The controller is under the obligation to implement appropriate technical and organisational measures to protect personal data against loss or any form of unlawful processing. The measures guarantee, having regard to the state of the art and the costs of implementation, an adequate level of security. The adequacy of the level of security depends on the risks involved in the processing and the nature of the data. The measures also aim at preventing unnecessary collection and further processing of personal data.

The CBP has developed a separate document in which the normative framework that results from Article 13 of the WBP has been worked out. This is Study No. 23, ”Security of Personal Data” [vBB01]; see also Section 2.6.2. The measures to be taken, depending on the identified risk classes of personal data, have been grouped in the following 14 categories in this study:

1. Security policy, security plan and implementation of the system of procedures and measures;
2. Administrative organization;
3. Security awareness;
4. Personnel’s requirements;
5. Workplace design;
6. Administration and classification of ICT infrastructure;
7. Access management;
8. Networks and external connections;
9. Use of software;
10. Bulk processing of data;
11. Data storage;
12. Data destruction;
13. Contingency plan;
14. Contracting out the processing of personal data;

The CBP stimulates the use of technical measures (Privacy-Enhancing Technologies; PET). The Minister of Justice has shown a preference for a technical (PET) measure in the cases where the choice between a technical and organisational measure exists. The grounds for this are that a technical measure is more effective because it is harder to evade its effect.

Assess whether the measures taken are sufficient and determine whether they are appropriate taking the following into account:

- The state of the art in technology;
- The costs of implementation;
- The risks, both regarding the processing and the nature and amount of data.

Assess the weighing of organisational and technical measures by the controller in view of the statements of the Minister of Justice as phrased above.

V.8 Processing by a Processor

(Article 14)

> The controller does not process personal data, instead, this is (partly) contracted out to the processor. This has been laid down in a contract or other legal act to create a binding agreement between the processor and the controller.

Determine whether the controller contracts out work to a processor. Determine if there is a contract between the controller and processor and that the following has been arranged therein if this is the case:

- Everyone who acts under the processor's authority, as well as the processor himself/herself, in as far as they have access to personal data, solely process these on instructions of the controller, with the exceptions of divergent legal obligations;
- The processor complies with the obligations of the controller regarding the implementation of appropriate technical and organisational measures to protect personal data against loss or any form of unlawful processing. This concerns the measures and procedures described in Chapter V.9;
- That the processor complies with the WBP and the specific legislation of that country concerning general security requirements, if the processor is established in another country of the European Union;
- The parts of the contract or the legal act relating to data protection as well as the envisaged security measures shall be set down in writing or in another equivalent form for the purposes of having proof;
- The controller checks compliance with the contract and legal provisions applicable to the processing periodically. The processor may also periodically engage an (external) independent auditor to carry out an investigation of the processing. The controller shall receive a report of the investigation carried out.

V.9 Transfer of Personal Data Outside the EU

(Articles 76 and 77)

> Without prejudice to compliance with the Act, additional rules apply to personal data which are subject to processing or which are intended to be processed in another country outside the European Union after being transferred. These data transfers are subject to rules.

Determine whether data transfer to countries outside the EU takes place.

Attention must be paid in the investigation to the circumstances which affect the transfer of data or of a category of data if this is the case. Specific attention must be paid to the nature of the data, the purpose or purposes of the processing or processing operations and the intended storage period, the country of origin and the country of final destination, the general and sectoral legal rules in force in the third country in question, as well as the professional and security measures which are complied with in those countries.

Assess the following for those cases in which data is transferred outside the EU:

- The country concerned ensures an adequate level of protection; or
- If the country concerned does not ensure an adequate level of protection, one or more of the following conditions are met:
  - The data subject has given his or her unambiguous consent;
  - The transfer is necessary for the performance of a contract between the data subject and the controller or for taking pre-contractual measures which are necessary to conclude a contract at the data subject's request;

  - The transfer is necessary for the conclusion or performance of a contract which has been or shall be concluded in the data subject's interest between the controller and a third party;
  - The transfer is necessary on important public interest grounds or to establish, exercise or defend in law any right;
  - Transfer is necessary to protect the vital interests of the data subject;
  - Transfer is made from a register set down by legal provisions and which can be consulted by anyone or by any person who can demonstrate legitimate interest in as far as the legal requirements for consultation are met for the case concerned.

- The Minister of Justice, after consulting the CBP, has issued a permit for a transfer or category of transfers of personal data. Transfer shall take place in accordance with the provisions attached to the permit.

7.2 Common Criteria (PETTEP)

The conclusion one can draw from Section 7.1 is that the Privacy Audit Framework is a work plan a privacy auditor can use in order to evaluate compliance of a processing system for personal data within the applicable privacy legislation.

The scope has to be defined for all audits. A very important part of the scope is the exact definition of the object of investigation. The object can be defined as being a product or a process.

An audit evaluating the quality of coffee grinders can be performed as a product audit; take one randomly chosen sample and make a statement on its quality. It can also be evaluated through process audit if the auditor evaluates the making of this type of coffee grinders.

When the processing of personal data via a product audit is being evaluated, the auditor evaluates the data the controller processes by just examining the contents of the database. A process audit is required to evaluated the mode of operation within the organisation of the controller if the auditor must also establish the circumstances used to collect that data.

A number of organisations evaluate products that can be used as part of an information system processing personal data or as a tool to process personal data in the world of privacy auditing. One example might be to perform an evaluation of Microsoft Access and establish if the product has the functionality to process personal data according to privacy legislation. Such a judgment cannot guarantee that the processing is legitimate because it leaves the possibility to a secure processing of ”unlawful” personal data. The only advantage the privacy auditor may have of this ”product audit” is that he can rely on the correct product operation.

Currently, the Common Criteria contain FPR Class: Privacy. This is the only class relating to the evaluation of privacy compliance. This class specifies and gives evaluation criteria on a small, although not unimportant, subset in privacy regulation. The issues are: anonymity, pseudonymity, unlinkability and unobservability. Compliance to this class does not, in any way, guarantee that the information system is compliant with the Data Protection Directive.

The evaluation criteria related to the FPR Class is best compared with a product audit. Even if one or more of the options are correctly implemented, it does not guarantee that the personal data processed has been lawfully collected.

Compliance auditing must mean, according to the Data Protection Authorities, carrying out a process audit whereby the flow of personal data within the organisation is checked, including the possible processor, from the moment the data are to be collected up until their destruction.

The conclusion is that the Common Criteria, although their methodology is correct, cannot at this moment in time be used as the framework for a Privacy Audit. Privacy obligations require that the system design process and security standards incorporate privacy requirements.

Activities are being considered, within the PETTEP project, to define a number of extensions to the FPR Class. Current thinking is to enhance this class with 11 Fair Information Practices. These practices are very easily matched with the nine principles of the Privacy Audit Framework (see Section 7.1). The Fair Information Practises are:

1. Accountability;
2. Identifying purposes;
3. Consent;
4. Limiting linkability;
5. Limiting collection;
6. Limiting use, disclosure and retention;
7. Data quality;
8. Safeguards;
9. Openness;
10. Individual access;
11. Challenging compliance.

The Privacy Audit shall be defined as an international standard if this extension to the Common Criteria is achieved, if, at the same time, the methodology is changed into a proper process audit and if these new CC again are recognised as an ISO standard.

Chapter 8

Privacy architecture for agents

G.W. van Blarkom
gbl@cbpweb.nl
CBP, The Netherlands

P. Verhaar
verhaar@fel.tno.nl
TNO-FEL, The Netherlands

M. van Breukelen
breukelen@tpd.tno.nl
TNO-TPD, The Netherlands

An intelligent agent has been designed in order to remain within the boundaries of the EU data protection directives (DPD) to protect the personal data of the user of the agent adequately. The privacy-related legal issues like consenting, the processing of personal data and the role of the controller that has an impact on the architecture of the agent shall be explained in this chapter. The ISA has to tackle the privacy threats in order to protect the ISA against attacks, security risks and privacy intrusion. A privacy model for the agents should be developed. A privacy-safe agent may offer two privacy limit options: an ISA operation under anonymity or under the DPD. The built-in legal know-how by ontologies and other mechanisms to enforce privacy legal aspects are significant.

8.1 Introduction

Creating an Intelligent Software Agent (ISA) that protects privacy shall mean two things. First, the ISA shall protect its own user's privacy by protecting:

- The personal data of its user that is stored within the ISA;
- Controlling the internal processing of these personal data;
- Controlling, and regulating the exchange of these personal data with participants in the environment, according to the applicable legal regulations. This also means that the fundamental privacy/interface principles must be supported. For example, transparency of processing, traceability, control over information sharing.

Secondly, the ISA shall protect the personal data of others that are collected by the ISA.

8.2 Privacy Related Issues

As we all know by now (see the previous deliverables of the different work packages), an Intelligent Software Agent (ISA) is a piece of software that is capable of performing tasks autonomously on behalf of its user. In doing so, it shall create a profile of its user. This profile shall contain personal data of the user. Part of the profile shall be used to make decisions about which activities need to be executed to perform one or more tasks. When performing tasks for its user, the ISA may exchange personal data of its user with its environment, see Figure 8.1 for a schematic drawing of the ISA and its environment. The ISA shall also collect, and create new personal data of its user during the execution of the tasks. From a privacy perspective, this profile (containing personal data) needs to be protected against privacy violations. Privacy violations can be initiated by the ISA itself or by the environment in which the ISA is active. For a more detailed description of these violations (threats), and who may initiate them, see deliverable D7 and Annex A.

8.2.1 Consent

Article 2 (h) gives the legal definition of ‘the data subject’s consent’ as:

> Any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed;

Article 7 (a) states:

> Personal data shall only be processed if the data subject has unambiguously given his consent;

On the processing of special categories of data, Article 8, Paragraph 2 (a) of the Directive states that the data subject must give his explicit consent to the processing of the data to cancel the fact that the processing of such data is prohibited (Article 8, Paragraph 1).

The legislator uses several adjectives describing consent: ‘freely’, ‘specific’, ‘informed’, ‘unambiguously’ and ‘explicit’. This leads to the conclusion that the legislator gives the data subject definite possibilities to consent before making the processing of personal data is lawful. Article 7 gives more possibilities for the processing of personal data, for instance the requirement of the performance of a contract or some kind of legal obligation, but if none of these apply, processing is lawful only if the data subject gives his consent before the processing is allowed to take place.

This paragraph shall give further information on the meaning of the adjectives mentioned earlier:

- ‘freely’, Article 2 (h)
  The data subject shall not be put under any sort of pressure to disclose personal data, for instance it shall not be compulsory to disclose certain excessive personal data items before a required service shall be provided;
- ‘specific’, Article 2 (h)
- ‘informed’, Article 2 (h)
  The controller must inform the data subject the purpose of collection and the possibly intended compatible further processing before the data subject has disclose any of his personal data;

- ‘unambiguously’, Article 7 (a)

- ‘explicit’, Article 8 Paragraph 2 (a)
  The controller must make it absolutely clear.

8.2.2 Processing Personal Data

Article 2 (d) defines the ‘controller’ as the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data.

Article 17 of the Directive states that the controller must implement appropriate technical and organisational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing;

8.2.3 Processing Personal Data on Behalf of a Controller

Article 2 (e) defines the ’processor’ as the natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller;

If a controller uses a processor of any kind, Article 17 Paragraph 2 obliges the controller to choose a processor that provides sufficient guarantees with respect to the technical security measures and organisational measures governing the processing to be carried out, and must ensure compliance to those measures.

The conclusion of the above is that the controller remains responsible for the lawful processing, even if (parts of) the processing is performed by a processor.

8.2.4 ISA and Privacy Related Issues

Violations initiated by the environment all come down to participants/components in the environment that can force the ISA to exchange personal data with or without noticing that this is happening. The ISA, therefore, needs to be protected against these malicious participants or components. To avoid privacy violations that are initiated by the environment the profile needs to be protected by security measures against malicious participants in the environment. Solutions could be the encryption of the profile, or, in case the ISA code may be manipulated, the ISA needs to be wrapped during transportation or periods of inactivity, and the ISA can only execute actions in secure environments. The problem arising when protecting an ISA using encryption is the fact that the ISA is made of software. This means that the ISA shall not be able to safely encrypt and decrypt the profile without help from trusted hardware components in the environment.

8.2.5 Privacy Threats for Intelligent Software Agents

With violations initiated by the ISA itself, different situations can occur, and, therefore, different solutions need to be considered. Examples of these situations and possible solutions are:

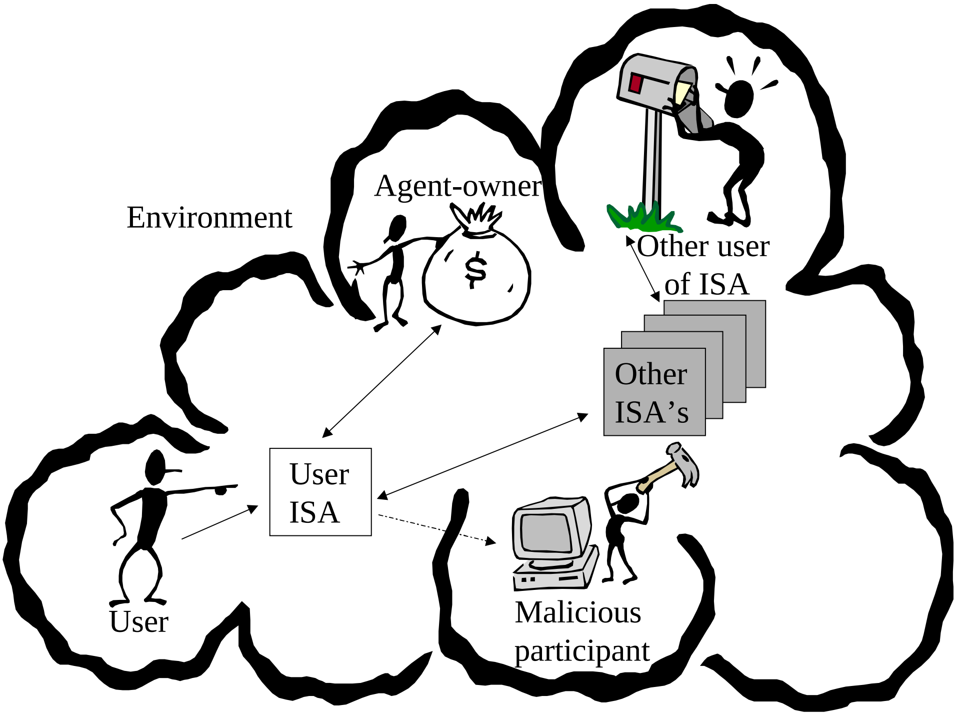


Figure 8.1: The ISA and its environment.

- The ISA can be owned by somebody else than the user. This so-called ISA-owner is then the controller, as stated in the European Directive on Privacy, of the user's personal data. Possible solutions for this situation are:
  - To prevent the exchange of personal data of the user between ISA and ISA-owner, and prevent manipulation of this personal data by the ISA-owner, by protecting the user's profile.
  - The ISA-owner must handle the profile of the user in accordance to the Privacy Directive, by implementing privacy mechanisms, based on the Privacy Directive, that shall enforce this.
- The mutual trust between the user and the ISA needs to be established. First, the user needs a way to verify the correct functioning of the ISA. Second, there must be an identification and authentication mechanism to make sure both the user and the ISA are sure they are communicating with each other.
- Because the ISA is autonomous, it can make automated decisions.
- The ISA communicates data from the profile to other participants in the environment, without regarding the privacy demands and wishes of its user. In most cases the ISA shall not have the means to handle the privacy demands and wishes in a proper manner. Firstly, the ISA must at least know the privacy regulations that apply. In other words: the privacy regulations must be translated in such a way that they can be added to the agent rule-base. Secondly, the user must be able to formulate his or her privacy demands and wishes. Based on these demands and wishes the

ISA needs to be able to take decisions about which personal data may be exchanged. The demands and wishes are limited to 'no privacy at all' and maximum privacy by using PET. Using PET it is possible to provide anonymity or pseudo-identities, both at different levels. Between these two limits all types of variations of enforcement of the privacy regulations are possible, e.g. a decision taking mechanism for the exchange of personal data based on the sensitivity of the personal data. Thirdly, the environment must have a means, e.g. P3P, to exchange privacy statements or privacy policies, so that the ISA can take decisions whether or not to trust certain parties in the environment and to communicate data from the profile with that party. Fourthly, if the ISA is not satisfied with the provided privacy policies, it must have means to create anonymity or pseudo-identities (PET) to make sure the privacy of its user shall not be violated when communicating with its environment, and to give the ISA control over the personal data that is exchanged, and the amount of personal data that is exchanged. It is also desirable to have the ISA come back to the user for further instructions in cases where it is not clear what decision must be taken. For example, if the ISA is not sure whether to share certain private information or not, the user must be consulted before the data is shared. In some cases, the user shall always want to be consulted to allow them to make case-by-case decisions depending on the identity of the recipient of the information.

- The integrity of the personal data that is collected or created is violated. Existing personal data can be manipulated, intentionally or unintentionally, in such a way that it is no longer representative for the user. The ISA can create new personal data that is not representative for the user. To make sure the agent shall only create new personal data that is representative it needs to consult the user to check the validity of the newly generated personal data.

As mentioned before the ISA shall also collect other data. Said data could include personal data of other persons. If that is the case then the ISA needs to make sure that it handles these personal data according to the applicable privacy legislation. For this project, the legislation shall be the European Directive on Privacy. Deliverable D7 provides a description of the privacy principles that are derived from the European Directive on Privacy and the OECD guidelines on Privacy. The privacy principles that are derived are:

- Reporting of processing;
- Transparent processing;
- 'As required' processing;
- Lawful basis for data processing;
- Data quality conservation;
- Rights of the parties involved;
- Data traffic with countries outside the EU;
- Processing personal data by processor;
- Protection against loss and unlawful processing of personal data.

The ISA needs to have a vision on how to use these privacy principles (adaptation to the world model of the agent, see Intelligent Software Agents and Privacy, a report published by the Dutch Data Protection Authority). It also needs to know what actions it can or must take (mental model) while processing these personal data. Together with the rules, the ISA needs to have mechanisms to perform the correct actions.

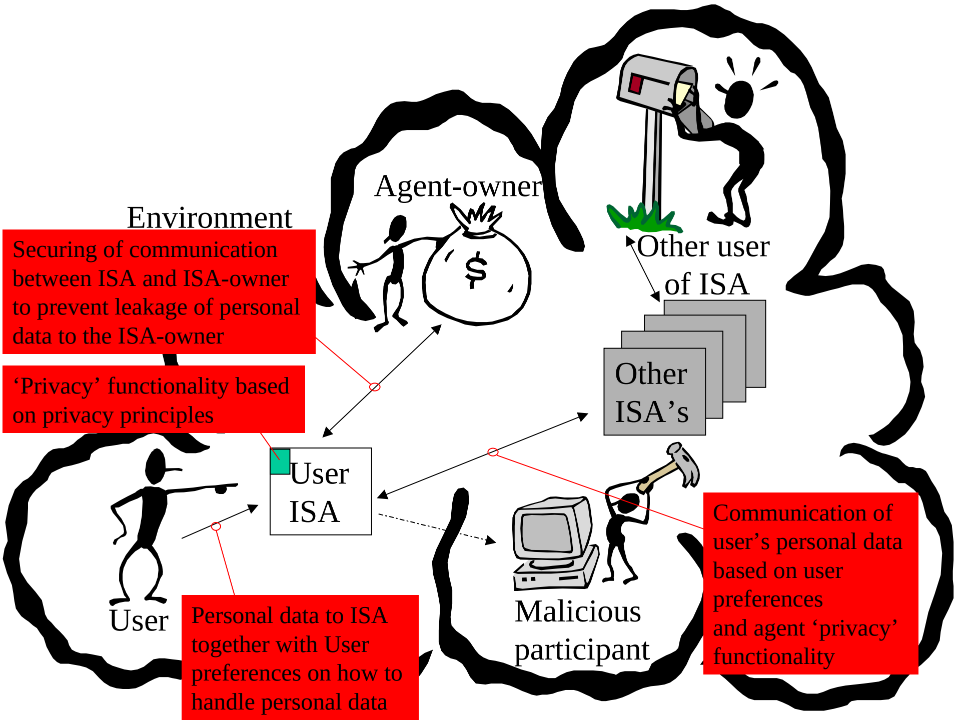

Agent-owner
Environment
Securing of communication between ISA and ISA-owner to prevent leakage of personal data to the ISA-owner
'Privacy' functionality based on privacy principles
Other user of ISA
Other ISA's
User ISA
User
Personal data to ISA together with User preferences on how to handle personal data
Malicious participant
Communication of user's personal data based on user preferences and agent 'privacy' functionality

Figure 8.2: ISA and its environment: the privacy issues.

8.3 Privacy Incorporated Software Agent (PISA)

An Intelligent Software Agent (ISA) has been described [BvES99] as software and/or hardware that is capable of acting autonomously in order to accomplish a task on behalf of its user in a complex network environment. Thus it can act on its own without interference of the user. The ISA may have properties and attributes like mobility, deliberative behaviour, and interaction with other ISAs, possibilities for identification and learning. The fact that it can act independently of its user while representing him or her, and the ability to exchange high-level information with other ISAs or potentially expose Personal Data (PD) in other ways to its environment, is relevant for the legal qualification of an ISA. ISA can appear in the role of a benevolent digital butler or a malicious digital criminal or as a means for constant surveillance. ISA mimics human roles and behaviour in the real world and ISA designers analyse and adopt human ways of acting while humans are trying to achieve their goals. This is reflected in the MAS[1] architecture. In order to perform properly, ISAs need personal data, of which a great part will be Personal Data (PD), and privacy preferences (a profile of personal preferences and aversions for goods and services) as well as non-personal data. The ISA gathers relevant task information, and personal data of both its user and others. The ISA will reveal personal data of the user according to his or her privacy protection level preferences. As the ISA processes personal data automatically, it

[1] MAS: Mulitagent System. In MAS is embedded decision trees that can learn from sample data and constraint satisfaction programming based on a set of constraint rules. Logic frameworks can extend a set of 'action selection strategies'.

obliges the controller[2] of the ISA to build-in (or to have built-in by a developer) a tracking system for proof of reasoning of the results ISA has achieved according to the Directive 95/46/EC. This approach would avoid liabilities[3] through ISA transactions.

The objective of the Privacy Incorporated Software Agent project (PISA project) is not to prove how ISAs behave, but to demonstrate that the privacy of the user may be protected in many kinds of processes by incorporating privacy-protecting features into an ISA. In the PISA project there are three types of ISAs: generic personal agents, specific task agents and service agents to segment responsibility with regard to the protection of the personal data.

8.4 Personal data

Personal data means any piece of information regarding an identified or identifiable natural person5. Whether data can be qualified as ’personal data’ depends on a number of elements. Within the scope of this essay, ’identification’ is the only significant element. According to Article 2 of the EU Directive 95/46/EC, a natural person can be identified ’directly or indirectly’. Direct identification requires basic details collected in PD. PD is name, address, a personal number, a widely known pseudo-identity, a biometric characteristic such as a fingerprint, etc. Indirect identification requires other unique characteristics or attributes or a combination of both, that would sufficiently identify a person[4].

Non-identification is assumed if the amount and the nature of the indirectly identifying data are such that identification of the individual is only possible with the application of disproportionate effort[5]. Disproportionate effort relates to the nature of the data and the size of the population as well as the resources (time and money) one is willing to spend in order to identify the person[6]. Internet identifiers such as an IP address, browsing activities of a user, session login data and the listing of web sites visited by an Internet user are classified as PD.

To better deal with protection of personal data, within PISA we divide it into three categories:

Level 1: ’Contact Information’. This set of personal data is transferred when direct communication between data subject and the other party is needed. Level 1 may contain items like: Name and Address information, telephone number and e-mail address, so therefore level 1 can be seen as PD. It is irrelevant whether one uses a real identity or a pseudo-identity. The sole use one can make of this data is when the flow of communication between the ISA of the data subject and the ISA of the other party have created the desired result and direct human contact between the user and the

[2] A controller is defined in the DPD in article 2 (d) as the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of processing of personal data.

[3] Article 15 section 1 of the Directive 95/46/EC grants the right to every person not to be subject to a decision which produces legal effects concerning him based solely on automated processing of data intended to evaluate certain personal aspects relating to him. The autonomous decision capacity of the ISA raises legal questions. Especially when the agent starts to build the profile of the user based on acceptance or rejection of actions performed by the agent. A rejection might be recorded as a dislike and an acceptance as a preference. A long list of likes and dislikes will be made as a basis for further decisions of the agent.

[4] See [Byg01]. IP addresses and email addresses may be qualified as personal data if they can be linked to a particular person.

[5] See Recital 26 of the EU Directive 95/46/EC. When processing takes place without any personal data the DPD is not applicable

[6] We can only note in passing here that the concepts of ’identification’ and ’identity’ are essentially contested in theory and often arbitrary and ambiguous in practice. See [Raa99].

other party is needed. For this reason this type of personal data is referred to as ’contact information’. In another MAS application, for instance an ISA looking for a special MP3 file, Level 1 may also be the credit card number the data subject uses to purchase an MP3 file for download.

Level 2: All others items of personal data, excluding the special categories of personal data as defined in Directive 96/46/EC Article 8 paragraph 1.

Level 3: Special categories of personal data as defined in Directive 96/46/EC Article 8 paragraph 1. Level 3 personal data are only to be processed under the conditions specified in Article 8 paragraph 2 through 7.

One difference between level 2 and level 3 personal data is that level 3 requires stronger security measures (technical and organisational) to protect these personal data.

Personal data can be divided in data directly under control of the User or data that will be generated by the Infrastructure. Level 1 personal data is assumed to be always Personal Data (PD).

Privacy-Enhancing Technologies (PET) will protect Level 1 by encapsulating these data within the PISA directly at the time of collection. Only the recipient entitled to receive level 1 will have access to the means for unwrapping the PET encapsulated level 1.

It is common knowledge that it cannot be ruled out that the identity of the data subject is identifiable from level 2 (and 3) personal data items. There is already a considerable research on the record regarding how to protect these data so as to prevent identification of the data subject. The PISA project does not intend to redo this research. In the PISA project the designers assume that identification of the data subject from level 2 and 3 is not possible.

8.5 Structure of PISA

In order to protect the ISA against attacks, security risks and privacy intrusion the ISA needs to have:

- built-in privacy protection-related functions, and mechanisms or interfaces to mechanisms like Privacy-Enhancing Technologies (PET) to act on these privacy protection-related functions;
- built-in legal know-how, and mechanisms to enforce legal aspects to protect the personal data according to the Data Protection Directive.

The ISA collects, processes and disseminates personal data and other transaction relevant information. PISA can have 2 privacy limit options: an ISA operation under anonymity or maximum-attainable pseudonymity, or under the Directive 95/46/EC (DPD). During the operation the PISA could be free to execute on Level 2 items only. That means no PD is processed, like in complete anonymity. PET and legal measures will protect level 1 and level 3. Once a contact is made and PD needs to be exchanged, arrangements can and need be made about the protection of the PD. Both, privacy protection-related functions and legal know-how are needed to create correct arrangements. When encryption is used a Trusted Third Party (TTP) could be involved to exchange encryption keys, for encryption and decryption of personal data of level 1 or level 3. The legal know-how should assure that the whole process is within the boundaries of the DPD.

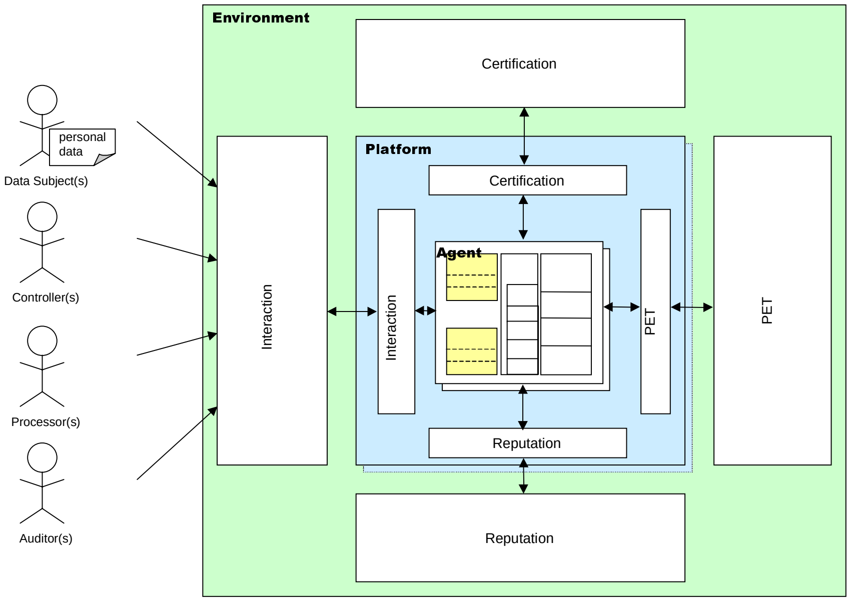


Figure 8.3: PISA Model in its environment.

ISAs are logically divided according to roles of the ’data subject’, the ’controller’ or the ’processor’. Taking into consideration the definitions of these entities in article 2 (a), (d) and (e) of the Directive 95/46/EC, it is only fair to state that an ISA can’t be either of these. The definition of the ’data subject’ refers to a ’natural person’ and those of ’controller’ and ’processor’ both refers to a ’natural or legal person’. Albeit that there is no exact legal definition of an ISA, it is in no way a ’natural’ or ’legal’ person.

The facts described above do not mean that no relation is possible between an ISA and one of the actors mentioned. It is fair to say that an ISA may do whatever it is supposed to do ’on behalf’ of such an actor. It is the natural or legal person for whom it is acting who has and holds complete responsibility for the behaviour of the ISA. The organisation putting an ISA on the Internet to process personal data is the controller of the system processing personal data. Likewise if a data subject develops an ISA himself for his own use, the ISA takes the role of both data subject and controller. This observation has influenced considerably the design of PISA and is reflected in the privacy environment model. This model also describes the legal and factual environment an ISA has to operate in.

In PISA, Privacy Protection is a combination of PET by e.g. pseudonyms, trust and secure environments and ”act according the EC directive”, Certification, Interaction and Reputation as shown in Figure 8.3.

8.6 Privacy Supporting Measures

Apart from understanding the terminology and semantics of privacy, a number of measures need to be taken for building a privacy protecting ISA. The ISA-platform and the ICT-environment need to provide certain provisions, such as mechanisms to create secure communication, and certificate authorities. It is important that the PISA can communicate

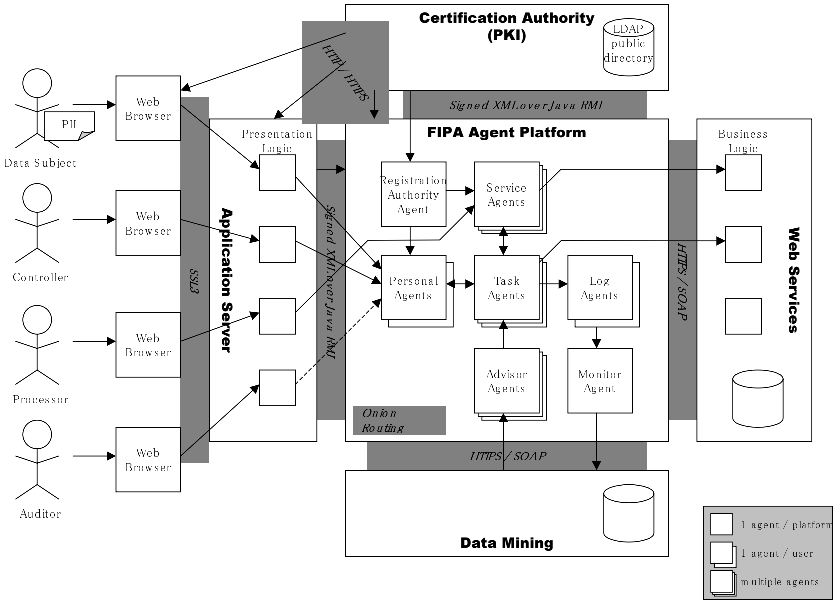


Figure 8.4: PISA architecture.

with these provisions in a way that it does not have to give up the privacy of its user.

8.6.1 Anonymity and pseudo-identities

In the PISA taxonomy a distinction of proxy agents is made between personal and task agents. Personal agents are permanent and contain all known personal information. However, personal agents delegate once to be used task agents that have a limited scope and lifetime and, thus, need to contain only the information that is relevant for the given task. Therefore, the personal information is hidden whenever it is not explicitly required.

Both the personal agents and the task agents that perform the dedicated tasks for a user (data subject) are not anonymous since they conform to the FIPA specifications and thus have a name by which they can be identified. It should not be possible that the identity of the data subject can be derived from the agent name. They can thus be seen as pseudo identities of the data subject. Only the personal agent of the data subject knows for whom the task agent is acting. Performing communication analyses within the ISA platform could reveal the relationships between user agents and task agents. A networking approach, which is based on onion routing, should solve this problem.

A task agent carries only the personal data it needs for the task it is performing. Before sending personal information to another ISA it makes sure that the receiver conforms to the user's preferences and the privacy principles before sending any information. Typically, a service agent would receive only personal data level 2. Another task or personal agent might get involved to transfer more sensitive information only after consent is reached at this level.

Task agents are also a conditio sine qua non due to the fact that the DPD requires explicit consent when personal data are exchanged and disseminated to third parties. A general purpose or implicit consent once given to a generic personal agent is not sufficient to fulfill the legal requirements.

Even when the task agent discloses Contact Information, this information might still hide the real identity of the data subject. The Contact Information might only contain an e-mail address for example, by which the data subject is not directly identifiable.

8.6.2 ISA certificates

In order to let ISAs authenticate each other, each ISA needs its own certificate. The certificate also forms the basis of other security solutions including securing communication and data storage. A special Registration Authority Agent issues the agent certificates. This ISA uses a registration authority component that is provided by an ISA-PKI. Apart from issuing certificates it also revokes and checks the validity of ISA certificates.

8.6.3 Agent Practises Statement (APS)

The user of the PISA will address his or her preferences on how the PISA should handle his or her personal data during the initiation of the PISA. These preferences will be added to the Agent Practises Statement (APS). The APS states the privacy policy of the ISA. The default level for the ISA in the EU is the DPD. The APS is built of policies for all the principles and, thus makes statements about how to handle personal data with respect to these principles. Both the APS and the preferences are described using the privacy ontology that is derived from the privacy principles. Chapter 10 describes how the ontology for the principles V 2, V 3, V 4, V 6, V 7 and V 9 (partly) is developed.

Signing the APS Before an ISA can send personal data to another ISA it must first request the APS of the personal data receiving ISA. The ISA that intends to send the personal data must make sure that the APS it receives is really the APS of its communication partner. This can be accomplished if the controller of the personal data receiving ISA has signed the APS and the certificate of the controller is known by a Trusted Third Party, the Registration Authority -ISA. The ISA that intends to send the personal data will have to check the signature of the APS with the Registration Authority ISA. An overview of the different PISA components is given in Figure 8.5.

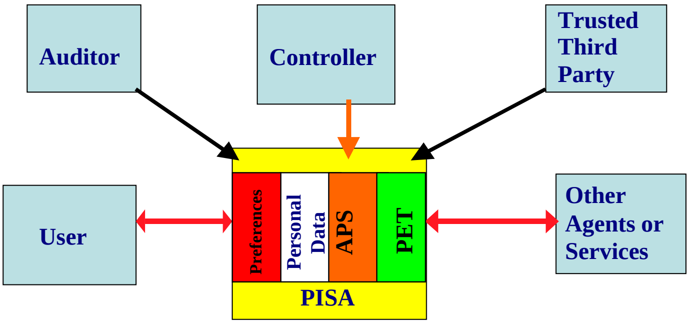


Figure 8.5: Overview of PISA components.

Chapter 9

Trust model and network aspects

L. Korba
Larry.Korba@nrc-cnrc.gc.ca

A. Patrick
Andrew.Patrick@nrc-cnrc.gc.ca

R. Song
Ronggong.Song@nrc.ca

NRC, Canada

9.1 Trust model – Agents and Trust

It is clear that a trusting relationship must develop between the user and the agent [PK03]. Users must be confident that the agent will do what they have asked, and only what they have asked. Moreover, to be effective the agent must be trusted with sensitive information, and use it only in appropriate circumstances. Since the trust between a user and an agent is so important, it is useful to examine the nature of trust in detail.

9.1.1 What is Trust?

Most generally, trust can be defined as "a generalized expectancy. . . that the word, promise, oral or written statement of another individual or group can be relied upon" (Rotter, 1980, p. 1) [Rot80]. In the context of software agents, this means that the agent can be relied upon to do what it was instructed to do. But trust is more than that; it is "the condition in which one exhibits behavior that makes one vulnerable to someone else, not under one's control" [Zan72]. Without the vulnerability, there is no need for the trust. In the context of software agents, it means no longer controlling the software directly, letting the process act on one's behalf, and accepting the risks that this may entail. Bickmore and Cassell [BC01] go on to describe trust as "people's abstract positive expectations that they can count on [agents] to care for them and be responsive to their needs, now and in the future" (p. 397).

This concept of making oneself vulnerable in order to accomplish a goal is essential for understanding trust. Without trust virtually all of our social relationships would fail and it would become impossible to function normally. If we can not trust the oncoming driver to stay in their lane, then it would become impossible to drive. If we do not trust the shopkeeper to deliver the goods we pay for, then simple purchases would become very awkward. We make ourselves vulnerable to others every day, but we are usually comfortable in doing so because we trust that their actions will not be inappropriate or harmful. Bickmore and Cassell (2001) describe trust as a process of uncertainty reduction. By trusting others to act as we expect them to act, we can reduce the things we have to worry about.

Taking a computer science approach, Marsh [Mar94] has defined trust in terms of the behavior of the person doing the trusting. Thus, trust is "the behavior X exhibits if he believes that Y will behave in X's best interest and not harm X". In the context of agents, this means behaving in a way that is appropriate if the agent will always have your best interests in mind, and cause you no harm.

For our purposes, then, trust can be defined as *users' thoughts, feelings, emotions, or behaviors that occur when they feel that an agent can be relied upon to act in their best interest when they give up direct control.*

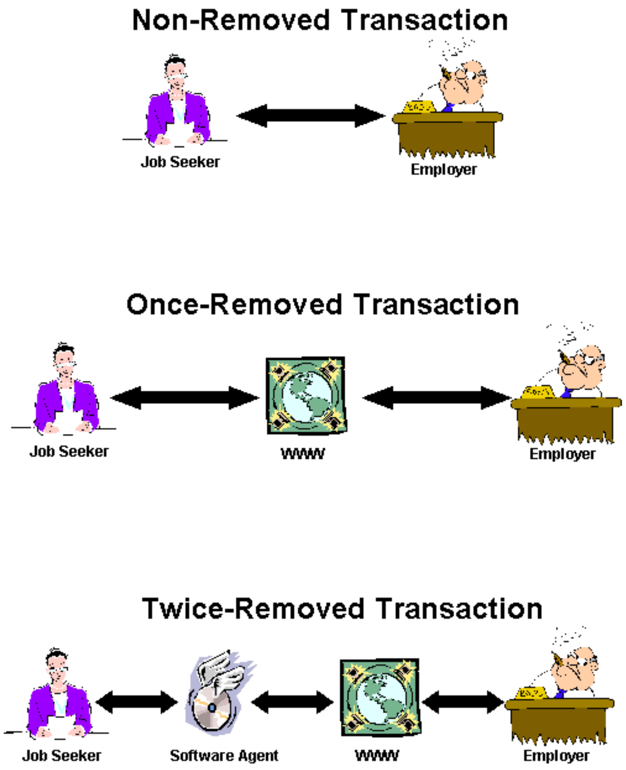
9.1.2 The Problem of Trusting Agents: Interactions Twice-Removed

Users may have difficulty trusting software agents because the user ends up working on a task that is twice-removed from the interface, see Figure 9.1. Consider the example of a user who is using a job-searching agent. A traditional, non-removed method of searching for a job would be to talk to employers directly, perhaps by visiting their offices. Here the job seeker is interacting directly with the potential employer to get information about the position (the top panel in Figure 9.1). A more modern method of searching for a job is to work in a computer-mediated fashion where the job seeker interacts with a computer program, perhaps a WWW browser, to view information that has been created by the employer (the middle panel in Figure 9.1). Thus, the interaction between the job seeker and the employer is once-removed. Riegelsberger & Sasse [RS01] refer to this as a dis-embedded transaction. With a job-searching agent, the job seeker would interact with a computer program, perhaps an agent control interface, to provide instructions to the agent. The agent, in turn, would search the Internet and gather information that has been provided by the employer. There is no direct connection between the user and the job-seeking activities; the bottom panel in Figure 9.1. Thus, the interaction between the job seeker and the potential employer is twice-removed (or dis-dis-embedded).

Research has shown that developing trust during once-removed interactions can be difficult, let alone trusting in twice-removed interactions. For example, Rocco [Roc98] showed that interpersonal trust is reduced markedly when communication is computer-mediated. Also, a numbers of studies, to be summarized below, have found that it can be quite difficult to develop trust during once-removed e-commerce interactions.

There are many valid reasons why users may be hesitant to trust software agents. Cheskin [Arc] argued that disclosing personal information might involve more personal risk than financial interactions because personal assets like self-respect, desirability, reputation, and self-worth can be more valuable than money. Also, since agents operated autonomously outside of the user's vision and control, things may go wrong that the user does not know about, or cannot correct.

Youll [You01] has also described the issues involved in trusting agents. First, the user must make their instructions clear to the agent. This instructing phase could fail for a number of reasons: (1) the user does not clearly define the instructions, (2) the agent does not fully understand the instructions, or (3) the user and the agent interpret identical instructions differently.



Non-Removed Transaction

Once-Removed Transaction

Twice-Removed Transaction

Figure 9.1: Explanation of twice-removed transactions.

Second, if the instructions have been understood, the user must be confident that the agent will execute its instructions properly, and only perform the tasks that the user intended. Third, the user must be confident that the agent will protect information that is private or sensitive. Finally, regarding the confidentiality of the information entrusted to the agent, the user must have confidence that the agent is not attacked or compromised in some way, such as through "hacking" or "sniffing". With all of these concerns, developing a trusting relationship between users and their agents is a difficult task.

On the other hand, there are also valid reasons why users might make the choice to trust agents. Again Youll [You01] describes the advantages that agents can bring to a task. Due to the twice-removed nature of the interactions between the end-user and the task, agents are well suited for tasks that require high degrees of privacy. An agent can establish its own identity on the network, and protect the identity of the end-user. An example of how this can be done was seen in the Lucent Personalized Web Assistant (LPWA; Gabber, et al. [GGMM97]), which acted as a proxy for users who wanted to navigate the WWW without revealing their true identities. Such services can even go so far as to establish new pseudonyms for each and every transaction, making it very difficult to establish a link back to the user.

Agents are also well suited for situations where interaction policies need to be established and followed. Since software agents are embodied in explicit computer code, it is possible to establish and follow clearly defined privacy policies, rather than relying on heuristics or emotions.

9.1.3 Building Successful Agents: A Summary Model

Most of the research to date on privacy and trust has been focused on (once-removed) e-commerce interactions. However, the lessons are very relevant and extendable to agent interactions, and they provide a good starting point until more research is conducted on agent technologies. An important contribution to research on e-commerce trust is a path model of e-commerce customer loyalty proposed by Lee, Kim, and Moon [LKM00], as is shown in Figure 9.2. These authors describe how attitudes towards e-commerce will be determined by the amount of trust instilled in the user, and the amount of cost perceived by the user. Trust and cost combine together, in opposite directions, to determine the overall acceptance. In addition, Lee et al. identify a number of factors that contribute to trust, such as shared values and effective communication. They also identify factors that lead to perceived cost, such as the level of uncertainty.

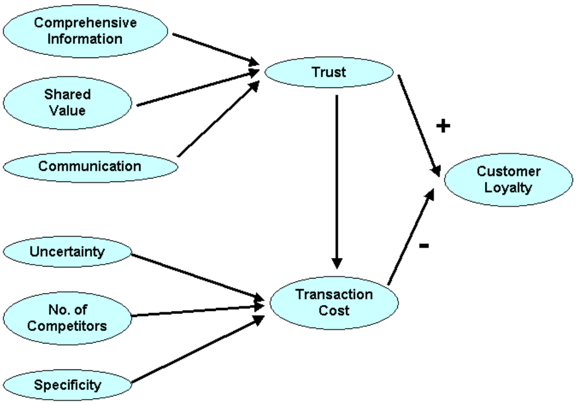


Figure 9.2: Lee, Kim, & Moon’s model of e-commerce loyalty.

An extended model of agent acceptance developed for this paper is shown in Figure 9.3. Here acceptance of the agent technology is determined by the combination of trust and perceived risk. The contributing factors identified by Lee et al. are included, along with factors identified by other researchers. This section reviews this model of agent acceptance in detail.

An important feature of Lee et al.’s e-commerce model, and the model of agent acceptance proposed here, is the separation of trust from perceived risk. The idea is that feelings of trust and risk can be established quite independently, and together they determine the final success of the agent technology. Trust contributes to the acceptance of the agent in a positive direction, while risk contributes in a negative direction. The effect is that the two factors interact with each other, so that agents instilling a low degree of trust may still be successful if there is also a low perceived risk. On the other hand, in very risky situations it may be that no amount of trust will offset the risk perceived by the user, and the agent will

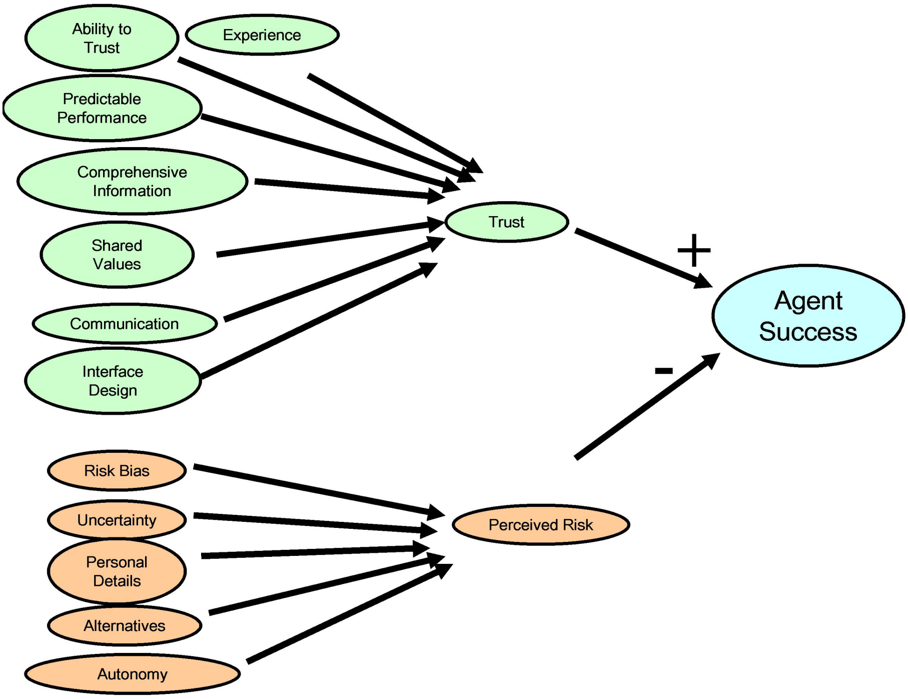


Figure 9.3: A model of agent success.

never be accepted. Rotter [Rot80], in his review of the social psychology of interpersonal trust, supports this idea that trust and risk are separate concepts, and both contribute to the final behavior of an individual. Grandison and Sloman [GS00] also describe trust and risk as opposing forces that combine during decision making about a service or an e-commerce transaction.

Another important feature of the model is that the risk being described is the risk perceived by the user. This perception may, or may not, be related to the actual risk of the technology employed in the agent system. For example, the job-seeker's personal information might be encrypted with a very strong encryption technique, but if the user believes that the information will be disclosed inappropriately, this fear contributes to the perceived risk, and works against acceptance of the agent technology.

9.1.4 Factors Contributing to Trust

As is shown in Figure 9.3, trust is a complex, multifaceted concept that is influenced by a number of factors (e.g., Grandison & Sloman [GS00]). In this section a number of factors contributing to feelings of trust are described, and specific design recommendations are made for building trustworthy agents.

Ability to Trust

The first factor that contributes to the trust a user may place in an agent service is their ability to trust. A number of researchers have proposed that people have a general ability to trust that forms a kind of baseline attitude when they approach any trust situation, and some people have a higher baseline level of trust than others. For example, Marsh [Mar94] describes "basic trust" as a person's general propensity to trust or not trust. This basic trust is part of their personality, and is one of the factors that contribute when making decisions about trust. Similarly, Rotter [Rot80] showed that there is a generalized trust that is "a relatively stable personality characteristic" (p. 1). Rotter also demonstrated that high and low trusters had markedly different opinions and behaviors (e.g., high trusters were less likely to cheat or lie, were seen as happier, and more attractive).

Directly related to the issue of trust on computer networks, Cranor et al. [CRA99] surveyed Internet users about their attitudes towards privacy and trust. The survey respondents were then classified into groups that differed in their concerns about online privacy, following a scheme originally proposed by Westin [Wes91]. The first group (27%) was only marginally concerned with online privacy and was quite willing to provide personal information when visiting WWW sites. This group did have some concerns, such as the desire to remove themselves from marketing mailing lists, but they were generally quite trusting. The second group (17%) was at the opposite extreme, and was labeled "privacy fundamentalists". These users were extremely concerned about privacy and were generally unwilling to provide any information to WWW sites, even when privacy protection measures were in place. The third and largest group (56%) was labeled the "pragmatic majority" because they had some concerns about privacy, but also had developed tactics for dealing with those concerns. For example, they would often look for privacy protection methods or statements when navigating the WWW.

Thus, we have abundant evidence that people differ in their basic tendency to trust. Perri 6 (yes his surname is the numeral 6) [601] cautions, however, that basic trust can be misleading because people's perceptions are heavily modified by the context. He suggests the people's trust can change quickly depending on the context and their experience, and it is important not to overemphasize the role of general personality characteristics.

When building agent systems that users will have to trust, developers should take into account the fact that users may differ in their general ability to trust. Some users may willingly trust an agent system with little reassurance of the privacy protection measures in place, while others may be very reluctant to give their trust. This means that interfaces must be flexible and be able to provide more information and reassurance for users that require it.

Experience

The second factor that contributes to trust is experience. It is clear that users can change their willingness to trust based on their experiences [Mar94, 601]. If they have been harmed in some way, for example, they may be less willing to trust in the future. This change in trust may be specific to the situation or it may be a change in their general ability to trust. Changes in trust can also come about indirectly because of the experiences or recommendations of others [GS00]. This means that trust can be "transitive", being passed from user to user.

Designers of agent systems should ensure that users are able to have positive experiences so they can develop trust. This means providing ample information on the operation of the agent (feedback). In addition, designers should support a sharing function so users can relate their experiences and trusting attitudes can be shared and spread (assuming the experiences are positive ones). This may mean collecting testimonials or anecdotes that can be shared with other users.

Predictable Performance

Another factor that can lead to agent trust is predictable performance. Systems and interfaces that perform reliably and consistently are more likely to be trusted by users. Bickford [Bic97] describes three important principles for predictable performance and its role in building trust:

1. consistency: The interface and system behave the same way each time they are used. For example, certain functions are always accessed in the same way, and always lead to the expected result.
2. aesthetic integrity: The interface has a consistent look and feel, throughout the entire system. This includes the page design, buttons, text styles, etc.
3. perceived stability: The system should appear stable to the user. It should not crash. There should be no changes without users' knowledge, and users must be kept informed about any operational issues, such as upgrades or downtimes.

Another aspect of predictable performance is response time. Users prefer response times that are consistent and predictable, rather than variable and unpredictable [Shn97].

The resulting recommendation is that developers should ensure that the interface is consistent and predictable. This may mean adopting a style guide or interface guideline that is used in all parts of the system. Developers should also ensure that the system behaves consistently, and appears to be stable.

Comprehensive Information

Another important factor in determining users' trust of a system is the amount of information provided. Systems that provide comprehensive information about their operation are more likely to be understood, and more trusted. Norman [Nor01] suggests that agent systems must provide an image of their operation so that users can develop a mental model of the way the system works. It is through this model that they will develop expectations and attitudes towards the system. Norman agues that users will develop mental models and assumptions about the system even when no information is provided, and these models may be wrong. To prevent this, developers should explicitly guide the model development by showing the operation of the system.

The importance of internal models of system operation was recently demonstrated by Whitten & Tygar [WT99]. This study tested users ability to use a PGP system to certify and encrypt e-mail. The results showed that the majority of the users were unable to use the system to perform the task. In fact, 25% of the users e-mailed the secret information without any protection. An analysis of the errors and an evaluation of the interface led these researchers to conclude that the major source of the problems was that users did not understand the public key model used in the PGP system. The PGP interface that was tested failed to provide the comprehensive information about how public key encryption works, and the roles and uses of public and private keys. Without this information, users often developed their own ideas about how the system worked, with disastrous results.

Another example of a system that does not provide comprehensive information is the "cookies" module used in WWW browsers [Bic97]. Cookies are small files that are assembled by WWW sites and stored on users' computers. Later, they can be retrieved by the WWW sites and used to identify repeat visitors, preferences, and usage patterns. The problem with cookies is that they can store a variety of information about users (including sensitive information) and yet their operation is invisible. Unless users explicitly change their browser options, they do not know when cookies are created or retrieved. In addition, most WWW browsers do not provide any way of viewing the cookies. They omit such simple functions as listing what cookies are stored on a system, and an ability to view the information stored within them. The P3P initiative [RC99] is an attempt to give users more control over cookies.

Developers of agent technologies must provide comprehensive information about how the system works. The role of the agent must be explained, and its operation must be obvious. This may mean allowing users to observe and track the actions performed by an agent, both in real-time and after the fact. In addition, effective interfaces should be developed for viewing and altering the information stored by agents.

Shared Values

Another factor that can lead to users trusting agents is the establishment of shared values between the user and the agent. That is, to the extent that the user feels that the agent values the things that they would, they will have more trust in the agent. In interpersonal relationships, these shared values are often built through informal interactions, such as the small talk that occurs in hallways or during coffee breaks. Bickmore and Cassell [BC01] tested the role of small talk (informal social conversation) in building trustworthy agents. These researchers included small talk capabilities in a real estate purchasing agent called REA. REA was a life-sized conversational agent embodied as an animated figure on a large computer screen. REA was able to engage in small talk conversations designed to increase feelings of closeness and familiarity. For example, REA conversed about the weather, shared experiences, and her laboratory surroundings. In one experiment, a condition that included small talk interactions was compared with another condition that only involved task-oriented interactions. The task in the experiment was to determine the users' housing needs, and this included gathering personal information about how much the user could afford to spend, and how large a house was required. When measures of trust and willingness to share personal information were examined, the results showed that the condition that involved informal social dialogues led to higher levels of trust among extroverted users (it is not clear why this effect was not found for introverted users).

Values between agents and their users can also be shared explicitly. For example, privacy policies can be clearly articulated so that users can compare their concerns with the policies in place [Arc].

Communication

Another factor that determines the amount of trust is the amount and effectiveness of communication between the agent and the user. Norman [Nor97] argues that continual feedback from the agent is important for success. This feedback should include having the agent repeat back its instructions so it is clear what the agent understood. Also, error messages should be constructed so that it is clear what was understood, and what needs to be clarified. In addition, through communication it should be made clear what the capabilities and limits of the agent are.

Interface Design

The final factor that can contribute to trust of an agent is the design of the interface itself. This means the look and feel of the software that is used to control the agent. This area includes such factors as appearance, functionality, and operation. Many of the generic attributes of good interface design also apply to designing agent interfaces. So, Norman's [Nor90] recommendations about "visible affordances" are relevant here, which means that whenever possible the function of an interface component should be clear from its visible appearance.

9.1.5 Factors Contributing to Perceived Risk

The other side of the model of agent success is perceived risk. Other things being equal, users will be more willing to use agent systems if they perceive the risks to be lower. The amount of perceived risk is influenced by a number of factors.

Risk Perception Bias

Similar to basic trust, users may have a basic or baseline level of perceived risk. This is probably best described as a bias to perceive situations as being risky or risk free.

Uncertainty

Another method to reduce risk perception is to reduce uncertainty. The more users know about a system and how it operates, the less they worry about taking risks (assuming all that they learn is positive). This is highly related to the "comprehensive information" and "communication" factors for building trust.

Personal Details

An obvious factor in risk perception is the amount of sensitive information being provided. If more personal details are being provided to the agent, perceptions of risk are likely to increase. System developers should only ask for information that is necessary to do the job, and avoid where possible information that may be especially sensitive. Exactly what information the users consider sensitive may require some investigation. For example, Cranor, Reagle, & Ackerman [CRA99] found that phone numbers were more sensitive than e-mail addresses because unwanted phone calls were more intrusive than unwanted e-mail messages.

Alternatives

Another factor that can lead to feelings of risk is a lack of alternative methods to perform a task. For example, if the only method to search for a job is to use a new agent technology, users may feel they are taking more risks than situations where there are multiple methods (i.e., non-agent WWW interfaces, phone calls, employer visits).

Specificity

Similarly, if there is a sole supplier of a service, users may feel they are at more risk from exploitation than situations where there are multiple suppliers. In the job-searching example, it means that users may be more comfortable if there are multiple job searching agents to choose from.

Autonomy

Perhaps the most important factor in determining users' feelings of risk towards an agent technology is the degree of autonomy granted to the agent. As discussed previously, agents can range from low risk advice-giving systems to higher risk, independent acting agents. Lieberman [Lie02] advocates developing advice agents and avoiding, for now, agents that truly act on their own. Advice systems have the advantage that they can stay in close contact with the user and receive further instructions as they operate. Further, advice agents can learn by example as they monitor what advice their users accept. In the job-searching example, it may be most appropriate for the agent to suggest possible jobs that the user should apply for, rather than completing the application autonomously.

9.2 Network aspects

This section discusses several aspects related to the use of a communication network by agents: scalability, traffic analysis, and anonymous communication.

9.2.1 Scalability

The aspect of scalability is not directly related to privacy or software agents but is sufficiently important to be mentioned. In the context of this handbook scalability can be defined as the potential of a PET solution to scale within a network. When a PET solution is scalable it means that when the size of the problem is increased, e.g. more users or agents take part, the PET solution does not require a disproportional amount of network resources.

To illustrate this: a solution where each agent is required to communicate with all other agents in a system would not be scalable. With 10 agents there would be $9 \times 9 = 81$ communications, with 1000 agents this would be already 998001! The solution has an order of N^2 and thus does not scale. This is obviously a simple example. Real PET solutions might be more difficult to analyze with respect to scalability. Yet, a non-scalable PET solution might not necessarily prove problematic as other system limitations, or non-scalability aspects, might pose a worse problem. Obviously PET demands more system resources than a solution without PET would but as long as the additional amount of resources is not disproportional, the solution will be scalable.

9.2.2 Traffic analysis

Traffic analysis is a serious menace to agent-based applications. An adversary can monitor and compromise certain parts of a distributed agent system by matching a message sender with the receiver. Protecting the disclosure of communication partners or the nature of communication between partners is an important requirement for confidentiality in an e-business context. It is also a property desired by agent users who want to keep their agent lives and relationships private. On the other hand, since most agent platforms use global name service in order to provide global tracking service, it makes traffic analysis attacks simple. The major attacks are described as follows:

- **Communication Pattern Attack:** An adversary may discover considerable useful information simply by tracking the communication patterns when agents send and receive messages.
- **Message Coding Attack:** An adversary can easily link and trace some messages if the messages do not change their coding during transmission.
- **Timing Attack:** An adversary can observe the set of messages coming into the network and the set of messages going out of it, to obtain some useful route timing information by correlating the messages in the two sets.
- **Packet Volume Attack:** An adversary can observe the amount of transmitted data (e.g. the message length, number of messages).
- **Metadata Attack:** An adversary can find the identity of an agent from metadata even if the data itself is not accessed in any way.
- **Message Delaying:** The adversary can delay messages to obtain some information regarding how data is handled within a communication network.
- **Intersection Attack:** An adversary may trace some agents by observation over a long period searching for special distinguishable behavior.
- **Collusion Attack:** A corrupt coalition of agents or parts of the system may be able to trace some agents.
- **Denial of Service Attack:** An adversary may obtain some information about the routes used by certain agents by rendering some nodes inoperative.
- **Replay Attack:** An adversary, who observes the incoming and outgoing messages, would capture and replay a message to the node to try to take it over.

Privacy for e-commerce has been recognized as a vital requirement for many years. However, TCP over IP version 4 is designed to allow computers to easily interconnect and to assure that network connections will be maintained even when various links may be damaged. This same versatility makes it rather easy to compromise data privacy in networked applications. For instance, networks may be sniffed for unencrypted packets, threatening the confidentiality of data, or using the attacks listed above, wherein the nature of a communication or information about the communicators may be determined. Research has led to techniques that provide varying levels of private communication between parties. The next section describes some of the more commonly known network privacy technologies concisely.

9.2.3 Anonymous communication

The primary goal of an anonymous communication network is to protect user communication against traffic analysis. In [Sim96] Simon proposes a formal model for an anonymous communication network. It is assumed that parties can communicate anonymously. In the simplest of such models, parties can send individual messages to one another anonymously. A stronger assumption is that parties receiving anonymous messages can also reply to them. An intermediate model allows one or more parties to broadcast messages efficiently and thus to reply to anonymous ones without jeopardizing that anonymity. However, Simon's model assumes that reliable, synchronous communication is possible. While this simplifying assumption may be unrealistic, it is not actually exploited in his proposed protocol. Rather, the assumption of synchrony serves to discretize time, abstracting out the issue of communications delays without preventing adversaries from taking advantage of them, since messages arriving during the same time period are queued in arbitrary order, to make it appear as though any one of them might have arrived first.

Anonymous communication has actually been studied fairly extensively. For example, in order to enable unobservable communication between users of the Internet, Chaum [Cha81a] introduced MIX networks in 1981. A MIX network consists of a set of MIX nodes. A MIX node is a processor that accepts a number of messages as input, changes their appearance and timing using some cryptographic transformation, and outputs a randomly permuted list of function evaluations of the input items, without revealing the relationship between input and output elements. MIXes can be used to prevent traffic analysis in roughly the following manner.

1. The message will be sent through a series of MIX nodes, say $i_1, i_2, \dots, i_d$. The user encrypts the message with an encryption key for MIX node i_d, encrypts the result with the key from MIX node i_{d-1} and so on with the remaining keys.

2. The MIX nodes receive a certain number of these messages, which they decrypt, randomly reorder and send to the next MIX node in the routes.

Based on Chaum's MIX networks, Wei Dai has described a theoretical architecture that would provide protection against traffic analysis based on a distributed system of anonymizing packet forwarders. The architecture is called Pipenet [Dai00]. Pipenet consists of a cloud of packet forwarding nodes distributed around the Internet; packets from a client would be encrypted multiple times and flow through a chain of these nodes. Pipenet is an idealized architecture and has never been built. Pipenet's serious disadvantage is that its packet loss or delay would be extremely bad.

Like the Pipenet architecture, the Onion Routing network [GRS99] has been proposed and implemented in various forms. It provides a more mature approach for protection of user anonymity against traffic analysis. The primary goal of Onion Routing is to provide strongly private communications in real time over a public network at reasonable cost and efficiency. In Onion Routing, instead of making socket connections directly to responding machine, initiating applications make connections through a sequence of machines called onion routers. The onion routing network allows the connection between the initiator and responder to remain anonymous. These connections are called anonymous socket connections or anonymous connections. Onion Routing builds anonymous connections within a network of onion routers, which are, roughly, real-time Chaum MIXes. While Chaum's MIXes could store messages for an indefinite amount of time while waiting to receive an adequate number of messages to mix together, a Core Onion Router is designed to pass information in real time, which limits mixing and potentially weakens the protection. Large volumes of traffic can improve the protection of real time MIXes. Thus with Onion Routing, a user directs his applications to contact application proxies that form the entrance to the cloud of nodes. The application proxy will then send an onion packet through a string of Onion Routers in order to create a route through the cloud. The application proxy will then forward the application data along this route through the cloud, to exit on the other side, and be delivered to the responder the user wishes to connect.

The Freedom network [BGS01, BSG00] was an anonymity network implemented on a worldwide scale and in use as a commercial privacy service from early 1999 to October, 2001. It was composed of a set of nodes called Anonymous Internet Proxies (AIP) that ran on top of the existing Internet. It not only used layers of encryption, similar to the MIX network and Onion Routing, but it also allowed users to engage in a wide variety of pseudonymous activities such as multiple pseudonyms, hiding the users' real IP address, e-mail anonymity, and other identifying information, etc. A key difference between the Freedom Network and Onion Routing is that the last node replaces the missing IP source address, which was removed by the sender, with a special IP address called the wormhole IP address.

As a lighter weight alternative to MIXes, Reiter and Rubin propose Crowds system [RR98b] in 1998. The goal of the Crowds system is to make browsing anonymous, so that information about either the user or what information he or she retrieves is hidden from Web servers and other parties. The Crowds system can be seen as a peer-to-peer relaying network in which all participants forward messages. The approach is based on the idea of "blending into a crowd", i.e., hiding one's actions within the actions of many others. To execute web transactions in this model, a user first joins a crowd of other users. The user's initial request to a web server is first passed to a random member of the crowd. That member can either submit the request directly to the end server or forward it to another randomly chosen member, and in the latter case the next member independently chooses to forward or submit the request. The messages are forwarded to the final destination with probability p and to some other members with probability $1 - p$. Finally, the request is submitted to the server by a random member, thus preventing the end server from identifying its true initiator. Even crowd members cannot identify the initiator of the request, since the initiator is indistinguishable from a member that simply passed on a request from another. Crowds system can prevent a Web server from learning any potentially identifying information about the user, including the user's IP address or domain name. Crowds also can prevent Web servers from learning a variety of other information, such as the page that referred the user to its site or the user's computing platform.

Recently, Freedman and Morris [FM02] propose a peer-to-peer anonymous network called Tarzan. In comparison with the Onion Routing and Freedom network, Tarzan uses the same basic idea to mix traffic, but achieves IP-level anonymity by generic and transparent packet forwarding, and also sender anonymity like Crowds system by its peer-to-peer architecture that removes any notion of entry-point into the anonymizing layer. In Tarzan, the system is design to involve sequences of MIX relays chosen from a large pool of volunteer participants. All participants are equal peers, i.e., they are all potential originators as well as relays of traffic. The packets are routed through tunnels involving sequences of Tarzan peers using MIX-style layered encryption. One of the ends of the tunnel is a Tarzan peer running a client application; another is a server-side pseudonymous network address translator to change the private address to a public address.

The above network-based approaches could be used to protect the users' privacy against the traffic analysis attacks, and satisfy the requirements of the privacy protection for network transaction in the agent-supported distributed learning environments. One example is the anonymous communications for mobile agents [KSY02] proposed in MATA'02. It may appear that anonymity networks are quite a dramatic approach to take for a distance learning environment. Yet consider the growth in the number of company's providing and taking advantage of outsourced technical training. Companies use this training to retool their workforce for new initiatives. Competitors would gain information regarding the plans of others if they could pinpoint the key people taking courses from distance learning providers. In this situation it is easy to see that anonymous networking would be a value-added service that would be provided by the distance learning service provider.

As a closing remark, threats against these anonymous communication networks have been discussed in [Ray00, SK02]. Generally it is a hard problem to achieve unconditional untraceability and anonymity for real-time services on the Internet, especially when we assume a very strong attacker model.

9.3 Publications

This section lists the publications of NRC related to the PISA project.

9.3.1 Journal Papers

- Patrick, A. Building Trustworthy Software Agents, IEEE Internet Computing, Nov./Dec. 2002.
- Korba, L., El-Khatib, K., Patrick, A., Song, R., Xu, Y., Yee, G., Privacy and Security in E-Learning, International Journal of Distance Education Technology, Idea Publishing Group. NRC Paper Number: NRC 44786
- Kenny, S., Korba, L. Adapting Digital Rights Management to Privacy Rights Management, Computers & Security, Vol. 21, No. 7, November 2002, 648-664.
- R. Song and L. Korba. Cryptanalysis of Scalable Multicast Security Protocol. Journal of IEEE Communication Letters, 2003. NRC 45826.
- R. Song and L. Korba. Pay-TV System with Strong Privacy and Non-repudiation. IEEE Transactions on Consumer Electronics, Vol. 49, No.1, May. 2003. NRC 45832.
- Song, R., Korba, L., Privacy and Mobility of the Internet Key Exchange Protocol, Computer Communications: Special Issue on Network Security, Spring, 2002.

9.3.2 Book/Chapters

- Korba, L., El-Khatib, K., Patrick, A., Song, R., Xu, Y., Yee, G., Privacy and Trust in Agent-Supported Distributed Learning, in the book: Designing distributed Learning Environments with Intelligent Software Agents, Idea Group, Inc. Publishers, 2003 (In print).

9.3.3 Conference proceedings

- El-Khatib, K., Korba, L., Ronggong, S., and Yee, G.: Secure Dynamic Distributed Routing Algorithm for Ad Hoc Wireless Networks. Accepted for the First International Workshop on Wireless Security and Privacy, Kahosiung, Taiwan, October 6-9, 2003.
- Korba, L., Song, R., A Reputation Evaluation System for Mobile Agents, Proc. for the 5th Int. Workshop on Mobile Agents for Telecommunications Applications, Marrakech, Morocco, October 8-10, 2003 (Invited Paper).
- Song, R., Korba, L. Security Communication Architecture for Mobile Agents and E-Commerce, Ninth International Conference on Distributed Multimedia Systems, Miami, Florida, Sept. 24-26, 2003.
- Yee, G., Korba, L., Feature Interactions in Policy-Driven Privacy Management, Proc. Of 7th Int. Workshop on Feature Interactions in Telecommunication and Software Systems, June 11-13, 2003, Ottawa, Canada.
- Korba, L., El-Khatib, K., Patrick, A.S., Song, R., Xu, Y., Yee, G., & Yu, J., Agent-based systems and privacy enhancing technologies. Paper to be presented at 18th IFIP International Information Security Conference, 26-28 May 2003, Athens, Greece (Conference Proceedings).

- Yee, G., Korba, L., The Negotiation of Privacy Policies in Distance Education, Proc. 14th IRMA International Conference, May 18-21, 2003, Philadelphia, USA (Conference Proceedings)
- Yee, G., Korba, L. Bi-Lateral E-Services Negotiation Under Uncertainty, The 2003 International Symposium on Applications and the Internet, January 27-31, 2003, Orlando, Florida.
- Korba, L., Song, R., Patrick, A.S., Yee, G., Xu, Y., & El-Khalid, K., Developments in Privacy Enhancing Technologies. 15th Annual Canadian Information Technology Security Symposium (CITSS) , May 12-15, Ottawa (Conference Proceedings).
- Patrick, A.S., & Kenny, S., From Privacy Legislation to Interface Design: Implementing Information Privacy in Human- Computer Interfaces. Paper presented at the Privacy Enhancing Technologies Workshop (PET2003), Dresden, Germany, 26-28 March, 2003 (Conference). To appear in a future issue of Lecture Notes in Computer Science (LNCS).
- Korba, L. Privacy technology for Distributed Computing: A Research Perspective. Canadian Security Symposium, Ottawa, ON, May 13-18, 2002 (Invited)
- Korba, L., Song, R., Network Based Approaches for Privacy, Canadian Security Symposium, Ottawa, ON, May 13-18, 2002
- Korba, L. Privacy in Distributed Electronic Commerce, Proc. 35th Hawaii International Conference on System Science (HICSS), Hawaii, January 7-11, 2002.
- Korba, L., Cullins, M., Ronggong, S., and Kenny, S. Canadian and European Union Joint Projects: Experience from Canadian and European Perspectives. Proc. of the Electonic Imaging and The Visual Arts Conference, EVA 2001, Montreal, October 4, 2001. (Invited)
- Korba, L. Towards Distributed Privacy for CSCW, Proc. 6th International Conference on CSCW in Design, London, Ontario, July 12-14, 2001.
- Korba, L. Kenny, S. Towards Meeting the Privacy Challenge: Adapting DRM, DRM 2002, Washington, D.C., November, 2002.
- Korba, L., Song, R., Yee, G. Autonomous Communication for Mobile Agents, Mobile Agents for Telecommunication Applications, October 22-24, 2002, Barcelona, Spain.
- Xu, Y., Korba, L. A Trust Model for Distributed E-Learning Service Control, World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education (E-Learn 2002), Montreal, Canada, Oct. 15-19, 2002.

9.3.4 Talks

- Korba, L. Rights Management for Privacy, 12th Annual Canadian Association for Cryptographic Research Conference, Toronto, Ontario, November 6-7, 2003 (Invited).
- Patrick, A.S., Building usable privacy protection. Invited address to the U.S. Federal Trade Commission Workshop on Technologies for Protecting Personal Information: The Consumer Experience, Washington, DC; May 14 (Invited).
- Korba, L., Privacy in a Peer-to-Peer World, CITO TechTalk: Peer-to-Peer Technologies, Ottawa, Ontario, February 25, 2003. (Invited)
- Korba, L., S. Kenny, Design Embedded Privacy Risk Management, 11th Annual Canadian Association for Cryptographic Research Conference, Toronto, November 22-23, 2002. (Invited)
- Korba, L., Changes in Technology and Law and the Impact upon Research in Security and Privacy, CITO e-Applications TechTalk, Toronto, November 21, 2002. (Invited)

9.3.5 Reports

- Korba, L., Song, R., Yu, Jiong, Report D24-2 Testing of PISA Scalability (WP5.4-2), August 30, 2003, 42 pages.

- Patrick, A.S & Holmes, C. Report D24-1 Testing of User Interface for the Privacy Agent, June, 2003, 24 pages.
- Patrick, A.S., & Kenny, S., Agent user interfaces and documentation. Report D.23 by the Privacy Incorporated Software Agent (PISA) Research Consortium submitted to the European Union and published electronically, January, 2003, 138 pages.
- Song, R. Korba, L. Modeling and Simulating Scalability of A Multi-Agent Application System, July, 2002, NRC Technical Report ERB-1098, NRC No.44952, 28 pages.
- Korba, L., Song, R. PISA System Scalability, WP5.2 Deliverable for PISA Project, May, 2002, 67 pages.
- Korba, L., Song, R. Network Threats and Opportunities for Privacy, WP5.1 Deliverable for PISA Project, August 2001, 76 pages.
- Song, R., Korba, L. Anonymous Internet Infrastructure based on PISA Agents, October, 2001, NRC Technical Report ERB-1090, NRC No.44899, 21 pages.
- Korba, L. Song, R. Investigation of Network-Based Approaches for Privacy, November, 2001, NRC Technical Report ERB-1091, NRC No.44900, 32 pages.
- Korba, L., Song, R. Network Threats and Opportunities for Privacy, WP5.1 Deliverable for PISA Project, August 2001, 76 pages.

9.3.6 Papers in Preparation

- El-Khatib, K., Korba, L., Ronggong, S., and Yee, G.: A Distributed Anonymous Path Computation Algorithm. in preparation.
- Patrick, A. Kenny, S., Implementing Information Privacy in Human-Computer Interfaces, IEEE Computer (Journal Submission)
- Korba, L., El-Khatib, K., Patrick, A., Song, R., Xu, Y., Yee, G., Privacy and Security in E-Government (Invited Book Chapter contribution)

Chapter 10

Design method

G.W. van Blarkom S. Kenny J.J. Borking
gbl@cbpweb.nl ske@cbpweb.nl jborking@euronet.nl
CBP, The Netherlands

M. van Breukelen A.P. Meyer A.K. Ahluwalia
breukelen@tpd.tno.nl meyer@tpd.tno.nl ahluwalia@tpd.tno.nl
TNO-TPD, The Netherlands

This chapter discusses the design method. A privacy-safe agent offers two privacy limit options, operating under anonymity or under the DPD, as already mentioned in the previous chapter. Privacy threat analysis has consequences for the design and has to be dealt with. The privacy-safe agent ontology distinguishes between several types of agents, i.e., personal agents, task agents, service agents and other agents. The Common Criteria are used for the anonymity operation mode of the agent. The privacy regulations are either included in the agent architecture or the data models. Building legal rules into agents is a challenging task and this is performed by the use of privacy ontologies. Ontologies are formal machine understandable descriptions of terms and relations in a particular domain, i.e., the encapsulation of knowledge about the data protection domain in an unambiguous standardisation for the privacy field. A method has been developed to realise privacy knowledge engineering (PYKE) in order to build the condensed privacy rules into an ISA (Intelligent Software Agent). This method is called Design Embedded Privacy Risk Management (DEPREM). Interaction protocols also need to be designed. These interaction protocols show how the communication between agents is organised. Privacy preferences and privacy policies determine the transfer of personal data. Monitoring and auditing of agents shall be cumbersome if audit trails have not been built in into the design of agents. The auditors of the data protection authorities and the data subjects can only access these trails. Human-Computer Interface (HCI) is an important topic for interface design as has already been said in Chapter 4. This chapter discusses HCI requirements.

10.1 Prevention or Minimisation

Section 3.1.5 contains a description of the seven PET principles. The first principle (limitation in the collection of personal data) is not really a technical measure as laid down in

Article 17 of the Data Protection Directive 95/46/EC. It must be the wish of both the data subject and the controller not to provide and not to collect, respectively, more personal data than required in relation to the purposes of the intended processing of personal data.

Only after the process of prevention or minimisation in the collection of personal data can the controller start thinking about appropriate security measures because it is not until that moment that the controller can evaluate the potential risks represented by the processing and the nature of the data to be protected.

The Common Criteria, as described in Section 3.2, define the class privacy as four different but, nevertheless, strongly related security measures.

10.1.1 Anonymity

The PISA ontology distinguishes between three types of agents: a ”personal agents’, a ”task agent” and a ”service agent’.

A personal agent communicates directly with a data subject or with the organisation the data subject wishes to contact. The consequence of this is that personal data carried by this type of agent needs to be able to identify the data subject. Security measures need to be in place to guarantee that only data subjects can gain access to their personal data or that personal data shall only be disclosed to the organisation after the data subjects have given their explicit consent.

The task agent operates on behalf of the data subject but the technical infrastructure must be designed in such a manner that, because there is no need to, there is no possibility whatsoever that a task agent (or the task agent’s controller or processor) is able to identify the data subject initiating the task agent.

The service agent operates on behalf of task agents. The identification of a service agent is in no way related to the identity of task agents. This identity can, therefore, not be used to establish the identity of the data subject. Service agents carry, however, the data from the initiating task agent. The (identifiable) data must be secured against unauthorised access or disclosure.

10.1.2 Pseudonymity

Using agent technology, data subjects have expressed the requirement that the tasks they want the agent(s) to perform are executed in absolute anonymity. The task agents a data subject initiates shall be identified by a pseudo-identity (the fourth PET principle). The data subject shall be the only one able to consent in disclosing his real identity. Each task agent shall have a unique identifier (pseudo-identifier).

10.1.3 Unlinkability

A data subject may wish to start several task agents while these agents may co-exist within the same time frame. These agents may perform the same type of task (find me a job) or they may perform different tasks (find me a job and find me a friend who also collects stamps). Every task agent shall be identified by a pseudo-identity that can be traced back to the data subject. The pseudo-identities shall have unique values that cannot be linked by any of the involved controllers or processors.

10.1.4 Unobservability

The agent's activities are performed in such a manner that no processor, controller or hacker shall be able to observe if all of the above measures are put in place. That which is possible in the off-line world, must also be possible in the on-line world: the data subject, that is, the individual, can go to the High Street job agency and browse the vacancy files without talking to any of the staff members.

10.2 Privacy Regulations

10.2.1 Consent, Privacy Policies and Preferences

Much value is given to the concept of ”consent” in privacy legislation within the European Union. It is not always easy to establish in the real world whether data subjects gave their consent to the processing of their personal data. Life is not really any easier on the electronic highway. A real signature could be proven genuine and an electronic signature, at least within the EU, is legally recognised in the real world although it requires a Trusted Third Party.

10.2.2 Categories of Personal Data

It has already been well accepted that security measures may differ depending on the nature of the personal data involved. The term ”Level n personal data” is introduced for this purpose, which, in fact, must be ”Group’. We have, nevertheless, chosen to use ”Level” because each group requires that a different level of security measures be put in place.

The following three levels of personal data have been identified:

Level 1 Deal Closing Information. This group of personal data is transferred at the very end of the scenario. Level 1 personal data may contain items such as name and address information, telephone number and e-mail address in a Job market case. It is irrelevant whether one uses a real identity or a pseudo-identity. The sole use one can make of this data is to send ”communication” between the data subject and the company to ensure they can get in direct contact. This type of personal data is sometimes referred to as ”contact information” for these reasons. In another MAS application, for instance an agent looking for a special MP3 file, Level 1 personal data may be the credit card number of the data subject to be debited for the MP3 file to be downloaded.

Level 3 Special categories of personal data as defined in Directive 96/46/EC, Article 8, paragraph 1[1]. Level 3 personal data are only to be processed under the conditions specified in Article 8, Paragraphs 2 to 7 inclusive.

Level 2 All others items of personal data.

[1] These special categories of personal data (Article 8 of Directive 95/46/EC) are: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and the processing of data concerning health or sex life.

Personal data related to Levels 1 and 3 shall only be transferred to other agents if it is strictly necessary. Level 2 personal data can be used more frequently, for example, to get in contact with other agents via the use of brokers or other service agents. Level 1 personal data requires stronger security measures (technical and organisational) to protect these personal data in relation to security. The data subject has the legal right to define different privacy preferences for the different levels of personal data.

10.2.3 Preferences the User Could Provide to the (P)ISA

A list is given in Section 3.3 on ”act according to the directive’: ”Privacy by Design” that represents the translation of the Data Protection Directive into nine ”areas of attention” (Privacy Principles)[2].

These areas of attention are to be used by three categories of interested persons or organisations. They are:

1. **Controllers**, when defining the technical and organisational measures to secure the processing of personal data;

2. **Data subjects** at the time they want to evaluate the legitimacy of the system or at the time they want to specify their privacy preferences to a controller before disclosing their personal data to the controller;

3. **Supervisory Authorities**, when performing a compliance audit.

The data subject has a right and the possibility to specify his preferences for three of these principles. This specification of privacy preferences is not applicable to the following privacy principles:

1. Intention and notification. This is an obligation placed upon controllers to inform the supervisory authorities of their intention to process personal data;

2. Transparency. Controllers must have the required functionality (procedures, leaflets, privacy statements, etc.) to inform the data subject on the intended processing of the personal data they are about to collect;

4. Legitimate processing grounds. Controllers must have grounds for the processing collection. A user of an agent has selected this agent because the controller specified a purpose and that purpose corresponds to the tasks the user wants the agent to perform;

5. Quality. Controllers are obliged to do their utmost to ensure data quality in relation to the data subject;

7. Security. Controllers are obliged to put in place appropriate security measures; the data subject cannot influence the system put in place by the controller;

8. Processing by a processor. Controllers have the right to use one or more processors to process personal data on their behalf. The legitimacy of the processing does not depend on the use of a controller; the responsibility is and shall remain with the controller even when a processor is used.

[2] See Section 7.1 on Privacy Audit Framework for an extensive description of these principles.

An automated system of data subject privacy preferences can be designed for the remaining three principles. The preferences, thus, specified must be matched with the Agent Practises Statement (APS) that contains the privacy policy of the agent (the electronic consent) before a disclosure of any personal data may take place (see Section 10.2.6).

3. Finality principle. This principle gives data subjects the right to specify, in detail, the retention period. Once specified, a technical measure must be in place to automatically delete their personal data if that period expires. This principle also enables data subjects to perform those actions they judge to be compatible with the purposes for which the data are collected;

6. Data subject's rights. This is a mixture of different issues for which the data subject can specify his preferences in relation to a few. The data subject has, for example, absolute rights that the controller has to make provisions for. These are the right of access, rectification, deletion and blocking. The next right is the relative right the data subject has to object to processing. Unless the controller can justify the processing, it shall be stopped upon the data subject's request. Next, the data subject may give an absolute objection against processing for commercial or charitable purposes. The data subject can specify in detail the organisations or the type of organisations to whom his personal data may or may not be disclosed in relation to this. The last right is the right not to be a subject of a decision that significantly affects the data subject if that decision is based solely on the grounds of automated processing of personal data intended to gain insight in certain aspects of a person's personality. The ”electronic consent', as described in Section 10.2.6, is without any doubt an automated decision. Whether it also affects the data subject significantly depends on the functionality of the agent in question. The controller has to evaluate the consequences of the electronic consent and decide whether the level to which it affects the data subject and, thus, whether the consent is unlawful, for each type of task agent.

9. Transfer of personal data outside the EU. Firstly, the controller must have legal grounds to disclose personal data outside the EU. Once this has been established, the data subject's right to give his consent to disclose his personal data must be dealt with.

10.2.4 Ontologies and System Design

Problems due to changes in the law are well known in conventional data processing; changes in tax law, for example, must be announced well in advance of coming into effect to allow time for the considerable task of altering programs which have to apply these laws in payroll and other applications. The privacy principles, however, as described in Section 3.1.5 have been accepted worldwide over a period of more than 20 years and, therefore, are considered stable enough to be implemented in information systems.

Ontologies have to be formulated and implemented into the design of ISAs in order to guarantee a system design against privacy risks as discovered in the privacy threat analysis as described in Section 2.10 to ensure that the nine privacy principles can be built into an Intelligent Software Agent (ISA).

Ontologies

What is an ontology in the context of ISA and PETs In a certain respect, every data model and XML Schema is an informal ontology. What is missing, however, is the level of formal semantics which allows you to:

- Perform reasoning about concepts;
- Provide interoperability based on formally documented consensus;
- Accurately reflect conceptual models of end users (where necessary).

The following are also crucial in defining an ontology in this context:

- An ontology must be based on a consensus process;
- An ontology must define semantics and not just a scheme of identifiers or names;
- As a guiding principle, an ontology must be seen as a computer's model of reality through semantics to ensure that the computer has no other source of knowledge, derivation or logic whatsoever except for that ontology itself.

The XML Schema, for example, is an informal ontology but it does not have a formal relationship to the world as semantics and, therefore, XML Schemas fall short of what can be achieved by implementing a more rigorous semantic framework.

The aspect of consensus is the most important defining factor of an ontology. It is trivial to express a single person's conceptual model as a formal ontology and, in many domains, this could be done within a matter of hours, whereas achieving consensus for the same domain may be a process which could take years.

What are the use case scenarios for Ontologies within PETs? Ontologies are useful in a context where machine reasoning and semantics is required. This is also true in the case where they are used to aid applications in visually representing semantics in a structured way to users.

The following specific use cases could be identified. While many, particularly those described in the privacy domain, are thought experiments, the use of ontologies as defined above has only become common practice within the last 3 years.

P3P P3P (Platform for Privacy Preferences) can be seen as an existing five-year consensus process for creating what is now an informal ontology. P3P would be an ideal arena for using ontologies as a basis for RDF statements about data transfer events. Using a well-defined ontology and interoperable syntax would allow P3P to interoperate cleanly with other policy expression languages such as the EP3P language proposed by IBM.

Audit systems While P3P may offer the ability to publish accurate and expressive privacy policies, it does not ensure any real enforcement. Another use case of ontologies within PETs is in creating a way of standardising database event logging in order to create an audit trail which can be analysed by a rule-based automated audit tool. A standard conceptual model for database event logging in conjunction with visual mapping tools may be used to transform Enterprise database (and other data transactions) logs into a standardised audit trail. Enterprises can use their own data model for recording events but this is, subsequently, related via a one-time mapping (using a visual tool) to an industry standard privacy audit ontology. Next, an automated tool is applied to map the enterprise logs into the audit standard and perform an automated trail analysis.

This would, subsequently, lead on to the question of how such analyses could be used to provide seals and certificates, which would suggest that seal providers and user groups would be an important group of stakeholders in the creation of the ontology and specification of which attributes to include in the mapping. The key issue is to provide organisations with tools, methods and reference models for making data protection an accountable process in the same way as standardised accounting procedures are. Ontologies provide a consensual basis for this, which could even, if necessary, be legally required for Enterprises to use in the same way as in accounting.

Access control systems Normally on a web server there is a security layer that mediates in authorising access to data based on security properties, but recent research has added an additional layer for privacy on top that mediates in authorising access to data based on privacy policies attached to data entities. Transactions, in this layer, are based on a set of rules encapsulated in an XML policy and users use these rule sets to decide whether to release their personal data. Ontologies can be used for capturing user preferences in a standardised way, ensuring legal enforcement, and for decoupling preference rule systems from application architecture. An example scenario is user-submitting data to a web server with the accompanying preference statement (or XML equivalent) ”I do not want this data to be released to any third parties”. Using ontologies, this statement can be expressed using the user’s conceptual model and it can, subsequently, be translated into a machine readable, interoperable statement that can follow the data wherever it goes, that can be used for reasoning, such as negotiation processes, addition of legal requirements, access control, etc. The use of an ontology must be seen to create a three-way communication process between the law, users and technology.

This type of system must also be seen in the context of federated identity management and authentication systems such as MS Passport where ”sticky” policies with associated semantics are equally vital.

Ontologies in user interfaces for data protection It is highly desirable for all browsers to represent privacy information in a semantically uniform way in an environment where there may be high-levels of liability attached to the way in which a privacy policy is represented to users. Browsers must, for example, state what is the lowest and the highest level of protection, who the intended recipients of information are, etc. EU Directive (95/46/EC) also requires that certain information about a data transaction be presented to the user before the user can actually make the transaction. Many enterprises are concerned about their liability in this context since they use XML to describe a policy, which may be translated into human readable form in many different ways. We suggest that ontologies may be used to standardise the way XML privacy policies are presented to the users.

Medical ontologies There are now several ontologies in the medical domain which are in active use in the mainstream market, such as, LSE datalink. Usage is mainly to validate data that depends on medical information and to standardise terminology. In this way, for example, what is described in a patient record can at least be partially validated against what is contained in the medical ontology and, thus, a certain amount of nonsensical data can be avoided. They are, in this sense, used in a proactive mode to ensure that one can only use terms and connections of terms under the constraints or as they are declared in the ontology. This is an example of how ontologies can be used to improve user interface. Another application is in linking hospital databases. Basically, it is a method to ensure, in both cases, that databases are more semantically structured.

Search and retrieval

- One of the classical use cases for ontologies is in using semantics for improving the efficiency of web searches. This may be summarised by the following example:

  > *If, for example, I am going to a meeting in Rome, but I have forgotten everything except the three facts that it is in a church on a certain day in Rome. If I search today using the Google search engine I would not obtain sensible results. This is a Google search which has access to pages which express the semantics of the page in a machine readable format. The date it was talking about, the place and the type of event, would, however, find this event.*

  This, of course, would assume a certain degree of semantic markup processing which could be tapped into by Google. It is not particularly appropriate in the context of PETs as it is dealing with open world semantics which is not necessary within the scenarios envisaged for PETs.

  This is effectively what was done in the ”ontoweb” project using an annotating tool to annotate pages according to this ontology. If the annotators discover unknown terms on these pages they feed them back into the ontology. This was found to work in enhancing queries, reasoning capabilities about structured data, and in automatically creating links between the pages and the databases behind the pages. Human users, in this case, were required to create the ontology and semantic markup. The query engine, subsequently, uses that ontology to look at the pages that are annotated in the ontoweb community. Clearly it is not ideal that human intervention is required but, in the same way as HTML layout requires human intervention, it has a huge pay-off in terms of interoperability and scalability; it may be seen as inevitable at least for the moment.

- The European Commission also indexes all documents with a general purpose ontology such as Eurobook and Agrobook which are used by many national parliaments. These are generally used to, firstly, summarise the contents according to standardised and appropriate keywords and to aid search and retrieval interfaces. Certain ontologies are multilingual to ensure you can search for each term and, subsequently, you can use the same ontology to search for the same term in other languages whereby the same underlying semantics is used.

- Another example is the use of ontologies in car manufacturing. This does, however, include the problem of analysing the different relations of different parts of the cars. This is not important in the product area but is very important in test and engineering cases where there are test cars and a high-level of prototyping. There are, in this case, a lot of dependencies which an ontology helps to manage. This is an example taken from the much wider application area of product catalogues.

- Ontologies are also used in back-office systems to make unified queries over heterogeneous database systems. Often customers want to be able to search several systems with one query, including several legacy systems. Ontologies are used to create a loose coupling between systems.

Capturing and building ontologies

Recommendations for ontology capture processes and consensus mechanisms There are two basic approaches to achieving consensus:

- A top-down approach where an authoritative body defines an ontology in a normative way, which is, subsequently, forcibly adopted by stakeholders (this being the easier approach to implement). This is possible within a strict legal framework or a near monopoly such as the SAP system.
- A bottom-up approach using a standard process to build consensus between stakeholders.

Below a description is given of these approaches in terms of specific methodologies of this type.

- University of Aberdeen / Shaw & Gaines: method elaborated by Adil Hameed's paper reconciling experts' ontologies [HSP01]. This paper is based on the principle that it is not possible to automate the entire process of ontology capture. Face-to-face user testing is required in order to incorporate background and procedural knowledge, which is not recorded in writing anywhere else. The techniques for:
  - Capturing ontologies from non-technical experts using techniques from cognitive science and psychology. In particular, interviews, context-based elicitation and conceptual graph analysis (see the Paper for further details). This has the effect of absorbing procedural and background knowledge into the ontology.
  - Capturing knowledge from documents using conceptual graphs analysis.
  - A Schema for analysing and resolving mismatches according to four categories. (See Section 10.2.4.)
  - Analysis of legal text in the PISA project (PYKE and DPREM). See Section 10.2.5.
  - Analysis of legal texts to extract higher level concepts or principles;
- To create a set of candidate ontologies among a small group of experts by informal discussion and to, subsequently, test these on a larger set of stakeholders to see which would be the most appropriate;
- A general overview of capture processes can also be found in [FL];

The following general principles were identified:

- An ontology capture methodology must achieve a consensus in a way that is transparent and manageable in steps through time.

  There are natural processes for resolving ontological mismatches, such as those which have occurred in the top-down approach of SAP or those which occur in discussing terminology in a group such as the eponymous meeting. Achieving such consensus may be time-consuming and problematic. For example: trying to achieve consensus on the concept of film. As the film goes through the manufacturing process, the way it is conceptualised is radically different. In the manufacturing process of film, at the stage where the emulsion is applied, film was seen, for example, as a large flat surface made from a synthetic material, but, later in the process, as a small roll. This illustrates how concepts change over time and in differing contexts and the importance and challenge of systematically capturing this.

- An ontology capture methodology must allow for alignment and reconciliation between complementary conceptual models. Often there shall exist complementary ontologies which can, subsequently, be aligned by various matching methods. As the base of partially overlapping ontologies increases, and the importance of reuse increases, this shall become a crucial process. It would be advisable to reuse, in the privacy domain, for example, existing geographical ontologies to describe certain aspects of data recipients and provenance. There may, however, be a need for some alignment activity with these ontologies to determine the relationship between, for example, legal-safe zones and geographical zones.

- An ontology capture methodology must use cognitive methods to capture and analyse ontologies from end users. It is important, in this case, not to influence the subjects by suggesting preconceived models. This includes, for example, modelling synonymy and levels of abstraction. (I.e. to determine when the same instance is being referred to at a different level of the ontological hierarchy.)

- There must be an iterative process of rechecking models with test subjects with smaller groups of stakeholders and simpler ontologies;

- Such techniques may also have the spin off of modelling expert strategies, which may, subsequently, be used in rule-based reasoning applications;

- Textual analysis methods must be used to capture and analyse ontologies in the legal domain, where human sources are perhaps less relevant;

- A combination of both types of techniques must be used to capture ontologies in expert domains, for example, for data commissioners' models;

- An ontology capture methodology must be able to detect and deal with inconsistencies;

- An ontology capture methodology must produce clear documentation for change tracking and reconciliation purposes;

- An ontology capture methodology must take into account cultural differences between groups (e.g. EU/USA). There may also be cultural effects caused by linguistic peculiarities, such as, for example, whether the word ”bridge” is masculine or feminine;

- An ontology capture methodology must clearly separate the identification from the concept (i.e., its name from its identifying characteristics).

Change, Context and Inconsistency. The following principles may, in general, be applied to change and inconsistency:

- The application behaviours which are dependent on an ontology must be decoupled from the rest of the application and encapsulated in a rule system. This rule system includes all behaviours which are dependent on the conceptual model to ensure that they may be changed as the ontology changes. The question, subsequently, arises as to what is left of the application, given that an application is essentially a bundle of behaviours conditional on certain facts fed to the application. User interface and system interface behaviours shall, however, be invariant in the context of such architectures and can, therefore, be said to remain within the core application.

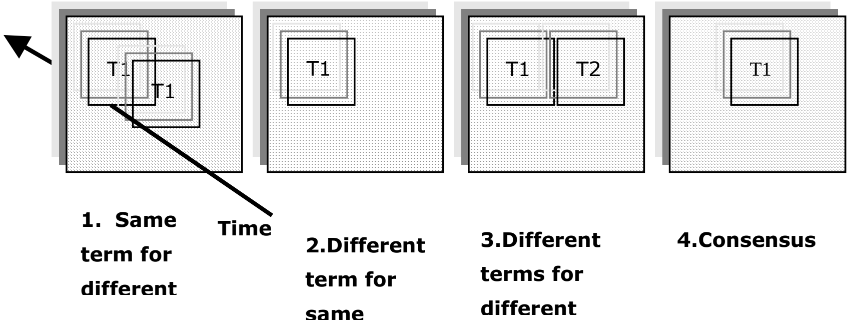

T1
T1
Time
1. Same term for different
T1
2.Different term for same
T1
T2
3.Different terms for different
T1
4.Consensus

Figure 10.1: Graphical representation of schemes for representing inconsistencies.

- The outer branches of an ontological hierarchy are likely to change more than the higher level concepts. Ontology design must attempt to define certain stable concepts which, barring major paradigm shifts [Kuh62], shall remain stable in the ontology, thus, minimising changes required to the rule system. It was suggested that in the domain of privacy, the set of principles outlined by the PISA project, may be used for such a stable basis. Principles, of course, would map into rules rather than concepts per se, but a set of core concepts could be extracted from the principles.

- Change must be documented according to a clear and traceable process to ensure that versioning can be effectively controlled. There are certain tools and processes emerging for this purpose. A good example is KAON [SMMS02] as well as the method described in [sem].

We shall inevitably have to resolve inconsistencies in conceptual models when developing ontologies and trying to reach a consensus. The following scheme for identifying inconsistencies has been identified.

1. The same term for different concepts (*Conflict*), such as, for example, whether ”Controller” means one responsible party to one person and one or several responsible parties to another;

2. Different terms for the same concept (*Correspondence*), such as, for example, ”Controller” and ”Responsible Party”;

3. Different terms have different concepts(*Contrast*), such as, for example, ”Controller” and ”Responsible Parties”;

4. The same term used for the same concept(*Concensus*).

We must also add temporal inconsistencies to the inconsistencies found.

5. Same term for different concepts at different times (*Temporal Conflict*)

6. Different terms at different times for same concept. Etc…

Schemes for classifying ontological mismatches are arbitrary in that they are in themselves conceptual models. As they may, however, be expressed as set theoretical assertions, it can easily be proved that they provide a logically complete description for any case.

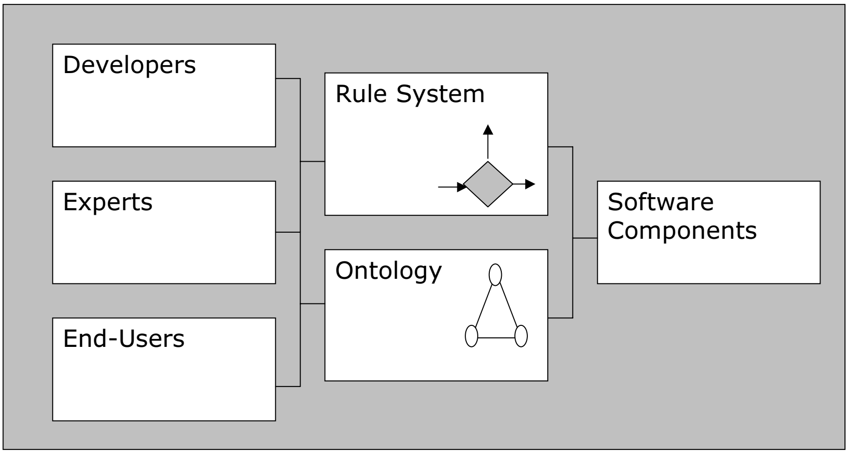


Figure 10.2: High-level view of ideal ontology architecture.

The conflict can be traced, in all these cases, to insufficient contextual information. In other words, a conflict arises either:

- If two terms are used for the same thing in exactly the same context; or
- If the same term is used in slightly differing contexts.

Resolution of conflict is, therefore, a matter, in most cases, of elaborating the context. This means either collecting enough contextual information at the collection stage or of going back to the collection process and collecting further contextual information.

It was also pointed out that any method which involves an iterative process of identifying mismatches and, subsequently, going back to stakeholders to resolve mismatches is not scalable because it involves exponential time in resolving complexity. Care must, therefore, be taken to match the level of iteration of the process with the complexity of the conceptual map (i.e., how many round trips to human subjects for consensus). This may not be an issue in domains where the number of terms is relatively limited such as the privacy domain.

Deploying ontologies

Architectural principles for ontology based applications. The most important principles can be identified as being:

Decoupling of Rule Systems (or Commitments), Conceptual Models (Ontologies) and User Interface Code. Reusability and change management are major drivers in modern system design. These principles are supported by one of the principal objectives for ontology design. That is, the decoupling of architectural components into Concepts, Rules and Application Architecture.

The objective of this principle is that conceptual models and procedural models (rule sets) can be put under the control of their direct stakeholders, they can be managed over time without requiring a restructuring of the underlying software components and they can be effectively standardised.

Documented consensus in ontology modelling. See Section 10.2.4. Clearly one of the main advantages of using ontologies is the ability to apply agreed standards across heterogeneous platforms and application types. Consensus is, therefore, vital and must be bought into from application developers and end users.

Design for change. Ontologies are going to change over time and need to be designed with this in mind. This means structuring the architecture to ensure that the conceptual models are cleanly decoupled from other components and have clearly documented processes for agreeing change. Mechanisms must be put in place for identifying, incorporating and documenting these changes.

Design for multiple conceptual models. The aim must not be to create a one-fits-all model since the unification of multiple conceptual models may not always be possible. Instead, complementary ontologies must be used in parallel with translation mechanisms giving the mapping between them where possible.

Design the system in such a manner that the ontology provides the application's model of the outside world. Any information needed by the application about the outside world needs to be included in the ontology. It must, for example, be a principle of designing the ontology that the ontology provides the complete semantics for an application.

Try to use existing paradigms. Uptake of an architecture or application shall be much easier if existing and simple paradigms such as XML, RDF or RDBMS are used.

Scalability. Use scalable technologies and architectures when, for example, a text file needs to be referred to each time an ontology is referenced. This would be bad design. Parsers used must be fast enough to cope with envisaged loads; an example of this is the fact that PROLOG encountered problems because of scalability problems.

The DOGMA approach, for example, creates a clean separation between concepts and procedures. It consists of sets of context-specific binary conceptual relations called lexons and a separate set of so-called ontological commitments. Each commitment corresponds to an explicit instance of an intentional task in terms of the ontology base [SMJ02].

Ontology Formats and Languages. The formalism for expressing the ontology is, in some respects, an arbitrary matter. Many ontologies can be expressed equally well in several different formats. Different experts present at the meeting held different views. There was a certain consensus, however, on criteria for choosing a particular formalism. The consensus was that the choice of formalism was immaterial as long as it was:

- Sufficiently expressive;
- Scalable (based on XML/RDF data stores);
- Interoperable (RDF/XML/OIL);
- Based on a simple standardised base syntax (RDF).

Common serialisation formats are: RDFS, DAML+OIL, OWL and KIF.

Reconciling end user models with expert models Ontological models aimed at end users are usually of a very different nature from those aimed at experts. This is particularly true in the area of privacy where there is a very technical document base in the legal domain and very little understanding of this among general end users.

Clearly ontology capture from end users shall follow the procedures that are similar to those outlined by Adil Hameed involving cognitive modelling, interviews and contextual testing, whereas legal modelling shall be more valuable if based on documents.

The question, subsequently, is how to reconcile these two seemingly opposing conceptual models to ensure that there is a clear mapping between the two. There are several important considerations in solving this issue.

- It cannot be hoped that there shall be much similarity between the user and legal models as most users are entirely aware of legal concepts of data protection.

- The domain may be divided into a coarse grained model and more fine-grained concepts. It may be expected that the legal domain would be interested more in the finer grained concepts and that the 2 domains would be mostly orthogonal on these. Both groups shall, for example, require fine grained concepts but these are unlikely to overlap. Users may, for example, wish to divide data types into fine categories, for example, e-mail may be more sensitive than date of birth, whereas the law shall only be concerned about the distinction between personal and non-personal data.

- It can be hoped that it shall be possible to align the two with a mapping ontology based on inconsistency resolution techniques as outlined above. Figure 10.1 incorporates such a mapping ontology; in this case between a technical system and legal system rather than a user system and a legal system.

System Design

From a systems design point of view, the privacy ontology can be implemented in the ISA as follows.

The following legal privacy principles (see heading 4) have been currently implemented in the ISA in order to prove the objectives of the PISA project (see heading 1) in the PISA project (due only to a lack of resources). The principles involved are 2, 3, 4, 6, 7 and partly 9. Both privacy preferences and privacy policies refer to a privacy principle. The generic abstract principle concept is detailed for each included principle (transparency, finality, data subject rights, legal processing and transfer). A statement can be given in order to specify the requirements or provisions (with respect to preferences and policies, respectively) that are specific to the given principle for each of the principles. All statements are subconcepts of a generic statement concept. Depending on the principle, the statements can have different values as its content. How the content of the statements is evaluated and preferences are compared to policies is described under the privacy transfer rules below.

The sender of PD declares preferences for groups of PD (PD Group). The PD can be of level 1, 2 or 3. Levels 1 and 3 may be encrypted and anonymised using PET with the help of a data protection authority. Preferences are given as statements on the requirements with respect to the principles in a PD group.

The agent practices statement (APS) contains policies for all the principles and, thus, makes statements about what is provided by the receiving party with respect to the principles.

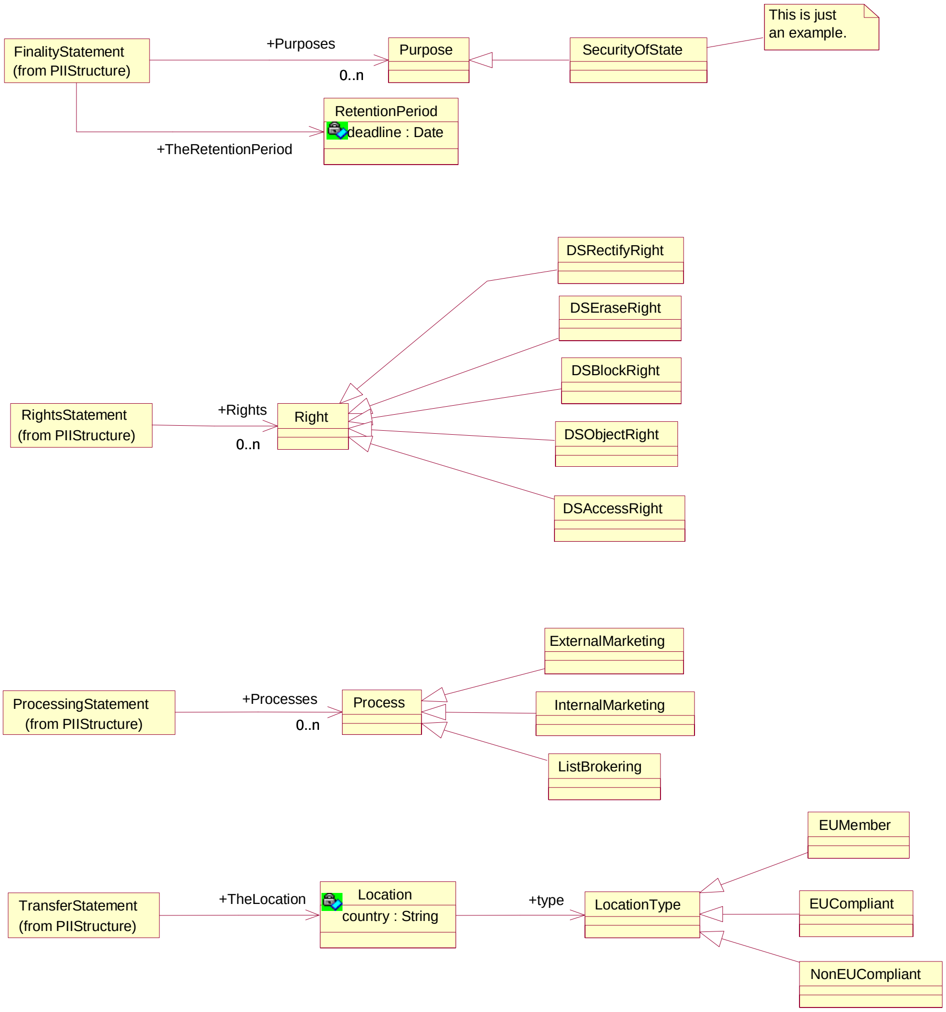


Figure 10.3: Privacy Ontology Concepts (Statements).

The specification of the location of the agent poses a difficulty in the above approach since the APS is constructed by the agent provider before the agents are even instantiated. This is because in some cases the agent provider cannot know beforehand where the agent shall be instantiated. The problem becomes even bigger if the agent is a mobile one and can travel from one host to another. Specifying the location in the certificate of the agent instead of in the APS can solve this problem. This shall even solve the problem for mobile agents if mobile agents get a new certificate each time they arrive at a new agent platform, which seems a logical approach since a platform agent provides the certificate.

Thereafter interaction protocols specify the privacy communication types between agents in the PISA demonstrator. Interaction protocols are extensions to UML sequence diagrams and serve as patterns for agent communication.

Interaction Protocols

The interaction protocols show how the communication among senders and receivers in the multi-agent system is organised. The protocols shall need to be combined for executing the given tasks. It is likely that more protocols appear to be needed when the implementation details become clearer. The basic protocols related to the transfer rules and data subject rights are explained below.

The sender of personal data (PD) must be able to determine whether the PD can be sent and, for this purpose, it needs both information on the agent that would receive the PD and privacy metadata on the PD. It needs to know, from the receiver, the privacy policy. The metadata of the PD is important because it contains the privacy preferences of the data subject. The sender of the PD must compare the privacy preferences with the privacy policy using the privacy transfer rules as stated below. The privacy transfer rules consist of one or more rules per privacy principle. If the transfer rules are evaluated positively and the PD is sent to the receiver, the metadata on the PD containing the privacy preferences of the data subject must be sent along with the PD. This way the receiver can act as a sender of PD in its turn.

The expressions that are evaluated positively (true) in the processing of the rules result in an agreement for transfer. All parts of all principles must be matched with positive results before information is sent; see Table 10.1 for more information.

T = True (positive result)
F = False (negative result)

Principle 2: Transparency Data subjects can specify in their privacy preferences whether they request transparency or not (Y/N). The receiver can offer no transparency (N) or the legally required transparency applicable in the EU (Y). If the data subject requests transparency the receiver must offer it, otherwise it does not matter, thus, three out of four possible combinations return a positive result.

Principle 3: Finality Principle The principle of finality has two aspects: purpose binding and retention period.

Purpose Binding The data subject can specify in his privacy preferences for what purpose information may be used. The receiver may not use the information for any purpose

Table 10.1: General consent model.

t^{pref}/t^{pol}	**N**	**Y**
N	*T*	*T*
Y	*F*	*T*

$$(\neg t^{pref} \wedge \neg t^{pol}) \vee (t^{pref} \wedge t^{pol}) \vee (\neg t^{pref} \wedge t^{pol})$$

Table 10.2: Data subject's rights.

r_i^{pref}/r_j^{pol}	**N**	**Y**
N	*T*	*T*
Y	*F*	*T*

r_1 = access, r_2 = rectify, r_3 = erase, r_4 = block, r_5 = object

that is not specified by the data subject. Thus, all the purposes that are specified in the receiver's policy must also be specified in the data subject's preferences.

$P_{pref} = \{p_1, \ldots, p_n\}, n \geq 0$; data subject's preferences for purposes (required)
P$_{pol}$ = $\{q_1, \ldots, q_m\}$, m>0 receiver's policy for purposes (provided)

$\forall_{1 \leq i \leq n, 1 \leq j \leq m} q_j \exists p_i : p_i = q_j$

Retention Period The data subject can specify a deadline in his privacy preferences by which the information must be deleted (retention period t_{pref}). The receiver's retention period (t_{pol}) must be shorter or equal than what the data subject requests.

$t_{pref} \geq t_{pol}$

Principle 4: Legitimate Grounds for Processing Data subjects can specify in their privacy preferences for what other processing their PD may be used[3]. The receiver may not use the information when data subjects have not specified any processing in relation to their PD.

P$_{pref}$ = $\{p_1, \ldots, p_n\}$, n>0 data subject's preferences for processing (required)
P$_{pol}$ = $\{q_1, \ldots, q_m\}$, m>0 receiver's policy for processing (provided)

$\forall_{1 \leq i \leq n, 1 \leq j \leq m} q_j \exists p_i : p_i = q_j$

Principle 6: Data Subject's Rights Data subjects can specify in their privacy preferences the rights that they have in relation to their PD. All the rights that the data subject requests (R_{pref}) must be provided by the receiver (R_{pol}). Five rights are defined by the privacy legislation. Three out of four possible combinations return positive results. See Table 10.2.

$R_{pref} = r_1^{pref}, r_2^{pref}, r_3^{pref}, r_4^{pref}, r_5^{pref}$
$R_{pol} = r_1^{pol}, r_2^{pol}, r_3^{pol}, r_4^{pol}, r_5^{pol}$
$\forall_{1 \leq i \leq 5} r_i^{pref}, r_i^{pol} : (\neg r_i^{pref} \wedge \neg r_i^{pol}) \vee (r_i^{pref} \wedge r_i^{pol}) \vee (\neg r_i^{pref} \wedge r_i^{pol})$

[3] For example, internal or external marketing, list brokering.

Table 10.3: Transfer outside the EU.

t^{pref}/t^{pol}	**EU-only**	**EU-compliant**	**Non-EU**
EU-only	*T*	*F*	*F*
EU-compliant	*T*	*T*	*F*
Non-EU	*T*	*T*	*T*

Table 10.4: Summary of privacy supporting measures.

Measure	**V2** **Transparancy**	**V3** **Finality**	**V4** **Legitimate grounds for processing**	**V6** **Data subject rights**	**V7** **Security**	**V9** **Transfer outside EU**
Anonymity & pseudo-identities	—	—	—	—	—	—
Agent certificates				√		
Secure communication				√		
Encryption				√		
Onion routing approach				√		
Signing the APS				√		
Deleting and updating PD			√			
Logging	√					
Monitor services	—	—	—	—	—	—

Principle 9: Transfer of Personal Data Outside the EU Data subjects can specify in their privacy preferences whether their PD may be transmitted outside of the European Union Member States (EU). The specification indicates the boundary of transmission as either EU-only, EU-compliant[4] or non-EU. The receiver specifies his location with respect to these boundaries. The receiver must reside within the boundary specified by the data subject. It must be noted that non-EU *policy* excludes EU-only and EU-compliant, while non-EU *preference* includes the others! See Table 10.3.

Privacy Supporting Measures

Apart from understanding the terminology and semantics of privacy, a number of measures need to be taken for building a privacy protecting system. Table 10.4 gives an overview of the measures and their relation to the privacy principles.

[4] The EU-compliant countries: Personal data can flow freely from the fifteen EU MS and three EEA member countries (Norway, Liechtenstein and Iceland) or third country without any further safeguards. The Commission has, so far, recognised Switzerland, Hungary and the US Department of Commerce's Safe Harbour Privacy Principles as providing adequate protection.

10.2.5 Privacy Knowledge Engineering and DEPRM

Problems that are due to changes in the law are well known in conventional data processing: changes in tax law, for example, must be announced well in advance of coming into effect to ensure there is sufficient time for the considerable task of altering programs which have to apply these laws in payroll and other applications. The privacy principles as described under heading 4 have, however, been broadly accepted over a period of more than 20 years and, therefore, considered stable enough to be implemented in information systems.

A method has been developed to realise privacy knowledge engineering in order to build the nine privacy principles[5] into the ISA. The method has been named Design Embedded Privacy Risk Management (DEPRM)[6]. The basic proposition of DEPRM is that, in order to assure a system design against privacy risks as discovered during the privacy risk analysis, the privacy rules must have a strong influence on the design forcing the system designer to include the data protection rights and duties in a mandatory fashion in the system. Legislation is often complex and laced with many exceptions and, therefore, the solution can be found in working with a simplification of the data protection legislation and using relatively stable privacy principles while retaining the capacity to integrate more aspects of the law when circumstances require that additional aspects are applied. Medical personal data are, for example, considered privacy-sensitive data and the law requires, therefore, a strict regime while processing these data[7]. The legal instantiation procedure in DEPRM works as given below.

- Firstly, privacy principles as, for example, known from the Convention 108 of the Council of Europe and the 1980 Organisation for Economic Cooperation and Development (OECD) guidelines[8] are determined. Next, a simplification of the law is realised through the linking ('chaining') of selected articles of the DPD that belong to the chosen privacy principles. Take for example the principle of transparency that defines that everyone must be informed about what is done with their personal data and that the data subject must also be informed of this as the data subject is mostly not aware of the many parties that take part in the processing and in the data flows. This principle is an amalgamation of Articles 10, a, b and c and 11, Sections 1, a, b, c and 11, Sections 2 and 13, Sections 1, a, c, d, e, f and g and 13 and Section 2[9] for recommendations for on-line collection of data. The principle of finality and purpose limitation (personal data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes) can be found in Article 6, Section 1, b and e, and Article 6, Section 2.

- This type of representation, therefore, clusters the DPD articles into logical components that are sufficiently generic to encapsulate the general meaning of the articles from which they derive. Certain articles are not "translated" into principles and are not implemented into de MAS design but play a role in relation to explaining and interpreting, such as, Articles 2, a to h (Definitions), 3, Sections 1 and 3, Sections 2

[5] The privacy principles have been broadly accepted over a period of more than 20 years and, therefore, are considered stable enough to be implemented in information systems.

[6] See [KB02].

Kenny presented the concept of DEPRM for the first time at the second CEN/ISSS data privacy IPSE (Initiative on Privacy Standardisation in Europe) open workshop in Paris on 27 September 2001. See Final IPSE Report IPSE-SG #11, Doc.no 7 – 13 February 2002.

[7] See Article 8 of the DPD, which defines the special categories of data, of which personal health data is one.

[8] 'Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data'.

[9] Article 29 of the DPD Working Party states in Privacy on the Net, WP Working Document 37, how transparency must be applied in on-line systems. See www.europa.eu.int/.comm/privacy.

and 4, Sections 1, a, b, c and 4, Sections 2, 5, 9 and 15, Section 1, etc., and influence the architecture. Certain articles are not relevant for the MAS representation, such as, the requirements for the supervisory authorities, the legal remedies, etc.

- Having formulated the privacy principles, the next step is splitting the principles into elements, that is, pieces of text that are structured subsets requiring orthogonality (completely independent factors in the mathematical sense)[10] and regularity. The principles can, thus, be deconstructed into a set of elements that relate to the articles they are derived from and emphasise the context and what has to be achieved. Next, each element is given a work item identifier. For example: 1. Principle: Transparency; 1.1. Data subject (DS) is aware of transparency opportunities; 1.1.1. Personal Data (PD level 1 till 3)) collected from DS; 1.1.1.1. Prior to DS PD capture: DS informed of: controller identity (ID) and Purpose Specification (PS) 1.1.1.2. Prior to DS PD capture: DS informed of: controller ID/PS and re Article 10 c whether PD contains special categories of data and whether PD have to be treated accordingly to the sensitivity of these data Further more elements have to be formulated around information duties, PD distribution, explicit consent, controller obligations (for example 2.2. A controller must process according to his PS), about finality (DS presented with controller PS & a proposed retention period (RP), prior to being presented with an option in consent decision for having PD processed, except where PS is statistical, etc.) All principles are treated as stated above. This is the final level of granularity that can be reached in order to subdivide the privacy principles before formulating these into technical specifications. The next step, in terms of logic, is to find the ontology[11] a mechanism to achieve shared understanding, neutralising the problem of two databases using different identifiers for what is the same concept, such as, postal code. Ontology and logic are different aspects of a system design which have mutual dependencies but are separate complementary models (a summary of the types of entities and, specifically, the type of abstract entities that are to be admitted to a language system leading to related concepts and terms for achieving that all ISAs act in the same way while transferring policies) and data model for a principle as a representation of the relevant work item identifiers [KB02].

  These ontologies with taxonomies and a set of inference rules providing the relational aspects of the taxonomies lead to a simplified conceptual model of the principle and may, subsequently, be implemented into the ISA and used as a backbone in conversations between agents about privacy preferences, the matching of privacy policies and exchanging PD. The sender agent and the receiver agent, therefore, must understand the same privacy ontology in which a ”privacy policy” has been defined. The agent needs this to be able to automatically give the same interpretation to the information in messages they exchange in their communication. The ontology must also define how a privacy policy is structured, that is, the information it may contain and how it is composed in the privacy policy. The ontology must also define how the system of privacy preferences is structured because these preferences also need to be sent from one task agent to another and, therefore, has to be interpreted by a receiving agent. The logic has to be described in a format understood by the application, that is, the agent platform, to realistically reflect the simplified legal model. The ontology has now been enriched with a knowledge base to enable interpretation of the queries between the agents, such as, if PD is labelled with high

[10] Webster’s Third New International Dictionary 1986 p.1594.

[11] Webster op. cit p.1577. See for further information: [New82, Lev84, GN87, Sow00, Gru, Gru93], http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html, http://www.w3.org/2001/sw/, http://www.w3.org/TR/webont-req/, http://www.semanticweb.org/knowmarkup.html#ontologies, http://protege.stanford.edu/publications/ontology_development/ontology101.html.

constraints, the ISA shall request the Agent Practices Statement (APS), that is, the privacy policy under which it operates, from the interacting agent. Transfer rules are embedded in ISA to decide when PD can be transferred to other agents. The general transfer rule is: IF APS-1 MATCHES privacy-preference AND PD level 2/3 MATCHES PD level 2/3 THEN ALLOW disclosure/exchange PD level 2/3. Many other items are taken into consideration. Security characteristics have to be implemented. What is the identity of the other agent, who is he or she representing? Can the ISA trust the other agent? Controller identity, purpose specification credentials have to be exchanged and interfaces have to be designed to prevent detection of the ISA before the analysis of the situation shows a safe environment.

10.2.6 Negotiation

General

The Data Protection Directive states that personal data shall be processed fairly and lawfully. The controller is the natural or legal person determining the purposes and means of the processing of personal data. If a data subject wants to dispute the processing of his personal data within a personal data filing system, the data subject shall have to address the controller because the controller is the only party accountable for that processing.

Even if the controller has appointed one or more processors to process the data on his behalf, it is still the controller that shall be held accountable for the possibly unlawful processing. If an investigation shows that the unlawful processing is caused by a processor's malfunctioning, the claim by the data subject must still be settled with the controller, who, in his or her turn, may start a legal dispute with the processor. The data subject must have no involvement in the latter action. The conclusion of the above is that any processing system of personal data has one and only one controller. All processing of personal data is the responsibility of the controller. Disclosing personal data to a processor, does not have any consequences as to the role and responsibilities of the controller. Disclosing to a processor does not change the liability concerning the data. As far as the data subject is concerned, the processor does not exist. Disclosing personal data from one personal data filing system to another (where both systems have their own controller) is only lawful if there are legitimate grounds for this processing as specified in Articles 6 and 7 of the Directive.

Definitions of Agents

Three notions of agents are distinguished within PISA:

1. Class of agent. The PISA environment has defined the following as well as others: the Personal Agent, this is an agent in direct contact with a real person or a real organisation; the Task Agent, this is an agent that shall be used to perform only one task on behalf of a personal agent;

2. Agents. Within any class one can define a number of agents each dedicated to perform a specific task;

3. Instance of an agent. One unique occurrence of a specified agent is being identified with an instance of an agent.

Application to the World of Agents; Electronic Consent

Each instance of an agent in a Multi-agent System (MAS) carries personal data. Just as is the case with data processing in the real world, there must always be a controller accountable for the processing of personal data in the world of an MAS and in the environment of the Internet.

It is important to note that the controller of an instance of an agent is neither the developer who designed or coded the agent nor the owner of the platform on which the instance of the agent resides. The controller is the natural or legal person facilitating the use of the instances of agents and who is responsible for the processing of personal data by means of these agents or through information systems behind these agents. All the roles mentioned here may be performed by one and the same organisation. The responsibility for the processing of personal data, however, designates who the controller is (purposes and means of the processing of personal data). The legal basis for disclosing personal data, inside or outside the environment of a MAS, is consent. According to the Data Protection Directive, consent must be specific, unambiguous and freely given. Disclosing in the context of an MAS means the passing on of (personal) data from:

- A data subject to an agent (collection from the data subject);
- An agent to a next agent (transfer within the MAS environment);
- An agent to a natural or legal person (external disclosure).

Initiating data subjects are not physically present to give their unambiguous consent in the majority of the processing steps, in the sequence of agents required to perform the task. Human consent can only be achieved for the steps in which data subjects themselves disclose their personal data to an agent (collection).

A concept, that is, electronic consent, has been introduced to comply with the requirements of the data subject from start till finish within the PISA project. Electronic consent can be defined as *the conditions under which it is assumed that the data subject has given his consent*.

Privacy Policy and Privacy Preference

The concept of electronic consent is based on the privacy policy of the controller and the privacy preference of the data subject.

Privacy policy. This is a statement about the rules the controller uses when processing personal data. This policy is shown to the data subject to ensure a comparison can take place with said data subject's privacy preferences no later than the moment the controller starts to collect the personal data from said data subject;

Privacy preference. Privacy preference defines the conditions under which data subjects are willing to give their personal data to a controller.

The rules for the two, the collection and transfer of personal data, are the same. Disclosure is only lawful if the privacy policy of the receiving party respects the privacy preferences of the sending party. This respect must be interpreted as follows: the privacy policy of the controller must be the same or stricter than the privacy preferences of the data subject. Data disclosed under these conditions must be accompanied by their related privacy preferences. These preferences need to be reused if the personal data is to be transferred to another agent. Each controller specifies for each agent the privacy policy that is applicable to that agent. Every time it must be established that the policy respects the privacy preference. The third type of passing on data (external disclosure) is of a more complex nature. Three cases can be identified within this context:

1. The controller at the end of the chain is the target of the original MAS operation (a company, organisation or individual);
2. The controller operates on behalf of many targets, but is not the target itself and each instance of this agent works for one target only;
3. The controller operates on behalf of many targets, but now each instance of the agent works for many targets.

These three cases are best illustrated by an example from the PISA Demonstrator environment:

1. Bank A is the controller of its own agent advertising a job as bank clerk;
2. Job Agency has bank clerk vacancies for the job at several banks. The agents" controller is the Job Agency. In instance 1 of the agent, the job is advertised for Bank A, in instance 2 for Bank B and so on;
3. Job Agency has bank clerk vacancies for the job at several banks. The agents" controller is the Job Agency. There is one instance of the agent and that one advertises for Bank A and Bank B, etc.

The electronic consent mechanism to be invoked, before any transfer to the agent happens, must function as described below for these three cases:

1. The mechanism is the same as described earlier. Consent for external disclosure was given at the time the personal data were transferred to the agent because the agent and the receiving target have the same controller;
2. The electronic consent mechanism must be performed twice before the personal data may be transferred to the agent. The agent shall have to carry two privacy policies, namely the one of the controller of the agent and the one of the target. Each of these two privacy policies must now respect the privacy preference of the data subject;
3. The electronic consent mechanism must be performed many times before the personal data may be transferred to the agent. The agent shall carry a number of privacy policies, namely the one of the controller of the agent and the ones for each of the targets it operates for. Each of these privacy policies must now respect the privacy preference of the data subject.

The latter case could even be more complex. Personal data may not be disclosed externally to all of the targets for which an agent operates because of the difference in privacy policies of the different targets. The preferences a data subject has may well differ depending on the type of data to be processed when designing an electronic consent mechanism. Developing a mechanism, that is, establishing a privacy preference statement, may be linked to:

- All items of personal data;
- Each group of personal data items;
- Each individual personal data item.

The checking process shall be more complex depending on the choice made, but the concept remains the same.

Warning

An agent must run on a so-called platform in an MAS environment. There is always an organisation providing the means (infrastructure) for that platform. A number of issues must be kept in mind in relation to the legal aspects:

- Different types of agents may run on the same platform;
- The organisation responsible for the infrastructure of the platform is not, by definition, the controller of the personal data filing system;
- Different types of agents, each having their own controllers, may run on the same platform;

The controller is, as has already been mentioned, the organisation that collects and does any further processing on personal data. This organisation may or may not be the same organisation that runs the agent platform. It may very well be the case that the controller is not the organisation responsible for the platform. In this case, the latter must be seen as a processor on behalf of the controller.

10.2.7 Processing Personal Data

Three levels of personal data are distinguished in Section 10.2.2. This distinction is being made because privacy legislation has ruled different levels of security measures depending on the nature of the data to be protected. Article 17 of the Directive states that the appropriateness of the security measures is, among others, dependant on the nature of the data to be processed. The levels of personal data introduced in PISA are an implementation of the nature of personal data. The strongest security measures must be placed upon level 1 personal data in an MAS environment. There may be an involvement of agents before the ”answer” to the data subject’s question can be given in an MAS environment. All these agents perform their tasks under the responsibility of one or more controllers. The basic privacy concept of PISA is that only the controller of the agents ultimately providing the answer must be able to reveal the real identity (level 1 personal data). The real identity is hidden behind the pseudo-identity and is used to identify the data subject initiating the question. Strict security measures must be in place to protect level 1 personal data because unlawful access to these data jeopardises the whole PISA system. An agents carries personal data linked to the pseudo-identity derived from the data subject’s personal agent. As the real identity is safely protected in level 1 personal data, simpler measures can be in place for levels 2 and 3. Because of the difference in nature of level 2 and level 3 personal data, Article 17 demands stricter measures for level 3 personal data.

10.2.8 Auditing Requirements

Audit Criteria

This section describes the functionality controllers of information systems must provide to supervisory authorities[12] performing a compliance audit to the EU directive 95/46/EC[13]. An auditor is bound to a professional code of conduct. This code forbids auditors to give

[12] See EU Directive 95/46/EC, Article 28.

[13] EU Directive 95/46/EC has the title: The directive on ”the protection of individuals with regards to the processing of personal data and on the free movement of such data’.

advice to an organisation and subsequently to perform the audit on the very same object of investigation. It is not unlikely that the author of this document shall be responsible for the compliance (privacy) audit towards the end of the PISA project. This document shall describe the audit trail information required in order to avoid the danger of collision. That description shall give no indications whatsoever on the technical implementation the developer shall adopt to generate the information. This document shall, therefore, give no advice as to the number of agents required to do the job and on the number and physical location(s) of the database(s) that shall be used to store the required information. The generic term AuditAgent shall be used in this document.

Audit Trails and Monitoring

The statements made in this document on the required information for the compliance audit covers only one aspect of the privacy audit. The audit trail only addresses the flow of personal data through the agent environment. It does not cover, among other issues, the quality, adequacy, relevance, excessiveness and security originating from the Directive which are also part of the object of investigation during the compliance audit. The process descriptions must, of course, also be part of the audit; that is, the description of the process to decide on the disclosure of personal data to any third party. It has been decided that all required audit trail information would be created at runtime without any explicit action being required from the data subject within the ”PISA Demonstrator” information system. It is the controller’s responsibility to supply all the necessary information to the supervisory authorities to ensure said authorities can perform their supervisory duties.

The audit trails must, primarily, provide the necessary input for the privacy auditors to perform their compliance audit. The system shall be judged on the following four audit quality aspects when a Data Protection Authority subjects any system, for instance the PISA Demonstrator, to a privacy audit:

- Confidentiality;
- Integrity;
- Availability;
- Audit ability.

The audit activities on controllability could benefit from a system of audit trails maintained during the life cycle of all agents within the MAS. The system of audit trails means that, while tasks are being performed within an information system, automatic logging of all activities within that system are being logged for future analysis. From a simplified point of view, an MAS is just a special form of an information system. The code to be developed for the PISA Demonstrator shall, for this reason, include code to generate audit trails; the AuditAgent.

There are a number of rules directly connected to a system of audit trails. They are created:

- Automatically without the intervention of the data subject;
- Without the option of switching it off;
- To a secure medium;
- In such a manner that any alterations to an audit trail are traceable.

There is no legal obligation that this type of agent informs a Data Protection Authority on the detection of possible malicious agents. If this legal obligation were there, it would cause an enormous problem in relation to deciding which Data Protection Authority (that is, in which EU Member State) to inform. Namely, the PISA Demonstrator shall have a controller in one Member State. At the same time, however, it shall probably process personal data from data subjects from all countries worldwide (regardless of whether they are a EU Member State).

Additional Functionality Provided by the Audit Trails

One of the privacy principles that shall be implemented in the PISA Demonstrator is the item on the ”Data subject's rights'. One of these data subject rights is described in Article 12 (a) of the Data Protection Directive.

> *”Each data subject is guaranteed the right to obtain from the controller confirmation as to whether data relating to him are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed”.*

The audit trails created by the controller on behalf of the supervisory authorities contain the data to enable the data subject to exercise this right. There needs to be one additional provision. This provision must filter the information disclosed to the data subjects in such a way that they can only access personal data related to them.

Caveat

The data processed within the realm of the audit trails are personal data. This is, in itself, a personal data processing system on which the provisions of the Data Protection Directive must be applied. The Directive recommends including these files within the definition of the information system when making the notification to the supervisory authority[14].

The Starting Point of the Investigation

The audit trail(s) provided must supply information on all data subjects that used a UserAgent and still have personal data processed there (privacy principle V.3 on ”retention period'). The information supplied must show the identification of the UserAgent used if the controller controls more than one type of UserAgent.

When privacy auditors want to analyse the audit trails created by the PISA Demonstrator, they shall start the research process at the premises of the controller of the UserAgent where the data subjects register their personal data. The audit trail itself shall show the information (Level 1 personal data) on all data subjects of the UserAgent(s) or the auditor shall have access to the information (Level 1 personal data) on all data subjects of the UserAgent(s) from the audit trail. It goes without saying that PET protection on Level 1 personal data must be lifted, if so required, for an official investigation. Cooperation with a possible Trusted Third Party may be required[15]

[14] See Directive 95/46/EC Section IX: Notification, Articles 18 and 19.

[15] The privacy auditor has the required authority to gain access to personal data without the consent of the data subject, Directive 95/46/EC Article 28, Paragraph 3.

The next step shall be that the auditors shall randomly select a number of data subjects on which they shall focus this part of the investigation. The last step is that the auditors shall focus on the selected data subjects. The controller, mentioned above, must facilitate access to the audit trail information if the controller has involved processors or if secondary controllers shall be processing the information.

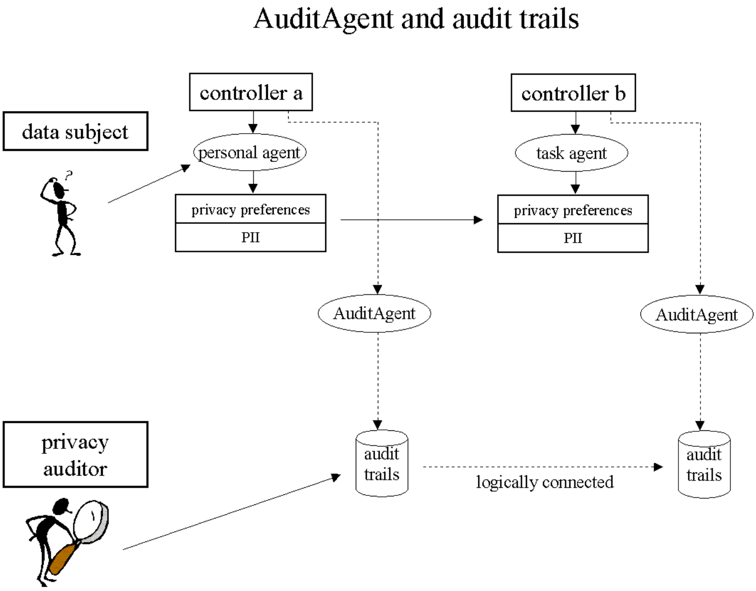


Figure 10.4: Schematic representation of the AuditAgent and the audit trails within a Multiagent System.

Why Does the Privacy Auditor Need to Establish Compliance?

Information is required from the PISA UserAgent and JobSeekAgent (in general terms, from the PersonalAgent(s) and TaskAgent(s)).

The system of AuditAgents must record the activities initiated by:

- The data subject when starting a PersonalAgent or TaskAgent;
- An agent starting a TaskAgent;
- An agent starting the first ServiceAgent within a sequence of ServiceAgents to perform a task;
- The start of a TaskAgent from a ServiceAgent.

Within the scope of the PISA Demonstrator, a decision has been made to assume that Levels 2 and 3 personal data are constructed in such a way that identification of the data subject is impossible. This means that a ServiceAgent does not process personal data. It is only required to log the start of the first ServiceAgent in the sequence in order to be able to check the contents of Level 2 (and 3) personal data using these data items in the initiating TaskAgent for this reason.

The auditor needs at least the following items:

- ID of the Agent;
- The link between the UserAgent and the TaskAgent it started. Which TaskAgents were started from an identified UserAgent;
- Type of the Agent;
- Date/timestamp of the action;
- Description of the action;
- Authentication of the agent (if the action is to start a new agent);
- Identification of the controller;
- Identification of the processor;
- Old versus new contents of personal data;
- ”Consent” for the disclosure to the next agent in line (privacy preference versus privacy policy). The APS including the purpose(s);
- The personal data that are being disclosed at the start of the first ServiceAgent in a chain;
- Proof that the ”noise” added to Level 1 personal data is unique (a string related to the unique agent ID?).

It is not unlikely that within an MAS a number of controllers shall be responsible for the creation of audit trails. The structure of the different audit trails must be defined in such a way that the contents are logically connected to facilitate the privacy auditor to follow the chain of activities initiated by a selected data subject.

Elsewhere in the PISA documentation it is stated that standardisation is required on the purpose descriptions. A similar remark is made here on the content and structure of the audit trails.

The list given above is not supposed to be a complete list of all items required. It is a non-limitative list of items to ensure the developer, knowing the exact details of the PISA Demonstrator, can decide on other items that the auditor may need during the audit process to establish conformity to the Directive.

The contents of the personal data as disclosed by the data subject must also be logged in relation to those agents that are collecting personal data. Similarly, if the data subject modified or deleted his personal data, the old and new contents must be visible. The controller must, under certain conditions, notify third parties, to whom the data have been disclosed, of any rectification as a result of Article 12, Paragraph c. The audit trails must log the agents that have complied with said Article.

The items requested may, of course, be represented by a code if the auditor is provided with the appropriate code table.

Chapter 11

Data Mining

M. den Uyl R. van der Veer P. Bison
denuyl@smr.nl rob@sentient.nl pbison@sentient.nl
Sentient Machine Research, The Netherlands

J. Giezen
giezen@tpd.tno.nl
TNO-TPD, The Netherlands

This chapter discusses two aspects of data mining and intelligent software agents: ”What is the actual and potential impact of data mining and ISAT on personal privacy concerns, and what are the opportunities for privacy protection?”.

11.1 Introduction

So, ”What is the actual and potential impact of data mining and ISAT on personal privacy concerns, and what are the opportunities for privacy protection?” is the question investigated in this chapter. Regarding this central question, three topics are crucial:

- Privacy, as intuitively experienced by people and the way it is condensed in regulations and legislation.
- Data mining, in its expression of web mining, for the collection and linking of personal data. (The term web mining addresses not just the World Wide Web, but all internet related data mining activities, especially those with intelligent agents).
- The accumulation of data as a consequence of the expansion of the number of data registrations and the expansion of the amount of information per registration plus the increase of information derivable from these expansions.

These three topics together are the background of the privacy threats to be discussed in this document. This is depicted in a triangle in Figure 11.1. To keep in mind that the main purpose of data mining will, generally speaking, be that beneficial effects will be pursued, the privacy threats are a subset of all the potential results of the mining activities.

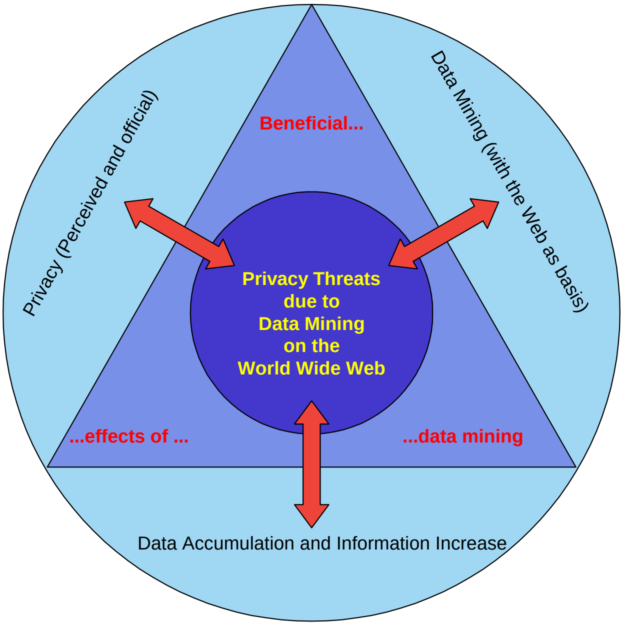

Beneficial...
Privacy (Perceived and official)
Data Mining (with the Web as basis)
Privacy Threats due to Data Mining on the World Wide Web
...effects of ...
...data mining
Data Accumulation and Information Increase

Figure 11.1: Three factors contributing to privacy threats.

Apart from privacy threats that may emerge from the use of data mining technology, the same technology can also be used to identify possible privacy violations.

Concerning the threats: a priori, three types of possible impact can be distinguished: Attack, Defence and Incorporation, i.e. the impact of data mining technology in the hands of certain parties may be offensive to legitimate privacy concerns, and data mining technology might conceivably be used as protection against threats to these concerns. The third type of impact, Incorporation, occurs to the extend that data mining practices will in fact conform to privacy enhanced standards and hence are neither vulnerable to attack, nor require defensive efforts from individuals. In this deliverable the focus is on data mining, the impact of ISAT as such is the subject of another deliverable. However some attention needs to be paid to the relation between data mining and intelligent agent technology.

In Section 11.2 a few conceptual and theoretical issues are discussed concerning the meaning and significance of privacy. In Section 11.3 current technologies, methodology and practices of data mining are outlined. In Section 11.4 the phenomenon of the increase of information on the World Wide Web is investigated and the potential of new information to be extracted from that. In Section 11.5 the possible threats of data mining to privacy concerns are discussed. At present, the conclusion is that current data mining as a set of technologies and methods actually poses a limited threat to privacy. However, this may change as technology progresses. In Section 11.6 the use of data mining as weapon against privacy intrusion is discussed and reasons are given why use of such weapon will be limited to specific situations and contexts. Section 11.7 provides a window on how to incorporate privacy protection issues in data mining activities and tools. This includes some particular approaches to privacy incorporated data mining practices. Finally, Section 11.8 contains the final conclusions to be drawn from the body of information and conclusions in the deliverable.

11.2 Privacy

In this section the issue of privacy is explained, in its official and unofficial interpretations with the emphasis on informational privacy. Inherent to the issue of privacy is privacy protection and also threat to privacy. The concept of threat will be explained; in brief, it is the manner in which privacy can be affected in a negative way. There are many ways, and the ones that may occur by means of data mining will be highlighted.

11.2.1 What is privacy?

Privacy can be regarded in two ways: the issue of privacy as experienced by the citizen and the official one. The first version, the unofficial one, can be summarised as:

> ”The claim of individuals to be left alone, free from surveillance or interference from other individuals, organisations or the state.”

Ideally, this should also be the definition adopted by the legislators and in some documents this is suggested. However, to consolidate the idea of privacy in legislation and practical ways to protect the privacy of the citizens, a somewhat different interpretation of privacy has been introduced. This has also resulted in the fact that, although people in the democratic countries will differ little in opinion about their view on privacy, the legislation about this subject and the measures to protect the citizens actively, differ widely across the democratic countries: there are distinct differences between the privacy regimes. And the interpretation of privacy in the remaining countries will differ even more.

Traditional privacy theories have tended to fall into one of two broad definitions, which are described as “non-intrusion”, and “exclusion” theories. The non-intrusion theory views privacy as “being let alone” or “being free from unauthorised intrusion”. On the other hand, the exclusion theory equates privacy with “being alone” [Tav99, Tav00].

Critics point out that the non-intrusion theory tends to confuse privacy with liberty. They argue that it is possible for one not to be let alone but still have privacy, and for one to be let alone and yet not have privacy. They claim that the exclusion theory also tends to confuse privacy with solitude [Tav99]. For example, can we say that the more alone one is, the more privacy one has? Instead, as critics point out, it can be possible for one to have privacy while not necessarily having complete solitude.

Both theories are often called “psychological privacy”, in that they seem to focus on psychological harms to a person that result from either physical intrusion into one’s space or interference with one’s personal affairs [Reg95]. However, the concept has been changing. Moor [Moo97] says: Rather than on psychological concerns, he argues, recent theories of privacy have tended to centre on issues related to personal information and to the access and flow of that information.

As information privacy became a distinct category of privacy concern, two theories of privacy closely related to private information have been discussed, which are the “control” and “restricted access” theories [Tav99, Tav00]. According to the control theory of privacy, one has privacy only if one has control over information about oneself. However, on a practical level, it is impossible for one to completely control over every piece of information about oneself. Another weakness of the control theory is that it confuses privacy with autonomy in focusing too much on the aspect of control [Tav00]. On the other hand, viewed from the restricted access theory, privacy consists in the condition of having access to information about oneself limited or restricted in certain contexts. One of advantages of this theory is that unlike the control theory, it recognises the importance of contexts or “zones” of privacy [Tav00]. However, this theory is criticised for underestimating the role of control or choice that is required in privacy. For example, someone who has privacy must be able to choose to grant, limit, or deny others access to information about oneself.

As viewed above, it seems that none of the theories - “non-intrusion”, “exclusion”, “control”, and “restrict access” - sufficiently explain enough the concept of privacy. Nonetheless, each theory somehow provides an important insight in understanding the concept of privacy comprehensively. If so, how can we combine those theories and successfully articulate the concept of privacy?

Restriction on the interpretation of privacy within the PISA project

In addition to the above, and to restrict the area of attention, within the PISA project only informational privacy will be studied. The reason for this is twofold. The first reason is primarily theoretical: the targets of attention of PISA are intelligent agents, data mining, databases and the like and these entities are about information and information processing, and have no obvious link with social privacy, physical privacy and other aspects of privacy. The second reason is of a more practical nature: the limitation is necessary to keep the work feasible within the budget and time scale of the PISA project.

The Global Privacy Principles

Global privacy principles are an abstraction of information privacy and are introduced to have a set of principles as a foundation for the methodology for its universal applicability, hence independent on the elaboration and interpretation of what privacy actually embodies in a certain legislation or environment. It has to be borne in mind, that, there are a number of ”privacy regimes” in the world, which differ in legislation, rigour of interpretation, forcefulness of operation and so forth; the one regime is simply more permissive than the other.

Although by definition formulated at a high level of abstraction, presentation of the global principles for informational privacy makes sense for a number of reasons. First, it provides a comprehensive framework of thinking about the subject. Second, it introduces some essential terms. Third, as for the methodology, consideration of these principles will prevent the view on the subject to be narrowed and will promote wider applicability of the results of the research.

The global privacy principles comprise the following four rules:

Principle of existence of privacy A data subject possesses an identity and other pertinent and variable information that this person may consider to belong to his or her privacy domain. This collection of information is called personal data.

Principle of withholding The data subject has the right and should be equipped with the ability to withhold some or all of his or her personal data to other persons and organisations at this person’s private choice. This, however, can be thwarted by legislation and regulations of national and international legislators.

Principle of controlled dissemination The data subject has the right to disclose some or all of his or her personal data to other persons and organisations, the collectors of the personal data, at this data subject’s own choice. This data subject may issue constraints on the further dissemination of the personal data and further processing of this data and has the right to change the personal data, to extend and restrict it, to withdraw this information and to change the constraints.

Principle of trusted usage The person or organisation that receives the personal data and stores it is called the collector. This collector of information that is marked to be one data subject's personal data has the obligation to keep to the constraints on dissemination and processing. Furthermore, this collector has the obligation to inform the person involved of its possession of personal data and to provide the opportunity for change. If so permitted, the collector may copy the personal data to one or more processors for further processing. A processor is obliged to obey the constraints. A collector can also act as a processor.

11.2.2 Articulating privacy

Moor [Moo97] suggests an advanced concept of privacy, called the "control/restricted access theory". In Moor's definition, the notion of a situation is emphasised to distinguish the loss of privacy from a violation of privacy. In his notion, a situation can be an "activity", a "relationship", or a "location", such as the storage, access, or manipulation of information in a computer database. Moor distinguish "naturally private situation" from "normatively private situation". In the former situation, one can be protected by natural means – e.g., physical boundaries in natural settings – from observation or intrusion by others. In the latter situation, one's privacy can be protected by ethical, legal, and conventional norms.

In a naturally private situation, privacy can be lost but not necessarily violated because there are no ethical, legal, and conventional norms according to which one has a right to be protected. In a normatively private situation, however, individual can be protected by norms (Moor, 1997: 30). Thus, when considering Moors view, it is not necessary for individuals to have absolute or unlimited control in order to have privacy. Although Moor conceptualises a "situation" vaguely, a situation paradoxically can be applied to broader concept of privacy. In this sense, Moor's theory seems to satisfy most of the specific conditions for an adequate theory of privacy. However, even though we have discussed that privacy related to personal information depends on a situation in Moor's term, one more question must be assessed. That is information itself. Much of the personal information (data), collected and manipulated by business companies, seem to be considered public in some senses. If they are literally public, it is needless to evoke privacy issues. Thus it is important to consider whether Moor's theory also provides a good means for determining when personal data is private or public.

What is personal information?

"Personal information is the individual's telephone number which is generally publicly available, as well as sensitive information about [the] individual's age, sex, sexual orientation, medical, criminal, and education history, or financial and welfare transactions. Personal information would also include biometric information, such as blood type, fingerprints and genetic makeup" (Canadian Information Processing Society, 1997).

Within the scheme of definitions in the European Union legislation document 395L0046, personal information is termed personal data and the individual involved is termed data subject. To avoid confusion, the European definitions will be applied as much as possible. For more information refer to the European document or to PISA deliverable 7: 'Methodology for Threat Analysis', in which document the most important definitions are listed and explained.

According to [Moo97], in the process of determining whether certain information is private or public, it is neither the kind of information, nor the content of information itself. Rather, it is the situation or context in which the information is used that can determine it. Taking

Moor's example, we can see that at private colleges faculty salary schedules are often kept confidential, whereas at larger state colleges faculty salaries are sometimes open in public. Along this line, Moor's control/restricted access theory can be applied to the process of determining private vs. public concerns of privacy protection.

Within PISA, in accordance with the European privacy regulations dating from 1995 and 1997, personal data is all data the data subject wants to consider private. So it is fully at the discretion of the data subject, and hence irrespective of the circumstances, whether privacy protection is to be demanded.

Privacy concerns

Why do people care about privacy? In terms of psychological theory on emotion and valuation [Fri87, OCC88] the question is what concerns are on stake, what sort of states or outcomes do people aim to achieve or avoid when they are being concerned about their own privacy?

Five different concerns can be identified, each with its own notion of privacy.

Security concern The security concern is about the protection of the self, and the related ones, in the physical, mental, self-esteem, social, reputation and other senses against threats from the outside world. Here the risk of losing something valuable is at stake, be it health, physical integrity, family relationships, financial well-being, freedom of moving around, etc. This concern includes misuse and abuse of personal data in the widest sense: not just static information like address, social and marital status, education, job information, religion, credit card information, health condition, etc. but also information of a dynamic nature like information about whereabouts, activities, purchases, encounters with other people, subjects of discussions, medical treatment, etc.

This security issue not only embodies the abuse of personal data, but also covers the reception of threatening and unwanted information. Some information can be threatening and can still be needed, e.g. medical information; some people want to be informed about the seriousness of their condition, others won't or try to deny it when they receive it notwithstanding. A typical example is a threatening letter, which also is likely to come from an anonymous source.

The reciprocity concern, also 'fairness concern' This concern is about the loss of a certain advantage when interfering with another person or with an organisation. In the modern world reciprocity in the commercial context is usually covered by financial compensation. The feature of this type of compensation is that it is known beforehand, or can be acquired in time and with satisfactory accuracy. But in the social, religious and informational contexts, and possibly other ones, reciprocity is not immediately self-evident, is not necessarily based on a financial compensation but on another compensation scheme or not at all. And accurate knowledge about the compensation and valuation of what it is worth for both producer and receiver is not necessarily available. This may provoke an uneasy feeling in case of innocuous cases, but in more serious cases the sensation of privacy intrusion can be real.

The concern of being judged without retort Now and then one is put in the position that his or her feeling of self-respect or social respect is challenged by another person, an organisation, a publication, or an occasional event. This is embarrassing if this happens and the person ('data subject') involved is in the position to retort the challenge by means of adding or correcting information. A typical example is when a person is erroneously arrested. An interesting case happened recently in Greater Manchester. An impostor had stolen money from a cash machine. The police detected this and used the video tape to identify the criminal. Unfortunately the police took the wrong person and the video was broadcasted in the Crime File programme. The person displayed immediately reported to the police that he had legally taken £20 from his account. Despite this, he was arrested. Because he was arrested he was suspended from his job. Some corrections were later made by the police, they apologised and he got his job back. But he remained too shocked to be able to work. Who can blame him? Source: [McC01].

But more embarrassing, sometimes even devastating, is the case that the person involved is unable to clear the case and is arrested or even convicted. People have been released from prison, in some cases after many years of imprisonment, after it turned out that they could not have committed the crime - or that the evidence turned out too flimsy to justify the conviction. Naturally, these cases are exceptional, but also in minor cases the person involved can experience it as a serious infliction of his or her privacy.

Disruption of mental or social quietude and solitude Privacy violations may have a temporary or persistent negative effect on the victim, but the effect may also be brief and superficial. Nonetheless, if these occurrences happen often enough, they are at least annoying, but in more serious cases or persistent repetition, they may disrupt concentration, hence are a cognitive burden, may lower temper, and now and then may result in an outburst. Noise is a typical example, but also visual cues can be intrusive. Stalking, as the persistent behaviour of an individual obtrusively following or otherwise bothering a victim, hence active stalking, may belong to this category if the stalker really manages to disrupt the mental or social quietude of the victim. Sadly, the police can do little against it; as a rule it requires a court verdict to keep the stalker away. Yet, it is a serious and undeniable violation of privacy.

Continuous or intermittent observation of a person Even if the privacy of a person is not really affected, but that person is nonetheless continuously or at some intervals intentionally observed by another person, this can be experienced as an intrusion of privacy, this case of the secrecy of the dynamics of life: encounters, activities, places visited, articles bought etc. The experience of the victim can mainly be at the level of uneasiness and annoyance, but in more serious cases, of which passive stalking is an example, the series of observations can really disturb the mental stability of a person. The awareness of being followed or watched alone may cause the effect. Despite being a privacy intrusion, albeit seemingly a minor one (except for the victim!) it is almost impossible to get a court order to the stalker to change his or her behaviour. On the other hand, in other circumstances, being observed can be a motivator as well, as is the case in the sport arena, the theatre and even the classroom.

The distinctions given above are necessary to establish the right view on the potential threats, as seen from the viewpoint of the subject, the person experiencing the privacy intrusion. As a consequence, to protect against these threats, and certainly effectively, different privacy enhancement measures are needed. But not all privacy threats hidden in the concerns outlined here are linked with informational privacy, the focus of attention of PISA. So, once there is general understanding of the privacy issue, the focus has to be put on all the points where information is used to the detriment and disfavour of the person involved.

11.2.3 The Implications of Data Mining in the Context of Fair Information Practices

Around the world, virtually all privacy legislation, and the policies, guidelines, or codes of conduct used by non-government organisations, have been derived from the set of principles established in 1980 by the Organisation for Economic Co-operation and Development (OECD). These principles are often referred to as "fair information practices", and cover eight specific areas of data protection (or informational privacy). These are: (1) Collection Limitation; (2) Data Quality; (3) Purpose Specification; (4) Use Limitation; (5) Security Safeguards; (6) Openness; (7) Individual Participation; and, (8) Accountability. Essentially, these eight principles of data protection or fair information practices codify how personal data should be protected. At the core of these principles is the concept of personal control: the ability of an individual to maintain some degree of control over the use and dissemination of his or her personal information.

Concerns about informational privacy generally relate to the manner in which personal information is collected, used and disclosed. When a business collects information without the knowledge or consent of the individual to whom the information relates, or uses that information in ways that are not known to the individual, or discloses the information without the consent of the individual, informational privacy may be violated. Data mining is a growing business activity, but from the perspective of fair information practices, is privacy in jeopardy? To determine this, we reviewed data mining from a fair information practices perspective. The principles are presented and discussed in the light of privacy issues with data mining.

Collection Limitation Principle

There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

Data mining can be about the combination of personal data from different sources. Even if for each of these sources the data subject has given his consent, this does not necessarily imply consent for the combination, not even the temporary combination. It is even possible that the data subject is unaware of the potential consequences of such combination and has to be informed about it before realistically be able to express his consent.

Data Quality Principle

Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete, and up-to-date.

Any form of data analysis is only as good as the data itself. Data mining operations involve the use of massive amounts of data from a variety of sources: these data could have originated from old, current, accurate or inaccurate, internal or external sources. Not only should the data be accurate, but the accuracy of the data is also dependent on the input accuracy (data entry), and the steps taken (if in fact taken), to ensure that the data being analysed are indeed "clean".

This requires a data mining operation to use a good data cleansing process to clean or scrub the data before mining explorations are executed. Otherwise, information will be inaccurate, incomplete or missing. If data are not properly cleansed, errors, inaccuracies and omissions will continue to intensify with subsequent applications. Above all else, consumers will not be in a position to request access to the data or make corrections, erasures or deletions, if, in the first instance, the data mining activities are not known to them.

Purpose Specification Principle

The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.

The Purpose Specification Principle is closely associated with the two surrounding principles, i.e. the Data Quality Principle and the Use Limitation Principle. It implies that before, and in any case not later than at the time of data collection it should be possible to identify the purposes for which these data are to be used, and that later changes of purposes should likewise be specified. Such specification of purposes can be made in a number of alternative or complementary ways, e.g. by public declarations, information to data subjects, legislation, administrative decrees, and licences provided by supervisory bodies.

New purposes should not be introduced arbitrarily; freedom to make changes should imply compatibility with the original purposes. Finally, when data no longer serve a purpose, and if it is practicable, it may be necessary to have them destroyed (erased) or given an anonymous form.

The reason is that control over data may be lost when data are no longer of interest; this may lead to risks of theft, unauthorised copying or the like. Data mining can be in conflict with the purpose of the data, because, if unprotected by a clear purpose specification and purpose negotiation before data mining starts, the data can be accessed freely.

Use Limitation Principle

Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with the Purpose Specification Principle except: a) with the consent of the data subject, or b) by the authority of law.

Purpose Specification means that the type of personal data an organisation is permitted to collect is limited by the purpose of the collection. The basic rule is that data collected should be relevant and sufficient, but not excessive for the stated purpose. In other words, restraint should be exercised when personal data are collected. Use Limitation means that the purpose specified to the data subject (in this case, the consumer) at the time of the collection restricts the use of the information collected. Hence, the information collected may only be used for the specified purpose unless the data subject has provided consent for additional uses.

Data mining techniques allow information collected for one purpose to be used for other, secondary purposes. For example, if the primary purpose of the collection of transactional information is to permit a payment to be made for credit card purposes, then using the information for other purposes, such as data mining, without having identified this purpose before or at the time of the collection, is in violation of both of the above principles. The primary purpose of the collection must be clearly understood by the consumer and identified at the time of the collection. Data mining, however, is a secondary, future use. As such, it requires the explicit consent of the data subject or consumer.

The Use Limitation Principle is perhaps the most difficult to address in the context of data mining or, indeed, a host of other applications that benefit from the subsequent use of data in ways never contemplated or anticipated at the time of the initial collection. Restricting the secondary uses of information will probably become the thorniest of the fair information practices to administer, for essentially one reason: at the time these principles were first developed (in the late 70s), the means by which to capitalise on the benefits and efficiencies of multiple uses of data were neither widely available nor inexpensive, thus facilitating the old "silo" approach to the storage and segregated use of information.

With the advent of high speed computers, local area networks, powerful software techniques, massive information storage and analysis capabilities, neural networks, parallel processing, and the explosive use of the Internet, a new world is emerging. Change is now the norm, not the exception, and in the quickly evolving field of information technology, information practices must also keep pace, or run the risk of facing extinction. Take, for example, the new directions being taken intending to replace the information "silos" of old, with new concepts such as "data integration“ and "data clustering". If privacy advocates do not keep pace with these new developments, it will become increasingly difficult to advance options and solutions that can effectively balance privacy interests and new technology applications. Keeping pace will enable us to continue as players in this important arena, allowing us to engage in a meaningful dialogue on privacy and future information practices.

The challenge facing privacy advocates is to address these changes directly while preserving some semblance of meaningful data protection. For example, in the context of data mining, businesses could easily address this issue by adding the words "data mining" as a primary purpose at the time of data collection - but would this truly constitute "meaningful" data protection? Take another example: when applying for a new credit card, data mining could be added to the purposes for which the personal information collected on the application form would be used. But again, would this type of general, catch-all purpose be better than having no purpose at all? Possibly, but only marginally so.

The quandary we face with data mining is what suggestions to offer businesses that could truly serve as a meaningful primary purpose. The reason for this lies in the very fact that, at its essence, a "good" data mining program cannot, in advance, delineate what the primary purpose will be - its job is to sift through all the information available to unearth the unknown. Data mining is predicated on finding the unknown. The discovery model upon which it builds has no hypothesis - this is precisely what differentiates it from traditional forms of analysis. And with the falling cost of memory, the rising practice of data warehousing, and greatly enhanced processing speeds, the trend toward data mining will only increase.

The data miner does not know, cannot know, at the outset, what personal data will be of value or what relationships will emerge. Therefore, identifying a primary purpose at the beginning of the process, and then restricting one's use of the data to that purpose are the antithesis of a data mining exercise. This presents a serious dilemma for privacy advocates, consumers, and businesses grappling with the privacy concerns embodied in an activity such as data mining. To summarise, the challenge lies in attempting to identify as a primary purpose, an as yet, unknown, secondary use. We offer some suggestions on how to address this issue in the next section.

Security Safeguards Principle

Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.

This is not specifically related to data mining. It applies to all cases of personal data.

Openness Principle

There should be a general policy of openness about developments, practices, and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.

The principle of openness or transparency refers to the concept that people have the right to know what data about them have been collected, who has access to that data, and how the data are being used. Simply put, it means that people must be made aware of the conditions under which their information is being kept and used.

Data mining is not an open and transparent activity. It is invisible. Data mining technology makes it possible to analyse huge amounts of information about individuals - their buying habits, preferences, and whereabouts, at any point in time, without their knowledge or consent. Even consumers with a heightened sense of privacy about the use and circulation of their personal information would have no idea that the information they provided for the rental of a movie or a credit card transaction could be mined and a detailed profile of their preferences developed.

In order for the process to become open and transparent, consumers need to know that their personal information is being used in data mining activities. It is not reasonable to expect that the average consumer would be aware of data mining technologies. If consumers were made aware of data mining applications, then they could inquire about information assembled or compiled about them from the business with which they were transacting - “information” meaning inferences, profiles and conclusions drawn or extracted from data mining practices.

Ultimately, openness and transparency engender an environment for consumers to act on their own behalf (should they so choose). Consumers could then make known to the businesses they were transacting with, their expectations about the collection, re-use, sale and resale of their personal information.

Individual Participation Principle

An individual should have the right: a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him; b) to have communicated to him, data relating to him i) within a reasonable time, ii) at a charge if any that is not excessive, iii) in a reasonable manner and iv) in a form that is readily intelligible to him; c) to be given reasons if a request made under subparagraph (a) and (b) is denied, and to be able to challenge such denial; and d) to challenge data relating to him and, if the challenge is successful, to have the data erased, rectified, completed or amended.

Data mining operations are extremely far removed from the point of transaction or the point of the collection of the personal information. As data mining is not openly apparent to the consumer, then the consumer is not aware of the existence of information gained through a data mining application. This prevents any opportunity to 1) request access to the information, or 2) challenge the data and request that corrections, additions, or deletions be made.

Accountability Principle

A data controller should be accountable for complying with measures which give effect to the principles stated above.

The accountability principle is added to the list to put responsibility in the hands of the collector and processor of personal data. Hence this principle binds the collector or processor of data to follow the other principles.

11.2.4 Threats

A threat is a potential violation of the privacy of a person or group of people. Hence a threat in itself does not necessarily affect privacy, but it might. The reason that privacy protection reasons from the phenomenon of threat, rather than from an observed privacy violation, is that the removal of threats, hence the removal of potential causes, also will remove unwanted consequences. The same way of reasoning is followed in safety analysis, where the term hazard takes the role threat has in the arena of privacy.

Privacy threats may exist in many forms, especially if the various aspects of privacy are considered: physical privacy, mental privacy, medical privacy, informational privacy, etc.. Fortunately, in the area of informational privacy the number of threats becomes more delineated. The main threats to be considered for informational privacy in the European Union can be derived from the rules stipulated by the EU and consolidated in national legislation by the member states. These threats are formulated in PISA deliverable 2.1: ”Methodology of Threat Analysis”. However, in other privacy regimes the threats will be different, in general less in number and perceived severity, because, at the moment, the EU regulations are the most demanding called into existence.

11.2.5 Summary

In this section we discussed the Global Privacy Principles and looked into personal information of people and their concerns of it being exposed. Then, data mining is set in the context of the eight principles of the Fair Information Practices.

11.3 Data mining

This section provides background information about data mining in its various guises. The topics to be highlighted are:

- what are the possible objectives;
- what kind of people or organisations are going to use it;
- what are the techniques that can be applied;
- what are the operational conditions necessary to make data mining feasible;
- how is it performed in actual practice.

This section does not introduce the potential threats of data mining (or web mining as it is done with intelligent agents); that will be presented in another chapter, when the consequences of the availability of all the data on the web is explored.

11.3.1 What is data mining?

Some definitions:

> “Data mining is the use of automated tools to examine and analyse data that has been stored in a database in order to find new, previously unknown relationships.”
>
> “Data mining is a set of automated techniques used to extract previously unknown pieces of information from large databases. In technical terms, data mining is the application of artificial intelligence (AI) and other intelligent techniques such as neural networks, fuzzy logic, genetic algorithms, decision trees, nearest neighbour method, rule induction, and data visualisation, to large quantities of data to discover hidden trends, patterns, and relationships.” (Kurt Thearling, [The95])

Pilot Software (1998) explains the origin of the term as follows:

> ”Data mining derives its name from the similarities between searching for valuable business information in a large database...and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides.”

Simply put, data mining is the process of exploring large quantities of data in order to discover meaningful information about the data, in the form of patterns and rules. In this process, various forms of analysis can be used to discern such patterns and rules in historical data for a given business scenario, and the information can then be stored as an abstract mathematical model of the historical data, referred to as a data mining model. After a data mining model is created, new data can be examined through the model to see if it fits a desired pattern or rule. From this information, actions can be taken to improve results in the given business scenario.

Data mining is not a ”black box” process in which the data miner simply builds a data mining model and watches as meaningful information appears. Although Analysis Services removes much of the mystery and complexity of the data mining process by providing data mining tools for creating and examining data mining models, these tools work best on well-prepared data to answer well-researched business scenarios-the GIGO (garbage in, garbage out) law applies more to data mining than to any other area in Analysis Services. Quite a bit of work, including research, selection, cleaning, enrichment, and transformation of data, must be performed first if data mining is to truly supply meaningful information. Data miningdoes not guarantee the behaviour of future data through the analysis of historical data. Instead, data mining is a guidance tool, used to provide insight into the trends inherent in historical information.

Though the term data mining is relatively new, the technology is not. Many of the techniques used in data mining originated in the artificial intelligence research of the 80s and 90s. It is only more recently that these tools have been applied to large databases. Why then are data mining and data warehousing mushrooming now? IBM has identified six factors that have brought data mining to the attention of the business world:

1. A general recognition that there is untapped value in large databases;
2. A consolidation of database records tending toward a single customer view;

3. A consolidation of databases, including the concept of an information warehouse;
4. A reduction in the cost of data storage and processing, providing for the ability to collect and accumulate data;
5. Intense competition for a customers attention in an increasingly saturated marketplace;
6. The movement toward the de-massification of business practices [Cav98].

With reference to point six above, “de-massification” is a term originated by Alvin Toffler. It refers to the shift from mass manufacturing, mass advertising and mass marketing that began during the industrial revolution, to customised manufacturing, advertising and marketing targeted to small segments of the population.

11.3.2 Mining data or growing knowledge?

Data mining is connoted to mining. Unfortunately, the data mining metaphor has its limitations and misleading connotations. It is not just that it suggests that data mining involves spending long hours in confined spaces, digging deep in dirty data, with more than average health risks. That just makes it harder to find qualified personnel. Indeed, before data mining can be exploited proficiently, the data masses have to be organised and cleaned. But once that has been done, the mining activities are about the unearthing of nuggets of information but also the excavation of raw ore which would remain buried if the mining activities would not have happened.

The main difference between statistical analysis and data mining has to be understood. Though the two are not that far apart in most cases, and statistical analysis can be a component of data mining tools, the difference can most clearly be explained at its extremes. Statistical analysis is most proficient when a research investigation is started and the researchers can determine prior to the start of their work which data elements will be measured and stored for further analysis: purposeful selection of data. At the other extreme is a database of information, some part of it collected for a certain purpose, now no longer necessarily relevant, some of it just for the sake of gathering data. Data mining is to bring facts to the fore or to find especially valuable collections of data. Facts or ’golden rules’ or even ’rules of thumb’ are especially looked for, because that will really help the organisation forward. But a collection of data with a special flavour can also be advantageous, for instance to approach with a special offer, to develop a dedicated product or to identify a potential market segment.

The serious problem with the mining metaphor is that it feeds false assumptions about the nature, methods and results of processes of automated data analysis. First, while minerals and ores are static and have stable value, knowledge is a much more dynamic and evasive thing. There is no intrinsic value to a given fact or statistic, value resides only in the possibility of use. The value of information is determined by relevance to users. Moreover, it requires knowledge to evaluate relevance. Knowledge grows on knowledge. That is to say, by running all sorts of algorithms on lots of data, many patterns may be found, which are significant in a statistical sense. It is almost impossible for people to appreciate how many different ‘statistically significant patterns’ even a modest sized database may contain. For example, Sentient’s DataDetective-TDD contains data from an extensive consumer survey (SUMMO DGO) collecting about 2000 elementary facts for each of about 13.000 respondents. Current estimates are that the total number of statistical significant patterns – i.e. group X significantly differs from group Y with respect to feature Z – in this database is some orders of magnitude higher than the total number of elementary particles in the known universe. But in order to interpret and evaluate the value and relevance of these patterns, someone, or something must apply the required specific knowledge to determine relevance and make sense out of bewilderment.

The activity, not to say the art of excavating information from data to produce knowledge, can also be captured in another metaphor. People always have a mental model of the world they work and participate in; in fact, each person has a collection of such models, which are by no means necessarily consistent and congruous internally and mutually. Such models offer facilities to their fosterers: they are necessary for the understanding of their world, for the management of it (so that it can be used in a beneficial manner to the fosterer of the mental model), but also for prediction and other extrapolations and to assess the value and trustworthiness of new information. So, to illustrate this point, occasionally, people will deny a fact because it is inconsistent with their mental model; outspoken examples occurred when new scientific discoveries clashed with existing ’truths’. But such a model needs confirmation and refinement but also correction: the model is by no means necessarily complete and correct.

Confirmation does not seem to be necessary, but people need confirmation now and then that they have the right idea. Also: an attempt for confirmation may also reveal false assumptions and, on top of this, a mental model is not necessarily fully conscious and to make it conscious is also part of the confirmation process. This three-fold process: confirmation, refinement and correction can be supported by data mining. Sitting on top of the organisational pyramid, or any part of it, a manager will attempt to confirm his or her mental model of some kind and may use the data masses to do this. ’Just try to find out whether this or that is true and to what extend.’ Also models with a more mathematical and statistical flavour can be treated in the same manner; from a philosophical point of view there is no difference: the model is based on the conscious knowledge of the researchers and now confirmation is needed. Or correction of course.

11.3.3 Alternative metaphor: cultivating knowledge by feeding models with data

A mental model is supported by world knowledge which helps to understand and interpret the world and the information it offers. The world knowledge is the basis and from this the mental model is created, which model provides the superstructure to bring the knowledge together and to offer the facilities of the mental model. This collection of world knowledge, be it used in models or not, is the apperceptive mass. It can be extended by adding new rules and other knowledge.

The creation of a mental model from the apperceptive mass can also be seen as a data mining experience: to add understanding to otherwise unrelated knowledge. Also this metaphor can be used to explain the meaning of data mining. Also the transition of unconscious knowledge into conscious knowledge is part of the metaphor: an organisation does have knowledge in its database, but it does not know it, yet. So the creation of a mental model has its equivalent in the formalisation of knowledge in a mathematical, statistical or other type of model.

11.3.4 The issues of location and data recognisability

Data mining has a number of different manifestations. These can be discerned in two dimensions: location of the data and recognisability of the data.

The dimension of location indicates where the data can be found or be expected. On the one extreme this location or set of locations are precisely known and the desired data is not to be expected somewhere else. At the other extreme it is entirely unknown where the data is and also unknown is the number of locations where the total of the desired data might be found; because there is no certainty about where the data can be found, the data may remain invisible and might even be unavailable indeed.

In the former case the list of locations to be visited can be exhaustive list, of which one hit can be enough; possibly more hits are necessary to complete the desired collection of data. No feedback with the originator of the search will be necessary, because the functionality can perform all of the actions out of itself. This situation will occur when an organisation has constructed a data warehouse containing all the relevant enterprise data; the locations are definitely known, and if the wanted enterprise data is not in the data warehouse, it is not available. The term retrieval is applicable here, because the data is stored to be retrieved and measures have been taken to make that possible.

In the latter case some sort of search engine or search algorithm will be needed to identify the necessary data. It is likely that in this case a lot of communication with the originator is necessary to define the candidate data more precisely and to deselect the irrelevant data. This is a more common situation encountered in web mining: data mining on the internet and the WWW in particular; initially, there will be no knowledge or experience about where to go, but search engines may help. In this case the term retrieval is of little meaning. Maybe it is stored by somebody at some time, and if the location can be identified, and the data can be recognised, retrieval is possible, but this is not necessarily always the case. Accuracy and reliability are also terms to be mentioned. There is no mechanism that ensures that the data offered meets any level of accuracy. And in commercial offerings some sort of deliberate deception is part of the game to attract the attention of customers, for instance low prices for items that are 'out of stock' the moment they are requested, with more expensive replacements being offered; the internet equivalent of practices in everyday commerce.

The dimension of recognisability defines whether the desired data really exist, is available and can be identified, irrespective of location. Certain data will exist, be available and be identifiable, e.g. the various types of items for sale. In such a case a deterministic search can be organised: because of its known availability in a known form, the search is expected to be successful. But other data may not even exist, or if the requested data exists it may not be available. And if it is available in some form, the issue remains on how to describe the data and to recognise it when offered, because some facts may be buried in a pile of text. This is a typical case of a non-deterministic search. In criminal investigations by the police, this is a common occasion: the suspects will try to conceal the data that may corroborate the evidence and the police can be in the dark about what to search for.

11.3.5 The characteristics of web mining

Traditionally, data mining starts with the data given - e.g. customer and transaction data collected by a company - and aims to find relevant patterns and regularities in customer behavior. Web mining comes in two varieties: web usage mining and web content mining. Web usage mining starts with surfer data - e.g. in log files from sites, or sometimes directly produced by the browser used by the web surfer - and aims to find profiles for users relevant to site owners. Web content mining starts with the data content provided or accessible at sites and typically aims to find profiles for sites, relevant to users/surfers. An example is rating sites on adult content - to protect users from unintended interaction with these sites. However, web content mining can also be aimed at finding -factual- information concerning individuals. This then, is potentially the most threatening to privacy.

11.3.6 Data mining: which organisations and people will use it

Data mining can be of benefit to any organisation or person that has access to large databases or to the internet with its many data offerings. Especially in the cases of a data warehouse that is well-organised towards data mining or for the search for deterministic data at predetermined locations on the web data mining can be performed well. The technology may however be a barrier to successful application, even in the mentioned relatively simple cases. The situation becomes unpromising when the technology has to be applied to non-deterministic search at unknown locations. Even if the technology will be adequate to do it with some degree of success, which at the moment isn't, then the technicalities and the development cost of such advanced applications can be prohibitive.

11.3.7 The data mining process

The data mining process is captured in a number of operational models: 'the way of working'. Some are based on the observation of good practice, followed by a formalisation step; the result is comparable to the 'waterfall model' of project management, which emerged in the same manner.

A primitive view on data mining reveals three basic steps. The first processing step is data preparation, often referred to as “scrubbing the data”. Data is selected, cleansed, and pre-processed under the guidance and knowledge of a domain expert. Second, a data mining algorithm is used to process the prepared data, compressing and transforming it to make it easy to identify any latent valuable nuggets of information. The third phase is the data analysis phase where the data mining output is evaluated to see if additional domain knowledge was discovered and to determine the relative importance of the facts generated by the mining algorithms.

The CRoss Industry Standard Process for Data-Mining (CRISP-DM, or just CRISP for short) on the other hand is based on a more theoretical approach of the matter, though, needless to say, practical experience guided the theorists to their formulation. The CRISP model is the result of a European project to devise a methodology for data-mining, which became the CRISP methodology, described in the CRISP-DM. See [CD99].

The methodology comprises of six steps, brought together via a reference model. This reference model contains a feedback loop to cater for improvements of previous steps. A brief overview of the six steps, as taken from the original:

Business understanding This initial phase focuses on understanding the project objectives and requirements form a business perspective, then converting the knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

Data understanding The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.

Data preparation The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modelling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record, and attribute selection as well as transformation and cleaning of data for modelling tools.

Modelling In this phase, various modelling techniques are selected and applied and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often necessary.

Evaluation At this stage in the project you have built a model (or models) that appears to have high quality form a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model and review the steps executed to construct the model to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue hat has not been sufficiently considered. AS the end of this phase, a decision on the use of the data mining results should be reached.

Deployment Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organised and presented in a way that the customer can use it. It often involves applying ”live” models within an organisation’s decision making processes, for example in realtime personalisation of Web pages or repeated scoring of marketing databases. However, depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise. In many cases it is the customer, not the data analyst, who carries out the deployment steps. However, even if the analyst will not carry out the deployment effort it is important for the customer to understand up front what actions need to be carried out in order to actually make use of the created models.

11.3.8 Purposes of data mining

According to Cavoukian [Cav98], data mining is usually used for four main purposes:

1. to improve customer acquisition and retention
2. to reduce fraud
3. to identify internal inefficiencies and then revamp operations
4. to map the unexplored terrain of the Internet

This reflects the more operational view on the subject, especially the last item. A more theoretical view can be explored:

1. *Internally oriented:*
    - to improve the understanding of management of the working of the organisation, to optimise its functioning;
    - to identify the staff members with the best results (which members can be different from what the managers think);
    - to detect possible fraud and to take measures against it;
    - to detect possible commercial espionage and to take measures against it;

2. *Externally oriented:*

   - to improve the understanding of the expectations, desires and complaints of the existing customers to satisfy those in order to keep those customers;
   - to improve the understanding of the needs and desires of potential customers with the intention to meet these needs in order to acquire those customers;
   - to improve understanding of the market and the needs for new or improved products or services;
   - to improve understanding of the behaviour and performances of competitors, in order to improve the own position in the market, increase of the market share, or the effort to keep the current share;
   - to identify the potential for diversification, the development of entirely new products or the creation of a market share elsewhere.

The difference between the two groups, internally versus externally, is that the data about the first group can be acquired more easily, with more control, continuously and reliability. If so needed, more data can be gathered. The acquisition of the data for the externally oriented data mining investigation is more difficult and discontinuous, with lower accuracy and at higher cost. And the most obstructive part: the data remains elusive: parts of it remain to be guesswork and the interpretation is not always clear. The survey presented above is data mining from the commercial and business perspective. There are also governmental and scientific/research perspectives with objectives of its own. A long standing exercise of governmental data mining is the analysis of road accidents with the aim to improve road safety by means of the reconstruction of roads, changes in legislation, improved driver education and even improvements in vehicle construction. Also the unravelling of drug related criminality, which often occurs in a network of people and their activities, is a similar kind of work: an enormous amount of data is collected, from which the researcher (the detective) has to extract a pattern with the aid of a number of tools and techniques and also: plodding on. Many a scientific enquiry in the field of medicine, history, archaeology etc. follows a similar pattern. In contrast: research in physics and chemistry follows the more conventional way of purposely collection of data.

Data mining and data warehousing

Although not an essential prerequisite, data mining potential can be enhanced if the appropriate data have been collected and stored in a data warehouse - a system for storing and delivering massive quantities of data. “Data warehousing is the process of extracting and transforming operational data into informational data and loading it into a central data store or warehouse” [Cav98]. The promise of data warehousing is that data from disparate databases can be consolidated and managed from one single database.

The link between data mining and data warehousing is explained as follows: Data warehousing is the strategy of ensuring that the data used in an organisation is available in a consistent and accurate form wherever it is needed. Often this involves the replication of the contents of departmental computers in a centralised site, where it can be ensured that common data definitions are in the departmental computers in a centralised site, where it can be ensured that the common data definitions are in use. The reason data warehousing is closely connected with data mining is that when data about the organisation’s processes becomes readily available, it becomes easy and therefore economical to mine it for new and profitable relationships [Cav98].

Thus, data warehousing introduces greater efficiencies to the data mining exercise. "Without the pool of validated and scrubbed data that a data warehouse provides, the data mining process requires considerable additional effort to pre-process the data". Notwithstanding, it is also possible for companies to obtain data from other sources via the Internet, mine the data, and then convey the findings and new relationships internally within the company via an Intranet [Cav98].

There are four stages in the data warehousing process: The first stage is the acquisition of data from multiple internal and external sources and platforms. The second stage is the management of the acquired data in a central, integrated repository. Stage three is the provision of flexible access, reporting and analysis tools to interpret selected data. Finally, stage four is the production of timely and accurate corporate reports to support managerial and decision-making processes [Cav98].

Data mining and data warehouses complement each other. Well-designed data warehouses have handled the data selection, cleaning, enrichment, and transformation steps that are also typically associated with data mining. Similarly, the process of data warehousing improves as, through data mining, it becomes apparent which data elements are considered more meaningful than others in terms of decision support and, in turn, improves the data cleaning and transformation steps that are so crucial to good data warehousing practices.

Examples of data mining

Retailers, who utilise point-of-sale databases, use the records to send targeted promotions based on an individual's purchase history. By mining demographic data, retailers can develop products and promotions to appeal to segmented consumer groups. Large data mining companies such as HNC Software and IBM have used data mining techniques to detect credit card fraud and to evaluate real estate [Enn98].

The health care field is a quickly growing user of data mining applications. Here, the technique can be used to directly assist practitioners in improving the care of patients by determining optimal treatments for a range of health conditions. One highly beneficial use of data mining in this field is its use in assisting health care workers to distinguish patients who are statistically at risk for certain health problems so that those patients can be treated before their condition worsens [Cav98].

Data mining has also found a friend in the National Basketball Association (NBA). The Advanced Scout is a data mining software that can analyse the movements of players to help coaches devise plays and strategies [Pal96]. While the coaches were using data mining to analyse a January, 1995, game between the New York Knicks and Cleveland Cavaliers, they found that when Mark Price played Guard, each of John William's jump shots was successful. Imagine their surprise when the software informed them of this fact. The shooting percentage for the Cavaliers as a team during that particular game was only 49.3%!

Knowledge results of data mining

Generally, data mining yields four types of information: associations, sequential patterns, classifications, and clusters [Cav98, Pal96].

1. Associations
   Data can be mined to recognise associations, which happen when occurrences are linked in a single event. For example, when a certain item (beer) is purchased, other items (chips, nuts or pop corn) are likely to be purchased as well.

2. Sequential patterns
   We can foresee sequential patterns or trends by analysing events linked over time. For example, if a house is bought, then 45% of the time a new oven will be bought within one month and 60% of the time a new refrigerator will be bought within two weeks.

3. Classifications
   Classification can help discover the characteristics of a certain group, and provide a model that can be used to locate a relationship in predetermined groups. For example, a coffee shop could mine data to ascertain the periods of peak and low customer visitation. This information, in turn, can help determine what kinds of promotions can be effective.

4. Clusters
   By using data mining, data items can be grouped according to some predetermined and logical relationships of data. For example, cluster technique can be applied to pinpoint certain consumer groups, such as those who can afford and like to buy luxurious cars.

A fifth can be mentioned, though it can be thought to be included in the listed results. This is the identification of rules or patterns, which express some ’law of nature’ or ’behavioural law’. To the organisation this is certainly valuable, even if the rule is of a soft kind: ’rule of thumb’. Such a rule can be regarded to be an association, or even to be a sequential pattern, but sometimes there is simply no linking event or whatever, there does not seem to be a clear-cut causal relationship, because non-recorded other factors play a role as well. So, sales of a certain product may drop suddenly if a competitor is in the market with a special offer; data mining may lead to wrong conclusions. But a rise in sales at a certain day according to a repeating pattern may be caused by an unknown external factor. The existence of ice ages, or better: the fluctuation in global temperatures was identified in this manner; the explanation followed much later. Note that it is difficult to either group it under association (linked in a single event) or sequential pattern (events linked over time). Generally then, applications of data mining can generate outputs such as:

- Buying patterns of customers; associations among customer demographic characteristics; predictions on which customers will respond to which mailings;

- Patterns of fraudulent credit card usage; identities of “loyal” customers; credit card spending by customer groups; predictions of customers who are likely to change their credit card affiliation;

- Predictions on which customers will buy new insurance policies; behaviour patterns of risky customers; expectations of fraudulent behaviour;

- Characterisations of patient behaviour to predict frequency of office visits.

11.3.9 Observations on data mining practices

Data mining aims to abstract generalised knowledge from large amounts of specific data. The typical aim of data mining as performed by companies operating on consumer markets for instance is general knowledge, i.e. patterns and regularities in customer behaviour, preferences and habits, not specific knowledge about individual customers. Extensive specific and factual information concerning individual customers is the starting point, not the product of data mining.

The two types of data mining in marketing databases (cf. EW e.a. 96) are exploratory modelling, or segmentation and profiling, and predictive modelling, or scoring and selection. In exploratory modelling, a marketer aims to find groups of customer with common profiles of attributes, 'target groups' to which particular marketing efforts can be directed. In predictive modelling a marketer develops scoring models that predict what customers are most likely to show desirable or undesirable behaviour (e.g. buying a particular product vs., churn or not paying bills). The results of such efforts are theories - if highly local ones - not facts.

Nearly all data mining projects in CRM are only concerned with 'the anonymous part' of consumer data. For most rules and patterns to be found by data mining exercises, names and -exact- addresses are not relevant. A typical case of generalised knowledge is exemplified by the lack of interest in the exact address. The full address of the house number, street and town-name type is often more a burden than a help. The behavioural pattern of the inhabitants of a larger collection of houses is more of commercial interest than that of one particular address, which may change behaviour quite suddenly, because of a change in family composition, marital status of the owner or even complete removal. The larger collection can be expected to show a more stable pattern, which obviously gives the commercial organisation more to hold on than an unstable singularity. Additionally, the potential greater details of the individual case is often of little interest. This generalisation counts on various levels. It may have its value at the level of a couple of streets, a neighbourhood, a town, a county, a country etc. Some preferences or patterns are applicable to a country as a whole, while others apply to a town or to a smaller unit of generalisation.

It is a different story for 'web-mining'. With automated collection and processing of information from the internet, the aim may well be to gain specific knowledge about particular entities. This is rapidly becoming common practice in B-to-B markets, and may result in privacy risks if it becomes available for collecting information on individuals. The difference between the web case and the commercial, enterprise owned database is that a stalker-type individual can more easily collect information from the web than from the database. Naturally, comparable abuse may occur in an organisation, but then the individual can more easily be identified than in the anonymous wilderness of the internet, especially when harnessed with intelligent and anonymous agents. Hence, the web environment is more conducive to such undesired activities as other environments, which are better controlled and also may have an incentive to guard their data base and to detect abuse.

11.3.10 Data mining and ISAT

Data mining and intelligent agent technology may appear to clearly stand apart in their respective primary purposes and fields of application, yet these technologies maintain a quite intimate and intricate relationship on closer inspection. DM: deriving useful knowledge from 'data crunching', i.e. machine processing of large amounts of data, is different from ISAT: designing and developing software systems that can autonomously help us perform a variety of information processing -search, retrieval, selection, communication tasks; or is it? A first clue is that DM and ISAT draw from the same academic sources - Artificial Intelligence and surrounding disciplines- and are employed in the same arena of advanced ICT applications; mobile 24/7 E-commerce, or perhaps even the development of Internet into a persistent omnipresent Intelligence Matrix. What these technologies have in common is that they are both knowledge based, their success in applications depends to a large extend on the ability to represent and reason with relevant knowledge. Data mining is essentially concerned with the automation of statistical analysis -exploratory or predictive- on large amounts of data; the crux of getting agents intelligent -adaptive and resourceful- is to have them perform more analysis on more data.

The close relationship between data mining and intelligent agent technology then follows from their mutual supporting role in realising applications. Many agent applications could benefit from equipping agents with basic data mining skills, to detect exceptional events, predict the outcomes of alternative actions or profile users and other agents. Data mining, on the other hand can benefit from agent technology to the extent this helps in making data mining easier, for a wider range of users, by offering agent support for difficult tasks.

Data mining is still considered a demanding task for specialists. What makes it difficult? If we follow the steps of the CRISP process model discussed before, with particular attention to the knowledge involved, it becomes clear why the goal stated by many -both users and suppliers- of putting data mining tools directly in the hands of end-users -without specific skills or training- is very hard to achieve. In fact, this is the long term big challenge for DM and ISAT technologies together.

Business understanding. ‘understanding the project objectives and requirements from a business perspective’ clearly is a highly knowledge intensive task. The knowledge involved is not just an understanding of the specific business rules or logic -what are the aims and the means, what defines success or failure, profit or loss- which may be complex enough. This specific knowledge cannot easily be separated from much wider knowledge domains concerning markets and economics, society and psychology, which at any moment may become highly relevant for understanding particular business objectives or vulnerabilities. In fact common sense is required at this stage, and we know that is a lot of knowledge to represent.

Data understanding, requires first of all extensive technical knowledge of basic statistics, measurement theory, data modelling and databases. But then, in order to relate the data in records to the events and processes that generated these data - a necessary requirement for identifying data quality problems - some common sense again comes in handy.

Data preparation is a highly skilled technical job, requiring the same kinds of technical skills and knowledge as the preceding one, and then preferably extensive experience with data preparation practices.

Modelling. Interestingly, the task of actual model building, receiving by far the most attention in the literature -it is here that fancy algorithms for machine learning, evolutionary computing, neural nets- are employed, is the easiest one to automate, since the relevant knowledge for guiding this process -given adequate tools- is relatively limited.

Evaluation, again requires on the one hand solid knowledge of test methods and statistics, and on the other hand a much wider knowledge on business and ‘real world’ issues, while success in Deployment is almost entirely dependent on such softer, social and organisational skills and knowledge.

A question to be answered, then, is whether an intelligent agent can be equipped for data mining on the web. The answer is: if there is data on the web that can be selected and understood by the agent, and the agent is able to deliver the information at its home site unconditionally, then data mining on the web seems to be possible. The obstacles are in fact the same as those that would obstruct data mining success on any commercial database. The data have to be more or less in a data warehouse format, must be approachable and recognisable by an external actor like an agent, after which the useful data needs to be copied.

To enable this all, the agent has to be instructed about:

1. the type of data that is wanted
2. other selection criteria
3. the sites to visit
4. recognition of trustworthy sites (not all interesting sites are by definition trustworthy)
5. the route to take (possibly)
6. when and where to deliver the data.

This information has to be captured in scripts. For reasons of versatility, this information should not be packed in software code, but in data. Currently, there are no standards for this, so each manufacturer may produce scripts of his own making.

11.3.11 Conclusion

As outlined before, in brief, the data mining process may result in two different results: 'nuggets of knowledge' condensed in rules and facts and the extraction of a set of data with special and interesting characteristics. To extract nuggets, the intelligent agent data mining process is obviously less feasible, though not impossible. The main reason is that the amount of coding necessary to derive the rule or the fact is rather alien to the idea of the free roaming agent. To cater for this possibility the list of types of instructions has to be extended with one that catches the script to derive the rule. To collect data about a certain subject, or to select data with certain characteristics is an activity that is more in harmony with the idea of an agent.

11.4 Data Accumulation and Information Increase

The available data in files, databases, other storage media and the internet increases not only uninterruptedly but also with an enormous speed. The nature and backgrounds of these growths, and the opportunities that they offer for data mining will be illuminated. The main opportunity is record linkage: the bringing together of related, but as yet uncoupled information and the various ways in which this can be performed will be presented.

11.4.1 Data Accumulation and Information Increase

The modern world has various ways in which data can be acquired. By means of information technology this is also followed by an ever increasing amount of data: data accumulation. To be more precise: the technology to handle this amount is also conducive to the collection of data; if the data could not be stored (memory) and processed (processing power) data collection would be pretty senseless.

This accumulation of data also facilitates the increase of information, even further away, new knowledge. Raw data in itself is of low value, primarily because of the amount which prohibits human processing, but from raw data information can be extracted by virtue of purification, condensation, selection and other processing techniques. The question is: what kind of techniques can be applied to extract information and knowledge from the heap of data.

11.4.2 The Increase of Registrations - the drawbacks

The mentioned accumulation of data is however not in one single collection, but on a large and ever increasing number of separate registrations. This increase is caused by the many people and registrations responsible for such registrations, but also by the further penetration of information technology to deal with that. This development is expected to continue at least in the next decade or even decades.

However, this increase shows also some drawbacks, among which fragmentation, variation in representation and visibility to the outside world are the chief ones. If the facility of data mining was restricted to the same database, or the same representation of data, and data collections that are fully visible to the potential audience, then data mining in general and web mining would have little to offer.

Fragmentation means that data that have some relationship with other data remain separated; bringing such data sets together is likely to generate new information. Variation in representation means that similar data is packaged in different ways, with the result that the data cannot be used interchangeably. (A comparable phenomenon is that the same information can be conveyed via different languages, but the different languages cannot be understood without proper education.)

Lack of visibility means that there is no real obstacle from using the data other than that the existence of the data is hidden from view, that search processes cannot find them, that search engines think they belong to a different class. Naturally, the technology of web mining has to target these issues to make the applicability of the technology as wide as possible.

Data enrichment and Record linkage Record linkage could be involved in data enrichment, which is part of the data mining process as described in Section 11.3.2. The specific side of data enrichment aimed in record linkage is to extract new information by means of the bringing together of isolated pieces of related information. What type of new information is expected can be known beforehand, but not necessarily so.

Definition:

> Record linkage is defined as combining existing person-specific data with additional data that refer to the same persons, their family and friends, school or employer, area of residence or geographic environment.

Record linkage is included here to analyse the obvious threat to privacy, certainly if this linking occurs without the explicit consent of the data subject. The data subject may forbid processing of his or her personal data for this purpose. On the other hand, record linkage is the technique of choice to unravel activities of a criminal gang in smuggling, drug production and trafficking, the transport of illegal immigrants, false document production and delivery, forgery of pieces of art, money counterfeiting and all other activities that may happen illegally and hence may cause the police to overrule privacy legislation. Such a criminal gang is typically a network of people, with places to work, equipment, transport means, routes, contacts, methods etc. all of which will be of interest to the police. The police then can construct a pattern of operation, the network of people involved etc. Also more mundane activities that are in conflict with the law like tax dodging or social security related fraud can be analysed with the same techniques.

11.4.3 Generating new information

Linkages fall in two main types: person-by-person linkages (including multi-person, survey-archive, and multi-archive links) and person-by-context linkages.

Person-by-person linkages generally fit Newcombe and colleagues' basic definition of linkage as "the bringing together of two or more separately recorded pieces of information concerning a particular individual or family." We provide examples of three kinds of person-by-person linkage: multi-person links, survey-archive links and multi-archive links. Each represents a different kind of data combination.

Multi-person links create dyads or small groups based on a specified relationship between persons (e.g. husband-wife) and on data indicating which respondents or data subjects have this relationship to each other. Multi-person linkages traditionally combine data on members of a nuclear family (e.g., creating parent-youth pairs from survey data, combining SSA earnings records for husbands and wives, creating "tax families" by combining household members' tax returns). Survey-archive links match information provided by survey respondents with existing records on these same individuals. Multi-archive links combine two or more full sets of existing records.

Person-by-context linkages bring together information on a person and a larger entity related to him or her (e.g., a student and his or her school). Person-by-context links are unlike the foregoing in that they combine person-specific data with contextual information on larger entities (geographic areas, political subdivisions, schools, and employers). Linkages are based on knowing where each respondent or data subject lives, what school he or she attends, or who his or her employer is.

11.4.4 Privacy issues

Among the privacy issues that may arise with respect to record linkage for research and statistics are five examples:

- **Consent to linkage.**
  Although data subjects' consent to linkage is sometimes obtained, in other instances, data subjects may be unaware that, in essence, new information about them is being created. Some linkages require data sharing between agencies, and when this occurs, certain laws and policies concerning disclosure and consent are relevant. Notably, the Privacy Act generally requires consent for disclosure from one agency to another, but there are exceptions. Furthermore, a data subject may have been asked for consent, and may have given that, without being informed about what it may mean in practice and what the potential consequences can be. To the data-naive people, such a request may seem fairly harmless, because the isolated pieces of information are also harmless and occasionally, linking may be harmless indeed. Active education on this point of the general public is needed to prevent deceptive requests by large organisations.

- **Data sharing.**
  In order to compile the information needed for record linkage and "make the link", agencies must often share identifiable person-specific data. But traditionally, data have been kept separately, and various statutes have been enacted to prohibit or control certain kinds of data sharing. Privacy concerns stem from a desire to control information about oneself and a perceived potential for inappropriate government use, as explained below. Security risks could also arise during data transfer.

- **Re-identification risks.**
  Some datasets are linked using a code-number procedure or are stripped of explicit identifiers as soon after the linkage as possible; nevertheless, re-identification of at least some data subjects may be possible through a deductive process, so only controlled use would be appropriate. To facilitate broader access to statistical and research data, agencies have created more fully "de-identified" public-use datasets.

  Although many linked datasets are not made available for public use, some are - and concerns about the re-identification risks associated with these datasets are increasing. The point to remember is that, with sufficient detail of information, every person on earth becomes unique, even if the identity itself is not known.

- **Potential sensitivity.**
  The potential sensitivity of data (risk of harm to data subjects) cuts across all other privacy issues. This is true for linked data as well as for single-source datasets. However, as explained below, linkage may heighten the sensitivity of data that, taken by themselves, appear to be relatively innocuous.

- **Security of linked data.**
  Security is crucial to protecting stored data. For linked data, this is especially true because a linked dataset may be more detailed or more sensitive than its components.

Consent to linkage

The issue of consent to linkage derives from a core concept of personal privacy: the notion that each individual should have the ability to control personal information about him or herself. Perceptions about the need for consent may vary according to type of linkage. For example, consent requirements have been advocated for multi-archive links (because full sets of existing records often do not have a voluntary component) and for linkages that are not closely related to the original purpose of the data collection. Consent requirements have also been advocated when vulnerable populations are involved or when risks appear to be higher.

11.4.5 Data sharing

Currently, there are three sets of data-sharing issues relevant to record linkage:

Functional separation This concerns the issue of when it is proper for an agency to share data with another agency. Therefore we define the Principle of Functional Separation: Data collected for research or statistical purposes should not be made available for administrative action toward a particular data subject. According to this principle, individually identifiable information collected or compiled for research or statistical purposes, which logically would include survey data as well as records-research databases, may enter into administrative and policy.

Risks to confidentiality and security Other privacy issues in data sharing include risks to confidentiality (more organisations and more persons are privy to identifiable data) and certain security risks (e.g., risks during transfer). Some see data sharing as inherently risky and believe that increased safeguards may be needed - especially during transfer.

Different rigour of privacy constraints The existence of personal data constraints (or constraints on privacy related data) is explained in Deliverable 7. The essence of it means that restrictions on use and obligations are linked to the personal data, which are binding for the two parties data subject and collector of information. Also the processor of information, to which party the collector may send the information, is bound to meet the constraints. The problem arises when two sets are combined with different privacy constraints, because the data subject ordained two different constraints, e.g. to further distribution or retention. Even if consent is given for linking, there is still a limitation in that the most stringent personal data constraints is likely to apply to the linked set.

De-identification and re-identification

Federal agencies have a long history of creating public-use datasets with de-identified information. These de-identified data have provided researchers and members of the general public with maximum access to information; they have also helped agencies maximise the return on their investments in data collection. Growing concerns about re-identification risks have led to considerable efforts to develop methods aimed at minimising these risks. Such risks might be higher for linked datasets than for the component data. Agencies may face trade-offs between (1) attempting to meet the difficult challenges of minimising the re-identification risks associated with wider access and (2) providing more restricted forms of access to confidential data

Potential Sensitivity of Linked Data

The privacy issues discussed above - consent to linkage, data sharing to make the links, and the re-identification risk associated with dissemination of linked data - are all intensified when sensitive data are involved. And when sensitivity is increased, there is also a need for greater caution in releasing identifiable linked data to researchers outside the linking organisation(s). This is important because federal record linkage often involves sensitive information and we believe that the linkage itself can heighten sensitivity, as explained below. Although sensitivity is a subjective and relative concept, certain laws provide protection for what could be considered sensitive information. For example, the Computer Security Act of 1987 defines sensitive information as including any unclassified information that, if lost, misused, or accessed or modified without authorisation could adversely affect the privacy to which individuals are entitled under the Privacy Act.

Various data that appear to be of low sensitivity can become more sensitive when linked. For example, if a person’s survey report of income is linked to his or her tax return - and the results indicate disparate income reports - the linked data would be more sensitive than the original independent data (because there is a new implication about the individual). Even some context links could create sensitivity by, for example, identifying persons associated with residential areas, schools, or places of employment with negative characteristics (e.g., high rates of stigmatised diseases). In instances where negative contextual information is either not known to the public or difficult to access, linkage to a person-specific dataset might increase the sensitivity.

Overall, it seems fair to say that sensitivity is potentially increased whenever the “whole is greater than the sum of the parts”. And for a variety of reasons, certain questions -or linkages- may be perceived as sensitive by at least some data subjects even if there appears to be no risk of harm in the eyes of the researcher or other outside observer.

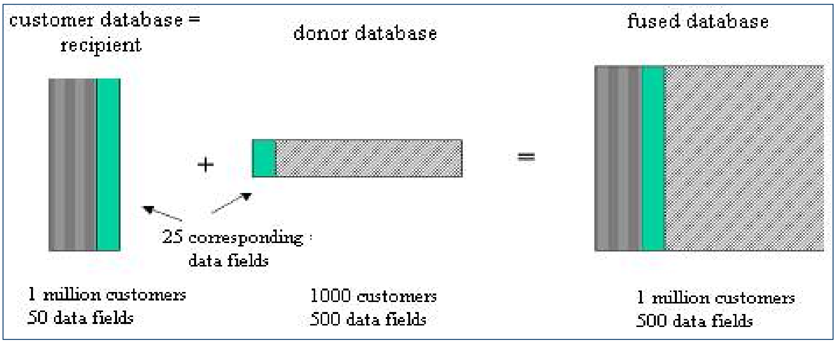


Figure 11.2: Data Fusion Concept.

Security of linked data

Security is important for all personal data and crucial for sensitive personal data. As noted above, even data that appear to be of relatively low sensitivity may become more sensitive when linked. At the same time, security has become particularly challenging as access to computers and the Internet has spread through the population, and agencies rely more extensively on computerised systems and electronically.

Associative data fusion and privacy

The amount of data that is collected about customers is generally growing very fast. However it is often scattered among a large number of sources. For instance, elementary customer information resides in customer databases, but market survey data depicting a richer view of the customer is not available for each individual. Simply collecting all this information for the whole customer database in a single source survey is far too expensive. A widely accepted alternative within database marketing is to buy external socio-demographic data which has been collected at a regional level. All customers living in a single region, for instance in the same zip code area, receive equal values. However, the kind of information which can be acquired is relatively limited. Furthermore, the underlying assumption that all customers within a region are equal is at the least questionable. Data fusion techniques can provide a way out. Information from different sources is combined by matching customers on variables that are available in both data sources. The resulting enriched data set can be used for all kinds of data mining and database marketing analyses.

The core data fusion concept is illustrated in Figure 11.2. Assume a company has one million customers. For each customer, 50 variables are stored in the customer database. Furthermore, there exists a market survey with 1000 respondents, not necessarily customers of the company, and they were asked questions corresponding to 1500 variables. In this example 25 variables occur in both the database and the survey: these variables are called *common variables*. Now assume that we want to transfer the information from the market survey, the donor, to the customer database, *the recipient*. For each record in the customer database the data fusion procedure is used to predict the most probable answers on the market survey questions, thus creating a virtual survey with each customer. The variables to be predicted are called *fusion variables*.

The most straightforward procedure to perform this kind of data fusion is *statistical matching*, which can be compared to k-nearest neighbour classification. For each recipient record those k donor records are selected that have the smallest distance to the recipient record, with the distance measure defined over the common variables. Based on this set of k donors the values of the fusion variables are estimated, e.g. by taking the average for ordinals or the mode for nominal.

Sometimes separate fusions are carried out for groups for which 'mistakes' in the predictions are unacceptable, e.g. predicting 'pregnant last year' for men. In this case the gender variable will become a so-called *cell variable*; the match between recipient and donor must be 100% on the cell variable, otherwise they won't be matched at all.

11.4.6 Conclusion

In this section record linkage has been discussed and its threats to privacy. New information can be created by person-to-person linkage or person-to-context linkage. Privacy issues that can arise are lack of consent given for linkage and data sharing, increased sensitivity of linked data and increased security risk by transporting data. These threats are enhanced by new techniques for coupling databases like associative data fusion that is described at the end of the section.

11.5 Data mining privacy threats

In the previous sections already a number of issues have surfaced that may have privacy implications and can indeed, under certain circumstances, result in a privacy threat. This section investigates this topic more fully and lists the main treats that emanate from data mining and especially the web mining variant.

11.5.1 The three main classes of privacy threats related to data mining

Data mining and the preservation of privacy of individuals can conflict in three different areas:

1. *Autonomous increase of personal data (privacy related information)*
   In order to facilitate data mining, more information than ever is needed. Therefore more information than ever is asked from people in situations where they are likely to disclose it. This information disclosure is transparent to the person concerned.

2. *The identification of previously unknown facts, rules and relationships*
   The main purpose of data mining is the finding of new, previously unknown relationships between the analysed data. The most privacy sensitive possible revelation is that of the personal identity (or 'almost' identity). This information disclosure is not transparent to the people concerned, and therefore important to discuss.

3. *The creation of more meaningful information*
   Another consequence of data mining is the striving towards data enrichment through record linkage or associative fusion. This makes the available information more meaningful, because the data is put in context. This is different from the previous type of privacy threat, where the result is more factual. The expansion of knowledge

about people who's identity is already known, is in itself a risk to privacy. This is not just about the data subject at the centre of investigation. The linking of a second person with the data subject can be a privacy threat to the second person.

11.5.2 Data mining, the Internet and privacy

Main issues

According to Slane [Sla98], privacy commissioner in New Zealand, there are four main privacy issues in data mining with the Internet (web mining):

1. **Security** Before the Internet, access to databases was reasonably limited to a few authorised people. However, the Internet makes it easier for more people to access them. Thus it is necessary to consider what access controls are needed so that only authorised people have access to appropriate portions of the database. Without strong access control, private information can be disclosed, illegally manipulated, and misused.

2. **Accuracy** With the growth of the Internet, data mining came to involve huge amounts of data from a variety of sources. The more databases involved, the greater the risk that the data is old or inaccurate and the more difficult it is to cleanse the data. Inaccurate data, which are not updated, unreliable, and incomplete, may lead to errors and misinterpretation.

3. **Transparency** At present, it is most unlikely that many people would understand what is involved with data mining. This means that they cannot correct the data about themselves if they are wrong and they cannot express any concern with the use or disclose of their private information. Most Web sites deny people's autonomy such as the right to opt-out or correct their data. In other words, people are not likely to have control over the use of their data once collected.

4. **Fairness** When data mining, no one can necessarily predict what kinds of relationships or patterns of data will emerge. However, most Web sites do not provide their customers a chance to choose to be included in any data mining process. It is questionable that data subjects are being treated fairly when they are unaware that personal data about them is being mined.

Conflicts between Business and Individual

When it comes to privacy concerns, Novak et al. [NHP97] develop a good model illustrating conflicts of interests between commercial Web providers and consumers. According to them, during the purchasing process, marketers have established that consumers pass through the following stages: 1) search; 2) purchase (negotiation, decision, payment); 3) fulfilment and delivery; and 4) post-purchase relationship. They found that in each of these stages, consumers and commercial Web providers are confronted with conflictive interests in terms of privacy.

As Table 11.1 indicates, the consumer's position is to preserve information privacy. On the other hand, the commercial Web provider's position is based on data collection for the direct marketing practices. From this, general conflicts of privacy concerns between businesses and individuals can be induced.

Table 11.1: Seller/Consumer Conflict of Interest at Stages of the Purchasing Process.

	Seller Interests	**Consumer Interests**
Search	Data mining	Anonymity
Negotiation	Buyer's identity	Seller's identity for authentication Anonymity
Decision	Buyer's certification Buyer's authentication	Anonymity Secondary use of information control
Payment	Confirmation of any third party (such as a credit card company)	Anonymity Integrity (protection against unauthorised payments)
Fulfilment/Delivery	Non-repudiation	Resource
Post-purchasing relationship	Development of an exchange relationship	Reliable relationship

11.5.3 Web mining privacy threats

In discussion of privacy sensitivity of data, the focus is on data explicitly and knowingly transferred by individuals to some organisation or company, e.g. by filling in questionnaires. More and more however, data can be found which have not been left intentionally, but which are traces left from activities. For example, the click patterns that can be found in log files. Also, historical data may remain accessible much longer than anticipated. Much of this information is not directly 'on the web' (i.e. contained in pages indexed by search engines) but is available through the web, by querying various databases. Below is a discussion of traces that are left by people on the internet, potentially posing a privacy threat, together with possible measures against the threats. The threats are described as consequences of a trace. If possible, a remedial measure is described.

Website visits

Trace: For most websites, every visit is logged by storing the IP address and the website address the user was linked to the current website. This information is stored together with the visit date.

Threat 1: Even though this log is invisible to the outside, the website owner can access it and misuse it. The IP address allows linking the user with other traces that include the IP address, including other visits to the same website. Furthermore, the IP address can be used to harass the user's computer by doing a denial of service attack or break into the computer (scanning ports, misuse file sharing if switched on by mistake). Furthermore, most providers are willing to reveal someone's true identity to law-enforcement organisations depending on the seriousness of the accusation.

Measures 1: The IP address problem can be dealt with from both sides: the user can decide to start using so-called 'IP spoofing': the IP address is made anonymous by letting all web traffic go through a Proxy server located on another computer on the internet. The website itself could adopt the policy to disable logging, but the disadvantage is that there would be no more measurements of the locations visitors are coming from, which is usually derived from the IP addresses. This can be solved by using an external trusted party, such as Nedstat, to do the logging and to provide the visit statistics to the website. Usually this is done by referring to an image file on the trusted party's server which leaves a visit trace there. This illustrates another measure the website can take: do not use link directly to images or other files on other websites. When such files are retrieved from the other sites, the user leaves an IP trace there. Therefore, it is good practice to make copies of the file on the local server.

Threat 2: The URL of the website the user came from may tell something about the person which the person would rather keep classified: for example a pornography website or the URL of a search website including the search criteria. If the previous website was protected using some login, the URL may include secret login information

Measures 2: The website can take the same measures as mentioned above to stop logging the previous URL, i.e. use an external logging service. The problem can also be addressed on the 'previous website' by not including the login information. At least the login information should be temporary, for example a temporary sessionID referring to a session that times out in a few minutes, making it impossible to abuse the URL long after it was logged. The best way to prevent misuse is to use a redirect page for every link to other websites: a simple page taking the new URL as an argument, without any session or login related data that puts the user through to the other website. This method only leaves a trace of the redirect page which cannot be misused.

Threat 3: The visit date is a threat to privacy if the previous URL and/or the IP address can be used to identify a visitor. That way, the trace is 'proof' that someone visited the website at that day and time, which can be incriminating if the person should have been doing something else at that time.

Measures 3: In order to prevent misuse of the visit date the URL and IP threats should be addressed as described above.

Threat 4: Internet Explorer from version 5 (covering 65% of the web users) requests a file FAVICON.ICO from the server if the user bookmarks the site. This is interesting information for the website owner which will probably lead to extra, possibly unwanted, attention to that user.

Measures 4: the website can adopt the policy to not log requests for FAVICON.ICO. The user can filter requests for the file using a local proxy program.

Usenet visits

Trace: Every time a user makes a posting on a Usenet newsgroup he/she leaves posting date/time, a (nick) name and possible an email address. This data is visible on the newsgroup servers, but also on web services that allow users to search Usenet data, such as www.dejanews.com (recently taken over by Google), which stores the discussion much longer than most newsgroup servers do.

Threat 1: The nickname can be used to link to other occurrences of that nickname and that way possibly to the same identity, or maybe even worse: to the wrong identity. For example: user A states his opinion in a news posting, and user B starts looking for other postings, also in other newsgroups and finds a quote that contradicts the opinion and confronts user A with it.

Measures 1: The strongest measure to prevent linking of nicknames would be to use a different nickname in every posting, but that would cause too many problems. Nicknames are necessary to have a discussion: you need to know who replies to whom. Instead, the user could decide to use a different nickname for every discussion thread. Mostly, people will not do this because it prevents them in creating a ‘presence’ on the internet and make themselves known. The third degree of nickname measures is to use a different nickname for different discussion groups. This would be more acceptable but still the social function of the nickname is a reason many users use the same nickname for different areas on the internet.

Threat 2: Just like the nickname the email address can be used to link to identities and it also allows people to contact the user on that address, and possibly harass by sending spam emails. Many internet advertising organisations use newsgroup postings to create an email advertising database.

Measures 2: Leaving out the email address makes it impossible for others to contact the user and continue the discussion in private. Email is also a way to inform users in case a discussion thread hasn’t changed for a while and the users are no longer monitoring the thread. A common measure is to use a non-existing email address and make clear in the address how it should be modified to become the right address. Example: johnREMOVEME@hotmail.com. This method is common knowledge to Usenet visitors and unfortunately also to the advertising organisations, resulting in increasingly complex ways of encoding the email address.

Threat 3: The posting date is a privacy threat as discussed in the website visit section above. In order to prevent linking the posting date to an identity, the measurements against the email and nickname threats should be taken.

Web discussion contributions

Trace: Web discussions are similar to Usenet postings and generally store more information than Usenet does. Often, the IP address of the sender is shown (mostly resolved to a more readable DNS name), and sometimes there is a link to the user’s profile in which the user has entered some info about his/herself.

Threat 1: The IP address poses the same threats as discussed in the web visit section. In this case the address is not only visible to the website owner, but also to all visitors of the website, increasing the threat.

Measures 1: The user can take the same measure as discussed in the web visit section (IP spoofing). The website could stop displaying the IP addresses, but the reason the addresses are shown is to check the identity of a person by linking the IP with other postings, in order to prevent users from impersonating others or to prevent them posting insulting messages anonymously. A good alternative is to encrypt the IP address using a proprietary encryption. This will allow users to identify a user only to the identity he or she has on that particular website. The disadvantage of this approach is that it is no longer possible to see what kind of provider the user has (often of interest in online-gaming communities) or from what country he or she is. If this information has an acceptable privacy level for the users of a website, the IP address can be displayed as the last two entries in the DNS name (mostly provider and country), preceded by a proprietary encryption of the rest. This allows for uniquely identifying the user locally as well as seeing what provider and country the IP is from.

IRC visits Internet Relay

Chat servers log nicknames and IP addresses and therefore the nickname and IP threats mentioned above apply. IRC servers normally display unencrypted IP addresses. Even worse: they keep track of all the nicknames an IP user has used in the past. The simple command WHO <IPADDRESS> returns all the Nicknames used together with that IP address. Hence, IRC servers do not apply any of the discussed measures against privacy threats.

Exposure through file-sharing

Trace: In the recent year, massive file-sharing has become very popular, with services like Napster (sharing MP3 music files) as initiators. File-sharing is basically a peer-to-peer virtual database of all kinds of files (video clips, programs, audio files) that are stored on the PC's of users. The nickname of a user and the names of the shared files on his/her computer are public information. Furthermore, the user's IP address can easily be tracked because file transfers (up- or download) are done directly to the user's PC.

Threat 1: The nickname can be linked to other users of that nickname on the internet and as a result link the file list to those identities. The collection of files can be incriminating as it may contain illegal material (e.g. MP3's or cracked software) or files revealing behaviour the user would rather keep a secret (e.g. pornographic material).

Measures 1: The user can take the same measurements as discussed in the nickname part of the Usenet visit section. Furthermore, the file sharing service itself could protect the user by suggesting the use of a unique nickname, or better force the user to have a unique nickname by generating one (e.g. a random number).

Threat 2: By looking at the filenames or file content of a user, one could make a link to the user's identity. If the researching person has knowledge of certain unique files sent to certain other identities (e.g. on a web community), these identities can be found on the file sharing service by looking for those files. Files could also contain references to an identity (e.g. an email or letter to the user accidentally added to the file sharing database).

Measures 2: One measure to protect against conclusions drawn from the file collection is to not expose the file collection of users, or at least make that an option for the users of the file sharing service. Furthermore, the sharing of certain files (e.g. emails, documents) could be prohibited or warned for as a possible privacy threat.

Threat 3: Tracking the IP numbers of file-sharing users (as has been done for Napster users in some countries) creates the IP threats as discussed in the web visit section. This is especially a threat for file-sharing users, as most files haring is copyright infringement and therefore liable to punishment.

Domain ownership

Trace: In order to gain ownership of an internet domain, the user must provide his identity or the identity of a representing organisation to the registration intermediary. This information, referred to as the WHOIS data, is stored on DNS servers and is publicly available using complicated DNS front ends or easy-to-use websites.

Often the DNS WHOIS data refers to a registration intermediary organisation and the real identity data has to be retrieved using a service of that organisation (e.g. http://www.networksolutions.com/)

The WHOIS database stores registrant name, address, creation data, expiration date, last change date, plus an administrative and technical contact with address and telephone number. The WHOIS search facility is used for various purposes, including:

- to identify the availability of a particular name in which someone is interested
- to determine if there are similar names already in use, and/or to contact domain owners to discuss transfer of the domain
- to identify and verify online merchants
- to identify online intruders for enforcement of intellectual property rights
- to source unsolicited email and sites that may be inappropriate for children
- to identify contacts in the investigation of illegal activity, including consumer fraud, child pornography, etc.

Threat: The privacy threat caused by WHOIS is obvious: the domain registrant can be harassed by direct contact, and links with other identities can be made.

Measure: The least thing a domain registrant can do is to register under the name of an organisation which he or she (partly) owns, with a business address and not a private address. The least thing a registration organisation can do is to clearly warn the user the data is made public. In the Netherlands, some individuals have scanned celebrity homepages for real celebrity addresses and published those addresses and telephone numbers on the internet. WHOIS search facilities could limit search capability by controlling the searches to for example a limited number per day for every user's IP address. However, there is too much competition to let search facilities survive after implementing such a search control. Even worse, owners of WHOIS databases, such as Verisign are thinking about selling their databases on CDROM. It will take major changes in international law and internet organisation to prohibit such privacy threats.

ISP traces

Trace: In order to connect to the internet, people use Internet Service Providers. Since all traffic (email, web, IRC, etc) is going through the ISP, theoretically it can be logged. Some of this traffic is encrypted and will normally remain secret - typically credit card transactions or login sessions over SSL.

Threat: Because everything can be logged, all the threats discussed above apply here. It is particularly easy to link the user's behaviour to his/her true identity since that is known to the ISP in order to communicate with and bill the user. An interesting recent discussion is around the subject of forcing ISP's to log traffic to facilitate law enforcement. The best known example of such a logging system is Carnivore, developed by the FBI, in which two levels of logging are used:

1. Full content logging, only if authorised by a federal district court judge:
    - capture all e-mail messages to and from a specific account
    - capture all the network traffic to and from a specific user or IP address

2. Partly logging, only for criminal suspects:
   - capture all the e-mail headers (including e-mail addresses) going to and from an e-mail account, but not the actual contents (or Subject: line)
   - list all the servers (web servers, FTP servers) that the suspect accesses, but do not capture the content of this communication
   - track everyone who accesses a specific web page or FTP file
   - track all web pages or FTP files that a suspect accesses

The FBI claims that Carnivore has been used roughly 25 times leading up to August, 2000. They used Carnivore only 10% of the time for such court orders: most of the time the ISP complies with the court order using their own facilities. The FBI claims that the majority of cases have been for counter terrorism, though they also mention hacking and drug trafficking (see the Carnivore FAQ).

The privacy threats of logging systems such as Carnivore are that it can be misused (for example track traffic of non-suspects by a Carnivore operator or someone who succeeds in breaking into the system), and bugs can lead to privacy violations (e.g. identifying an innocent user as a criminal).

Measurements: It is essential for ISP to have a policy that prohibits the logging of traffic. Users and user organisations demand such a basic level of privacy protection. Furthermore, it would require a large investment in hard- and software for ISP's to implement the logging for all users. An exception is logging for law enforcement where special care should be taken to avoid misuse (high security) or allow bugs to violate privacy (extensive testing plus control mechanisms).

11.5.4 Case of Privacy Threat: The FBI Wiretap discussion

Police work concerns at least in part the investigation of criminal activities for instance for drug related crime and, after the fact, to identify the culprit(s) of murderers. In the US this is allocated to the Federal Board of Investigation (FBI), which organisation covers the most important crimes. To unravel the network of criminal activities the Federal Communications Commission (FCC) granted the FBI the Communications Assistance for Law Enforcement Act (CALEA) during the year 1994, less officially called the wiretap law. This law was requested by the FBI based on the argument that it was losing ground against criminals because many of the emerging wireless telephone companies were not designing wiretapping capabilities into their networks and that continuation of existing wiretap capabilities was necessary to make police investigations possible.

However, the FBI interpreted the law to such an extend that telecommunications companies and civil liberties groups accused the FBI of overstepping the powers intended by the Congress. Naturally, the telecommunications companies fulminated against the law because of the technological burdens imposed on them, without real commercial advantage at their side. But the arguments of the civil liberties groups, among which the Centre for Democracy and Technology, is of more interest here. They were against the stretched interpretation of the law because of the possibility of credit-card information, pharmaceutical prescriptions and other sensitive personal data to be exposed in a wiretap.

So the case was taken to court and the most appropriate was the U.S. Court of Appeals for the District of Columbia in Washington. The court decided in its verdict of August the 15th 2000 that the FBI had requested too much. The arguments went twofold: one line was about the financial consequences for the telecom companies, the other line was about the privacy aspects. The court wrote: ”The statute requires the Commission to consider more than the burden on law enforcement. After all, any privacy protections burden law enforcement to some extent.” The argument by the FBI that it had considered the privacy aspects with ”a little bit of had wringing and worrying” was rejected out of hand by the Court of Appeals.

The verdict seems to have implications to the Internet wiretap system Carnivore, created by the FBI for use at Internet service providers. The expanded interpretation of the CALEA plus Carnivore would have created an enormous amount of data to get information from by means of data mining techniques. And because the FBI is interested in suspects and hence people in the first place, this would certainly have included personal data. There can also hardly be put a ring around an investigation to restrict it, because everyone in contact with suspects can, at least in theory, also be a suspect. Needless to say that such a contact might be entirely innocent. The reaction of James Dempsey of the Centre for D&T is to the point: ”That (the court order) means government cannot get its hands on what it’s not authorised to get just by promising it won’t read what it’s not supposed to read.”

For full information refer to [Wasa].

(Note: Sure enough, such an enormous collection of personal data would have interested many commercial organisations like insurance companies; the question is how long the FBI employees would be able to withstand the temptations.)

11.5.5 Case: The 2001 situation concerning identity theft in the US

Identity theft is a profitable branch of criminality. In its extreme form a certain person may fully impersonate as somebody else. In a milder form another person’s bank account can be plundered or other unpleasant activities can be undertaken. To make identity theft possible, some pertinent data about a private person has to be obtained. Favourite documents include credit card statements, employment information and other documents which reveal some of the aspects of a personal life; a social security number is an interesting part of it. Such documents can relatively easily be stolen from mailboxes; this may include new bank checks. Then a stepping-stone strategy can be followed, during which the impersonator gathers more and more information about the victim and starts to act more and more in the victim’s name. This includes the use of stolen cheques to pay for more information about the victim from external sources, like information brokers and via these brokers from banks or even directly [Wasb]. Such a broker may apply legal actions, but also trickery, among which pretext calling can be used. (Pretext calling is calling people without mentioning the real intentions, but a pretext instead to induce the target to reveal some information.) Following such practices, and using seemingly innocent information, ultimately a phoney driver’s licence can be obtained. And the bank account(s) can be plundered, leaving the victim with a ruined reputation and a life in chaos. A case is described in [Wasc].

In the past this took a long time to perform. But with the advent of on-line information this has changed. Legitimate businesses are using commercial on-line data brokers, which collect and sell personal information. Their intentions may be proper, but the same tactics and associated information can also be followed by people and organisations with less proper intentions.

Some law enforcement officials and regulators say identity theft has become one of their most pressing problems. The federal Office of the Comptroller of the Currency estimated that there are half a million victims of identity theft per year in the United States. And according to the Justice Department identity theft is one of the nation’s fastest growing white-collar crimes. The misuse of social security number has reached the level of ”a national crisis" (words of the inspector general of the Social Security Administration). Needless to say that intelligent agent technology will only facilitate such misuse. The search for information, where to find it and what, will become easier and more effective. A lot is available on the internet. And data mining can be the machinery on which such data collection is based.

11.5.6 Conclusions

There are many ways in which people leave traces on the internet and even more ways to let these traces become a privacy threat. Measures against these threats can be taken by both users and internet services. The users will take these measures if they are properly informed and it is therefore the task of user guiding organisations such as ISPs and internet services to educate the user. Better informed users will result in higher standards for the privacy protection by internet services, resulting in a competition between internet services that will automatically shake out organisations and technologies that have not established decent privacy protection.

The overall conclusion is that data mining indeed causes privacy threats. Especially the linking of data, which activity may give rise to the identification of entirely new facts about a data subject or the creation of richer, more meaningful information, is a threat that is particular to data mining.

11.6 Data Mining in defence of privacy

Whilst data mining can become a threat to privacy in the ways examined in the previous section, it could potentially also be used as a defensive weapon or shield, to protect individuals from privacy violations or detect and disarm perpetuators of privacy infringements. In what ways could data mining be used to serve and protect privacy? From the three theoretical possibilities of prevention, detection and correction, clearly detection is the defence mechanism that could most likely benefit from data mining. Preventive measures are oriented towards making infringements impossible by limiting access to privacy sensitive data, e.g. by encryption, or by withholding, i.e. not making personal data accessible.

Corrective measures are difficult, because private data, once being known, cannot be easily retracted or erased; people cannot forget by instruction or intent information they should not have had access to. As some of the case examples in Section 11.5 illustrate, it also becomes increasingly difficult for organisations to 'forget on demand' i.e. to ascertain that no copies of deleted data remain anywhere in the organisation. Perhaps a correction could consist in compensating for privacy infringements, but the problem to be solved then -what price to pay- is a normative problem more then a data analysis problem.

11.6.1 Detection of privacy infringements

Privacy infringement can loosely be defined as the gathering of personal data to affect the privacy of the data subject. On this issue relatively few instances are reported in the literature, but the real problem is much greater, refer to the cases reported earlier in the document. The low number of reports can be explained. Each privacy infringement usually affects an individual person and the real damage to that individual is, again usually, not enormous. What reaches the press and public attention are the really formidable cases; the cases that can be considered at the level of nuisance will not be reported, but these are high in number. This situation is comparable to the high number of victims, both lethality's and injuries, in road traffic. Unless it concerns a really big accident with a major number of people being injured or killed, it will not hit the front pages. This in contrast to the comparably low number of victims in air traffic accidents, which almost always are reported in detail. In a certain sense, the high incidence rate, especially in the sector of road traffic, makes the public immune to such accounts. Yet, in their totality, the numbers are staggering, in traffic and privacy infringements alike. Moreover, the privacy infringement problem is rising and agent technology will be one of its harbingers. The consequences have to be investigated.

The main problem is: in order to apply data mining methods you need representative data collected over many instances. There is a close resemblance with fraud analysis in financial sectors, namely: instances of fraud tend to be 1) rare, 2) highly varied in form, and 3) go unnoticed to a large extend. The reasons are clear: the fraudulent person wants to stay undetected and will apply every possible effort and method to keep it that way. Hence many an organisation will not be able to answer the first request of a data miner: 'give me a set of representative examples as training data'.

Intrusion detection in server systems

Some research has been done in the area of intrusion detection in server systems (e.g. [LS01]). Servers are frequent targets of hackers who try to infiltrate server systems using all their creativity and the most obscure backdoors available. It is very desirable to be able to detect patterns of intrusion automatically but intruders, of course, try to leave as few marks as possible. Nonetheless, such impostors leave some traces, which can be detected by the application of special means, though the number of traces can be low indeed. Alternatively, artificially generated traces might be used for training (cf. [LS01]).

A second best scenario is to collect samples of all 'normal' scenario's and regard all exceptions as intrusions. This can be compared to virus scanners that scan in two modes: looking for patterns of known viruses or detect patterns of possible new viruses (heuristic scanning). A key assumption is that the training data is nearly 'complete' with regard to all possible 'normal' behaviour of a program or user. Otherwise, the learned detection model can not confidently classify or label an unmatched data as 'abnormal' since it can just be an unseen 'normal' data.

Data mining for privacy protection is likewise only possible if there is a data collection of large numbers of actual infringements of sufficiently standard format to allow for data mining algorithms to work on. However, the collector of personal data has the obligation to protect the personal data. One of the things to be done is to log all data requests performed on the database with personal data. Requests done by unauthorised people will be rejected out of hand, but the request as such, especially when repeated, can be interesting. Furthermore, to log the activities of authorised people and systems is certainly interesting from a privacy protection point of view, both to work as a watchdog and, later on, to identify potential perpetrators. A good logging mechanism with tools to analyse the data may also function as a deterrent against people with malicious intentions. In addition, it is possible to force users of the database first to declare that they want to use personal data of a certain person, for what reason and for how long.

Lee and Stolfo propose a system architecture that includes two kinds of intelligent agents: the learning agents and the detection agents. A learning agent, which may reside in a server machine for its computing power, is responsible for computing and maintaining the rule sets for programs and users. A detection agent is generic and extensible. It is equipped with a (learned and periodically updated) rule set (i.e., a classifier) from the remote learning agent. Its detection engine executes the classifier on the input audit data, and outputs evidence of intrusions.

The biggest challenge of using data mining approaches in intrusion detection is that it requires a large amount of audit data in order to compute the profile rule sets. Moreover, this learning (mining) process is an integral and continuous part of an intrusion detection system because the rule sets used by the detection module may not be static over a long period of time. Given that data mining is an expensive process (in time and storage), and real-time agent-based detection needs to be lightweight to be practical, we can't afford to have a monolithic intrusion detection system.

The main advantages of such a system architecture are:

- It is easy to construct an intrusion detection system as a compositional hierarchy of generic detection agents.
- The detection agents are lightweight since they can function independently from the heavyweight learning agents, in time and place, so long as it is already equipped with the rule sets.
- A detection agent can report new instances of intrusions by transmitting the audit records to the learning agent, which can in turn compute an updated classifier to detect such intrusions, and dispatch them to all detection agents.

Concluding, data mining for privacy protection can profit from research in intrusion detection in two important areas: designing ways to train the models used and implementing a system that can be used in an agent-based environment.

11.6.2 Detection of potential privacy infringements through web mining

It may be very difficult to detect specific instances of misuse of personal data. However it may be possible to evaluate to what extent a given site interacts with users in such a way as to provide potential risks to privacy concerns. The behaviour of web sites is transparent. When one accesses a site, the HTML code generated from this site is loaded into one's browser and can be inspected, e.g. by using the 'View source code' option. Through inspection of this code many privacy sensitive site behaviours -such as use of cookies or other 'pins' or identifiers, redirection to other sites possibly leading to data linkage- can be established directly. An example of this approach can be found in a recent study on privacy risks in the use of various search engines, the category of sites most intensely visited by business and consumer web users [Roe01]. This study is described briefly in the next paragraph.

Search engines privacy risks

Search engines may be one of the most widely used and most important services the World Wide Web offers. Concerning privacy search engines could be described as a 'data bottleneck'. Companies usually need to analyse lots of access statistics to create a profile for a single user. Yet when using a search engine the user herself submits condensed and accurate information on her topics of interest. This makes search engines an interesting target for profiling and advertising companies.

Many users do not understand why leakage of information not containing any personal data such as email addresses or real names is considered a threat by privacy concerned users. However the tiny bits of information which get dropped every now and then form a trail which can be used to trace users, i.e. the user is not anonymous any more but pseudonymous. Being pseudonymous means that it is not known who the user really is, but it is known that it is her. The problem is that giving personal data such as an email address or a name only one time will make all past and future pseudonymous actions personal actions which can be linked to the user's real identity.

Also in the computer security sector it is common practice to disclose as little information as possible as every tiny bit of information might provide valuable information to an attacker trying to break into the system. There exist attacks exploiting bugs in web browsers yielding access to the user's system, and knowledge of the exact browser and OS version will simplify the task of the attacker.

Thirteen search engines were analysed, including most of the market leaders such as Altavista, Google, Excite, Lycos, Hotbot, Yahoo, as well as a few less well known engines. Instead of discussing every single search engine examined, it seems more relevant to describe the problems found while testing the search engines. They can be classified as shown in Table 11.2.

Conclusion: practices causing privacy risks are very common in current leading search engines. The most common problem is caused by various forms of redirection, which lead to information provided by the user of a search engine -a query, sometimes cookies or IP address- being shared with other companies or organisations without users being aware that this is happening. Particularly Doubleclick (www.doubleclick.com) receives in this way much information concerning users of search engines. Of course, the privacy risks involved are typically small, because the information concerning specific users provided in this way is very limited, and in most cases it will be very hard to link this information to particular individuals. But the principle risk involved is that even small fragments of factual behaviour patterns could be linked to specific individuals, by agencies having access to large and connected volumes of behaviour fragments, such as e.g. Echelon. In the next Chapter we will discuss how this risk can be eliminated, first we look at ways this analysis of privacy risks in web sites could be generalised and scaled up.

11.6.3 Web Mining with ParaBots

One obvious drawback of the approach illustrated with the analysis of search engine sites is that it requires rather intensive and detailed analysis of code and scripts that could become very laborious indeed when a larger range of e-commerce sites would be investigated. To be able to extend this type of analysis to categories of e-commerce that occur in much larger numbers on the web (e.g. job or labour market sites, financial services, dating, etc.) large parts of the analysis will need to be automated. This can be done by applying intelligent agent as well as data mining technologies to the task, as is the case for the ParaBots (PAge RAting Bots) system currently under development at Sentient Machine Research.

ParaBots is designed to perform three basic tasks:

1. Autonomous or interactive collection of potentially relevant sites from the Internet. ParaBots can use two approaches to this task: a) Crawling or spidering: given a startset of url's, ParaBots retrieves all the contents from the pages these url's refer to, finds all links -new url's- on these pages and then may retrieve these new pages and search them for new links. b) Meta-search: a topic for search can be specified by a -possibly long- list of keywords judged relevant. From this list of keywords ParaBots generates a list of queries that can be send to various search engines. The

Table 11.2: Problems found while testing the search engines.

Problem:	**IP leakage**
Impact:	If the user has a static IP address she can be identified and traced across several sessions; this is the case with all webpages although external pages (advertisers) have no business knowing the user's IP address.
Solution:	Use proxys or anonymisers
Problem:	**Cookies**
Impact:	Tracing the user across several pages and over several sessions is possible
Solution:	Disable cookies or use filtering proxy
Problem:	**The HTTP header contains plenty of information** such as language, OS and browser version
Impact:	Those informations will be leaked to the servers
Solution:	Use filtering proxy
Problem:	**The query string is part of the URL of the search results page** the query string will be sent to other servers using the referer. This has been a problem with all tested search engines. Using redirected links may (or may not) be an attempt of the search engines to prevent this, though redirects are another problem by themselves (see below).
Impact:	External sites jumped to from the search engine page will know the query string used for finding them
Solution:	Use filtering proxy
Problem:	**HTML-embedded cookies / user pinning**
Impact:	Tracing the user across several pages (but not over several distinct sessions) is possible
Solution:	None yet. Maybe content-filtering? This will become tricky and will probably not be failsafe. Another possibility is to not allow hidden fields in html-forms, thus all data to be sent can be seen by the user. Of course this will cause a lot of sites not to work any more.
Problem:	**Leakage of query strings** to other sites by passing them as parameters to external servers
Impact:	External sites get to know the query string.
Solution:	No definite solution. It seems the sites which the query strings are leaked to are only few, prominent among them doubleclick.net. By blocking those using a filtering proxy information leakage can be prevented.
Problem:	**JavaScript, possibly loaded from other sites** JavaScript is far too powerful and will enable the server to obtain information such as the local IP address (even if using a proxy!), local configuration information etc..
Impact:	Tracing of users, leakage of information such as OS, browser version, screen resolution, used plugins...
Solution:	Disable JavaScript.
Problem:	**Redirected links**
Impact:	The server will know which of the links presented the user chooses to follow.
Solution:	Do not use servers using redirected links or use a local redirector script which will strip the redirection before passing the URL to the webserver. Alternatively content-rewriting may be used.
Problem:	**Sharing identifiers by using 302-redirects**
Impact:	This can be used to share cookies or other user identifiers in a way very difficult to detect for average users.
Solution:	Difficult. In the case presented in this chapter disabling cookies will suffice, but this concept may be extended to do other mischief not thought of yet.
Problem:	**The x/y field is treated differently by some browsers**
Impact:	Lynx users can be detected quite reliably
Solution:	Patch lynx

results from these queries are integrated in one long list of potentially relevant sites. Of course, these two collection methods can be combined in many ways.

2. Autonomous analysis -or ‘rating’- of the content of pages or sites. In many cases this involves the integration of many pages into sites and classification of the sites based on relevant features. Analysis can be based on different types of content, first the HTML code, next text or even images. The first ParaBots system was designed to rate sites for the presence of pornographic image contents. The ParaBots training environment contains facilities for teaching the system new classifications for web pages simply by giving positive and negative examples for the new classification.

3. Gathering and examination of collections of classified sites (surveys) in a data mining environment. i.e. Sentients DataDetective visual data mining software, for instance to trace trends in successive surveys or to compare the frequencies of specific features in various surveys.

The tool can be used for various types of web mining applications. For instance:

1. Mining on HTML code to detect privacy sensitive interactions on websites. In this case, the HTML code that builds up a website is investigated for e.g. code that places cookies, reads IP addresses, leaks data to linked websites, etc.

2. Text mining on privacy policies that are provided by many sites. The results can be compared with policies as defined by the P3P standard. P3P defines 4 basic data structures, each containing a number of fields, that are suitable to classify upon.

In the next phase of the PISA project, the ParaBots web mining tool will be used to analyse the privacy policies and risks off labour market e-commerce site in a few European countries, to begin with The Netherlands and Italy.

System Architecture

The functional architecture of the ParaBel system, a ParaBots system configured by Sentient Machine Research for an undisclosed client, consists of 6 components.

1. URL DataManagement. This is the web interface where lists of URL’s are managed. The URL’s can be downloaded from the web or collected from internally managed databases.

2. Page & SiteParsing. This component parses the pages in their compound parts (HTML, Text, Banners, Images, etc) and stipulates the relations between the pages that form a site.

3. FeatureAbstractionModules. Here the various classifications are executed with respect to the compound parts of the pages. The classification can be based upon decision rules or upon trained ANN modules.

4. Sample & IndexIntegration. The automatically extracted site information is saved in files that can be combined or segmented to create surveys.

5. Sample Annotation. Subsets of site collections can be separated for specific examination. These subsets can be manually enriched, or through linkage to other databases containing information on sites or the companies that operate them.

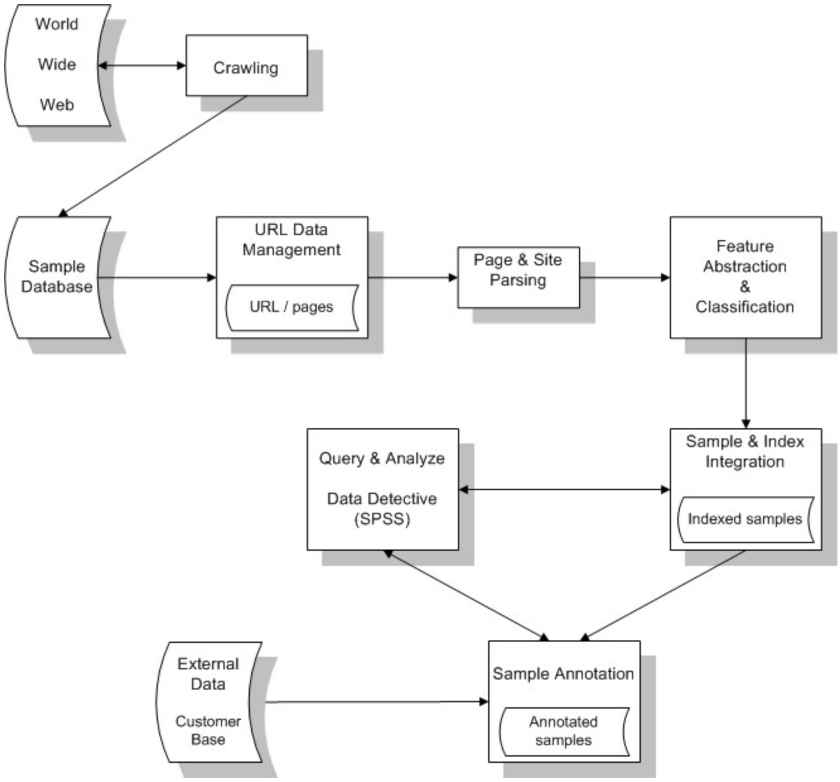


Figure 11.3: Parabot system architecture.

6. Query & Analyze. A user interface to the survey files. The Sentient Machine Research tool DataDetective searches the survey files and supports the identification of segments and profiles and the discovery of patterns and trends.

11.6.4 Collaborative monitoring of privacy policies and practices

The computational and bandwidth requirements of web analysis applications like ParaBots make it clear that it is outside the scope of an individual surfer to attempt to evaluate in this way the risks posed by various sites. The continuous monitoring of many sites on the web can only be managed by organising a collaborative effort, in which individual surfers share their experiences and resources to detect and signal sites that pose privacy threats.

11.6.5 Conclusion

Data mining for privacy protection can profit from research in intrusion detection in two important areas: designing ways to train the models used and implementing a system that can be used in an agent-based environment. A web mining system Parabots for analysing and classifying web sites is described.

11.7 Privacy Incorporated Data Mining

The increasing popularity of recommending systems (e.g. Amazon's 'buyers of this book also bought...) brings out a basic dilemma: for a recommending system to be optimally effective in its advice, it should have as much information as possible concerning the history of each individual (=data subject). To elaborate on the example of a public library, with the intention to inform clients about potentially interesting books, the library should have a profile of that reader to match it with the profiles of the readers that rated the book positively.

On the other hand, privacy concerns dictate that, in the case of a public library, only personal information required for the purpose of borrowing books - and getting them returned - should be stored. This limits the elbowroom for the library management to store profiles of subscribers, let alone to link those with books. Hence, as one library manager formulated it: 'If someday the police comes by to ask who borrowed that book on bomb making last year, I should be able to state truthfully that that information is nowhere registered in the library.'

The situation is akin to classical social dilemma's whose discussion dates back to Aristotle. E.g. in Garrett Hardin 'The tragedy of the commons' (1968) the example is that of a limited amount of common grounds, free for all to graze their cattle on, which rapidly leads to exhaustion of the common land, if all farmers optimise only their own outcomes by having as much cattle as possible graze there. Of course, the situation is mirrored in the sense, that with recommending it is a matter of bringing in personal data to fertilise the grounds for recommending, from which other people benefit.

In the case of recommending systems the dilemma is particularly clear. In such C-to-C (consumer to consumer) data mining applications the parties involved are in the same positions, both being potential contributors to and profiteers from data mining on pooled historic information. In standard B-to-C situations, it is quite possible that both sides gain from optimising the pool of accessible data. However, fairness is hard to evaluate because of the obvious asymmetries in pay off matrices between a business and a consumer.

How can the dilemma best be solved to obtain optimal knowledge from historical data, without compromising the concerns for privacy?

11.7.1 Mining anonymous data

Anonymous data can be seen as to consist of two different categories: data that is and never will be related to any person like traffic statistics, bestseller sales, weather records and the like. The other part is the part that is either made anonymous, hence in fact pertains to a certain person, or that has never been related to any person but can possibly be. The former part is of no concern here. The latter part can be privacy related, though not without supplementary activities. For instance, data made anonymous can be made identifiable again by investigation of lists of the people concerned. Originally anonymous data can be linked to a person by field investigation, as the police often do to identify a suspect. So, although data can be anonymous in every aspect at first sight, this is not an unbreakable record.

11.7.2 Unlinking identity and behaviour data

Behaviour data, such as the products a customer has bought or the web pages a user has visited, are essential for data mining tasks like recommending. However, privacy threats may prevent the storage of such data. These threats exist if and only if the behaviour data can be linked to an identity, so it is important to find ways to prevent or limit such linking. This section discusses the threats, step by step, how they can be prevented by rules, and how these rules can be enforced.

Threat 1: The main threat is that the user's identity can be linked to the behaviour storage to retrieve his/her behaviour, or vice versa.

Preventing the linking of identity with behaviour is a difficult challenge because the link itself is necessary for adding behaviour data to the storage. For example: when a customer buys some extra products, he or she has to be linked to the right behaviour storage in order to add these products to the list of total bought products.

One approach for a solution is based on the fact that creating the link is only required at clearly defined short moments, where the 'user' is present: buying a product, borrowing a book. The situation where only the user possesses the key to make the link will lead to the following procedure, referred to as the *dual-key procedure*:

1. The user is identified using one key, e.g. using a login, or member pass
2. The user exhibits some behaviour, e.g. buys a product
3. The system requests the second key from the user
4. The system finds the right behaviour record using the second key
5. The system adds the behaviour to the record
6. The second key is forgotten: not stored anywhere on the system

This way, the system's database cannot be used for practices that endanger privacy, simply because behaviour cannot be linked to an identity.

Let's look at the remaining privacy threats in this system and find possible solutions for them.

Threat 2: The user key can be stored and used later to link identity and behaviour. Therefore, the system should apply the following rule:

- *The forget-rule*: The user key is only present in the system during the user transaction

An organisation can say to do so, but users would have more trust if such a rule would be enforced automatically by the system's architecture. The rule above can be enforced by using a challenge-response authentication mechanism:

1. The storage system generates a unique sessionID
2. The user transfers his/her first key unencrypted together with the unique sessionID encrypted with the user's private second key. Added to this message is the behaviour data.
3. The storage system finds the user record using the first key
4. The storage system decrypts the sessionID using the public key stored in the records and verifies it.

5. If okay, the storage system adds the behaviour data.

This way, if the forget-rule is followed by the storage system, it is followed by the entire system, which makes it easier to control. The behaviour storage system can be a blackbox third party system, opaque to its users, or even better: a remote service, for example a trusted central organisation. The latter solution is the safest: the only communication containing identity and behaviour is a secret shared with the trusted organisation. Because the communication is based on challenge-response, it is no use to break into the system and tap/fake communication. The system is as safe as the key management.

Threat 3: The second key can be used during the transaction time to link identity and behaviour for other purposes. Therefore the system should apply the following rule:

- *The key-use rule*: The second key is only used for adding the behaviour data to the behaviour storage

This rule can be enforced by using the trusted communication link with a separate storage system, as described above.

The disadvantage of the trusted-party behaviour storage is inflexibility: the storage system may not meet the functional requirements of the client organisation. This can be solved by using an extension-architecture: the client can gain access to the behaviour data only (and not to the identity) through a database-like interface such as ODBC. There are three ways to realise this extension model:

1. the storage system is a locally installed service, accessed using the database-interface
2. the storage system is a remote service, extended with a plug-in build by the client, installed with the storage system. The plug-in service is remotely accessed, like the storage service
3. the storage system is a remote service that also allows remote access to the data. This approach is the most flexible, but may cause performance problems since all data must travel over the net

A remote storage system would typically be accessible over the internet using http based communication (e.g. SOAP). Of course, the storage system could include standard services for deploying the behaviour data, such as recommending.

Threat 4: The behaviour data and the user's identity can be stored separately from the trusted party's storage system, so the organisation can use that data to do everything they want, and the rule enforcement mentioned above has no effect.
Therefore the system should apply the following rule if the described rule enforcement is used:

- The behaviour data is only stored in the trusted party's storage system

This is a rule that is very hard to enforce, since the behaviour of a user is explicitly known to the system at the moment of the interaction. There are two situations in which it can be enforced:

1. Disconnection from the user's identity because it's not needed. In some cases, the user's identity is not essential to the organisation, so the organisation's system can be structurally disconnected from the user and the link between user and behaviour

storage. That way, the identity-behaviour data cannot be stored separately. Example: in a supermarket, the client shows his/her client card and the client number is only sent to the storage system, while the register also sends the list of purchases (behaviour) to the storage system. In the storage system the identity is combined with the behaviour based on the sender's address and the time.

2. Disconnection from the user's identity by boarding out identity-based services. The processes that require the user's identity, such as a purchase transaction, can be handled by trusted third parties instead (possibly the same party that takes care of the behaviour storage). That way, the organisation no longer needs to know the real identity of the user. For websites, users will receive a temporary identity to streamline the interaction, which is very common and does not pose a threat to privacy, since there is no connection to the identity.

Threat 5: Individual behaviour data can be used to create probability-links to people's identity. By matching the new behaviour of a user (e.g. a basket) to all behaviour records, a list of potential behaviour records can be retrieved for that identity. Example: the police investigates all behaviour records to look for suspicious patterns. After this, new behaviour (bought products, click streams) can be matched to the suspicious patterns in order to identify suspects. Therefore, in order to prevent this threat, the system should apply the following rule:

- The behaviour data is not used to make links to identities.

This rule can be enforced by using the third-party behaviour storage model, as discussed above. Furthermore, the suggested extension interface in the storage system would have to be restricted: no access to individual behaviour records is allowed. This restriction makes it even more interesting for behaviour storage parties to provide standard data deployment functions.

How do the discussed mechanisms map on systems in which the transactions are more conventional? Let's take a standard library as an example. In such a library users have simple passes, mostly containing a string of digits. These digits need to contain a key to identity-based data that the library needs to do business (name, address and preferably telephone number). To be able to keep track of behavioural data and respect privacy, the discussed dual key procedure can be followed: the digits need to contain a second key to the behaviour record. That way, the system is able to add new behaviour to the record of past behaviour, after which the key can be forgotten (see forget-rule). The first key is needed to store what books the user has borrowed, yet it does not threat privacy as it cannot be linked to the behaviour data. Because of the conventional limitations of a user pass the rules cannot be enforced, as it is easy to impersonate a user by stating the first and second key (see forget-rule). However, a smartcard allows enforcing the rules: it would contain the identity key and an algorithm to do the sessionID encryption as described above.

11.7.3 Virtual Panels

Given the different threats as discussed in the previous section, some degree of trust in the service or organisation receiving personal data is required from the individual providing these data. How can s/he be certain these data will not, under any circumstance, be brought as evidence against her/him? Clearly, threat 5 in the preceding section is the most difficult one to become reassured about. As long as somewhere specific and factual behavioral data patterns are stored, even if anonimised, there is a certain chance that someday someone will be able, by 'deduction and combination' to derive some harmful or embarassing information concerning a particular individual from these data. "On August 1 1984 a person borrowed three books: 'Zen and the art of motorcycle maintenance'; 'Sentex bombs for beginners' and 'Sensual Strangling part II' from the City Library. We've just found the Zen book you failed to return. Now about the other literature . . . "

In order to counter threat 5, anonimisation is not enough. Only by making certain that no specific factual behavioral data patterns remain stored anywhere in the system, deriving potentially harmful information becomes impossible. In this case, rather than storing factual behavior fragments, 'pseudo-behavioral fragments' are stored. The basic dilemma in recommending -collaboratively usefull information vs. individual privacy- may well be solved by collecting only fragmented data and reconstructing these data into histories for panels of virtual individuals. Rather than storing the transaction histories of say 100.000 clients, a library could build up virtual histories for say a few thousands of representative virtual clients. Together, this virtual panel of virtual clients, can be used, on the one hand, for optimally efficient recommending, while on the other hand, the true history of individual clients can not be reconstructed from the virtual panel database of stored information, consisting only of isolated and anonymous fragments of pseudo-behavior patterns. The main issue is that co-occurrences are not stored as they occurred, e.g. virtual panel 'members' may borrow twice as many books than real clients.

One way of 'privacy preserving data mining' is illustrated by Srikant: add random noise to the data in such a way that individual data can no longer be retrieved precisely, while aggregate information remains unaffected.

11.7.4 Conclusion

In this section five different threats are identified that can lead to the linking of identity to behavioural data. Several procedures can reduces these risks e.g. the third-party behaviour storage model and the dual-key procedure. These mechanisms are illustrated for a library setting. Lastly, the use of so-called virtual panels is discussed.

11.8 Conclusions

Data mining and intelligent agent technology are likely to have a complex, perhaps confusing, but very profound impact on our concept and personal experience of privacy. The data mining process may result in two different results: 'nuggets of knowledge' condensed in rules and facts and the extraction of a set of data with special and interesting characteristics. To extract nuggets, the intelligent agent data mining process is obviously less feasible, though not impossible. The main reason is that the amount of coding necessary to derive the rule or the fact is rather alien to the idea of the free roaming agent. To cater for this possibility the list of types of instructions has to be extended with one that catches the script to derive the rule. To collect data about a certain subject, or to select data with certain characteristics is an activity that is more in harmony with the idea of an agent.

Traditional data mining is much less a threat to privacy than often thought. The aim of data mining is to discover local theories starting with specific data and to produce general patterns and rules. While these rules can predict certain characteristics of, most often, anonymous data records they do not likely lead to new facts that can be threatening to the privacy of individuals.

Three activities related to data mining can potentially lead to treats to privacy though:

1. The biggest problem is that the possibilities of data mining induce organisations to collect and store increasingly many factual data. These data can be threatening to privacy if e.g. security is breached of if they are linked to other data stores.

2. Record linkage can lead to threats to privacy. New information can be created by person-to-person linkage or person-to-context linkage. These threats are enhanced by new techniques for coupling databases like associative data fusion.

3. Another potential danger to privacy comes from web mining. This is especially dangerous if applied to data that has not been left intentionally, but which are traces left from activities.

Data Mining can be performed though without many privacy concerns if a number of rules are followed e.g. the third-party behaviour storage model and the dual-key procedure.

Further, data mining can put to use in defence of privacy in systems like Parabots that can analyse and classify web sites with regard to their privacy policies. This knowledge can be used as a Privacy Protection Service.

Chapter 12

Human Computer Interaction

A.S. Patrick
Andrew.Patrick@nrc-cnrc.gc.ca
NRC, Canada

S. Kenny
ske@cbpweb.nl
CBP, The Netherlands

Cassandra Holmes
Carleton University, Canada

M. van Breukelen
breukelen@tpd.tno.nl
TNO-TPD, The Netherlands

Privacy principles, derived from privacy legislation, influence the Human-Computer Interaction (HCI). HCI is the study of mental processes and behaviour as they pertain to users interacting with computers (and other technical devices). HCI is often thought of as a blend of cognitive psychology and computer science, and the discipline is usually taught in one or both of these university departments. An important sub-topic in HCI is interface design, the discipline of creating easy-to-use and effective control systems. Interface design is important for privacy because the users interact with the software or service through the interface, so all the privacy features must be represented in the interface design.

Approaches to privacy protection are generally grouped into two sets. The first set consists of promise-based approaches based on assurance. The (perceived) privacy provided by these approaches is based on the credibility of the assurance, which is generally a product of the perceived corporate reputation of the data handler or auditor. As can be seen from recent US corporate events, reputation in this sense can be highly transitory. Practical applications of this promise-based approach to privacy are COBIT-based auditing approaches and Platform for Privacy Preferences (P3P) assurance services. The essence of these approaches is that identity is revealed and personal data is transferred on the reassurance that the environment is trusted.

The second set of approaches to privacy protection is based on self-determined anonymity. Here the user's identity is protected, although most approaches eventually involve at least one other party other than the user knowing the real identity. That is, in both practical and legal senses, actual anonymity is difficult or potentially impossible to achieve. Examples of this approach at the network level are mix nets, "anonymisers", and onion routing. The problem with anonymity approaches is that many advanced services, such as WWW portals, group discussions, or job-searching assistance, require personally identifiable information about the user in order to be effective. Usability, interoperability, and scalability challenges are also significant with these approaches. The result is that promise-based approaches to privacy protection are likely to dominate the privacy arena for some time to come. This means that users must rely on the assurances provided by the system operators, and any guidance regarding providing effective, believable assurance is especially important.

This chapter first, in Section 12.1, illuminates designing an interface taking the privacy legislation as a starting point. The second part of the chapter, Section 12.2, discusses the results of a test of an interface design prototype that was carried out to determine if the design concepts in the prototype were successful in constructing an agent technology that (1) users can use, (2) users can understand, and (3) users can trust with secure and private information.

12.1 From Privacy Legislation to Interface Design

There is increased awareness by the general public of their right to, and the value of, their privacy. Recent surveys indicate that Internet users are very concerned about divulging personal information online, and worried that they are being tracked as they use the Internet [Kob02]. Research has indicated that users are failing to register for WWW sites because they feel that they cannot trust the Internet with personal or financial information [Sau01]. In addition, information privacy is increasingly being associated with business issues such as reputation and brand value [KB02]. Moreover, governments within the European Union, Canada, Australia, and Switzerland have adopted privacy protection legislation that is enforced through independent governmental bodies with significant oversight powers. There has been little guidance, however, provided to system developers and operators on how to implement and comply with these privacy guidelines and rules, and how to soothe users' privacy concerns. The goal of this chapter is to document a process that begins with privacy legislation, works through derived privacy principles, examines the HCI requirements, and ends with specific interface design solutions. The approach taken is one of "engineering psychology" in which knowledge of the processes of the brain is used when doing system design [WH00].

In the sections that follow we explain how the European Privacy Directive 95/46/EC (EC, 1995)[1] has been analysed to produce a set of detailed privacy principles. The principles are then examined from a human factors point of view and a set of HCI requirements are developed. Figure 12.1 shows a schematic representation of the approach we have taken. The left side of the figure shows how the examination proceeds from the EU Privacy Directive through to a set of HCI requirements and categories, and this process is documented in this chapter. The right side of Figure 3.1 shows the a proposed "Privacy Interface Analysis" methodology that begins with a thorough understanding and modelling of the software or service and ends with specific interface solutions. Detailed instructions and examples for the Privacy Interface Analysis are presented in Chapter 5. Overall, our intent is to introduce the core concepts of privacy protection and HCI requirements, and then illustrate a Privacy Interface Analysis that other developers can follow.

To illustrate the Privacy Interface Analysis technique, we use the example application adopted by the PISA consortium. In the PISA Demonstrator, each user has a personal agent to which he can delegate tasks such as searching for a job or making an appointment with another person or a company. The personal agent in turn creates a dedicated agent for each task it is given. For example, a Job Search Agent (JSA) might communicate with

[1] European Commission (1995). Directive 95/46/EC of the European Parliament and the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data. *Official Journal of the European Communities* (1995), p. 31.

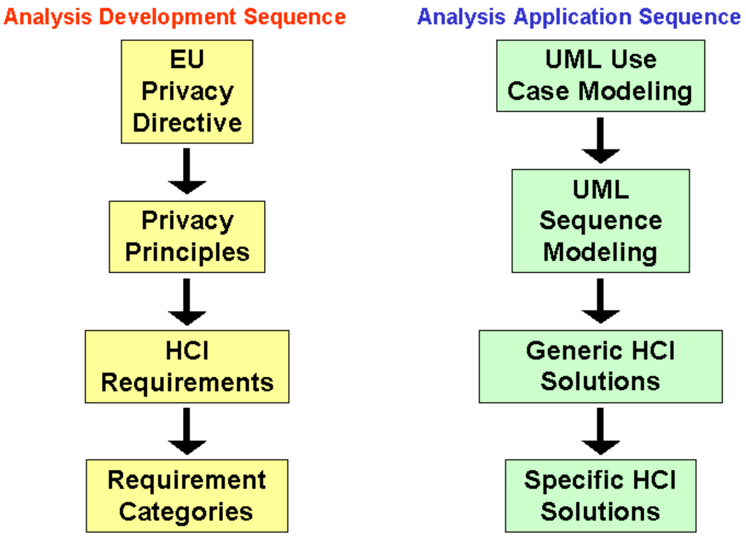


Figure 12.1: Schematic representation of our approach.

Market Advisor Agents to locate good places to look for jobs. A Job Search Agent may also interact with a Company Agent to get more information about a position. Maintaining privacy protection as the agents share information and make autonomous decisions is the challenge of the PISA project.

Privacy Guidelines and Legislation

Privacy can be protected in regulatory environments such as the EU and in self-regulatory environments such as the US. In a self-regulatory environment there is an implied behavioural model of privacy protection. That is, market mechanisms are most important and the behaviour of the various players will be determined by their motivations to succeed in a market, such as maximising profit and minimising user complaints. In the regulatory environment the model is one of compliance, although market factors may also play a role. The approach taken in this paper is to focus on the EU Privacy Directive and to facilitate “usable compliance” through privacy-enhanced interface design. This does not mean that the principles and recommendations outlined here will not be valuable in a self-regulatory environment because attention to privacy concerns and good system design will likely have rewarding results, even if they are not legally required [KB02].

Methods for Privacy Protection

Approaches to privacy protection are generally grouped into two sets. The first set consists of promise-based approaches based on assurance. The (perceived) privacy provided by these approaches is based on the credibility of the assurance, which is generally a product of the perceived corporate reputation of the data handler or auditor. As can be seen from recent US corporate events, reputation in this sense can be highly transitory. Practical applications of this promise-based approach to privacy are COBIT-based auditing approaches and Platform for Privacy Preferences (P3P) assurance services. The essence of these approaches is that identity is revealed and personal data is transferred on the reassurance that the environment is trusted.

The second set of approaches to privacy protection is based on self-determined anonymity. Here the user's identity is protected, although most approaches eventually involve at least one other party other than the user knowing the real identity. That is, in both practical and legal senses, actual anonymity is difficult or potentially impossible to achieve. Examples of this approach at the network level are mix nets, "anonymisers", and onion routing. The problem with anonymity approaches is that many advanced services, such as WWW portals, group discussions, or job-searching assistance, require personal data about the user in order to be effective. Usability, interoperability, and scalability challenges are also significant with these approaches. The result is that promise-based approaches to privacy protection are likely to dominate the privacy arena for some time to come. This means that users must rely on the assurances provided by the system operators, and any guidance regarding providing effective, believable assurance is especially important.

Related Work

Alfred Kobsa [Kob01, Kob02] has recently conducted analyses with goals similar to the current project. Kobsa is interested in personalization services, such as WWW sites that remember your name and preferences. Such personalised services are made possible because the sites collect personal information about the users, either explicitly by asking for the information, or implicitly by tracking usage patterns. Although the personalised services can be useful and valuable, the storage and use of personal information both worries some users, and falls under the auspices of privacy guidelines and legislation. Kobsa has examined the implications of the privacy laws and user concerns and developed design guidelines to help WWW site operators build privacy-sensitive systems. These guidelines include suggestions like: (1) inform users that personalization is taking place, and describe the data that is being stored and the purpose of the storage, (2) get users' consent to the personalization, and (3) protect users' data with strong security measures. The current analysis goes deeper to focus on the requirements necessary when complying with the European Privacy Directive, and includes a discussion of specific interface techniques that can be used to meet those requirements.

Privacy Principles

EU Legislation The right to privacy in the EU is defined as a human right under Article 8 of the 1950 European Convention of European Human Rights. The key privacy document is Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data, and the free movement of such data (hereafter referred to as The Directive) (European Commission, 1995)[2]. Also, Directive 97/66/EC (European Commission, 1997)[3] of the European Parliament and of the Council concerning the processing of personal data and the protection of privacy in the telecommunications sector applies and strengthens

[2] European Commission (1995). Directive 95/46/EC of the European Parliament and the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data. *Official Journal of the European Communities* (1995), p. 31.

[3] European Commission (1997). Directive 97/66/EC of the European Parliament and of the Council of 15 December 1997 concerning the processing of personal data and the protection of privacy in the telecommunications sector. *Official Journal L 024*, 30/01/1998 p. 0001-0008.

the original directive in the context of data traffic flow over public networks. These two directives represent the implementation of the human right to privacy within the EU.

The Directive places an obligation on member states to ratify national laws that implement the requirements of The Directive. This has resulted in, for instance, Wet Bescherming Persoonsgegevens 1999 in The Netherlands and The Data Protection Act 1998 in the UK. The national legislatures of EU member states must implement The Directive to substantially similar degrees. Such implementation includes sanctioning national enforcement bodies such as the Dutch Data Protection Authority with prosecutory powers.

The Directive defines a set of rights accruing to individuals concerning personal data (PD), also known as Personally Identifiable Information (PII), with some special exceptions, and lays out rules of lawful processing on the part of users of that information that are applicable irrespective of the sector of application. Specifically, The Directive specifies the data protection rights afforded to citizens or“data subjects, plus the requirements and responsibilities of “data controllers” and by association “data processors”. The Directive attempts to balance the fundamental right to privacy against the legitimate interests of data controllers and processors – a distinctive and central characteristic of the EU approach to data protection. Each of the roles is described in Table 12.1.

Since The Directive in principle prohibits the processing of EU citizens’ data in nations whose privacy laws are not as strong as those in the Union, an understanding was required between the EU and the US [Rei01]. The result is a “Safe Harbor” arrangement in which US companies voluntarily self-certify that they fulfil the privacy requirements as stated in the Safe Harbor agreement. The effect is that Safe Harbor signatories become equivalent to European processors.

Overview of the Resulting Principles As The Directive concerns itself with data processing, it must be implemented through a combination of information technology and governance initiatives. Privacy principles abstracted from the complexities of legal code have been developed to simplify this process. Table 3.2 shows a high-level summary of the privacy principles. Our research has focused on the privacy principles of (1) transparency, (2) finality and purpose limitation, (3) lawful basis, and (4) rights because these principles have the most important implications for user interface design. The remainder of this paper will be restricted to these four privacy principles.

HCI Requirements

The principles shown in Table 12.2 have HCI implications because they describe mental processes and behaviours that the Data Subject must experience in order for a service to adhere to the principles. For example, the principles require that users **understand** the transparency options, are **aware** of when they can be used, and are able to **control** how their PD is handled. These requirements are related to mental processes and human behaviour, and HCI techniques are available to satisfy these requirements. For example, an HCI specialist might examine methods for ensuring that users understand a concept, such as providing documentation, tutorials, and interface design characteristics.

Table 12.3 presents a more detailed summary of the four privacy principles under consideration in this paper. Included in Table 12.3 are the HCI requirements that have been derived from the principles. These requirements specify the mental processes and behaviour of the end user that must be supported in order to adhere to the principle. For example, the principle related to the processing of transparency leads to a requirement that users know who is processing their data, and for what purpose.

Table 12.1: Privacy Roles Defined by The Directive.

Data Subject	The Data Subject is a person who can be identified by reference to one or more factors specific to his or her physical, physiological, mental, economic, cultural or social identity. Even data associated to an individual in ambiguous ways may be deemed *reasonably* identifiable given a reasonable projection of technological development. Further, following Article 1 of the Council of Europe Convention 108/81 (Council of Europe, 1981)[4], the fundamental right to data protection applies not because of the nationality of the data subject, but as a result of a Controller or Processor operating in the EU.
Controller	The Controller is the custodian of the data subject's data and the party who determines the purpose and means of processing personal data. The Controller is defined as the holder of ultimate accountability as it relates to the correct processing of the subject's personal data. Though an organisational entity is itself legally accountable, in reality those actually responsible for assuring correct processing are those operating at the governance level, and frequently this is a company's board of directors.
Processor	The Processor is the entity that processes personal data on behalf of the Controller where the Controller determined that this is required. The Processor is accountable to the Controller, not to the Data Subject.

The HCI requirements outlined in Table 12.3 are not unrelated. The core concepts in the requirements can be grouped into four categories:

1. **comprehension**: to understand, or know;
2. **consciousness**: be aware, or informed;
3. **control**: to manipulate, or be empowered;
4. **consent**: to agree.

In the category of **comprehension**, the requirements can be summarised as building a system or service that will enable users to:

- comprehend how PD is handled;
- know who is processing PD and for what purposes;
- understand the limits of processing transparency;
- understand the limitations on objecting to processing;
- be truly informed when giving consent to processing;
- comprehend when a contract is being formed and its implications;

Table 12.2: High-Level Summary of Privacy Principles.

(bold principles are analysed in detail)

Principle	**Description**
Reporting the processing	All non-exempt processing must be reported in advance to the National Data Protection Authority.
Transparent processing	The Data Subject must be able to see who is processing his personal data and for what purpose. The Controller must keep track of all processing performed by it and the data Processors and make it available to the user.
Finality & Purpose Limitation	Personal data may only be collected for specific, explicit, legitimate purposes and not further processed in a way that is incompatible with those purposes.
Lawful basis for data processing	Personal data processing must be based on what is legally specified for the type of data involved, which varies depending on the type of personal data.
Data quality	Personal data must be as correct and as accurate as possible. The Controller must allow the citizen to examine and modify all data attributable to that person.
Rights	The Data Subject has the right to acknowledge and to improve their data as well as the right to raise certain objections.
Data traffic outside EU	Exchange of personal data to a country outside the EU is permitted only if that country offers adequate protection. If personal data is distributed outside the EU then the Controller ensures appropriate measures in that locality.
Processor processing	If data processing is outsourced from Controller to Processor, controllability must be arranged.
Security	Protection against loss and unlawful processing

- understand data protection rights and limitations.

In the category of **consciousness**, the requirements are to allow users to:

- be aware of transparency options;
- be informed when PD is processed;
- be aware of what happens to PD when retention periods expire;
- be conscious of rights to examine and modify PD;
- be aware when information may be collected automatically.

In the category of **control**, the requirements are to allow users to:

- control how PD is handled;
- be able to object to processing;
- control how long PD is stored;
- be able to exercise the rights to examine and correct PD.

Table 12.3: Privacy Principles (1.x), HCI Requirements, and Design Solutions.

	Privacy Principle	**HCI Requirement**	**Possible Solution**
1	Transparency: Transparency is where a Data Subject (DS) is empowered to comprehend the nature of processing applied to her personal data.	users must be **aware** of the transparency options, and feel empowered to **comprehend** and **control** how their Personal Data (PD) is handled	during registration, transparency information is **explained** and examples or tutorials are provided
1.1	DS informed: DS is aware of transparency opportunities	users must be **aware** of the transparency options	Opportunity to track controller's actions made **clearly visible** in the interface design
1.1.1	For: PD collected from DS. Prior to PD capture: DS informed of: controller Identity (ID) and Purpose Specification (PS)	users **know** who is controlling their data, and for what purpose(s)	at registration, user is **informed** of identity of controller, processing purpose, etc.
1.1.2	For: PD not collected from DS but from controller. DS informed by controller of: processor ID and PS. If DS is not informed of processing, one of the following must be true: DS received prior processing notification, PS is legal regulation, PS is security of the state, PS is prevention/detection/prosecution of criminal offences, PS is economic interests of the state, PS is protection of DS or rights of other natural persons, PS is scientific/statistical & PD is anonymized, or PD are subject to any other law governing their processing/storage	users are **informed** of each processor who processes their data, and the users **understand** the limits to this informing	- **user agreements** states that PD can be passed on to third parties - user agreement also contains information about usage tracking limitations - when viewing the processing logs, entries with limited information are coded to draw **attention**, and users are reminded about the tracking limitations
1.3	When PD are used for direct marketing purposes, DS receives notification of possible objection. This notification may occur every 30 days	users **understand** that they can object to processing of their PD for direct marketing, and the limitations on those objections	- during registration, users must **opt-in** to processing for direct marketing or charitable purposes - to ensure understanding and awareness, users are given examples and a **Just-In-Time Click-Through Agreement** (JITCTA) is used for final acceptance - users are also reminded of their opt-in/out option in a preferences interface screen

Table 12.4: Privacy Principles (2.x), HCI Requirements, and Design Solutions.

	Privacy Principle	**HCI Requirement**	**Possible Solution**
2	Finality & Purpose Limitation: the use and retention of PD is bound to the purpose to which it was collected from the DS.	users **control** the use and storage of their PD	interface elements for making privacy decisions are prominent and **obvious**
2.1	The controller has legitimate grounds for processing the PD (see Principle 3.1)	users give implicit or explicit **consent**	click-through agreement should obtain **unambiguous consent** for controller to process the PD
2.2	Obligations: A controller must process according to his PS, controller also ensures other processors present a PS to be considered a recipient of the PD. When assessing a processor, the controller considers PD sensitivity and the similarity of processor PS to agreed-upon PS and location of the processor. The processor can only go beyond the agreed PS if: the processor's PS is state security, or prevention/detection/prosecution of criminal offences, or economic interests of the state, or protection of DS, or rights of other natural persons, or scientific/statistical analysis	users **understand** that their PD could be used for other purposes in special cases	- **user agreements** states that PD can (must) be passed on in special cases - when viewing the processing logs, entries with limited information are coded to draw **attention**, and users are **reminded** about the special cases
2.3	Retention: the DS is to be presented a proposed retention period (RP) prior to giving consent, except where PS is scientific/ statistical. Controller ensures processor complies with RP, except where PS is scientific/statistical. When RP expires, it is preferably deleted or made anonymous. A record should be kept of processor's and controller's past adherence to RPs.	- users are **conscious** of RP prior to giving **consent** - users are **aware** of what happens to their data when the retention time expires	- When data is provided, a retention period entry field will be **highlighted** - Users are **informed** when information is deleted or made anonymous because of retention period expiry.

Table 12.5: Privacy Principles (3.x), HCI Requirements, and Design Solutions.

	Privacy Principle	**HCI Requirement**	**Possible Solution**
3	Legitimate Processing: Legitimate Processing (LP) is where the PD is processed within defined boundaries.	users **control** the boundaries in which their PD is processed	interface elements for making privacy decisions are prominent and **obvious**
3.1	Permission: To legitimately process PD, controller ensures that one or more of the following are true: the DS gives his explicit consent, the DS unambiguously requests a service requiring performance of a contract, the PS is legal obligation or public administration, the vital interests of the DS are at stake. When matching the PS agree to by the DS and the PS of the possible processor, any of the following will prevent processing: The controller/processor's actual PS differs from the PS consented to by the DS, the controller/processor intends passing the PD to a new processor, the controller/processor is not located in the EU, or the processor is violating a fundamental right to be left alone	- users give **informed consent** to all processing of data - users **understand** when they are forming a contract for services, and the implications of that contract - users **understand** the special cases when their data may be processed without a contract	- **JITCTA** to confirm unambiguous consent to data processing - **JITCTA** to confirm the formation of a contract, and the implications/limitations of the contract - in the tracking interface, include a **reminder** of special cases when data can be processed without a contract
3.2	Sensitive Data: The controller may not process any PD that is categorized as religion, philosophical beliefs, race, political opinions, health, sex life, trade union membership, or criminal convictions unless the DS has given their explicit consent or the processor is acting under a legal obligation	when dealing with highly sensitive information (religion, race, etc.), **users provide explicit, informed consent** prior to processing	if sensitive information is provided by the user, use a **double JITCTA** to obtain unambiguous consent for its processing

Table 12.6: Privacy Principles (4.x), HCI Requirements, and Design Solutions.

	Privacy Principle	**HCI Requirement**	**Possible Solution**
4	Rights: DS has the right to self-determination within the boundaries and balance of The Directive.	users **understand** and **can exercise** their rights	- at registration, use a **click-through agreement** to ensure that users know their rights - interface layout provides **obvious** tools for controlling the rights functions
4.1	Access: DS is conscious of her rights. The DS has right to retrieve this data on PD processing: (1) who has received it; (2) who gave them it; (3) when; (4) for what PS & (5) if a delete or anonymise operation has been acknowledged & authenticated. Items (1) (3) (4) should be disclosed if the proposed PS is any one of: state security, prevention/detection/prosecution of criminal offences, economic interests of the state, legal regulation, or protection of rights and freedoms (of other persons). If the DS is below the age of consent then access requests must be made by his/her legal representative (LR). In all cases, authentication should be proportional to the PD sensitivity	- users are **conscious** of their rights, which include right to know who has received their data, from whom, when, and why, and they **understand** the exceptions to these rights - users **understand** and **can exercise** their rights	- the tracking functions are displayed **prom i nently** - the exceptions to the rights are presented in the **user agreement**, and **reminders** are provided in the tracking interface
4.2	Control: DS may issue erase, block, rectify, or supplement commands on their PD. The DS is informed of the result of their command within 30 days. The communication is either: request accepted and executed, or request denied and an explanation. If the PD will not be editable due to the storage strategy applied, then DS is informed & asked to consent prior to providing any PD. Controller is accountable for the correct execution of DS requests for erase, block, rectify, or supplement the PD	- users are **conscious** of their rights, they can **exercise** control over their data, which ability to erase, block, rectify, or supplement the data - users are **informed** when data will not be editable and they provide **consent** to processing	- the tracking functions are displayed **prominently** - the exceptions to the rights are presented in the **user agreement**, and **reminders** are provided in the tracking interface - the **commands** to erase, block, rectify, and supplement are associated with the tracking logs and **obvious** to operate - a **JITCTA** is used when data will not be editable
4.3	Objections: if DS has not given direct consent to processing and the PS is public administrative or Legitimate Processing, the controller determines validity of the objection. If the PD is sensitive data and/or the PS is sensitive then the objection is accepted and the PD is deleted. If the PS is direct marketing then any objection is accepted and the PD is deleted.	users are **empowered** to object to processing for certain purposes	the tracking logs contain a **prominent function** to object to the processing
4.4	Derived Information: Certain PS supplied by processor to controller or controller to DS could be used to gain an insight into a person's personality, e.g., services of interest to the DS. This derived information shall not be processed unless: the DS is informed of the PS related to the derived information, he/she unambiguously requests a service requiring performance of a contract and has issued explicit consent. The DS can object to the processing of the derived information at any time, and the derived information must be deleted.	users **understand** and are **informed** that their behaviour may provide some information, and they have provided **consent** for the processing of this information. They are also **empowered** to object to this processing	- the concept of derived information is **explained** at registration, and an example is provided - a **JITCTA** is used to confirm consent to processing - processing logs or other results of derived information are always presented with an **obvious** interface for objection

Finally, the requirements in the area of **consent** are to build systems that allow users to:

- give informed consent to the processing of PD;
- give explicit consent for a Controller to perform the services being contracted for;
- give specific, unambiguous consent to the processing of sensitive data;
- give special consent when information will not be editable;
- consent to the automatic collection and processing of information.

This list represents the essential HCI requirements that must be met in order to build systems that provide usable compliance with the European Privacy Directive. System designers will be well served if they consider the dimensions of comprehension, consciousness, control and consent when building privacy-enhanced systems.

Interface Methods to Meet Requirements

The field of interface design has developed a set of techniques, concepts, and heuristics that address each of the requirement areas. It is beyond the scope of this paper to provide an exhaustive review of the field of interface design, and interested readers are encouraged to examine one of the many HCI books for more information, for example [Nie94, Nor90, PRS^{+}94, Shn97, WH00].

Comprehension The obvious method to support comprehension or understanding is training. Users can be taught concepts and ideas through classroom training, manuals, demonstrations, etc. Such methods can be very successful, but they can also be expensive, time-consuming, and inappropriate when learning computer systems that will be infrequently used. Today, much effort is devoted to supporting comprehension without resorting to formal training methods.

User documentation, especially online or embedded documentation, is often used as a replacement for training. Most computers and software come with manuals of some sort, and much is known about how to develop material that people can learn from effectively [Nie94]. Studies have shown, however, that most users do not read the documentation, and often they cannot even find the printed manuals [CC87]. As a result, designers often resort to tutorials and help systems to support comprehension. Help systems can be designed to provide short, targeted information depending on the context, and such systems can be very powerful. It is often difficult, however, to learn an overview of all the features of a system using built-in help. Tutorials are another method of supporting learning, and they can work well if they are designed with a good understanding of the needs of the user.

There are other methods for supporting understanding that do not rely on documentation. For example, research in cognitive psychology has shown that users often develop personal "mental models" of complex systems. These models are attempts to understand something to a level where it can be used effectively, and such models can be quite effective when faced with complex systems. HCI specialists can exploit the human tendency to create models by either guiding users to develop appropriate models, or by examining the models that already exist and accounting for them. For example, people often have a mental model of a furnace thermostat that is analogous to a water faucet. That is, the more that it is "turned on", the faster the water (or heat) will flow. This model is incorrect because most furnaces can only operate at one flow rate and the thermostat only determines the temperature where the heat flow will be shut off. It is interesting to note that this erroneous mental model has persisted for a long time, and thermostat interface designers would likely want to take it into account. Thus, a thermostat designer might add a feature to automatically return the setting to a normal room temperature some time after the thermostat was suddenly turned to an abnormally high setting.

A related interface technique is the use of metaphors. Most modern graphical computer systems are based on a desktop or office metaphor, where documents can be moved around a surface, filed in folders, or thrown in a trashcan. The graphical elements of the interface, such as document icons that look like pieces of paper and subdirectory icons that look like file folders, reinforce this metaphor. The metaphor is valuable because it provides an environment that users are familiar with, and thus they can use familiar concepts and operations when interacting with the system. The familiar metaphor decreases the need to develop new knowledge and understanding.

There are other, more subtle techniques that can facilitate comprehension. For example, the layout of items on the screen can convey some meaning or information. Items that are grouped together visually will likely be considered to be group together conceptually [Nie94], and interface designers can take advantage of that. Also, items that are ordered horizontally in a display will likely be examined from left to right, at least in North American and European cultures. Interface designers can use this sequencing tendency to ensure that users follow the recommended sequence of operations.

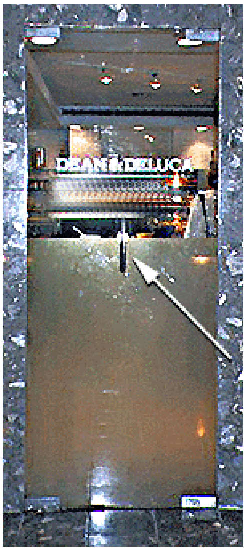
Feedback is also very important for supporting understanding [WH00]. Most complex systems require some experience and learning before they can be used effectively. Without feedback, users may not learn the consequences of their actions and understanding will be slow to develop.

Consciousness The requirement of consciousness refers to the user being aware of, or paying attention to, some concept or feature at the desired time. It is related to comprehension because the awareness may require some background knowledge before conscious attention is useful. Consciousness in this context can be thought of as bringing knowledge or understanding to the attention of the user so it can be used when required.

There are many interface techniques for making users aware of something. System messages or pop-up windows are an obvious technique for making the user aware of something. For important information, these windows can be constructed so the users have to acknowledge the message before they can continue using the system. A more subtle technique is to remind the user of something, without interrupting their work. This is sometimes seen in “help assistants” (such as the Microsoft Office Assistant) that make suggestions while users interact with the interface. Another way to remind users is through the arrangement of the interface. For example, if a particular option is available to a user at a certain time, placing icons or messages nearby in the interface layout can ensure that users are aware of the options.

Even more subtle methods use display characteristics to draw attention. Printing text in a certain colour, such as red, can draw attention. Changing the colour dynamically can be more effective. Sounds are also frequently used to make users aware of some event. The human factors discipline has a long history of designing systems that make users aware of certain things at certain times [WH00].

Control Control refers to the ability of the user to perform some behaviour. Control is related to comprehension because the user must understand the task and context to behave effectively. Control is also related to consciousness because users must be aware of the need to act before they can execute the behaviour. The issue of control, however,



The door is solid glass with a vertical handle in the middle.
(from http://www.baddesigns.com; reprinted with permission)

Figure 12.2: A door with poor affordances.

is that once the user knows that they are supposed to do something (awareness), and they understand what to do (comprehension), can they actually carry out the action.

An important concept for ensuring control is affordance, which means to provide naturally or inevitably. The classic example is door opener design. With some doors, users may approach the door, understand that it is a door, be conscious that they need to open the door, and still not be able to perform the action (see Figure 12.2 for an example). In contrast, a simple metal plate placed on the surface of the door tends to be a natural signal to push the door (in fact, these are often called "push plates"), whereas a metal loop placed vertically at the edge of a door tends to be a natural signal to pull the door. By using affordances, interface designers can make the door easy to control.

Another interface technique that supports appropriate actions is mapping. The idea is to map the appearance and function of the interface to the device being controlled. This might mean making a physical analogy of the real world in the interface, such as arranging light switches on a wall in the same order that the lights are arranged in the ceiling [Nor90].

Many of the subtle HCI techniques that can be used to support control are related to "obviousness". To the extent that the interface can be made obvious to the user, control (and understanding) can be smooth and effective. When interfaces are not obvious, users may have serious problems using the device or system (see http://www.baddesigns.com for some amusing examples of non-obvious designs). The goal of the interface designer is to build something that is so obvious to the user that comprehension, consciousness, and control will develop with little learning and effort.

Consent The final HCI requirement category is consent. Users must be able to consent or agree to terms or conditions that may be associated with a system or service. Moreover, the consent should be "informed", meaning that the users fully understand what they are

Table 12.7: Guidelines for Creating Click-Through Agreements.

1. Opportunity to review terms: users must view the terms of the agreement before consenting to the agreement. A recent case involving Netscape (Thornburgh, 2001)[5] established that it is important that there be no other method to obtain the product or service other than by clicking-through the agreement.
2. Display of terms: the terms have to be displayed in a "reasonably conspicuous" (Thornburgh, 2001)[6] manner. A recent case involving Ticketmaster (Kunz, 2002)[7] established that simply linking to the terms at the end of a long home page was not enough.
3. Assent to terms: the language used to accept the agreement must clearly indicate that a contract is being formed.
4. Opportunity to correct errors: there should be a method for users to correct errors, such as seeking a final confirmation before proceeding, or allowing the user to back-out of an agreement.
5. Ability to reject terms: the option to reject the terms of the agreement should be clear and unambiguous, and the consequences of the rejection should be stated (e.g., "if you do not agree, you will not be able to install this software").
6. Ability to print the terms: the interface should allow the user to print the terms for later reading.

agreeing to, and what implications this may have. Obviously, supporting informed consent is related to the requirements for comprehension and consciousness.

The most common method for supporting consent in computer applications is a "user agreement". When you have installed new software on your computer, or signed-up for an Internet service, you have undoubtedly seen an interface screen that presents a User Agreement or Terms of Service. In order to continue, you have had to click on an "I Agree" button or an equivalent label. These interface screens are commonly called "click-through agreements" because the users must click through the screen to get to the software or service being offered [Tho01]. (An alternative label is "click-wrap agreement", in parallel to more traditional "shrink-wrap" agreements attached to software packaging.) These agreement screens are an attempt to provide the electronic equivalent of a signed user agreement or service contract [Sla99]. By clicking on the "Agree" button, the user is confirming their understanding of the agreement and indicating consent to any terms or conditions specified in the accompanying text.

The legality of these click-through screens in forming the basis of a legal agreement or contract has been established, but with some qualifications. The Cyberspace Law Committee of the American Bar Association has recently reviewed the case law and developed a set of guidelines for creating click-through agreements [Kun02, KDDT01]. These guidelines have been summarised into six principles to be considered by system developers, and these are listed in Table 12.7 [HC02, Tho01].

Other factors that should be considered when creating click-through agreements [Sla99] are to redisplay the terms and conditions at product start-up (reminding), and to support the ability to review the terms at any time (e.g., in the "help" or "about" menus). In addition, developers should adapt the terms and conditions to local languages and requirements. If these principles and considerations are heeded, case law suggests that click-through agreements will likely be enforced, at least in US courts. (Some jurisdictions, such as Germany and China, are unlikely to enforce any of these agreements [Sla99].)

The text of many click-through agreements tends to be long and complex, often to ensure that all the points raised above are addressed. The result is that many users have difficulty reading and understanding the documents (a comprehension problem), and many users click the "Agree" button without considering the terms at all (a consciousness problem). The problems arise because people have limited cognitive capacity: we have limited attention spans, a restricted ability to process large quantities of detailed information at one time, and limited memories. Thus, using interface techniques that are sensitive to user characteristics is important. This observation may be particularly relevant if users are being asked to agree to a number of terms that will affect them substantially, such as the processing of their personal data.

Ensuring that users fully understand and unambiguously agree to the processing of their personal information is important for complying with privacy legislation and guidelines. Consider the definition of consent provided in the EU Directive 95/46/EC on privacy protection (European Commission, 1995)[8]:

> *'the data subject's [user's] consent' shall mean any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed. (Article 2-h)*

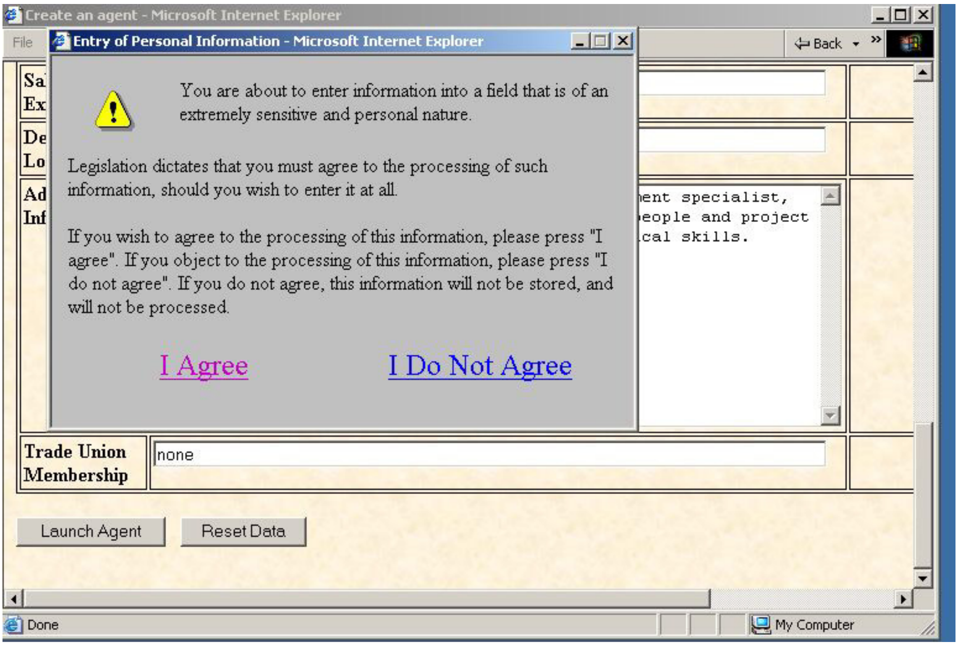
It is clear that a large, cumbersome, complicated User Agreement presented to the user only when they begin to use a product or service fails to live-up to the requirements for "specific" and "informed" consent, and yet these types of user agreements are the majority. These issues are of particular concern in relation to explicit consent. For example, the EU Directive states that when sensitive data (e.g., race, ethnic origin, religious beliefs) are processed, the user must give "explicit consent" (Article 8-2-a) to the processing of the sensitive data. Again, a single, large, click-through User Agreement does not meet the spirit of The Directive.

The solution to this problem proposed here is a new concept of "Just-In-Time Click-Through Agreements" (JITCTAs). The main feature of a JITCTA is not to provide a large, complete list of service terms but instead to confirm the understanding or consent on an as-needed basis. These small agreements are easier for the user to read and process, and facilitate a better understanding of the decision being made in-context. Also, the JITCTAs can be customised for the user depending on the features that they actually use, and the user will be able to specify what terms they agree with, and those they do not. The responses made by the user during the JITCTAs can also be recorded so there is a clear, unambiguous record of the specific agreements made with the user. In order to implement JITCTAs, the software will have to recognise when users are about to use a service or feature that requires that they understand and agree to some term or condition.

A sample screen capture of a JITCTA is shown in Figure 12.3. In this example a user has selected the Trade Union Membership information field in the Create Agent interface screen of the PISA interface. Since this would be considered sensitive information in the EU Privacy Directive, a JITCTA has appeared to obtain explicit, specific, timely, unambiguous consent to the processing of this data.

In summary, well-formulated click-through agreements are legally permissible in many countries, and Just-In-Time Click Through Agreements improve on this device by supporting more appropriate decision-making and control that is sensitive to human factors constraints.

[8] European Commission (1995). Directive 95/46/EC of the European Parliament and the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data. *Official Journal of the European Communities* (1995), p. 31.


Create an agent - Microsoft Internet Explorer
File
Back
Entry of Personal Information - Microsoft Internet Explorer
Sa
Ex
De
Lo
Ad
Inf
I Agree
I Do Not Agree
Launch Agent
Reset Data
Done
My Computer

You are about to enter information into a field that is of an extremely sensitive and personal nature.

Legislation dictates that you must agree to the processing of such information, should you wish to enter it at all.

If you wish to agree to the processing of this information, please press "I agree". If you object to the processing of this information, please press "I do not agree". If you do not agree, this information will not be stored, and will not be processed.

```
ent specialist,
eople and project
cal skills.
```

Trade Union Membership	none

Figure 12.3: An example of a Just-In-Time Click-Through Agreement (JITCTA).

Summary and Conclusions

This chapter introduced design guidance for privacy-enhancing technologies from a human factors point of view. For the first time, this work specified what must be included in human-computer interfaces to satisfy the spirit of European privacy legislation and principles, and satisfy the privacy needs of the users (“usable compliance”).

The current work has focused on European privacy legislation and, although the resulting principles, requirements, and solutions are general, one of the challenges that remains is to ensure that the knowledge is equally applicable in other legislative settings, such as Canada, and in areas operating in a self-regulatory fashion (e.g., the USA). For example, it is possible that the market forces operating in the USA will lead to privacy requirements and expectations that have not been anticipated. Even in regulated environments, the privacy legislation and guidelines will change and evolve, and thus the human interface guidelines will also have to be dynamic.

Privacy-enhancing technologies are also evolving and changing, and this will have an effect on the types of solutions that are available, and also the privacy needs and expectations of the users. For example, the P3P protocol, if implemented widely, may have a profound effect on the privacy domain by bringing privacy issues to the attention of millions of Internet users, and hopefully providing an easy-to-use privacy control interface (e.g. [CAG02].)

12.2 Testing of User Interfaces for a Privacy Agent

12.2.1 Introduction

This section the results of a preliminary usability test of portions of a design prototype. The purpose of the testing was to determine if the design concepts in the prototype were successful in constructing an agent technology that (1) users can use, (2) users can understand, and (3) users can trust with secure and private information. This testing involved laboratory sessions where volunteers interacted with the software system on a computer and answered questions about the features and performance that they experienced.

The PISA Interface Prototype

A stand-alone interface demonstration was developed at NRC in collaboration with the PISA partners. The goal was to develop a set of WWW pages and back-end applications that would demonstrate a ”look and feel” for the PISA demonstrator before the actual agent platform was available. This prototype was designed to be both stand-alone, so it could be demonstrated and tested, and also modular and well-structured, so it could be integrated into the final PISA Demonstrator. These design goals were met and the interface prototype can be demonstrated over the Internet.

The interface prototype was developed using manually-edited HTML code to facilitate fine-level control of the appearance and behaviour. In addition, using native HTML code, rather than a WWW authoring tool, allowed for detailed documentation of the interface components and design concepts. JavaScript was also used extensively to support dynamic and customizable WWW pages that reflected the individual usage of the system (e.g., whether a user had created an agent or not). Cascading Style Sheets (CSS) were also used to ensure that the entire prototype had a consistent look and behaviour, which is an important design consideration when building trust.

In order to give the interface prototype realistic agent-like behaviours, browser cookies were used to keep track of the actions of individual users. This allowed the prototype to track users’ privacy preferences and actions as they used the system. Also, a simulated agent platform was developed using the Perl programming language and the Common Gateway Interface (CGI) specification. This simulated platform produces random agent-like actions that appear in the interface as inter-agent communications or requests to the users for more information.

The most notable interface features of the prototype are the use of a certificate-like background, pastel colours, colourful icons, metaphors that represent the functionality (e.g., a detective looking at footprints for the tracking function), left-to-right sequencing of the menu bar, and a status display about waiting messages. The interface also contained rollover help, where terms are defined or information is provided when the users point to key words in the interface. Other design features of the prototype are summarized in Table 12.8.

Remote Usability Evaluation

This evaluation was conducted in partnership with the Human Oriented Technology Laboratory (HOT Lab) at Carleton University as part of a separate investigation of remote usability testing methods. An HCI Master’s degree candidate, Cassandra Holmes, examined different remote usability testing methods when evaluating the PISA interface. A diagram

Table 12.8: Notable HCI Design Features in the Interface Prototype.

1.	security/trust measure are obvious (e.g., logos of assurance)
2.	there is a consistent visual design, and visual metaphors supported by graphical elements
3.	there is a conservative appearance, with the background resembling a legal certificate
4.	there is a functional layout organized around user tasks
5.	the interface design provides an overview of the functions, the ability to focus and control individual components, and it provides detailed information on demand
6.	sequencing by layout is supported in the left-to-right ordering of ”create agent”, ”modify agent”, ”track agent”, etc.
7.	the interface is designed to support embedded help
8.	where appropriate, the interface requires confirmation of actions, such as agreeing to the processing of personal information
9.	user are reminded of their data protection rights, and controls for those rights are prominent and obvious (e.g., objecting to processing)
10.	a double ”Just-In-Time Click-Through Agreement” (JITCTA) is used for specially sensitive information (i.e., union membership)
11.	pop-up windows are used to present interim results or to ask the user for clarification
12.	obvious agent controls are included (start, stop, track, modify)
13.	controls are provided for setting, customizing, and modifying privacy preferences and controls (e.g., retention period)
14.	visual design is used to remind users of transparency limits (i.e., the agent tracking logs are colour-coded when complete transparency is not available)

depicting the setup of these tests is shown in Figure 12.4. Participants in the test interacted with a client computer, which ran a standard WWW browser. This computer retrieved the prototype WWW pages from a WWW server, which also acted as the experimenter’s console. The client computer was connected to the experimenter’s computer using a protocol called Virtual Network Computing (VNC; www.realvnc.com), which allowed a copy of the client computer’s display to be viewed and captured at the experimenter’s computer. Camatasia software (www.techsmith.com) was used to record all the actions made by the participant and experimenter, and microphones were used to capture all the comments spoken by the experimenter and participant.

Most of the testing was done remotely – the participant and the experimenter were in different physical locations, with the participant working at the client computer and the experimenter observing at her computer. During these remote sessions all interactions between the participant and the experimenter were done using remote collaboration tools such as Microsoft NetMeeting for text communication and the Reliable Audio Tool (RAT; http://www-mice.cs.ucl.ac.uk/multimedia/software/rat/) for audio communication. One of the goals of the research project was to test for differences when the interaction was done with a voice channel or a text-based ”chat” channel. Also, in some of the conditions the participant and experimenter interacted, while in others the participant simply spoke or typed their responses for later analysis. The experiment also included a control condition that used a standard usability test method. Here the participant and experimenter sat side-

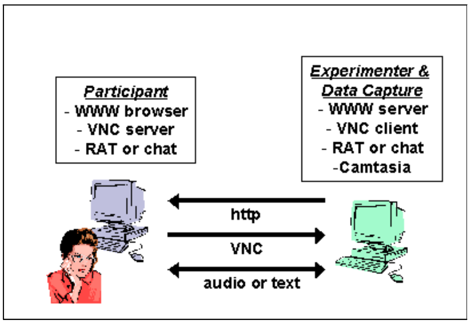


Figure 12.4: Configuration of the remote usability tests.

by-side in the same room and used one computer.

12.2.2 Preparing the Prototype for Testing

A number of changes were made to the PISA Prototype Interface after the release of the first report. Some of these changes were a result of further refinement of the PISA sample application done with the PISA partners. Other changes were done in reaction to casual reviews done by usability experts while the experiment was being setup. These changes in the prototype fall into two classes: changes to the appearance of the interface screens, and changes to the behaviour of the stand-alone demonstration.

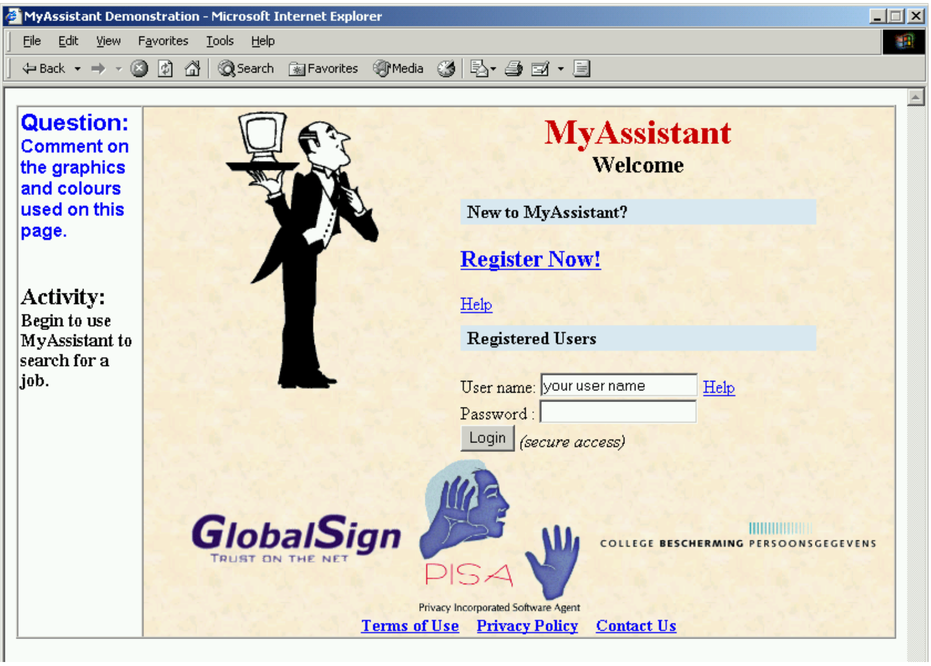
Changes to Appearance

The initial PISA interface prototype consisted of an application that was specific to searching for jobs on the Internet. The PISA team desired a more general-personal PISA Demonstrator, with job searching being only one of a few possible uses. For these reasons, the title of the prototype was changed to ”MyAssistant” to support the idea of a general-purpose service. In addition, a number of search tasks were displayed to the participants, including jobs, music, cars, software, and news, although job searching was the only active option.

Other changes were made to the vocabulary used in the interface in an attempt to make it easier to read and understand. Most important, there was no mention of “agents”, but instead “tasks” that could be created and given to the personal assistant service. Other changes involved wording changes to make the terms more appropriate for a young, Canadian test population, such as replacing the term “marketing” with “advertising”. Some difficult terms in the initial interface prototype, such as ”direct marketing” were also replaced with terms that might be easier to understand, such as ”advertising sent by us”.

Changes to Behaviour

The most notable change to the behaviour of the interface was suggested by the PISA partners during a review of the initial interface prototype. Instead of being separate steps, the entry of personal information and the setting of privacy preferences were integrated on the same page. Participants entered their personal information at the top of the screen, and



MyAssistant Demonstration - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media

Question: Comment on the graphics and colours used on this page.

Activity: Begin to use MyAssistant to search for a job.

MyAssistant

Welcome

New to MyAssistant?

Register Now!

Help

Registered Users

User name: your user name Help

Password :

Login (secure access)

GlobalSign
TRUST ON THE NET
PISA
Privacy Incorporated Software Agent
COLLEGE BESCHERMING PERSOONSGEGEVENS

Terms of Use Privacy Policy Contact Us

Figure 12.5: The two-column layout of the testing interface.

then scrolled down to set their privacy preferences. It was hoped that this would make a more intuitive interface that closely tied the personal information and the privacy controls.

Other changes were made because of the remote usability testing procedures. Since the experimenter and participant were communicating remotely, or not communicating at all, instructions for the participant to follow and prompts to elicit usability comments were integrated into the interface display. This was done by displaying short messages in a left-hand column, while the ”MyAssistant” service was displayed in the right-hand column (see Figure 12.5). In addition, dynamic usability prompts were programmed so that probe questions were displayed in the left-hand frame after certain actions by the participants. For example, upon completion of a form, a usability probe question might appear before the user could go on to the next step in the instructions.

Limitations of Testing

It is common practice to limit usability testing sessions to one hour in duration. So, this initial testing of the PISA interface prototype was limited to specific functions. Included in this usability test were the steps of: user registration, login, agent (task) creation, entering personal information, setting privacy controls, and launching the agent (task). Other testing will be necessary to evaluate the functions of tracking agent actions, objecting to the processing of personal information, viewing results, etc. This usability test did include questions about the overall visual design of the interface, however, along with probes about the understandability of the personal assistant service being used to search for jobs and the overall trustability of the service. Thus, this limited test is still of value.

12.2.3 Experimental Method

Participants

Fifty individuals were recruited from the Carleton University community through posters, the Introductory Psychology notice board, and also by approaching individuals in common areas. Other volunteers were acquaintances of the authors or individuals that were referred by other participants. Participants were given an honorarium of CAN$10.00 for their participation in the study. One participant's testing time substantially exceeded one hour, so she was paid an additional CAN$10.00.

The qualifications for participation were familiarity and experience with the Internet, English as a native language, and the ability to type. To assess typing ability, all participants completed a two-minute typing test using TypeMaster Typing Test 6.2. Participants were required to obtain a minimum of 20 words per minute (gross) on the typing test. All participants met the typing criteria.

One interesting characteristic of the participants was that they were late adopters of Internet services. At the end of the testing session they completed a questionnaire that asked if they agreed with the statement: ”Among my peers, I am usually the first to try out new Internet sites or services”. Only 42% of the participants agreed with this statement, with only 6% strongly agreeing. In contrast, 52% of the participants disagreed or strongly disagreed with the statement.

It should be noted that the participants were not representative of the general population, or of the likely user group for a real PISA service. Instead, the participants were young, Canadian college students, who may have very different attitudes towards WWW site design, Internet services, job searching, or privacy preferences than older adults or European citizens. This important limitation should be kept in mind when interpreting the results.

Materials

The PISA interface prototype was the object to be tested in this analysis. The interface was presented as the ”MyAssistant” service, and the following instructions were presented to the participants:

> You are going to help us test a new service on the Internet called *MyAssistant*. *MyAssistant* helps individuals search for many different things on the Internet, such as jobs, cars, or music. Today you will use *MyAssistant* to find a new job. *MyAssistant* asks you to enter information so that it can search the Internet for you to find jobs that you might be eligible for. *MyAssistant* then reports back if it finds anything relevant.
>
> You will do several things on the *MyAssistant* web site and answer questions about the job searching service and your experiences using it. The *MyAssistant* service that you are using is a mock-up and not all areas of the site are working. We are testing the web site, not you, or your abilities. If you have problems or do not like something, we will use that information to improve the *MyAssistant* site.

All testing was conducted at Carleton University's Human Oriented Technology Laboratory (HOT Lab; http://www.carleton.allowbreak ca/allowbreak hotlab/). Two separate rooms were used to simulate distance for the remote testing, and the participant was in one testing room while the experimenter was in another. The participant and experimenter workstations consisted of a computer using Microsoft Windows 2000 and 19-inch monitors. Each computer also had Internet Explorer, Microsoft NetMeeting, and the Reliable Audio Tool (RAT) installed. Each workstation was also equipped with a headset with microphone that was used during the voice communication sessions.

The VNC screen sharing software was installed on both computers to allow the experimenter to view the participant's screen in real-time. The experimenter also had Notepad (a text editor) open on the screen for taking notes during the sessions. Audio and all information from the experimenter's monitor were recorded into an AVI file by the Camtasia recording software for later data analysis.

Procedure

Upon arrival, participants were greeted and asked to read and complete the informed consent form. After completing the form, participants were given the two-minute typing test. Once the test was completed, participants were given a brief overview of what was to take place during the testing session. Participants were also directed to read the instructions thoroughly and were asked if they had any questions. Before the test began, the experimenter started the Camtasia screen recording software.

The first interface page encountered by participants was an instruction page. The instruction page provided information about the web site and its purpose, explained the purpose of the usability test, and outlined the importance of providing detailed thoughts, comments, and answers to questions. The instruction page also explained that the task instructions and questions would appear in a column on the left side of the screen and how participants were to provide their thoughts, comments, and answers to questions. After participants read the instructions, they began the usability test by clicking a link at the bottom of the page. Clicking the link brought participants to the greeting page of the MyAssistant web site, where they followed the instructions and answered the probe questions.

After completing the usability test, participants were given a 16-item questionnaire to assess the overall usability of the prototype. In addition, this questionnaire enquired about their attitudes towards the trustability of the Internet in general, Internet services, and the "MyAssistant" (PISA) prototype that they had just tested.

12.2.4 Results

Usability Test Method Results

There were differences in the number and nature of comments made in the different remote testing conditions depending on whether the communication was verbal or text-based, and depending on the presence or absence of live interaction with the experimenter. There were no differences, however, in the usability performance measures and usability ratings that are discussed below. Therefore, the results presented here come from all 50 participants regardless of the format of their usability test.

Ability to Understand and Use Interface

During the usability test, participants were asked questions to determine whether or not they understood the purpose of various interface elements, terminology, and procedures (e.g., "What is the purpose of naming your task?"). Answers were coded in terms of whether or not participants understood. There were also two behavioural indicators in the

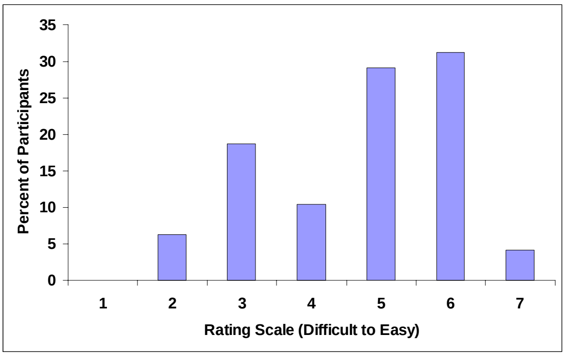


Figure 12.6: Frequency of ease of use ratings (1 = very difficult, 4 = neither easy nor difficult, 7 = very easy).

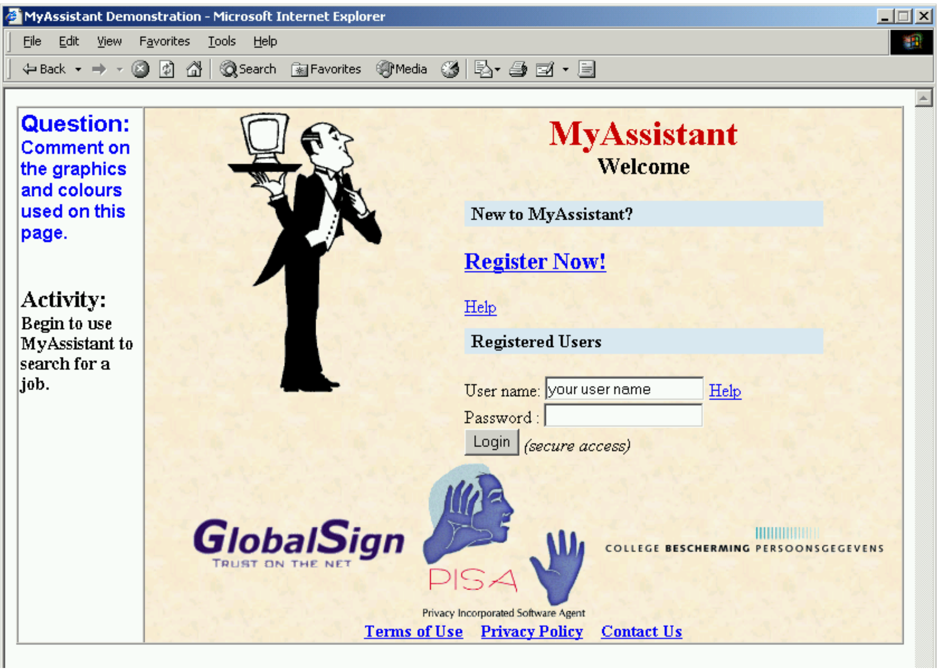
usability test, indicative of whether or not participants understood what they were doing. First, the experimenter observed which link participants initially chose on the greeting page: the “Register Now” link, which was for new users, or the “Login” link, which would not work if the user had not yet created a user account. The second indicator was whether or not participants were able to successfully register. To obtain total performance scores, the total number of “yes” designations for each participant was summed. The highest possible score was 27.

The average performance score was 17.50 (SD = 3.89), which indicates that only 65% of the concepts and functions were understood by the users. The most notable problem areas were that users did not understand the privacy preference settings, 30% of the participants attempted to login before registering, and that many people did not understand some of the key terms used in the interface (e.g., ”job sector”, ”company type”). More details about these problems are presented in the analysis of comments presented below.

Ease of Use Ratings

At the end of the testing session, participants were asked to rate the ease of creating their job search. They were asked to use a scale of 1 to 7, with 1 being “very difficult”, 4 being “neither easy nor difficult”, and 7 being “very easy”. The average ease of use rating was 4.70 (SD = 1.38), which represents a rating of ”slightly easy”. A frequency count of the ratings is shown in Figure 12.6, and it can be seen that the most common ratings were 5 or 6, but these were offset by a fair number of low ratings (2 or 3). In their comments, participants most frequently used terms like ”pretty good” and ”fairly easy” to describe the usability of the prototype. The most common complaints were problems in understanding the terms used in the interface.

The final questionnaire also included two questions on ease of use. First, participants were asked to indicate their agreement to the statement ”This site was simple to navigate” on a scale of: 1 = ”strongly disagree”, 2 = ”disagree”, 3 = ”undecided”, 4 = ”agree” and 5 = ”strongly agree”. The average rating was 3.64 (SD = 1.05), which again represents a slightly positive result. A count of the number of people who agreed or strongly agreed (ratings of 4 or 5) indicated that 76% agreed with the statement. Thus, the site was fairly easy to navigate, but it was not spectacular.



MyAssistant Demonstration - Microsoft Internet Explorer

Question: Comment on the graphics and colours used on this page.

Activity: Begin to use MyAssistant to search for a job.

MyAssistant

Welcome

New to MyAssistant?

Register Now!

Help

Registered Users

User name: your user name Help

Password :

Login (secure access)

Terms of Use Privacy Policy Contact Us

Figure 12.7: Greeting screen of the PISA prototype.

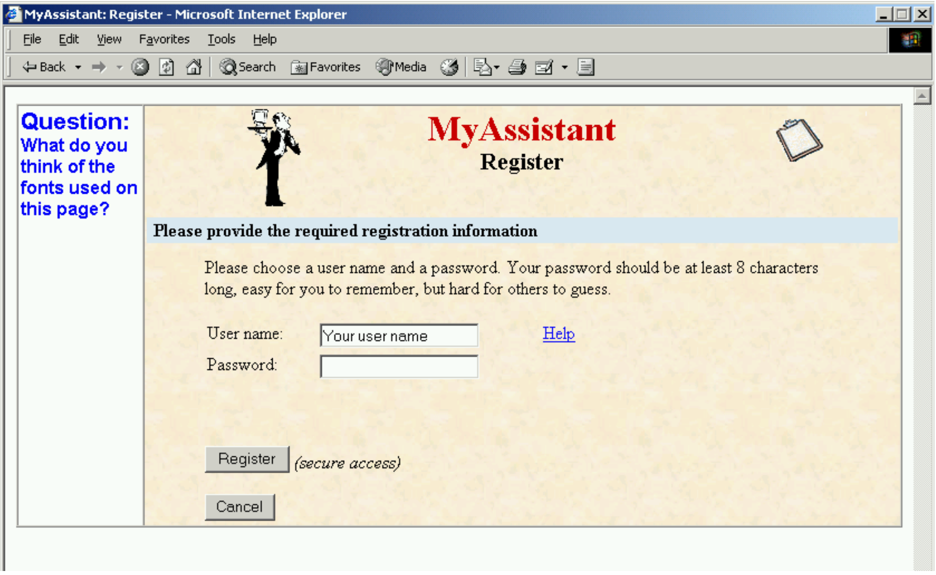
A second question asked users to indicate their agreement with the statement ”Overall, this site worked very well technically” on the same five-point scale. The average rating was 3.62 (SD = 1.14) and 72% of the participants agreed or strongly agreed with the statement. Again, these results indicate that the prototype worked fairly well, but that there was room for improvement.

Usability Opinions and Comments

The most important analysis is of the comments made by users while using the prototype interface. In this section an image of each of the interface screens is presented, along with a summary of the usability parameters and comments of the users.

Figure 12.7 shows the initial greeting screen that users experienced when they started interacting with the interface prototype. The two-column layout is evident, with the main PISA interface shown in the right-hand column and the instructions and probe questions shown in the left-hand column. The probe questions were presented in a blue san-serif font, while the assignment instructions were shown in a black Roman font. Participants were instructed to respond to the probe questions before they followed the instructions and interacted with the interface.

Concerning the colours and graphics, 38% of the participants commented that they did not like the graphics used in the interface. Participants who did not like the graphics felt that the butler graphic had been copied from the Ask Jeeves web site (www.ask.com), and one person commented that the PISA graphic at the bottom looked like someone had been push against a window trying to escape. Those that did like the graphics described


MyAssistant: Register - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media

Question: What do you think of the fonts used on this page?

MyAssistant

Register

Please provide the required registration information

Please choose a user name and a password. Your password should be at least 8 characters long, easy for you to remember, but hard for others to guess.

User name: Your user name Help

Password:

Register (secure access)

Cancel

Figure 12.8: The registration interface screen.

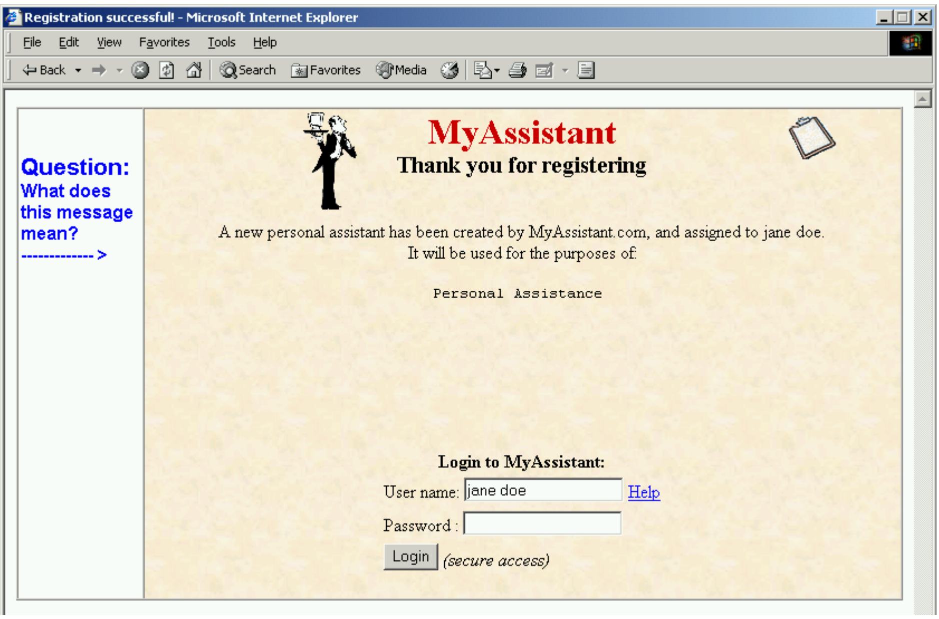
them as ”professional looking” and ”cute”. Similarly, 42% did not like the colours used in the interface. Comments were that the colours were bland and boring, and perhaps there were too many colours used in the interface. Those participants who did like the colours described them as soothing or relaxing.

In terms of usability, it was not always clear to the participants that they had to register before they could login to the service. One 30% of the trials, users attempted to login (and failed) before they completed the registration. The help information was often successfully used to determine that registration was required before login. These results suggested that a clearer set of instructions, such as asking a direct question ”Do you have a MyAssistant account?” may be necessary to make this interface more usable. On a positive note, some users did indicate that they liked the ”secure access” label beside the login button because it provided reassurance that the information was being transmitted securely.

Another finding related to the visual appeal of the interface comes from the questionnaire administered at the end of the testing session. Participants were asked their level of agreement with the statement: ”Visually, this site resembled other sites I think highly of.” Only 42% of the participants agreed with this statement, with only 4% strongly agreeing. A total of 50% of the participants disagreed with the statement, with 22% strongly disagreeing. It is clear from these results that the visual design of the interface needs to be improved.

Figure 12.8 shows the interface screen used for registering with the service. Users were asked to enter a username and password, using any text that they wished, and they had no problems doing that. In response to the probe question about the fonts, 88% of the participants like the fonts because they were standard Roman fonts that were familiar.

Figure 12.9 shows the interface screen that was displayed once the users had registered for the service. The probe question was intended to ascertain if users understood the message about creation of a new personal assistant. The results were that 92% of the users understood the message, although some felt that the message was redundant because it indicated that a personal assistant was created for providing personal assistance. Mostly,


Registration successful! - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media

Question: What does this message mean? ------------->

MyAssistant

Thank you for registering

A new personal assistant has been created by MyAssistant.com, and assigned to jane doe.
It will be used for the purposes of:

Personal Assistance

Login to MyAssistant:

User name: jane doe Help

Password :

Login (secure access)

Figure 12.9: Interface screen to acknowledge registration.

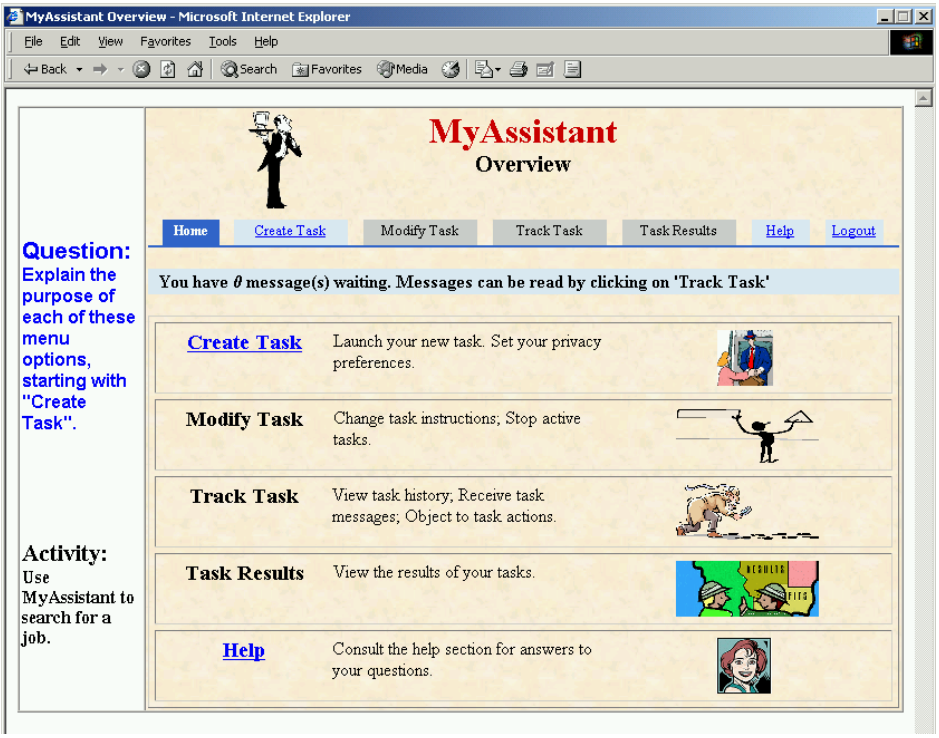
however, this page was useful because it reinforced the metaphor that a personal assistant had been created just for them.

A few users were confused by the need to login on this interface screen. These users thought they had already logged in when they registered for the service. This finding suggests that registration and login might be combined into one function to improve the usability.

Figure 12.10 shows the main interface screen of the prototype. It is designed to provide an overview of the major functions of the service, and a central point that users will return to when they use the service. Users were generally able to understand this page and build a good mental model of the assistant being given tasks that can be modified and tracked. This is an important finding that indicates that the main service model of the prototype is understandable and familiar to the users.

The probe question asked users to explain what they thought each of the menu options meant, and participants were able to do this easily (¿90% of users). The percentage of people able to accurately describe each function is shown in Table 12.9. Some users did comment, however, that the term ”task” was somewhat confusing but they were able to reason that it referred to creating a job search, which they knew to be the sample application for this test.

Users also continued to comment on the visual aspects of the interface screens. Here a few users commented that they did not like the images associated with each function. They felt that they were simple clip-art graphics taken from a standard library (which they were), and that they appeared ”cheap” or ”unprofessional”. A few users also felt that they did little to aid the understanding of the functions. One of the purposes of using these graphics in the interface was to tie together screens that are associated with the same function. So, screens that are involved with the more-detailed aspects of the function would contain the same graphic. This is seen in the interface figures below which all involve detailed interaction


MyAssistant Overview - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Home
Create Task
Modify Task
Track Task
Task Results
Help
Logout

Question: Explain the purpose of each of these menu options, starting with "Create Task".

Activity: Use MyAssistant to search for a job.

MyAssistant

Overview

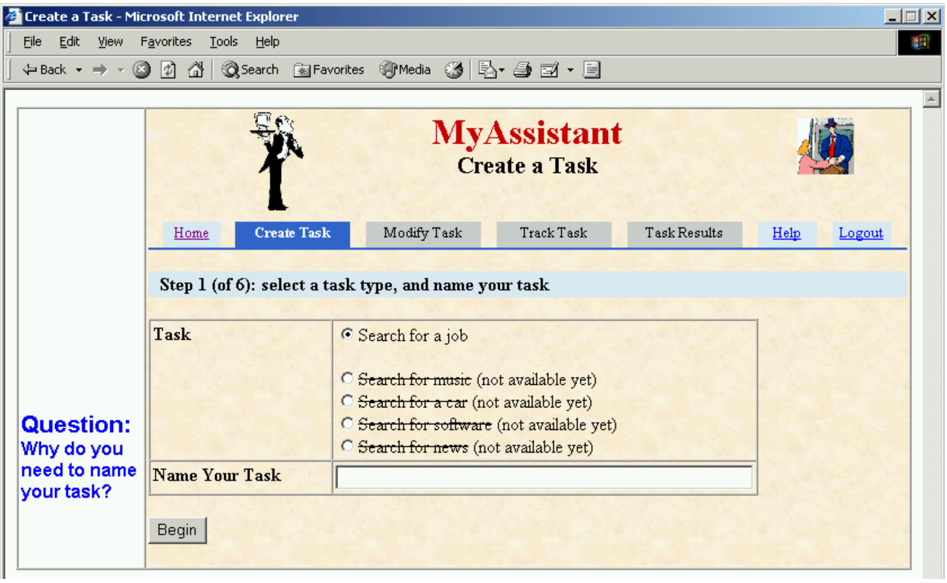
You have *0* message(s) waiting. Messages can be read by clicking on 'Track Task'

Create Task	Launch your new task. Set your privacy preferences.
Modify Task	Change task instructions; Stop active tasks.
Track Task	View task history; Receive task messages; Object to task actions.
Task Results	View the results of your tasks.
Help	Consult the help section for answers to your questions.

Figure 12.10: The main interface screen of the prototype, showing the major functions.

Table 12.9: Percentage of Users Who Understand each Function.

Function	Percentage of People Understanding
Create Task	92
Modify Task	92
Track Task	90
Task Results	96
Help	96


Create a Task - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Home
Create Task
Modify Task
Track Task
Task Results
Help
Logout
Begin

MyAssistant

Create a Task

Step 1 (of 6): select a task type, and name your task

Task	● Search for a job ○ ~~Search for music~~ (not available yet) ○ ~~Search for a car~~ (not available yet) ○ ~~Search for software~~ (not available yet) ○ ~~Search for news~~ (not available yet)
Name Your Task	

Question: Why do you need to name your task?

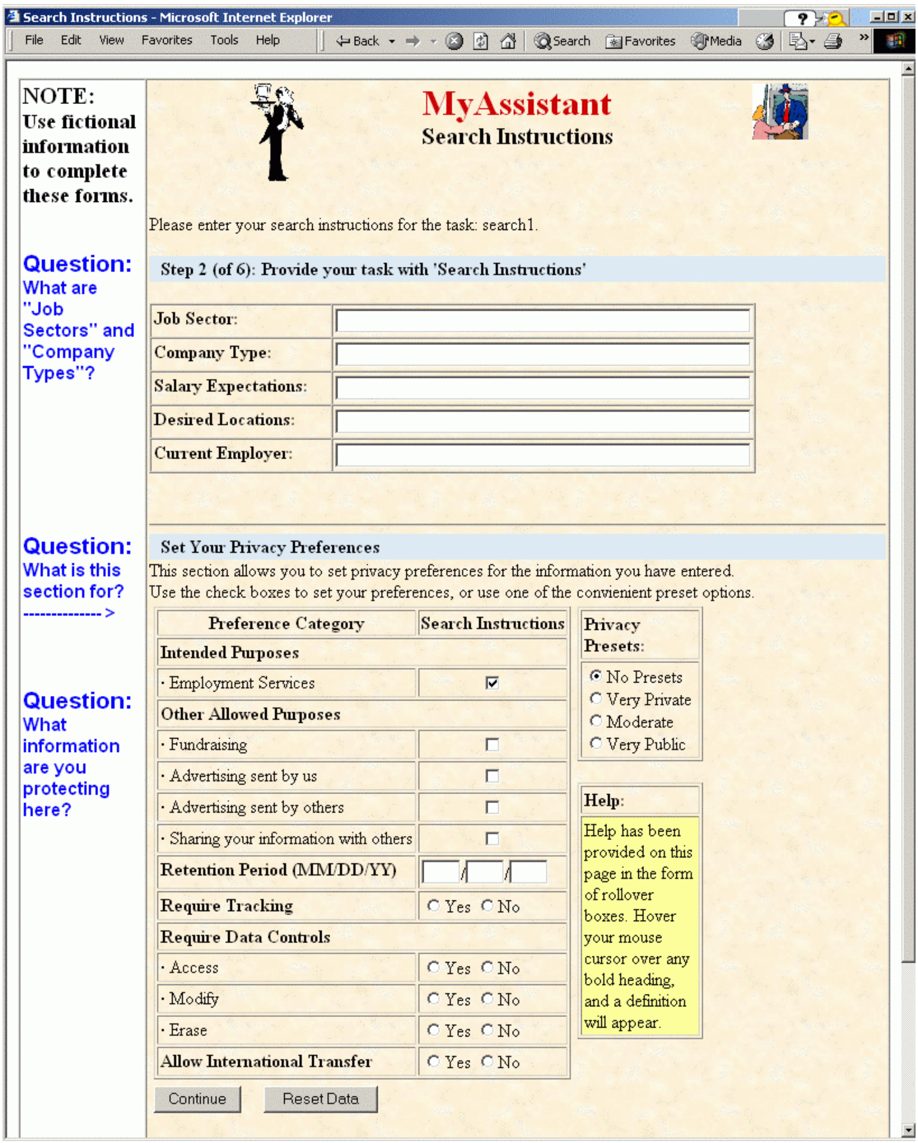
Figure 12.11: The interface screen used to create a new task.

related to the ”create task” function, so they all have the same graphic element in the upper right corner. These comments from the participants should lead to a questioning of this design element in the PISA interface, and/or an improvement in the quality and nature of the graphics.

Figure 12.11 shows the interface screen used to create a new task. The results of the probe question were that 50% of the participants did not know why they should name their task. Many people thought that the task type ”Search for a job” was the name of their task, so they did not know why they had to type that information into the name field. Others felt that the task name would somehow restrict their search to certain jobs of interest. Obviously, users did not realize that they could create many tasks of the same type, and would later have to identify them to check on progress and view results. This feature of the PISA service will have to be made clearer in the final implementation. Perhaps the task naming could be done as a final step, once the parameters of the search have been entered.

Figure 12.12 shows the interface screen that was used to enter parameters for the job search, as well as privacy preferences for that information. Both the entry of information and the setting of preferences are done on the same screen, which is a change form the initial interface prototype described in an earlier PISA report [PK03].

The results for the probe questions on ”job sectors” and ”company types” indicate some problems with these terms. The data show that 28% of the participants did not understand what was meant by job sector, and 60% did not understand the term ”company type”. It is not even clear to the experimenters what these terms were intended to mean, and clearly these must be reconsidered in a final implementation. Some users were also confused about how to complete the ”salary expectations” field because they did not know what unit of measure to use (hourly, monthly, yearly, etc.). Similarly, they were often not clear if ”desired locations” referred to countries, provinces, cities, or neighborhoods. Some users also did not understand why they were asked for their ”current employer”. The intention here is to allow users to enhance their privacy by not contacting their current employer,


Search Instructions - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Continue
Reset Data

NOTE:
Use fictional information to complete these forms.

Question: What are "Job Sectors" and "Company Types"?

Question: What is this section for? -------------- >

Question: What information are you protecting here?

MyAssistant
Search Instructions

Please enter your search instructions for the task: search1.

Step 2 (of 6): Provide your task with 'Search Instructions'

Job Sector:	
Company Type:	
Salary Expectations:	
Desired Locations:	
Current Employer:	

Set Your Privacy Preferences

This section allows you to set privacy preferences for the information you have entered.
Use the check boxes to set your preferences, or use one of the convienient preset options.

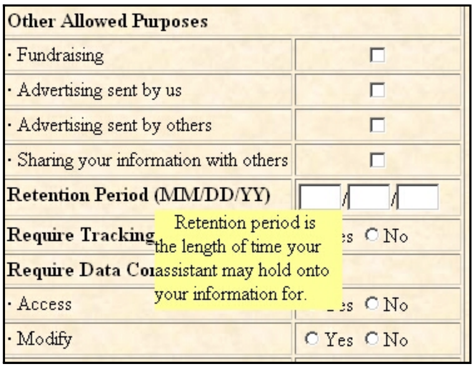
Preference Category	Search Instructions
Intended Purposes	
· Employment Services	☑
Other Allowed Purposes	
· Fundraising	□
· Advertising sent by us	□
· Advertising sent by others	□
· Sharing your information with others	□
Retention Period (MM/DD/YY)	__ / __ / __
Require Tracking	○ Yes ○ No
Require Data Controls	
· Access	○ Yes ○ No
· Modify	○ Yes ○ No
· Erase	○ Yes ○ No
Allow International Transfer	○ Yes ○ No

Privacy Presets:
● No Presets
○ Very Private
○ Moderate
○ Very Public

Help:
Help has been provided on this page in the form of rollover boxes. Hover your mouse cursor over any bold heading, and a definition will appear.

Continue Reset Data

Figure 12.12: The interface for setting job search parameters and privacy preferences.



Other Allowed Purposes	
· Fundraising	□
· Advertising sent by us	□
· Advertising sent by others	□
· Sharing your information with others	□
Retention Period (MM/DD/YY)	__ / __ / __
Require Tracking	es ○ No
Require Data Co	
· Access	es ○ No
· Modify	○ Yes ○ No

Retention period is the length of time your assistant may hold onto your information for.

Figure 12.13: Example of rollover help.

which could be embarrassing, but that intention is not clear in this interface presentation.

Concerning the probe about setting privacy preferences, 16% of the people failed to understand what was intended here. Some felt that this portion of the screen was a continuation of the search instructions. More problematic was the next probe question, which asked users what information they were protecting when setting the preferences. A total of 64% of the users were not able to correctly answer that it was their search parameters that they were protecting. The most common reason was that they did not feel that the search parameters were anything that needed protection.

Users often commented that this was a very complex screen with a lot of information to read and understand. One feature of this interface screen that was useful was rollover help, as is shown in Figure 12.13. When users pointed their mouse at some of the terms in the interface a yellow information box would appear to provide more information. This design feature was successful and 86% of the participants indicated that this form of help was useful, although many did not discover the feature immediately. In fact, without the rollover help, many of the privacy preference terms, such as ”intended purposes” and ”require tracking”, would not have been understood at all. So, while the rollover help was effective, improving on the labels and terms used in the interface would be a better solution than relying on the extra information.

Many of the users failed to notice the preset buttons that were available. Those that did were surprised when selecting a preset changed all of the privacy settings they had made previously. Other users commented that some of the preset values seem to be inconsistent, such as allowing ”advertising by us” when the most private preset was used (this was an implementation flaw in the prototype and this option should have been cleared when the most private preset was selected).

Another problem with this interface was the check boxes associated with some of the options, such as the ”other allowed purposes”. It was not clear to some users whether checking the box allowed that purpose, or disallowed it. A better design might be to use ”yes” and ”no” buttons, as was done for other parameters on this page (e.g., ”require tracking”). The ”retention period” parameter was also a source of confusion. Some users interpreted the field as a duration to be entered (i.e., number of months, days, and years) rather than a date. The label should probably be changed to something like ”retention date” or ”retain until” to make the parameter easier to understand.

Figure 12.14 shows the interface screen that was used to collect contact information, such as name, address, and telephone number. In this interface the participants were much more likely to associate the privacy preferences with the personal information, with only 16% not able to make the connection. Thus, people are able to understand protecting the privacy of their contact information, where they had problems understanding this for the search parameters. However, it was not clear to users why this privacy preference information was being collected again, and many commented that the preferences should only be gathered once for all the information collected.

The results for probe questions about some of the key privacy-protection terms were encouraging. For the label ”intended purposes”, 96% of the participants were able to accurately describe what was meant, although the rollover help was essential. Similarly, 92% were able to describe ”other allowed purposes”. The term ”retention period” was understood by 94% of the users, while ”allow international transfer” was understood by 98% of the users. There were some problem areas with the terminology, however, that the rollover help did not successfully prevent. Only 84% of the users understood what was meant by ”require tracking”, and 84% understood what was meant by ”require data controls”. Again, all of these terms should be clarified so the rollover help information is not required for understanding.

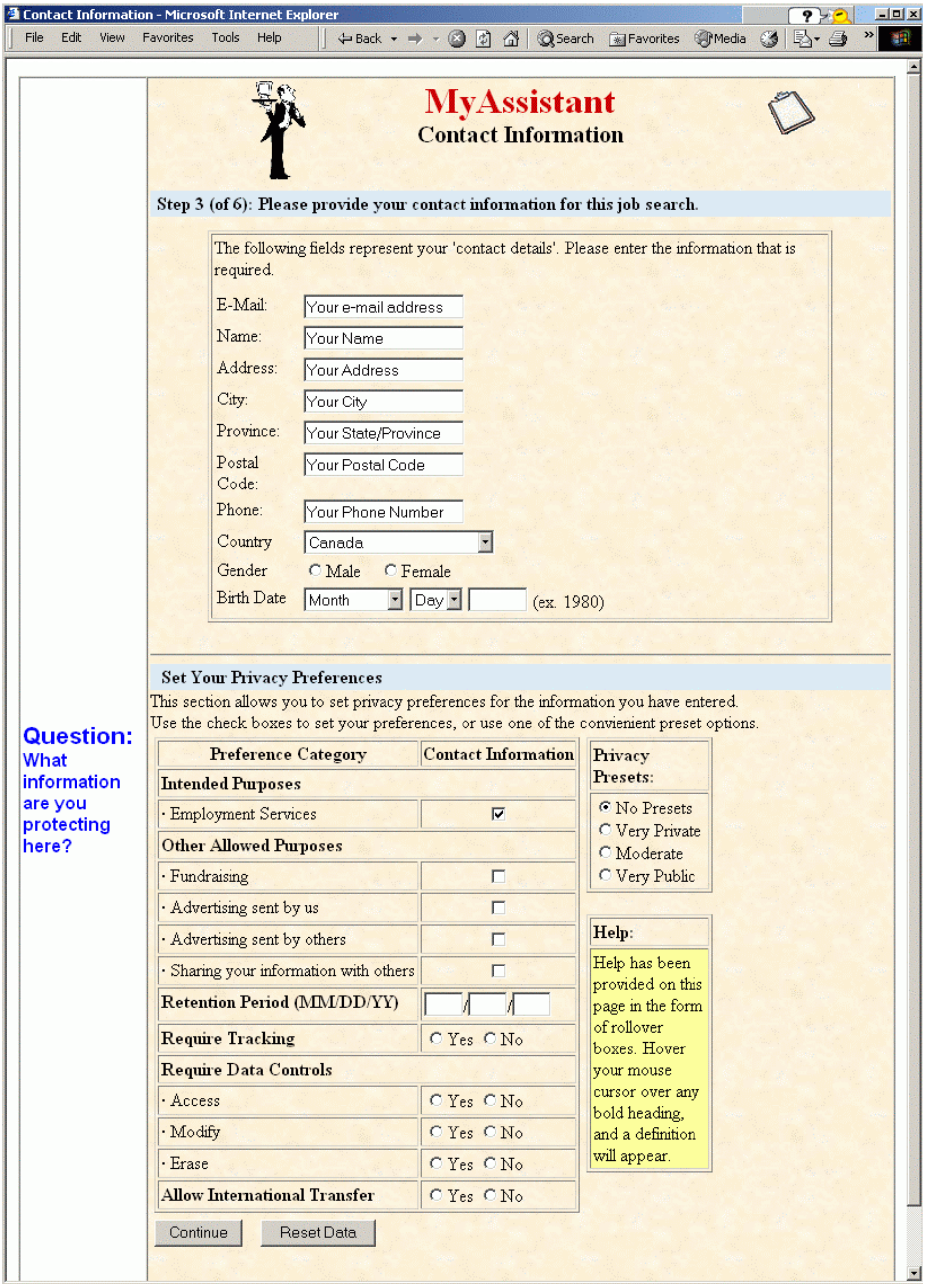
It is interesting to note that some users were aware of data-gathering issues and spontaneously commented that the interface should not ask for information about gender and birth date during a job search.

Figure 12.15 shows the interface screen for entering resume (CV) information, such as education and job history. This screen was fairly easy to understand, with 86% of the users stating that they knew what information to enter in the form. It was not clear, however, how much detail should be provided, such as dates of employment, how far to go back, etc.

The results for the understanding of the privacy preference controls were much better for this interface screen. Now 90% of the users understood what information was being protected by using the preferences. It is not clear if this increase in understanding was due to the fact that this was the third time users completed the privacy protection forms or that this is clearly personal information worthy of protection. One possible change to the final PISA implementation might be to move this interface to the first in the sequence so the need for privacy protection preferences might be clearer. Another, preferable, option is to only ask for the privacy preferences once for all the personal information that has been entered.

A unique interface concept that was tested in this screen was a ”Just-In-Time Click-Through Agreement” (JITCTA; see [PK03] for a description). In this example the JITCTA was a blue box that appeared whenever the user started to enter information into the ”trade union membership” field, and the pop-up message is shown in Figure 12.16. The intent was to follow the principles of the EU Data Directive and warn users that they were entering particularly sensitive information, and to obtain explicit consent for the processing of that information. Unfortunately, in this interface test the implementation of the JITCTAs was flawed such that the pop-up message sometimes failed to appear. When it did appear, for 48% of the users, the reactions were mixed. Some users appreciated the pop-up information, while others found them to be annoying or they ignored them completely think they were advertisements. A few users justified this by stating their belief that ”all pop-up windows are advertisements”. Other comments were that the message about not storing information was confusing (i.e., it was not clear if the information is stored if the user clicks ”I Agree”), and some users doubted the believability of the statement.

The implementation of the JITCTA in this prototype did not strongly associate the pop-up window with the data that was being entered (trade union membership). This was because the window appeared in its own frame, complete with a title bar and widgets to minimize


Contact Information - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media

Question: What information are you protecting here?

MyAssistant
Contact Information

Step 3 (of 6): Please provide your contact information for this job search.

The following fields represent your 'contact details'. Please enter the information that is required.

E-Mail:	Your e-mail address
Name:	Your Name
Address:	Your Address
City:	Your City
Province:	Your State/Province
Postal Code:	Your Postal Code
Phone:	Your Phone Number
Country	Canada
Gender	○ Male ○ Female
Birth Date	Month Day (ex. 1980)

Set Your Privacy Preferences

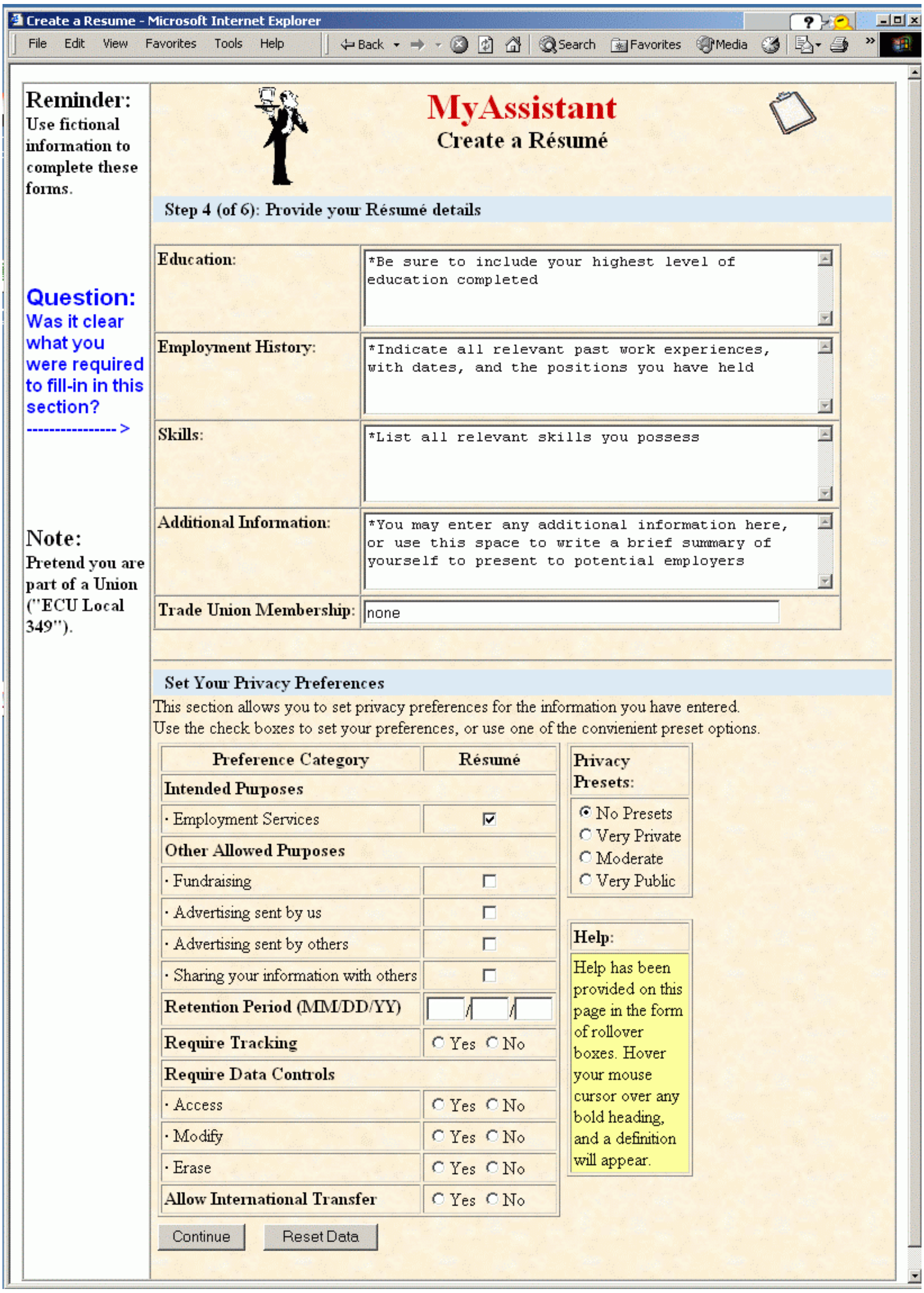
This section allows you to set privacy preferences for the information you have entered.
Use the check boxes to set your preferences, or use one of the convienient preset options.

Preference Category	Contact Information
Intended Purposes	
· Employment Services	☑
Other Allowed Purposes	
· Fundraising	□
· Advertising sent by us	□
· Advertising sent by others	□
· Sharing your information with others	□
Retention Period (MM/DD/YY)	/ /
Require Tracking	○ Yes ○ No
Require Data Controls	
· Access	○ Yes ○ No
· Modify	○ Yes ○ No
· Erase	○ Yes ○ No
Allow International Transfer	○ Yes ○ No

Privacy Presets:
● No Presets
○ Very Private
○ Moderate
○ Very Public

Help:
Help has been provided on this page in the form of rollover boxes. Hover your mouse cursor over any bold heading, and a definition will appear.

Figure 12.14: The interface screen for entering contact information.


Create a Resume - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media

Reminder: Use fictional information to complete these forms.

Question: Was it clear what you were required to fill-in in this section? ---------------- >

Note: Pretend you are part of a Union ("ECU Local 349").

MyAssistant

Create a Résumé

Step 4 (of 6): Provide your Résumé details

Education:	*Be sure to include your highest level of education completed
Employment History:	*Indicate all relevant past work experiences, with dates, and the positions you have held
Skills:	*List all relevant skills you possess
Additional Information:	*You may enter any additional information here, or use this space to write a brief summary of yourself to present to potential employers
Trade Union Membership:	none

Set Your Privacy Preferences

This section allows you to set privacy preferences for the information you have entered.
Use the check boxes to set your preferences, or use one of the convienient preset options.

Preference Category	Résumé
Intended Purposes	
· Employment Services	☑
Other Allowed Purposes	
· Fundraising	□
· Advertising sent by us	□
· Advertising sent by others	□
· Sharing your information with others	□
Retention Period (MM/DD/YY)	/ /
Require Tracking	○ Yes ○ No
Require Data Controls	
· Access	○ Yes ○ No
· Modify	○ Yes ○ No
· Erase	○ Yes ○ No
Allow International Transfer	○ Yes ○ No

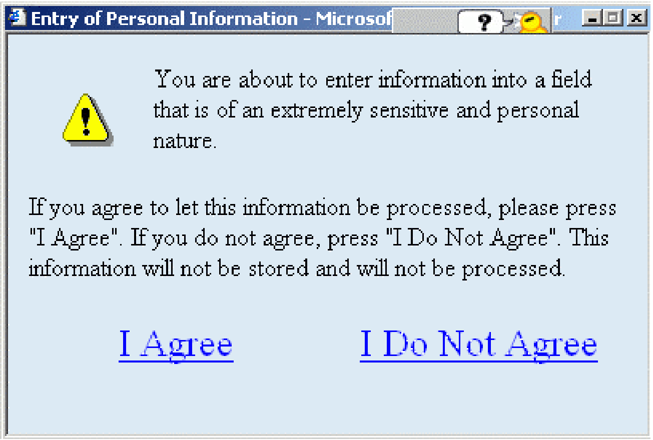
Privacy Presets:

● No Presets
○ Very Private
○ Moderate
○ Very Public

Help:

Help has been provided on this page in the form of rollover boxes. Hover your mouse cursor over any bold heading, and a definition will appear.

Figure 12.15: The interface for collecting resume information.



You are about to enter information into a field that is of an extremely sensitive and personal nature.

If you agree to let this information be processed, please press "I Agree". If you do not agree, press "I Do Not Agree". This information will not be stored and will not be processed.

Figure 12.16: A Just-In-Time Click-Through Agreement (JITCTA).

and close the window, and it had a distinct background colour. In addition, the window appeared fairly far from the input field that triggered it. These factors may have led some users to ignore the window. A better implementation might be to integrate the pop-up into the interface screen, perhaps using the ”iframes” feature available in modern browsers.

It is interesting to note that a few users recognized the sensitive nature of sharing trade union membership with potential employers, and appreciated the extra warning when entering this information.

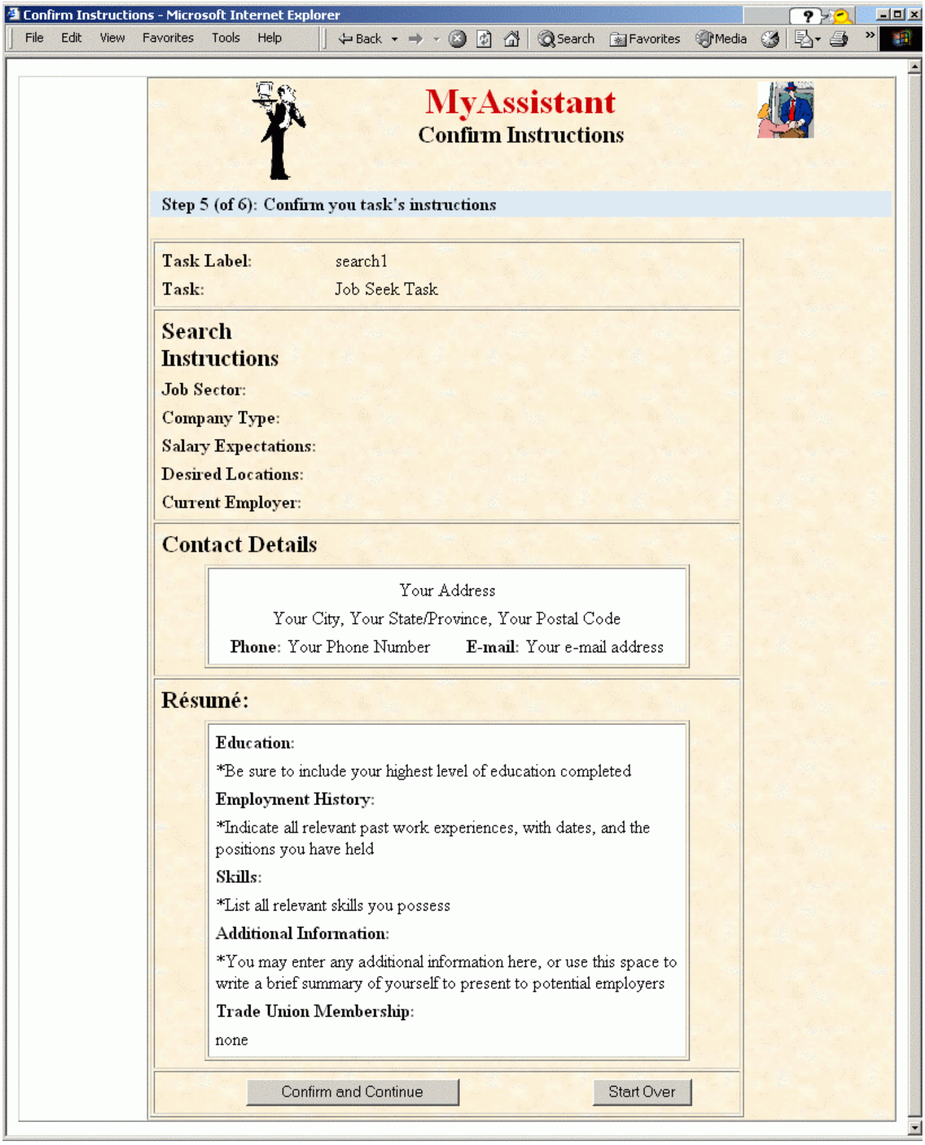
Figure 12.17 shows an interface screen that was intended to confirm all the information that the users had entered. Most of the participants greatly appreciated this feature, although some commented that the formatting of the information was not what they expected and they would have preferred to have more control.

Figure 12.18 shows the next interface screen that the users encountered. For this usability test the ”Start Over” button was a simple function that erased all the user information in all the previous screens. This was very frustrating for the one user who pressed this button, and any proper implementation should preserve the user information from the previous screens so that the users only need edit the information they wish to change.

A JITCTA was also used in this interface screen. Here the pop-up window appeared when the user pressed the ”Launch Task” button, and the message asked for final confirmation before the personal information was processed. This JITCTA appeared reliably and probing about the purpose revealed that 94% of the users understood the intended purpose. However, some of the participants felt that this pop-up was redundant and unnecessary since they were already agreeing to the use of personal information by using the MyAssistant service.

Figure 12.19 shows the final interface screen that confirmed that the search task had been started. The probe question associated with this screen gathered a usability rating and, as already mentioned, the average rating was 4.7, which corresponds to a ”slightly easy” rating. The most common reasons given for providing low ratings were problems in understanding the privacy preferences terms, and the repetition of the preference input forms.

In summary, the results of the usability analysis indicated that the prototype worked fairly


Confirm Instructions - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media

MyAssistant
Confirm Instructions

Step 5 (of 6): Confirm you task's instructions

Task Label:	search1
Task:	Job Seek Task

Search Instructions

Job Sector:

Company Type:

Salary Expectations:

Desired Locations:

Current Employer:

Contact Details

Your Address
Your City, Your State/Province, Your Postal Code
Phone: Your Phone Number **E-mail:** Your e-mail address

Résumé:

Education:

*Be sure to include your highest level of education completed

Employment History:

*Indicate all relevant past work experiences, with dates, and the positions you have held

Skills:

*List all relevant skills you possess

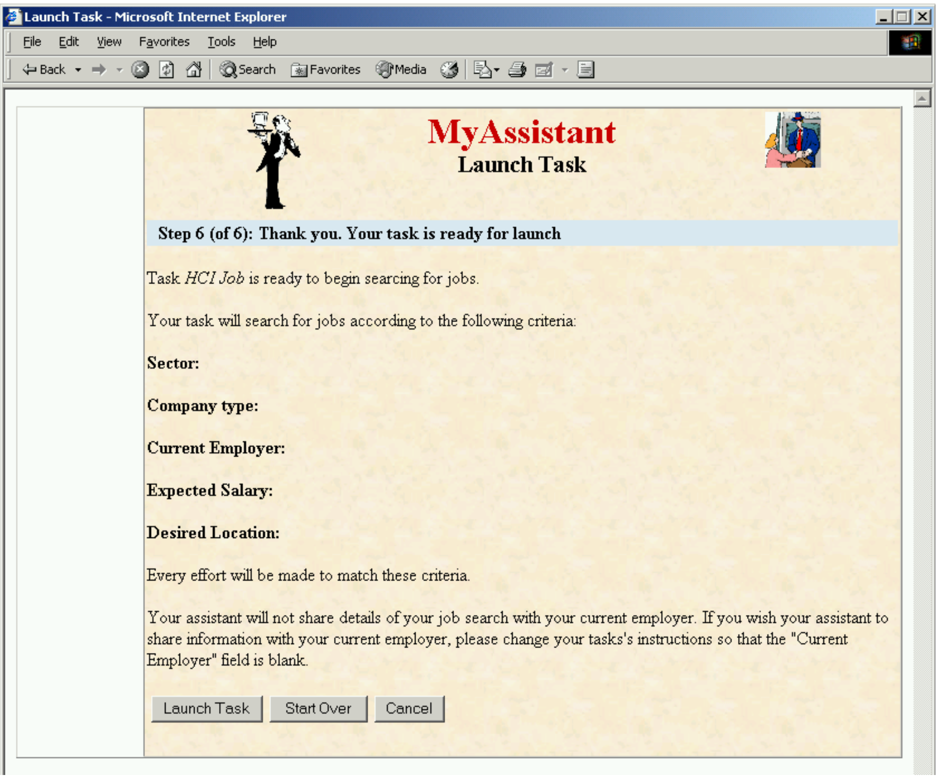
Additional Information:

*You may enter any additional information here, or use this space to write a brief summary of yourself to present to potential employers

Trade Union Membership:

none

Figure 12.17: An interface screen to confirm the information that was entered.


Launch Task - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media

MyAssistant
Launch Task

Step 6 (of 6): Thank you. Your task is ready for launch

Task *HCI Job* is ready to begin searching for jobs.

Your task will search for jobs according to the following criteria:

Sector:

Company type:

Current Employer:

Expected Salary:

Desired Location:

Every effort will be made to match these criteria.

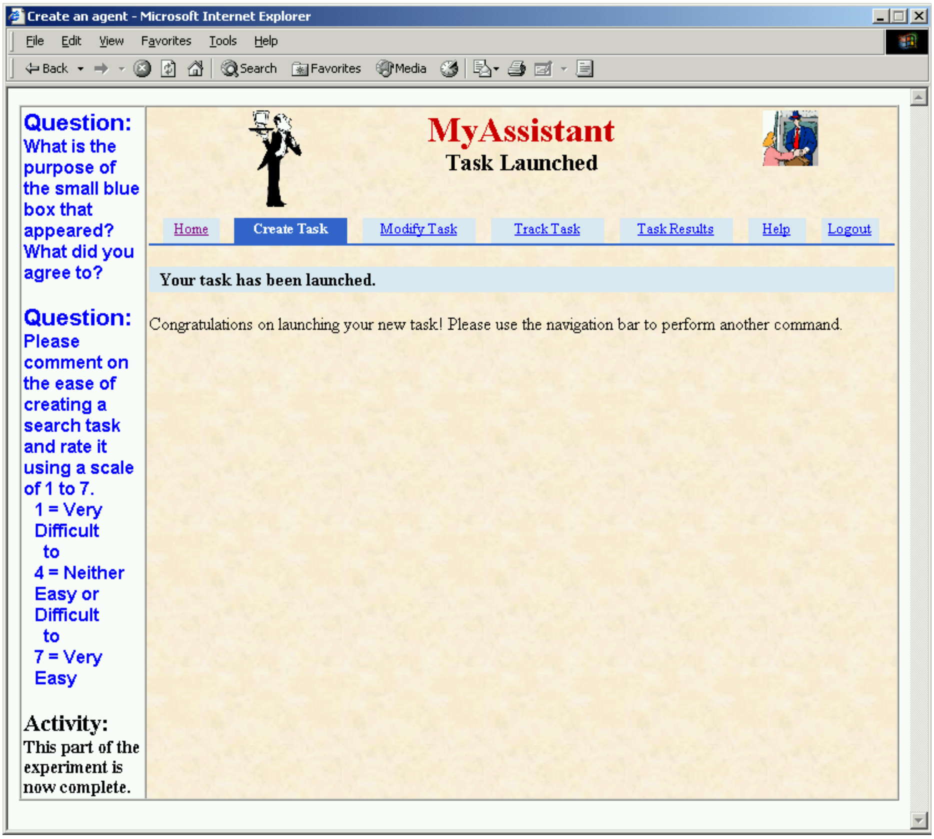
Your assistant will not share details of your job search with your current employer. If you wish your assistant to share information with your current employer, please change your tasks's instructions so that the "Current Employer" field is blank.

Figure 12.18: The interface screen used to launch the search task.

well and was reasonably easy to navigate, but it had poor visual appeal. Users generally understood the concept of a personal assistant who could provide services, and most of them understood the major functions (create, modify, track, get results). The most notable problem was that users had trouble understanding what information was being protected when they completed the privacy preferences portion of the data entry screens. Some of this is due to the ordering of the screens where the search instructions (type of job, salary, etc.) were completed first. It was not clear to users why they would need to protect this information. Users' understanding of the privacy preference features increased as they used the interface, especially on the third data input screen where they entered their contact information (name, address, etc.). The users had to rely heavily on rollover help information to understand the privacy preference terms, such as "retention period" and "require tracking", and they were frustrated by having to set their preferences three times.

Trustability Questionnaire Results

Data was also collected from a questionnaire given to the participants at the end of the session. This questionnaire contained some questions about the usability of the prototype interface, and these have already been discussed. The questionnaire also included some questions related to the trustability of the MyAssistant service, relative to other services on the Internet. The most notable finding was that, whereas only 54% of participants were willing to send personal information on the Internet at large, 84% would provide their


Create an agent - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media

Question: What is the purpose of the small blue box that appeared? What did you agree to?

Question: Please comment on the ease of creating a search task and rate it using a scale of 1 to 7.
1 = Very Difficult
to
4 = Neither Easy or Difficult
to
7 = Very Easy

Activity: This part of the experiment is now complete.

MyAssistant
Task Launched

Home
Create Task
Modify Task
Track Task
Task Results
Help
Logout

Your task has been launched.

Congratulations on launching your new task! Please use the navigation bar to perform another command.

Figure 12.19: The final interface screen, confirming launch of the task (agent).

resume to the prototype, 80% would provide their desired salary, and 70% would provide name, address, and phone number. A similar find was that, whereas only 34% thought that Internet services at large acted in their best interest, 64% felt that the prototype service would act in their best interest.

Another interesting finding was that 66% of the users agreed that they would ”feel comfortable depending on the privacy provided by MyAssistant”. On the other hand, only 48% of the people characterized the MyAssistant service as ”trustworthy”, with 44% of the users being ”undecided”. Overall, the results from the trustability questionnaire were encouraging because they suggest that the design features that were supposed to increase trustability had some positive effect. These results should be interpreted cautiously, however, because they were not uniformly strong and users may have answered the questions in a way that they thought would please the experimenter. Also, opinions on a questionnaire often do not translate into actual behaviour.

Summary of Results

This usability test of the PISA Interface Prototype has led to a number of results, which are summarized in Table 12.10.

Table 12.10: Summary of Notable Results.

1.	users were able to complete or understand an average of 17.5 out of a possible 27 (65%) major concepts or functions in the interface
2.	the average ease-of-use rating was 4.7 on a 7-point scale, corresponding to a rating of ”slightly easy”
3.	the average ease-of-navigation ratings was 3.64 on a 5-point scale, corresponding to a slightly positive rating
4.	the average rating of proper operation of the prototype was 3.62 on a 5-point scale, corresponding to a slightly positive rating
5.	30% of the users attempted to login without registering first
6.	38% of the users did not like the graphics, describing them as copied from another site. Others thought the graphics looked ”professional” and ”cute”
7.	42% of the users did not like the colours in the interface, describing them as ”bland” and ”boring”. Others described them as ”soothing” or ”relaxing”
8.	50% of the users thought that the visual appearance was unlike other sites that they think highly of
9.	88% of the users liked the fonts in the interface because they were familiar
10.	92% of the users understood the concept of a personal assistant created for them
11.	some users questioned the need to login as a second step after a successful registration
12.	users were generally able to understand the major functions presented in the prototype, such as ”create task”, ”modify task”, etc.
13.	users were sometimes confused by the term ”task” when used to describe a set of instructions given to the assistant
14.	50% of the users could not understand why they needed to name a task being created
15.	users frequently did not understand the terms ”job sector” or ”company type”
16.	users sometimes did not know what units to use for input fields like ”desired salary” or ”desired location”
17.	users sometimes did not understand why they were asked to provide their ”current employer”
18.	64% of users failed to associated the privacy preferences with the job search parameters
19.	users often did not feel that job search parameters required privacy protections
20.	users often had difficulty understanding the terms used in the privacy preference interface. Rollover help did provide assistance, but it should not have been necessary
21.	many of the users failed to notice or use the preset buttons available for setting privacy preferences
22.	the function of the check boxes in the privacy preferences was often unclear, such as whether checking ”other allowed purposes” had the effect of allowing the purpose or not
23.	users were often unclear about how to complete the ”retention period” field
24.	understanding of the privacy preference screens increased when contact information and resume information were entered
25.	a Just-In-Time Click-Through Agreement (JITCTA) to seek explicit confirmation when processing ”union membership” information failed to appear reliably, but when it did reaction was mixed. Some users appreciated the warning about the sensitive information, while others ignored the message completely.
26.	a JITCTA to seek final confirmation before the task agent is launched also had mixed reactions, with some users finding the pop-up box redundant and annoying
27.	results from the trust questionnaire revealed that, whereas only 54% of participants were willing to send personal information on the Internet at large, 84% would provide their resume to the prototype, 80% would provide their desired salary, and 70% would provide name, address, and phone number
28.	whereas only 34% thought that Internet services at large acted in their best interest, 64% felt that the prototype service would act in their best interest.
29.	66% of the users agreed that they would ”feel comfortable depending on the privacy provided by [the prototype]”
30.	only 48% of the people characterized the prototype service as ”trustworthy”, with 44% of the users being ”undecided”.

12.2.5 Recommendations

The usability testing of the prototype PISA interface revealed a number of strengths and weaknesses. Table 12.11 contains a set of recommendations resulting from the usability test.

12.2.6 Conclusions

It should be remembered that the participants were not representative of the general population, or of the likely user group for a real PISA service. Instead, the participants were young, Canadian college students, who may have very different attitudes towards WWW site design, Internet services, job searching, or privacy preferences than older adults or European citizens. This important limitation should be kept in mind when drawing conclusions.

The prototype PISA interface contained a number of design features that were attempts to enhance the usability and trustability. One notable design feature was the use of a certificate-like background, pastel colours, and colourful icons. Reactions to this design were mixed, and it may we worthwhile to experiment with a brighter, more professional looking visual design. Another design feature was the metaphor of a personal assistant who could be given tasks, and users were able to understand and use this metaphor. Users also reacted well to the functional organization of the interface based on user tasks.

Major problems were seen in the understanding of the terms used for controlling the privacy preferences. The feature of rollover help worked well, although it should not have to be relied on so heavily to understand the terms in the interface. It is interesting to note that [CAG02] also reported problems with users' understanding of privacy terminology, and one suggestion was to create a wizard-like interface to support understanding.

The design element of using JITCTAs to seek confirmation of actions, such as the processing of sensitive information, worked partially, but a better implementation of the pop-up menus is needed.

Overall, the purpose of the testing was to determine if the design concepts in the PISA interface prototype were successful in constructing an agent technology that (1) users can use, (2) users can understand, and (3) users can trust with secure and private information. The results indicate that users can use the major features of the interface, such as creating a job-searching agent. However, some of the specific features, such as controlling specific privacy preference parameters, are in need of more attention. Concerning understanding, the results clearly indicate that users have difficulty understanding the privacy preference terms used in the interface, and this is the most important characteristic to improve. Finally, users did find the service to be trustable, although it is clear that, with the problems in understanding the interface, the maximum possible trust was not created.

Table 12.11: Recommendations for Improving the PISA Interface.

1.	improve on the terms used throughout the interface. Small usability tests can be conducted on the terms to ensure that potential users share the meaning that was intended by the developers.
2.	consider using more polished and professional looking graphical images
3.	consider a brighter, more attractive color scheme
4.	keep to standard fonts, such as Times Roman
5.	improve the registration process so that it is clear that users must register before they can use the service
6.	integrate registration with login so that users are entered into the system automatically after they register
7.	continue to highlight when information is transferred securely
8.	the mental model of a personal assistant being created for the user is working, and users understand that the assistant can be given tasks
9.	look for a word other than ”task” to label the process of creating a job for a personal assistant to do
10.	move the step of naming a task to the end of the creation process, and make it clear why users must name their task (so they can differentiate them later)
11.	remove or replace the terms ”job sector” and ”company types”
12.	use selection menus instead of text boxes whenever a list of items is appropriate
13.	make units of measurement clear, such as for ”desired salary” or ”desire location”
14.	make the reason of entering the current employer clear at the time that it is entered, and make sharing the resume with the current employer a separate and explicit step
15.	integrate the privacy preferences screen into a single screen that is completed once, after all the personal information is collected
16.	collect the contact information first, since the need for privacy protection is clearest with this information
17.	continue to use rollover help, but make it more prominent and easier to discover
18.	change preference parameters so they are all yes/no choices, not check boxes whose function is unclear
19.	make the privacy presets more prominent and a clear alternative to customized choices for each parameter
20.	make it clear that using a privacy preset will erase any custom parameter settings
21.	fix the implementation of the presets in the interface code
22.	fix ”retention period” so it is either a duration for keeping the information or a date
23.	change the JITCTAs to be integrated into the interface screens instead of being distant, separate interface windows
24.	fix the implementation of the JITCTAs for the union membership field
25.	improve the wording for the sensitive information JITCTA
26.	fix the ”start over” feature so that previously entered information is saved and can be edited

Chapter 13

Conclusions and future outlook

Privacy by design One of the objectives of the PISA consortium is to support the design and evaluation of Information and Telecommunication systems so that the informational privacy of the user is protected in all kinds of processes. Therefore we studied the case of how to incorporate privacy-protecting features into an ISA[1]. The result is this handbook on Privacy and Privacy-Enhancing Technologies (PETs), including the references to all the PISA reports and publications. The purpose of this handbook is to identify what privacy laws and rules apply to ISA's, and what measures need to be implemented to create a Privacy Incorporated Software Agent (PISA).

Has the PISA consortium reached its goal? We claim that we have done so. We have developed:

- a new privacy threat analysis,
- a new cryptographic algorithm,
- privacy ontologies,
- a system for agent-based network privacy,
- a modelling approach for assessing the scalability for privacy components and architectures,
- an approach for building trustable human computer interfaces, and
- an overall architecture for privacy in agent-based applications.

There is a major limitation, however. We assumed that parts of the environment are trustworthy and realistically this is most often not the case. Further research is needed to tackle this difficult issue with new cryptographic solutions for untrusted mobile platforms, hardware approaches and privacy support tools as reputation schemes.

[1] See PISA contract no IST-2000-26038.

New legal privacy issues Legal research shows that although an ISA cannot be held liable for violations itself, as it lacks a legal personhood[2], it still must have built-in, privacy-protection features. This means not only in the expression of the privacy preferences of the user of the ISA, but also in the receiving of the Personal Data (PD) of others and the handling of privacy sensitive data. The PD of the user and the received PD of others must be managed and processed according to the law. Moreover, the ISA is a tool that may be used by a data subject, a controller, a processor, or a provider. Since these ISA users are informed of the capabilities and risks of the ISA, they are liable for violations with respect to privacy and other legislation. This implies that the ISA does not perform a role of his own but acts on behalf of its user. As we have seen, the developer is liable for privacy infringing code and, from a privacy point-of-view, any malfunctioning ISA intruding on the privacy of others or damaging the privacy of its user.

Research results and challenges Various interesting research challenges have been outlined, covering the whole spectrum of user and back-office systems, infrastructure and non-technical issues.

Security Protecting mobile software agents is difficult, While some solutions for problems in the software agent domain have been demonstrated in this work and elsewhere, many of the problems have not been solved. Critical for software systems is that execution does not occur in a trustworthy environment. The major open problems have been described in this handbook and some solutions are proposed. Many threats have been identified and, based on the system model and assumptions, several of these threats have been chosen to provide a solution for. The system model and assumptions made in Chapter 5 of this handbook are different than the model used in the demonstration and the remaining chapters. We did this because in the research we took a long-term perspective for providing privacy and security such that when the time has come to weaken the assumptions boundaries of security and privacy are already defined in the new situation. In the short term it is not feasible to assume that anagent platform can be trusted completely.

Instead of only looking at small problems and solving each of them individually, we examined the broader picture where the idea is that an agent owns a task and must be protected. Descriptions of a new model of agents and protecting tasks led to a new definition of perfect secrecy. Several dilemmas were derived such that the theoretical boundaries become clear. Since not all problems have been solved, much research is still necessary in this area. Especially important is the area of computing with encrypted data in an agent environment and making decisions on encrypted data. Our solutions depend on the possibility of computing with encrypted data. Currently this approach is only theoretical. Also there is an important initiative of pushing trusted computing platforms by the leading industries[3].

Datamining The overall conclusion is that data mining indeed causes privacy threats. Especially the linking of data, which activity may give rise to the identification of entirely new facts about a data subject or the creation of richer, more meaningful information, is a threat that is particular to data mining.

Data mining for privacy protection can profit from research in intrusion detection in two important areas: designing ways to train the models used and implementing a system that can be used in an agent-based environment. A web mining system for analysing and classifying web sites is described.

[2] See Solum, Legal Personhood for Artificial Intelligences (1992) 70 North Carolina Law review pp 1231-1287.

[3] TCPA, http://www.trustedcomputing.org

Five different threats are identified that can lead to the linking of identity to behavioural data. Several procedures can reduce these risks e.g. the third-party behaviour storage model and the dual-key procedure. These mechanisms are illustrated for a library setting. Lastly, the use of so-called virtual panels is discussed.

Traditional data mining is much less a threat to privacy than often thought. The aim of data mining is to discover local theories starting with specific data and to produce general patterns and rules. While these rules can predict certain characteristics of, most often, anonymous data records they do not likely lead to new facts that can be threatening to the privacy of individuals.

Further, data mining can put to use in defence of privacy in systems like Parabots that can analyse and classify web sites with regard to their privacy policies. This reputation knowledge can be used as a Privacy Protection Service.

Human computer interfaces For the first time design guidances for privacy-enhancing technologies from a human factors point of view are introduced. This work specified what must be included in human-computer interfaces to satisfy the spirit of European privacy legislation and principles, and satisfy the privacy needs of the users ("usable compliance").

The current work has focused on European privacy legislation and, although the resulting principles, requirements, and solutions are general, one of the challenges that remains is to ensure that the knowledge is equally applicable in other legislative settings, such as Canada, and in areas operating in a self-regulatory fashion (e.g., the USA). Even in regulated environments, the privacy legislation and guidelines will change and evolve, and thus the human interface guidelines will also have to be dynamic.

Privacy-enhancing technologies are also evolving and changing, and this will have an effect on the types of solutions that are available, and also the privacy needs and expectations of the users. For example, the P3P protocol, if implemented widely, may have a profound effect on the privacy domain by bringing privacy issues to the attention of millions of Internet users, and hopefully providing an easy-to-use privacy control interface

Research on privacy and human computer interfaces included:

- an examination of what it means to build a trustable interface to agents, and what would be involved in trusting agents with personal information,
- this resulted in a set of design guidelines for building trustable agent systems,
- an examination of privacy legislation and principles to determine HCI requirements for "usable compliance",
- this resulted in a set of requirements and design solutions to satisfy the spirit of the privacy directives,
- an experimental evaluation of how to effectively facilitate understanding of privacy concepts and terms,
- suggestions on how to maximise the trustable nature of the interface.

Networks In research of the privacy issues associated with networks and software agents, the question we tackled was: What is needed in order to make network communication for agents systems confidential?

The main result of the PISA research deliverables include was an onion routing protocol approach for network privacy for agents.

The research team working on the network privacy also have produced other research results as follow-on work. These include:

- a protocol design for network privacy using IPSEC,
- a protocol to make pay TV services more privacy preserving,
- secure routing and messaging for ad-hoc networking, especially useful for scatternets in short range wireless networks,
- a privacy preserving electronic currency protocol,
- network privacy applied reputation management systems as a partial solution for network-based exposures in network privacy systems.

We implemented and tested a prototype of the agent-based onion routing protocol for operation in the JADE agent platform. It was added as part of the demonstration software.

Scalability In the research on the scalability for privacy incorporated agent systems, the question that was asked was: What techniques can we use to assess the performance of security and privacy technologies for agent systems with any thousands of software agents in advance of their implementation?

We built JADE applications that use the cryptographic primitives and protocols that are commonly used to implement security and privacy primitives, tested the performance of the overall system in agent environments ranging from one to several thousand agents, and documented all results in terms of timing and processor load.

We analysed the expected performance of different possible PISA system configurations to determine the most effective system arrangement. Based upon our tests of subcomponents for security and privacy and modelled systems, our conclusions include the following results:

- Detailed comparisons of distributed and client-server approaches, showing the distributed approach is most efficient.
- We also offer guidelines for system design with different sizes of agents.
- Scalability improves with the use of a single agent container per machine.
- Increasing the number of user agents has an effect on the average processing time for the request-reply messages in the user agent side, but no effect on the average time for processing the request-reply messages in the job agent side. The average processing time for the request-reply messages on the user agent side increases linearly with an increase of the number of user agents.
- A public key system would have a severe impact on the system scalability if some operations such as digital signatures and decryption were used in the server side of the PISA demonstrator. But certificate-based authentication, the symmetric key cryptography, and integrity function only have little effect on the system scalability.
- Identity protection normally has only a small impact on system scalability, whatever it uses the anonymous or pseudonymous approaches.
- A self-control mechanism is necessary to manage performance of systems like onion routing for network privacy. The scalability for the onion routing network would improve with distributing only one onion routing agent to one machine and balancing the workload of each onion routing agent.

Privacy audits The Privacy Audit Framework is a work plan a privacy auditor can use in order to evaluate the compliance of a processing system for personal data within the applicable privacy legislation. The scope has to be defined for all audits. A very important part of the scope is the exact definition of the object of investigation. The object can be defined as being a product or a process.

When the processing of personal data via a product audit is being evaluated, the auditor evaluates the data the controller processes by just examining the contents of the database. A process audit is required to evaluated the mode of operation within the organisation of the controller if the auditor must also establish the circumstances used to collect that data. For systems based on intelligent software agents this means a need for designing and implementing new features to support the audit process.

Evaluation Currently, the Common Criteria contains FPR Class: Privacy. This is the only class relating to the evaluation of privacy compliance. This class specifies and gives evaluation criteria on a small, although not unimportant, subset of privacy regulations. The issues are: anonymity, pseudonymity, unlinkability and unobservability. Compliance to this class does not, in any way, guarantee that the information system is compliant with the Data Protection Directive.

The evaluation criteria related to the FPR Class are best compared to a product audit. Even if one or more of the options are correctly implemented, it does not guarantee that the personal data processed has been lawfully collected. Compliance auditing must mean, according to the Data Protection Authorities, carrying out a process audit whereby the flow of personal data within the organisation is checked, including the possible processor, from the moment the data are to be collected up until their destruction.

The conclusion is that the Common Criteria, although their methodologies are correct, cannot at this moment in time be used as the framework for a Privacy Audit. Privacy obligations require that the system design process and security standards incorporate privacy requirements.

Activities are being considered to define a number of extensions to the FPR Class. Current thinking is to enhance this class with 11 Fair Information Practices. These practices are very easily matched with the nine principles of the Privacy Audit Framework. The Privacy Audit shall be defined as an international standard if this extension to the Common Criteria is achieved and if, at the same time, the methodology is changed into a proper process audit.

Bibliography

[601] P. 6. Can we be persuaded to become PET-lovers? Presented at the OECD Forum Session on Privacy Enhancing Technologies, October 8 2001.

[ABN] D. Abdu and O. Bar-Ner. Software Agents: A general overview. http://t2.technion.ac.il/∼s3180501/agent.html. (Unfortunately, this link does not work anymore).

[AF90] Martín Abadi and Joan Feigenbaum. Secure Circuit Evaluation: A Protocol Based on Hiding Information from an Oracle. *Journal of Cryptology*, 2(1):1–12, 1990.

[AF96] M. Abe and E. Fujisaki. How to Date Blind Signatures. In *Advances in Cryptology – ASIACRYPT '96*, pages 244–251, 1996.

[Arc] Cheskin Research & Studio Archetype/Sapient. eCommerce Trust Study. http://www.cheskin.com/think/studies/ecomtrust.html.

[Ban00] D. Banisar. Privacy and human rights. Electronic Privacy Information Center, 2000. Washington/London.

[BC01] T. Bickmore and J. Cassell. Relational agents: A model and implementation of building user trust. In *Proceedings of SIGCHI'01*, pages 396–403, Seattle, WA, U.S.A., March 31 – April 4 2001.

[BFK00] O. Berthold, H. Federrath, and S. Kopsell. WebMIXes: A System for Anonymous and Unobservable Internet Access. In *Proceedings of the workshop on design issues in anonymity and unobservability*, pages 101–115, Berkeley, California, 2000. California International Computer Science Institute.

[BGS01] A. Back, I. Goldberg, and A. Shostack. Freedom 2.1 Security Issues and Analysis, May 2001.

[Bic97] P. Bickford. Human interface online: A question of trust. http://developer.iplanet.com/viewsource/bickford_trust.html, 1997.

[BMW98] Ingrid Biehl, Bernd Meyer, and Susanne Wetzel. Ensuring the Integrity of Agent-Based Computations by Short Proofs. *Lecture Notes in Computer Science*, 1477:183–194, 1998.

[Bor96] J. Borking. Der Identity-Protector. *Datenschutz und Datensicherheit*, 20(11):654–658, 1996.

[BPS00] O. Berthold, A. Pfitzman, and R. Standke. The disadvantages of free MIX routes and how to overcome them. In *Proceedings of the workshop on design issues in anonymity and unobservability*, pages 27–42, Berkeley, California, 2000. California International Computer Science Institute.

[BR01] J.J. Borking and C.D. Raab. ”Laws, Pets, and Other Technologies For Privacy Protection”. *Journal of Informatics, Law and Technology (JILT)*, January 2001.

[Bra94] Stefan Brands. Untraceable Off-line Cash in Wallet with Observers. In Douglas R. Stinson, editor, *Advances in Cryptology – CRYPTO ' 93*, volume 773 of *Lecture Notes in Computer Science*, pages 302–318. International Association for Cryptologic Research, Springer-Verlag, Berlin Germany, 1994.

[BS] Code of Practice for the Risk Analysis and Management Method (British Standards 7799).

[BSG00] P. Boucher, A. Shostack, and I. Goldberg. Freedom Systems 2.0, 2000.

[BvES99] J.J. Borking, B.M.A. van Eck, and P. Siepel. Intelligent Software Agents and Privacy. A&V study #13, The Hague, 1999.

[Byg01] L.A. Bygrave. Electronic Agents and Privacy: A Cyberspace Odyssey 2001. *International Journal of Law and Information Technology*, 9(3):275–294, 2001.

[CAG02] L.F. Cranor, M. Arjula, and P. Guduru. Use of a P3P User Agent by Early Adopters. In *Proceedings of Workshop on Privacy in the Electronic Society*, Washington D.C., November 21 2002.

[Can00] 2000 Canadian Personal Information Protection and Electronic Documents Act, 2000.

[Cav98] A. Cavoukian. Data mining: Staking a claim on your privacy. Available at: http://www.ipc.on.ca/english/pubpres/sum_pap/papers/datamine.htm, 1998. Information and Privacy Commissioner/Ontario.

[CC87] E.M. Comstock and E.A. Clemens. Perceptions of computer manuals: A view from the field. In *Proceedings of the Human Factors Society 31st Annual Meeting*, pages 139–143, 1987.

[CD99] CRISP-DM. Step-by-step data mining guide. Available at: http://www.crisp-dm.org/, 1999.

[Cen] Central Computers and Telecommunications Agency (CCTA). Information Security Handbook.

[CFN90] David Chaum, A. Fiat, and M. Naor. Untraceable Electronic Cash. In Shafi Goldwasser, editor, *Advances in Cryptology – CRYPTO ' 88*, volume 403 of *Lecture Notes in Computer Science*, pages 319–327, Santa Barbara, CA, USA, 1990. International Association for Cryptologic Research, Springer-Verlag, Berlin Germany.

[CH97] A. K. Caglayan and C. G. Harrison. *Agent Sourcebook: a complete guide to Desktop, Internet and Intranet Agents*. John Wiley & Sons; 1st edition, 1997. ISBN: 0471153273.

[Cha81a] D. Chaum. Untraceable Electronic Mail, Return Address, and Digital Pseudonyms. *Communications of the ACM*, 24(2):84–88, 1981.

[Cha81b] David L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *CACM*, 24(2):84–88, February 1981.

[Cha82] David Chaum. Blind Signatures for Untraceable Payments. In David Chaum, Ronald L. Rivest, and Alan T. Sherman, editors, *Advances in Cryptology: Proceedings of Crypto 82*, pages 199–203. Plenum Press, New York and London, 1983, 23–25 August 1982.

[Cha92] David Chaum. Achieving Electronic Privacy. *Scientific American*, pages 96–101, August 1992.

[CKAC01] J. Camenisch, G. Karjoth, J. Algesheimer, and C. Cachin. Cryptographic security for mobile code. In *"Proceedings 2001 IEEE Symposium on Security and Privacy"*, pages 2–11. IEEE, 2001.

[Cla96] R. Clarke. Identification, Anonymity and Pseudonymity in Consumer Transactions: A Vital Systems Design and Public Policy Issue. Invited presentation to the Conference on 'Smart Cards: The Issues, Sydney, October 18 1996. Available at http://www.anu.edu.au/people/Roger.Clarke/DV/AnonPsPol.html.

[CPS94] J. L. Camenisch, J-M. Piveteau, and M. A. Stadler. Blind Signatures Based on the Discrete Logarithm Problem. In *Proc. Advances in Cryptology – EUROCRYPT '94*, pages 428–432, 1994.

[CRA99] L.F. Cranor, J. Reagle, and M.S. Ackerman. Beyond concern: Understanding net users' attitudes about online privacy. Technical Report TR 99.4.3, AT&T Labs-Research, 1999. http://www.research.att.com/library/trs/TRs/99/99.4/.

[CT98] Christian Collberg and Clark Thomborson. On the Limits of Software Watermarking. http://www.cs.auckland.ac.nz/∼collberg/Research/Publications/CollbergThomborson98e, 1998.

[CT99] Christian Collberg and Clark Thomborson. Software Watermarking: Models and Dynamic Embeddings. In *Conference Record of POPL '99: The 26th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, pages 311–324. ACM Press, 1999.

[CvdL02a] K. Cartrysse and J.C.A. van der Lubbe. An agent digital signature in an untrusted environment. In *"Proceedings of the second international workhop on security of mobile multiagent systems (SEMAS-2002)"*, 2002. "held at the first international conference on autonomous agents and multiagent systems (AAMAS2002)".

[CvdL02b] K. Cartrysse and J.C.A. van der Lubbe. Privacy protection software design. Deliverable 12 of the PISA-project, September 2002.

[CvdL04] K. Cartrysse and J.C.A. van der Lubbe. Mobile code security: an information theoretic approach, 2004. Submitted to IEEE Symposium on Information Theory.

[CvdLL02] K. Cartrysse, J.C.A. van der Lubbe, and R. Lachman. Mechanisms for secure data exchange and storage with PKI. Deliverable 14 of the PISA-project, July 2002.

[Dai00] W. Dai. Pipenet 1.1, 2000. Available at http://www.eskimo.com/∼weidai/pipenet.txt.

[EC95] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data. Official Journal of the European Communities, 1995.

[Enn98] Mari-Anne Ennor. Data mining, a controversy in consumer privacy. Available at: http://www.digisys.net/users/m_ae/Datamining_2.htm, 1998. Research project paper for Management Information Systems at City University.

[FGS96] W. Farmer, J. Guutman, and V. Swarup. "Security for mobile agents: Authentication and State Appraisal". In *"In proceedings of the 4th European Symposium on Research in Computer Science (ESORICS'96)"*, pages 118–130, September 1996.

[FL] M. Fernández-López. Overview Of Metholdologies For Building Ontologies. http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-18/4-fernandez.pdf.

[FM02] M. J. Freedman and R. Morris. Tarzan: A Peer-to-Peer Anonymizing Network Layer. In *Proceedings of the 9th ACM conference on Computer and communications security (CCS'02)*, pages 193–206, November 2002.

[Fri87] N. Frijda. *The Emotions*. Cambridge University Press, New York, 1987.

[FSW] P-A. Fouque, J. Stern, and G-J. Wackers. Cryptocomputing with rationals.

[GGMM97] E. Gabber, P. Gibbons, Y. Matias, and A. Mayer. How to make personalizedf web browsing simple, secure, and anonymous. In *Proceedings of Financial Cryptography 97*, LNCS 1318. Springer-Verlag, February 1997. http://www.bell-labs.com/project/lpwa/papers.html.

[GN87] M.R. Genesereth and N.J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann Publishers, San Mateo, CA, 1987.

[GRS96] D. M. Goldschlag, M. G. Reed, and P. F. Syverson. Hiding Routing Information. In Ross Anderson, editor, *Information hiding: first international workshop, Cambridge, U.K., May 30–June 1, 1996: proceedings*, volume 1174 of *Lecture Notes in Computer Science*, pages 137–150, Berlin, Germany / Heidelberg, Germany / London, UK / etc., 1996. Springer-Verlag.

[GRS99] D. Goldschlag, M. Reed, and P. Syverson. Onion Routing for Anonymous and Private Internet Connections. *Communication of the ACM*, 42(2):39–41, 1999.

[Gru] What is an Ontology? http://www-ksl.stanford.edu/kst/what-is-an-ontology.html.

[Gru93] T.R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.

[GS00] T. Grandison and M. Sloman. A survey of trust in Internet application, 2000. http://www.comsoc.org/livepubs/surveys/public/2000/dec/grandison.html.

[GWB97] I. Goldberg, D. Wagner, and E. Brewer. Privacy-Enhancing Technologies for the Internet. In *Proceedings of IEEE COMPCON '97*, pages 103–109, 1997.

[HB98] R. Hes and J.J. Borking. Privacy Enhancing Technologies: The path to anonymity. A&V 11, Registratiekamer, The Hague, 1998.

[HC02] T.D. Halket and D.B. Cosgrove. Is your online agreement in jeopardy?, 2002. Retrieved January 9, 2003 from the CIO.com Web Site: http://www.cio.com/legal/edit/010402_agree.html.

[Hoh98] F. Hohl. Time Limited Blackbox Security: Protecting Mobile Agents from Malicious Hosts. In G. Vigna, editor, *Mobile Agents and Security*, volume 1419 of *Lecture Notes in Computer Science*, pages 92–113. Springer-Verlag, Berlin, 1998.

[HSP01] A. Hameed, D. Sleeman, and A. Preece. Detecting Mismatches among Experts' Ontologies acquired through Knowledge Elicitation, 2001. http://www.csd.abdn.ac.uk/∼apreece/research/download/es2001.pdf.

[ISO99] ISO 15408 Common Criteria for Technology Security Evaluation, 1999.

[KAG98] G. Karjoth, N. Asokan, and C. Gülcü. Protecting the Computation Results of Free-Roaming Agents. In K. Rothermel and F. Hohl, editors, *Proceedings of the 2nd International Workshop on Mobile Agents*, volume 1477 of *Lecture Notes in Computer Science*, pages 195–207. Springer-Verlag: Heidelberg, Germany, 1998.

[KB02] S. Kenny and J.J. Borking. The value of Privacy Engineering. *The Journal of Information, Law and Technology (JILT)*, (1), 2002.

[KDDT01] C.L. Kunz, J. Debrow, M. Del Duca, and H. Thayer. Click-through Agreements: Strategies for Avoiding Disputes on Validity of Assent. *Business Lawyer*, 57(1):401–429, November 2001.

[Kob01] A. Kobsa. Tailoring privacy to users' needs. In M. Bauer, P.J. Gmytrasiewicz, and J. Vassileva, editors, *User Modeling 2001: 8th International Conference*, pages 303–313, Berlin Heidelberg, 2001. Springer Verlag. (Invited Keynote), http://www.ics.uci.edu/∼kobsa/papers/2001-UM01-kobsa.pdf.

[Kob02] A. Kobsa. Personalized hypermedia and international privacy. *Communications of the ACM*, 45(5):64–67, 2002. http://www.ics.uci.edu/∼kobsa/papers/2002-CACM-kobsa.pdf.

[KSY02] L. Korba, R. Song, and G. Yee. Anonymous Communications for Mobile Agents. In *In Proceeding of the 4th International Workshop on Mobile Agents for Telecommunication Applications (MATA'02), LNCS 2521*, pages 171–181, Barcelona, Spain, October 2002. NRC 44948.

[Kuh62] T.S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 1962.

[Kun02] C.L. Kunz. Click-through Agreements: Strategies for Avoiding Disputes on Validity of Assent, 2002. http://www.efscouncil.org/frames/Forum%20Members/Kunz_Click-thr_%20Agrmt_%20Strategies.ppt.

[Lev84] H.J. Levesque. Foundations of a functional approach to knowledge representation. *Artificial Intelligence*, 23:155–212, 1984.

[Lie02] H. Lieberman. Interfaces that give and take advice. In J.M. Carroll, editor, *Human-Computer Interaction in the New Millennium*. ACM Press, New York, 2002.

[LKM00] J. Lee, J. Kim, and J.Y. Moon. What makes Internet users visit cyber stores again? Key design factors for customer loyalty. In *Proceedings of CHI 2000*, pages 305–312, The Hague, Amsterdam, 2000.

[LM99] S. Loureiro and R. Molva. Privacy for mobile code. In *"In proceedings of distributed object security workshop, OOPSLA'99"*, Denver, November 1999.

[LRSW99] Anna Lysyanskaya, Ronald L. Rivest, Amit Sahai, and Stefan Wolf. Pseudonym Systems (Extended Abstract). In Howard Heys and Carlisle Adams, editors, *Selected Areas in Cryptography*, pages 184–199. Springer-Verlag, 1999. Lecture Notes in Computer Science No. 1758.

[LS01] W. Lee and S. Stolfo. Data mining approaches for intrusion detection, 2001.

[Mar94] S. Marsh. Formalising trust as a computational concept, 1994.

[McC01] K. McCarthy. The Register, 2001. Available at: http://www.theregister.co.uk/content/8/18393.html.

[Moo97] J.H. Moor. Toward a theory of privacy in the information age. *Computers and Society*, 27(3):27–32, 1997.

[MR94] M. Minsky and D. Riecken. A conversation with Marvin Minsky about agents. *Communications of the ACM*, 37(7):23–29, 1994.

[Mül96] J.P. Müller. *The Design of Intelligent Agents: A Layered Approach*, volume 1177 of *Lecture Notes in Artificial Intelligence*. Springer Verlag, 1996. ISBN: 3540620036.

[NA97] H.S. Nwana and N. Azarmi, editors. *Software Agents and Soft Computing: Towards Enhancing Machine Intelligence, Concepts and Applications*, volume 1198 of *Lecture Notes in Computer Science*. Springer, 1997.

[New82] A. Newell. The knowledge level. *Artificial Intelligence*, 18(1):87–127, 1982.

[Ng00] S.-K. Ng. *Protecting mobile agents against malicious hosts*. PhD thesis, "Chinese University of Hongkong", June 2000.

[NHP97] T.P. Novak, D.L. Hoffman, and M.A. Peralta. Information privacy in the market space: Implications for the commercial uses of anonymity on the web. Available at: http://www2000.ogsm.vanderbilt.edu/papers/anonymity/anonymity2_nov10.htm, 1997. Discussion Paper prepared for the conference "Anonymous Communications on the Internet: Uses and Abuses" University of California Irvine.

[Nie94] J. Nielsen. *Usability Engineering*. Morgan Kaufmann, San Diego, CA, October 1994. ISBN 0125184069.

[Nor90] D.A. Norman. The Design of Everyday Things. Currency/Doubleday, 1990.

[Nor94] D.A. Norman. How Might People Interact with Agents. *Communications of the ACM*, 37(7):68–71, 1994.

[Nor97] D.A. Norman. How might people interact with agents. In J. Bradshaw, editor, *Software agents*. AAAI Press/MIT Press, Menlo Park, CA and Cambridge, MA, 1997. http://www.jnd.org/dn.mss/agents.html.

[Nor01] D.A. Norman. How might humans interact with robots? Human robot interaction and the laws of robotology. Keynote address to a DARPA/NSF Conference on Human-Robot Interaction, September 2001. http://www.jnd.org/dn.mss/Humans_and_Robots.html.

[OCC88] A. Ortony, G.L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, 1988.

[Pal96] B. Palace. Data mining: What is data mining? Available at: http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm, 1996.

[PK03] A.S. Patrick and S. Kenny. D23: Agent User Interfaces and Documentation. PISA Deliverable D23, January 2003.

[PRS+94] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey. *Human-computer interaction*. Addison-Wesley, Reading, MA, 1994.

[Raa99] C. Raab. Identity Checks - and Balances. In E. Bort and R. Keat, editors, *The Boundaries of Understanding: Essays in Honour of Malcolm Anderson*, pages 87–95. International Social Sciences Institute, Edinburgh, 1999.

[Raa02] C.D. Raab. Privacy-Enhancing Technologies in their context, May 23 2002. contribution to the documentation for the symposium 'privacy by design' organised by the College bescherming persoonsgegevens.

[Ray00] J. Raymond. Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems. In H. Federrath, editor, *"Anonymity 2000"*, volume 2009 of *Lecture Notes in Computer Science*, pages 10–29. Springer-Verlag, 2000.

[RC99] J. Reagle and L.F. Cranor. The platform for privacy preferences. *Communications of the ACM*, 42(2):48–55, February 1999.

[Reg95] P.M. Regan. Legislating privacy: Technology, social values, and public policy. University of North California Press, 1995. Chapel Hill, NC.

[Rei01] J.R. Reidenberg. E-Commerce and Trans-Atlantic Privacy. *Houston Law Review*, 38(3):717–750, 2001. http://www.law.uh.edu/Journals/hlr/downloads/%2038-3%20pdf files/HLR38P717.pdf.

[RF95] Registratiekamer/TNO-FEL. Privacy-Enhancing Technologies: the path to anonymity. The Hague, 1995. The Hague.

[Roc98] E. Rocco. Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In *Proceedings of CHI 98*, pages 496–502, Los Angeles, U.S.A., 1998.

[Roe01] M. Roessler. Search engines and privacy. Available at: http://www.franken.de/users/tentacle/papers/search-privacy.txt, 2001.

[Rot80] J.B. Rotter. Interpersonal trust, trustworthiness, and gullibility. *American Psychologist*, 35(1):1–7, 1980.

[Rot98] V. Roth. Secure recording of itineraries through cooperating agents. In *"Proceedings of the ECOOP workshop on distributed object security and 4th workshop on mobile object systems: Secure Internet Mobile Computations"*, pages 147–154, France, 1998. INRIA.

[RR98a] M. Reiter and A. Rubin. Crowds: anonymity for web transactions. *ACM transactions on information and system security*, 1:66–92, 1998.

[RR98b] M.K. Reiter and A.D. Ruben. Crowds: Anonymity for Web Transactions. *"ACM Transactions on Information and System Security"*, 1(1):66–92, 1998.

[RR99] Michael K. Reiter and Aviel D. Rubin. Anonymous Web transactions with crowds. *Communications of the ACM*, 42(2):32–48, February 1999.

[RS01] J. Riegelsberger and M.A. Sasse. Trustbuilders and trustbusters: The role of trust cues in interfaces to e-commerce applications. In Beat Schmid, Katarina Stanoevska-Slabeva, and Volker Tschammer, editors, *Towards The E-Society: E-Commerce, E-Business, and E-Government, The First IFIP Conference on E-Commerce, E-Business, E-Government (I3E 2001), October 3-5, Zürich, Switzerland*, volume 202 of *IFIP Conference Proceedings*, pages 17–30. Kluwer, 2001. Main Track - Part One: Security and Trust, http://www.cs.ucl.ac.uk/staff/jriegels/trustbuilders_and_trustbusters.htm.

[RSG98] M.G. Reed, P.F. Syverson, and D.M. Goldschlag. Anonymous Connections and Onion Routing. *IEEE Journal on Special Areas in Communications*, 16(4):482–494, May 1998.

[Sau01] C. Saunders. Trust central to E-commerce, online marketing. Internet Advertising Report, November 2001. http://www.internetnews.com/IAR/article.php/12_926191.

[Sch97] Fred B. Schneider. Towards Fault-tolerant and Secure Agentry (Invited paper). In *Proceedings of the 11th International Workshop on Distributed Algorithms*, Saarbucken, Germany, September 1997. Also available as TR94-1568, Computer Science Department, Cornell University, Ithaca, New York.

[sem] http://www.semtalk.com.

[Sha48] Claude E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July / October 1948.

[Sha49] Claude E. Shannon. Communication Theory of Secrecy Systems. *Bell System Technical Journal*, 28(4):656–715, October 1949.

[Shn97] B. Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interfaction*. Addison-Wesley, 1997.

[Sim96] D. R. Simon. Anonymous Communication and Anonymous Cash. In *"In Advances in Cryptology - CRYPTO '96, LNCS 1109"*, pages 61–73. Springer-Verlag, 1996.

[SK02] R. Song and L. Korba. Review of Network-Based Approaches for Privacy. In *Proceedings of the 14th Annual Canadian Information Technology Security Symposium*, Ottawa, Canada, May 2002. NRC 44905.

[Sla98] B.H. Slane. Data mining and fair information practices good business sense. Available at: http://www.privacy.org.nz/media/data_min.html, 1998.

[Sla99] K.H. Slade. Dealing with customers: Protecting their privacy and enforcing your contracts, 1999. http://www.haledorr.com/db30/cgi-bin/pubs/1999_06_CLE_Program.pdf.

[SMJ02] P. Spyns, R. Meersman, and M. Jarrar. Data Modelling versus Ontology Engineering. *SIGMOD Record*, 31(4):12–17, December 2002. http://lsdis.cs.uga.edu/SemNSF/SIGMOD-Record-Dec02/Meersman.pdf.

[SMMS02] L. Stojanovic, A. Maedche, B. Motik, and N. Stojanovic. User-driven Ontology Evolution Management. In *Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management EKAW*, Madrid, Spain, October 2002. http://kaon.semanticweb.org/Members/maedche/1023194568.pdf.

[Sow00] J.F. Sowa. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

[SPC95] Markus Stadler, Jean-Marc Piveteau, and Jan Camenisch. Fair Blind Signatures. In Louis C. Guillou and Jean-Jacques Quisquater, editors, *Advances in Cryptology—EUROCRYPT 95*, volume 921 of *Lecture Notes in Computer Science*, pages 209–219. Springer-Verlag, 21–25 May 1995.

[ST98] Tomas Sander and Christian F. Tschudin. Protecting Mobile Agents Against Malicious Hosts. In Giovanni Vigna, editor, *Mobile Agent and Security*, Lecture Notes in Computer Science No. 1419, pages 44–60. Springer-Verlag: Heidelberg, Germany, 1998.

[Sta03] William Stallings. *Cryptography & Network Security: Principles & Practice*. Prentice Hall International, 3rd edition edition, 2003.

[STRL00] P. Syverson, G. Tsudik, M. Reed, and C. Landwehr. Towards an analysis of onion routing security. In *Proceedings of the workshop on design issues in anonymity and unobservability*, pages 83–100, Berkeley, California, 2000. California International Computer Science Institute.

[SYY99] T. Sander, A. Young, and Moti Yung. Non-interactive cryptocomputing for NC^1. In IEEE, editor, *40th Annual Symposium on Foundations of Computer Science: October 17–19, 1999, New York City, New York,*, pages 554–566, 1109 Spring Street, Suite 300, Silver Spring, MD 20910, USA, 1999. IEEE Computer Society Press.

[Tav99] H.T. Tavani. *Ethics and Information Technology*, chapter KDD, data mining, and the challenge for normative privacy, pages 265–273. Kluwer Academic Publishers, The Netherlands, 1999.

[Tav00] H.T. Tavani. Privacy and the internet. Paper presented at the Ethics & Technology Conference, June 5, 1999 2000. Available at: http://www.bc.edu/bc_org/avp/law/st_org/iptf/commentary/.

[Tet00] O. Tettero. Intrinsic Information Security: Embedding Security Issues in the Design Process of Telematics Systems. Technical Report 6, Telematica Instituut, Enschede, The Netherlands, 2000.

[The95] K. Thearling. From data mining to database marketing, 1995. Available at: http://www3.shore.net/∼kht/text/wp9502/wp9502.htm.

[Tho01] D. Thornburgh. Click-through contracts: How to make them stick. Internet Management Strategies, 2001. http://www.loeb.com/FSL5CS/articles/articles45.asp.

[TW02] Wade Trappe and Lawrence C. Washington. *Introduction to Cryptography with Coding Theory*. Prentice-Hall, Inc., Upper Saddle River, NJ 07458, USA, 2002.

[vBB01] G. van Blarkom and J.J. Borking. Beveiliging van Persoonsgegevens (Information security of personal data). A&V study #23, The Hague, 2001.

[vdL98] J.C.A. (Jan C. A.) van der Lubbe. *Basic methods of cryptography*. Cambridge University Press, New York, NY, USA, 1998.

[Ver01] J. Versmissen. Keys of Trust: TTP Services and Privacy. A&V Study No. 22, The Hague: Registatiekamer, 2001.

[Vig98] G. Vigna. Cryptographic Traces for Mobile Agents. In G. Vigna, editor, *Mobile Agents and Security*, volume 1419 of *Lecture Notes in Computer Science*, pages 137–153. Springer-Verlag, Berlin Germany, 1998.

[vRGB$^+$95] H. van Rossum, H. Gardeniers, J. Borking, et al. Privacy-Enhancing Technologies: The Path to Anonymity. Technical report, Registratiekamer, The Netherlands, and Information and Rpivacy Commissioner Ontario, Canada, August 1995. Volume I and II, Achtergrondstudies en Verkenningen 5a/5b.

[Wasa] http://www.washingtonpost.com/wp-dyn/articles/A32193-2000Aug15.html.

[Wasb] http://washingtonpost.com/wp-dyn/business/specials/privacy/A41927-2001May3.html.

[Wasc] http://washingtonpost.com/wp-dyn/business/specials/privacy/A77996-2001May25.html.

[Wes67] A. Westin. *Privacy and Freedom*. Atheneum, New York, 1967.

[Wes91] A.F. Westin. Harris-Equifax Consumer Privacy Survey, 1991. Atlanta, GA: Equifax, Inc.

[WH00] C.D. Wickens and J.G. Hollands. *Engineering psychology and human performance*. Prentice Hall, Upper Saddle River, NJ, (3rd ed.) edition, 2000.

[WT99] A. Whitten and J.D. Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *Proceedings of the 9th USENIX Security Symposium*, August 1999. http://www.cs.cmu.edu/∼alma/johnny.pdf.

[Yee99] B. S. Yee. A Sanctuary for Mobile Agents. In J. Vitek and C. Jensen, editors, *Secure Internet Programming: Security Issues for Mobile and Distributed Objects*, volume 1603 of *Lecture Notes in Computer Science*, pages 261–273. Springer-Verlag, Berlin Germany, 1999.

[You01] J. Youll. Agent-Based Electronic Commerce: Opportunities and Challenges. In *5th International Symposium on Autonomous Decentralized System with an Emphasis on Electronic Commerce*, Dallas, Texas, U.S.A., March 26–28 2001. Position statement for panel discussion on Agent-Based Electronic Commerce: Opportunities and Challenges.

[YY97] Adam Young and Moti Yung. Sliding Encryption: A Cryptographic Tool for Mobile Agents. In Eli Biham, editor, *Fast Software Encryption: 4th International Workshop*, volume 1267 of *Lecture Notes in Computer Science*, pages 230–241, Haifa, Israel, 20–22 January 1997. Springer-Verlag.

[Zan72] D.E. Zand. Trust and managerial problem solving. *Administrative Science Quarterly*, 17(2):229–239, 1972.

Appendix A

Applying the handbook: The Job Market Case

M. van Breukelen
breukelen@tpd.tno.nl
TNO-TPD, The Netherlands

A. Ricchi
alfredo.ricchi@finsa.it
FINSA, Italy

P. Bison
pbison@sentient.nl
SMR, The Netherlands

The PISA consortium has decided to build a job market application to demonstrate Privacy-Enhancing Technologies (PET) as a secure technical solution to protect the privacy of the citizen when he/she is using intelligent agents. The users of the application will consist of job seekers as well as employers looking for new employees. This chapter is based on the descriptions of a possible scenario from deliverables D1, D2 and D10.

A.1 Case description: Ideal situation

Somebody wants a new job, and hires a PISA to search the web to find a suitable job, and while doing so, protect the privacy of this person. We first describe a situation in which an ISA is used, so the privacy aspects of searching for a job can be addressed.

Let's assume the person who is looking for a new job is called Applicant. Applicant consults an ISA to help him look for it. The ISA will be provided with Applicant's background and what kind of job he is looking for. This information is collected in the user-profile of the ISA. The user-profile contains a long list of personal data related to Applicant, such as his name, birthday, marital status, number of publications, recent jobs, objectives about new job, etc. Applicant's ISA accesses the Internet and starts searching for different Job-MarketAgents, and companies which it will contact (visit). The ISA is able to decide itself, which user information to exchange with the environment. For example if the ISA is dealing with a company for a management-job, the ISA might reveal information about the management qualities of Applicant. On the other hand if the ISA is negotiating with a university, it will reveal Applicant's number of publications. It can also negotiate with other ISAs about the different companies and retrieves a lot of information about different jobs.

A.2 PISA Demonstrator case: the Job Market Case

Applicant will go to a site where a personal ISA is already created or will be created to perform the task. This ISA is called ’the UserAgent’. In order to perform the task, the UserAgent will delegate tasks to other agents. Delegating tasks could mean that personal data is transported from one ISA to another, and a privacy issue is created. The ISA should have means to decide whether or not to disclose the data, and to use PET if disclosure will take place. For every new job search a JobSeekAgent will be created by the UserAgent to delegate this task to. The JobSeekAgent will contact the JobMarketAdvisorAgent to find out which JobMarketAgents are of interest for this specific job search. Then the JobSeekAgent will contact a JobMarketAgent to find out if there are any vacancies available. If there are, the JobMarketAgent will tell the JobSeekAgent how to contact the VacancyAgents. Every VacancyAgent represents a single vacancy. If a vacancy fits Mr. Johnson’s preferences (so far), the VacancyAgent will initiate contact between the JobSeekAgent and the CompanyAgent. The CompanyAgent represents the company that published the vacancy. A company can have more vacancies, and so an CompanyAgent can have more VacancyAgents. If JobSeekAgent and CompanyAgent come to an agreement the JobSeekAgent will contact the UserAgent to make sure both the UserAgent and the CompanyAgent can negotiate the final aspects of the process. As a final action, both the Applicant and the Company are contacted by their ISAs with the appointment to sign the contract.

All agents can be on the same platform, but they could also be on different platforms. The ISAs will not leave their platform.

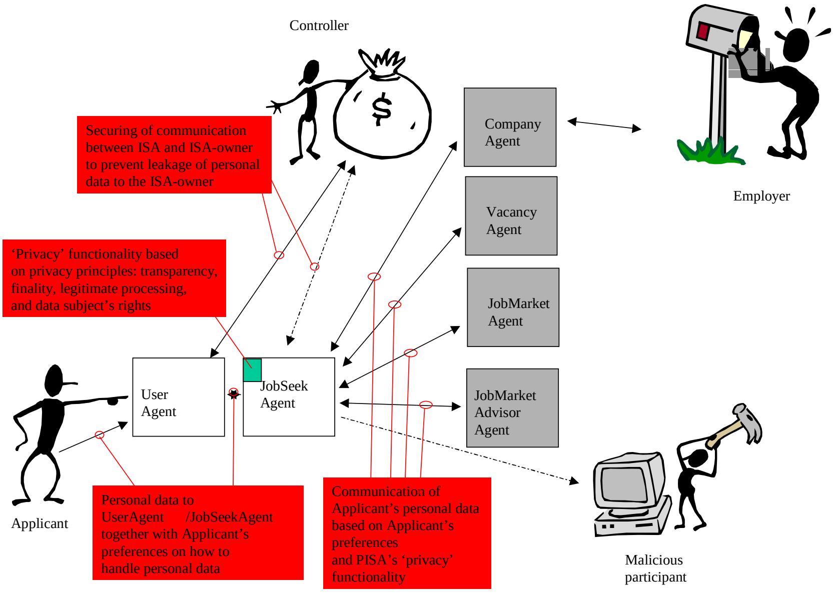

Controller
Securing of communication between ISA and ISA-owner to prevent leakage of personal data to the ISA-owner
Company Agent
Employer
Vacancy Agent
‘Privacy’ functionality based on privacy principles: transparency, finality, legitimate processing, and data subject’s rights
JobMarket Agent
User Agent
JobSeek Agent
JobMarket Advisor Agent
Applicant
Personal data to UserAgent /JobSeekAgent together with Applicant’s preferences on how to handle personal data
Communication of Applicant’s personal data based on Applicant’s preferences and PISA’s ‘privacy’ functionality
Malicious participant

Figure A.1: The job market case.

The main problems with this situation can be stated as:
- How does the agent decide what privacy information is given at which stage and to whom?
- How can it be guaranteed that the information given by the agent to another entity, will not be given to a third party who has not received permission from the agent to read this information?
- How does the agent negotiate with other participants?
- What information does the agent store and how does it do this?
- What way does the agent inform the user (log file, email, . . .)?

A.3 Scenario

The scenario is described in Table A.1.

A.4 Privacy relations

A.4.1 Assumptions

Focus on the protection of own user's personal data for the job market case. In other words: the ISA has only personal data of its user, and not of other persons.

A.4.2 Elaboration of the privacy and security solutions

- [1] All information that is applied to the ApplicantAgent by the Applicant needs to be considered as personal data.
- [2] The ApplicantAgent needs to know the sensitivity-level of all the personal data that the Applicant has applied. Therefore sensitivity-levels need to be defined. For each item of personal data the Applicant gives the required level. WP2 advises to create 3 levels (LOW - meaning Personal Data that is not directly related to the Identity of the Applicant, MEDIUM - meaning the personal data is directly related to the identity of the Applicant, HIGH - meaning highly identifiable personal data also related to sensitive information about e.g. religion, race and sexual preferences), with a default setting of MEDIUM for each item.
- [3] The ApplicantAgent shall never exchange personal data of the Applicant with the Agent-owner.
- [4] Communication between Applicant and ApplicantAgent should always be based on mutual identification and authentication.

Table A.1: Description of the scenario.

	What	How	General privacy & security issues	General privacy requirements
1	A user (Applicant) is looking for a job and therefore gives personal data to ApplicantAgent (registration). If the ApplicantAgent already excists the Applicant will update the Applicant's profile	The Applicant visits the portal, clicks on 'register' and fills in the necessary information. An ApplicantAgent is created and assigned to this Applicant. (The Applicant will update profile) Personal information is sent to ApplicantAgent.	● exchange of personal data is subject to the level of sensitivity ● authentication between parties ● data integrity must be provided ● ApplicantAgent integrity must be provided ● Data confidentiality during exchange and storage in agent.	● ApplicantAgent needs knowledge of privacy (/privacy level) and rules to decide which mechanisms to use about interaction with other entities in the environment. ● Protection mechanisms for personal data (encryption). ● Authentication mechanism ● Integrity mechanism for ISA
2	User gives a task to the ApplicantAgent. Task: find a job, which is specified according to my profile.	Communication is set up between user and ApplicantAgent.	● Authentication ● Decision on what security and privacy level communication may take place.	● Authentication mechanism ● Protection mechanisms for personal data (encryption). ● Exchange of privacy policy ● Privacy-Enhancing Technologies
3	ApplicantAgent is delegating task to JobSeekerAgent.	JobSeekerAgent is created. Necessary information to complete the task is exchanged.	● Authentication ● Decision on what security and privacy level communication will take place	● Authentication mechanism ● Protection mechanisms for personal data (encryption). ● Exchange of privacy policy ● Privacy-Enhancing Technologies
4	JobSeekerAgent seeks a JobMarketAgent.	- JobSeekerAgent consults JobMarketAdvisorAgent. - JobMarketAdvisorAgent gives advise. - JobSeekerAgent decides who to contact (which JobMarketAgent).	● Authentication ● Decision on what security and privacy level communication will take place ● Traffic flow analysis	● Authentication mechanism ● Protection mechanisms for personal data (encryption). ● Exchange of privacy policy ● Unobservability needs to be provided ● Privacy-Enhancing Technologies
5	JobSeekerAgent makes a first selection of appropriate vacancies.	- JobSeekerAgent contacts JobMarketAgent and asks for vacancies matching a query (query = subset of profile). JobSeekerAgent receives list of VacancyAgents.	● Authentication ● Decision on what security and privacy level communication will take place ● Traffic flow analysis	● Authentication mechanism ● Protection mechanisms for personal data (encryption). ● Exchange of privacy policy ● Unobservability needs to be provided ● Privacy-Enhancing Technologies
6	JobSeekerAgent tries to find an agreement with a VacancyAgent.	JobSeekerAgent contacts VacancyAgent on received list (continues to do this for all entities on list). JobSeekerAgent negotiates with VacancyAgent. Agreement is reached for both parties.	● Authentication ● Decision on what security and privacy level communication will take place ● Traffic flow analysis	● Authentication mechanism ● Protection mechanisms for personal data (encryption). ● Exchange of privacy policy ● Unobservability needs to be provided ● Privacy-Enhancing Technologies
7	JobSeekerAgent informs ApplicantAgent about match. VacancyAgent informs EmployerAgent about match.	Exchange of messages according to interaction protocol.	● Authentication ● Traffic flow analysis	● Authentication mechanism ● Protection mechanisms for personal data (encryption). ● Exchange of privacy policy ● Unobservability needs to be provided
8	ApplicantAgent and EmployerAgent continue the negotiations. ApplicantAgent and EmployerAgent make appointment for applicant and employer.	Exchange of messages according to interaction protocol. Exchange of personal information to come to the appointment.	● Authentication ● Decision on what security and privacy level communication will take place ● Traffic flow analysis	● Authentication mechanism ● Protection mechanisms for personal data (encryption). ● Exchange of privacy policy ● Unobservability needs to be provided ● Privacy-Enhancing Technologies
9	ApplicantAgent informs applicant about appointment. EmployerAgent informs employer about appointment.	Exchange of messages according to interaction protocol via user interface.	● Authentication ● Decision on what security and privacy level communication will take place	● Authentication mechanism ● Protection mechanisms for personal data (encryption). ● Exchange of privacy policy

A.4.3 ApplicantAgent needs knowledge of privacy (/privacy level) and rules to decide which mechanisms to use about interaction with other entities in the environment

• [5] Personal data of sensitivity MEDIUM or HIGH shall only be exchanged by the ApplicantAgent (When delegated to JobSeekingAgent this also applies to this ISA):
- [5.1] when the party the ApplicantAgent is communicating with, handles the personal data with respect to the legislation (to be checked by e.g. an appropriate privacy policy(/statement)), or
- [5.2] when the ApplicantAgent uses PET. For sensitivity MEDIUM pseudoidentities can be used, while for sensitivity HIGH strong pseudoidentity mechanisms, or anonymity should be used.
• [6] if the ApplicantAgent initiates or is initiated to transfer personal data then prior to these personal data being transferred the ApplicantAgent shall ask the party the ApplicantAgent is communicating with for the ID, location, and purpose specification.
• [7] if purpose specification of the party the ApplicantAgent is communicating with does not meet the ApplicantAgent's expectations then the ApplicantAgent shall not exchange personal data of sensitivity MEDIUM and HIGH, unless the personal data is protected by PET, see [5.2].
• [8] securely log all interaction data concerning the communication with other party
• [9] the Applicant can demand an inspection of the log-file. In order to do so the Applicant and the ApplicantAgent need mutual identification and authentication, based on at least something that's owned (token) by the Applicant.
• [10] if the expiration date of the personal data is reached then delete or anonymise the personal data

A.4.4 Protection mechanisms for personal data (encryption)

• [11] All personal data of the Applicant should be protected against unwanted disclosure to the environment. Therefore the personal data shall be encrypted using keys of:
- [11.1] small length to protect LOW sensitive personal data;
- [11.2] moderate length to protect MEDIUM sensitive personal data;
- [11.3] long length to protect HIGH sensitive personal data.
• [12] The encryption and decryption of personal data may only be executed in a trusted part of the environment
• [13] When exchanging personal data the personal data may only be transported in an encrypted form, therefore both the ApplicantAgent (JobSeekerAgent) and the other party should assign the session encryption keys to use.

A.4.5 Authentication mechanism

Requirements [4] and [9] already address authentication.

A.4.6 Integrity mechanism for ISA

• [14] It should not be possible for the environment to change the knowledge-base, rule-base, and the personal data held in the profile. Therefore the integrity of the

- [14.1] ISA's code shall be protected by wrapping technology
- [14.2] profile shall be protected by the encryption of the profile

A.4.7 Exchange of privacy policy

• Specification by WP1, WP3 and WP5 have to add here.

A.4.8 Privacy-Enhancing Technologies

Requirements [5.2] and [7] already address the use of PET

A.4.9 Unobservability needs to be provided

• [15] To prevent participants in the environment to do traffic flow analysis, e.g. with the use of data mining the network architecture should provide means to scramble the communication information by using unobservability mechanisms.

A.5 Conclusions

There are 2 limits to the way the ISA can provide privacy for its user. The first way is to handle personal data of its user according the privacy regulations. Therefore the ISA needs to be told what the privacy regulations are. The second way is using PET to avoid the exchange of personal data or to control the quality and quantity of the exchanged of personal data.

PISA will give its user a combination of these 2 ways. Before the exchange of personal data is necessary PISA will protect the privacy of its user by avoiding the exchange of personal data by using PET. It will also use PET when its possible to control the quality and quantity of the personal data. When more personal data needs to be exchanged the ISA shall use the privacy regulations to make sure the personal data will be handled properly.

Within the Job market case PISA will only protect the personal data of its user. A draft decision list has been made for the development of the PISA-demonstrator:

1. In the demonstrator four privacy principles will be built-in. These are; transparency (ref. V.2); finality principle (ref. V.3); legitimate ground of processing (ref. V.3); Data Subject's rights (ref. V.6); Security (ref. V.7) and partially Transfer of personal data outside the EU (ref. V.9). See reference document J.P. Leerentveld, G. W. van Blarkom, WBP Raamwerk Privacy Audit. Samenwerkingsverband Audit Aanpak, The Hague 2000 - Privacy Audit Framework under the Dutch Data Protection Act (WBP);
2. Three choice levels for the user for ref. V.2, V.3, V.6 will be built-in. Levels are yes, specified; No, conditional and More than EU directive requires;
3. Three levels of PD will be implemented. Level 1: Contact Information. This information enables the identification of the data subject without the assistance of third parties. Level 1 examples are: Name and Address information, telephone number, and email address either using a real identity or a pseudo-identity. Level 3: Special categories of

personal data as defined in Directive 96/46/EC Article 8 paragraph 1.Level 2: All others items of personal data. Level 3 personal data are only to be processed under the conditions specified in Article 8 paragraph 2 through 7;
4. As the goal of the PISA project is not to prove that level 3 personal data can be processed in a secure way, the PISA design team has decided that level 3 personal data won't be processed in the PISA demonstrator;
5. Included in the PD will the Bio (CV) of the user, his music and movies likes and dislikes
6. The Bio will contain 1. Know-how and skills; 2. job preferences, desired terms and fringe benefits; 3. contact information. These data will be split according to PD levels;
7. The agent will always start functioning in the anonymous mode. PD level 1 will be protected by encryption and or other PET measures until conditions are safe to exchange or disclose personal data;
8. Level 2 and 3 in as far as indirect identification is concerned will not be demonstrated as the discussion is well known and the used datasets in the bio are incomplete (without PD level 1);
9. The location of the third party agent will be split into EU countries (list), Countries in conformity with EU Directive (Canada, Hungary, Iceland, Norway and Switzerland) and Countries non compliant with EU Directive (ref. partly V.9);
10. If non compliant EU country then demonstrator will request user for explicit consent to process data under sub EU standards;
11. Agent privacy policy (APS) will be assumed to be on the EU level within EU compliant countries. Basic development policy is that APS and preferences have to match before transfer of PD (see ref V.4);
12. Retention period (see ref. V.3) will be added to privacy preferences (PISA-PT) and PD, i.e. 14 days storage life. If no job appointment, then receiving agent has to delete PD and task agents will be killed;
13. An task agent will receive only a one-time pseudo-identity per task in order to prevent trace-ability. Noise will be added to prevent identification of platform. PET measure will be PD level 1 + noise field. Service agent will not receive PD level 1;
14. Controller and processor through controller will request TTP to release crypto key for PD level 1 on for use by user agent and company agent;
15. Consent will be built into task agent. Consent is implemented in consent field = true for all controllers while processing personal data for one task; For a processor a Service Level Agreement (SLA) (ex article 17 paragraph 2 Directive) indicating the APS of the controller will be required;
16. Purpose specs will be listed and need worldwide standardisation (FIPA) for ontologies. In the demonstrator jargon the purpose for finding a job will be mirrored by the description "employment services";
17. Billing will be ignored in the demonstrator. Solutions might be digital cash (D.Chaum) and or a special payment/billing task agent;
18. The monitoring agent will be used for all similar transactions/job seeking tasks of different users. The monitoring agent will log agent and data paths for auditing purposes ((list requirements Gilles);
19. Transparency will be implemented through monitoring agent;
20. Data Subject rights will be fully implemented, i.e. access to own data, change, delete,
21. Two schemes will be added to this development decision list i.e. the structure of agents used and the process flow chart of the agents.

A.6 Final Decisions for development of PISA Demonstrator 24 July 2002

When PISA agents are implemented within MAS environments as commercial products, obviously the full privacy legislation has to be implemented by means of PET measures.

The privacy threat analysis made clear that violation of privacy legislation constitutes a serious threat. The objective of the PISA project is to prove that the informational privacy of the user (data subject) is protected in all kinds of agent processing by incorporating privacy protecting measures into an ISA.

The PISA Demonstrator however is set up as a project to prove that legislation can be built into agents as PET measures. The proof does not necessitate all legal provisions to be implemented. The Dutch Data Protection Authority has clustered the articles of the Data Protection Directive (95/46/EC) in such a way that the Directive is divided into nine privacy principles.

- **V.1 Intention and Notification**
- V.2 Transparency
- V.3 Finality
- V.4 Legitimate ground of processing
- V.5 Quality
- V.6 Data subject's rights
- V.7 Security
- V.8 Processing by a processor
- V.9 Transfer of personal data outside the EU

In order to prove the objective of the PISA project a number of decisions has been taken.

The demonstrator will contain the following built-in privacy principles:

- Transparency (V.2); Finality (V.3); Legitimate ground of processing (V.4) and Data Subject's rights (V.6) will be implemented completely. Partially implemented will be Security (V.7) and Transfer of personal data outside the EU (V.9). [1]
- User will be able to indicate in the User- and subsequent TaskAgent(s) his privacy preferences (PISA-PM) for ref. V.2, V.3, V.4, V.6, V.9;
- All agents within the PISA Demonstrator environment will have their privacy policy statement (Agent Practices Statement - APS) coded within themselves.

The user will state his PISA-PM before submitting his personal data (PD) to the first agent in the sequence of agents to perform the task. Before any PD will be transferred to a successor, i.e. a receiving agent:

- the APS of the receiving agent will be collected and read and interpreted by the sending agent. If the APS matches the PISA-PM then the PD will be transferred to the receiving agent;
- Agent privacy policy (APS) will be assumed to be on the EU level within EU compliant countries. Note that in the PISA Demonstrator project TaskAgent(s) will be confronted with APS that provide less protection than EU standards or no APS at all;
- The PISA-PM is inspired by P3P development of W3C. However, as long as P3P is not on EU level, P3P will not be used;
- For a processor a Service Level Agreement (SLA) (ex article 17 paragraph 2 of the Directive) indicating the APS of the controller will be required.

[1] For V see reference document J.P. Leerentveld, G. W. van Blarkom, WBP Raamwerk Privacy Audit, Samenwerkingsverband Audit Aanpak, The Hague 2000 - Privacy Audit Framework under the Dutch Data Protection Act (WBP).

It is well recognised that security measures may differ depending on the nature of the PD. Within the PISA project three types of PD are distinguished.
• Level 1: Contact information. This information enables direct identification of the data subject without the assistance of third parties. Level 1 examples are: Name and Address information, telephone number, and email address. It is of irrelevant if one uses a real identity or a pseudo-identity;
• Level 3: Special categories of personal data as defined in Directive 96/46/EC Article 8 paragraph 1. Level 3 personal data are only to be processed under the conditions specified in Article 8 paragraph 2 through 7.of the Directive;
• Level 2: All others items of personal data.

It is not the goal of the PISA Demonstrator project to prove that different security measures can be put in place. Therefore:
• No level 3 PD will be processed in the PISA Demonstrator.

According to the Directive, data items from which it is possible to identify the data subject are considered to be items of PD. It is already proven that PD Level 2 and 3 can, by using PET technology, be transformed in such a way that these items together no longer can be regarded as PD. Thus:
• it is assumed (for the PISA Demonstrator) that the PD Level 2 (and 3) cannot lead to identification of the data subject.

Within a MAS environment a user may well want to use different agents to perform different tasks; e.g. he wants to find a job, he is looking for a particular book, he is in search for a holiday destination etc. Each of these TaskAgents needs a subset of the complete set of PD the user submitted to his UserAgent.
• Included in the PD at the UserAgent will be groups of PD, e.g. the Bio (CV) of the user, his job preferences, and PD items like his music and movies preferences;
• PISA-PM (user preferences) will be allocated to the level of each group of PD (see above);
• The TaskAgent: “job seeking agent” will contain 1. Know-how and skills; 2. Job preferences, desired terms and fringe benefits; 3. Contact information. These data will be split according to PD levels and will be copied from the UserAgent;
• Any agent will receive only a one-time pseudo-identity per task in order to prevent traceability.

Because the ServiceAgents do not need knowledge on the identity of the user:
• The UserAgent will store PD Level 1 as a set of encrypted data items;
• The PD Level 1 will be disclosed to the Company in encrypted format. The method of encryption will be designed in such a manner that a lawful recipient, e.g. the company with whom the data subject (user) made the appointment, is the only party able to decrypt. It may be necessary to add a ‘noise data item’ to PD level 1 to prevent identification and possible linkability at any agent platform;
• Service agents will not receive PD level 1.

It has been stated the privacy principle V.9 will only be implemented partially.
• The location of the controller will be split into EU countries (list), countries having an adequate protection level, i.e. in conformity with the EU Directive (Canada, Hungary, Iceland, Norway, New Zealand and Switzerland) and countries non compliant with EU Directive;
• The concept of Safe Harbour will not be implemented in the PISA Demonstrator.

One of the important aspects of principle V.3 (Finality) is that PD should be kept no longer than is necessary for the purpose for which they were collected.
• Retention period will be added to privacy preferences (PISA-PM) and PD. The user has to fill in the desired period, e.g. 14 days retention period. If the storage period has expired,

the user can no longer, departing from the UserAgent, initiate new TaskAgents, unless he changes the retention period;
• The UserAgent and TaskAgent will automatically delete the received PD;
• The data subject will be able to kill task agents and change or delete PD at any time.

One of the key issues of privacy legislation is the consent the data subject must give for processing his PD.
• Consent will be built into every agent. Consent is implemented within the system of PISA-PM and the APS. Both of these will contain the purpose(s) of the data processing by the agent. If the user starts a TaskAgent he has to specify the purpose(s) for which that agent will work. This is considered to be the explicit consent the data subject gives. Transfer to a successor agent is only allowed if the APS specified purpose(s) are in exact conformity with the user specified purpose(s);
• Purpose specs will be listed and need worldwide standardisation (FIPA) for ontologies. PISA-PM needs standardisation as well (W3C). In the PISA Demonstrator jargon the purpose for finding a job will be mirrored by the description “employment services”.

The privacy principle Transparency will be implemented together with the requirements for the creation of audit trails. All activities from the PISA Demonstrator environment will be logged in a system of audit trails (MonitorAgent).
• Transparency will be implemented through audit trails. Placing a filter on the audit trail, allowing the selection of the items related to a specific data subject, is an implementation of the data subject right to have access to his personal data and the right to know how these are being processed;
• Agents will be used for all similar transactions/job-seeking tasks for different users. MonitorAgents, details to be decided upon, will log for controller, processor and data subject agent movements and data paths for auditing purposes;
• The other data subject rights, e.g. access to own data, change, delete, will be fully implemented.

Billing will be ignored in the PISA Demonstrator.
• Solutions might be digital cash (as proposed by D. Chaum) and or a special payment/billing TaskAgent.

Appendix B

Aspects of the PISA demonstrator

M. van Breukelen A.P. Meyer
breukelen@tpd.tno.nl meyer@tpd.tno.nl
TNO-TPD, The Netherlands

The PISA demonstrator was developed to show a working proof of concept of how agents can protect the privacy of users in a multi agent system and how the privacy principles can be incorporated into such a system. This appendix shows some aspects of the PISA demonstrator.

B.1 The demonstrator scenario

Figure B.1 shows the users of the PISA demonstrator, their agents and the data flow between the agents. The general user in the demonstrator is shown at the left side and can use his UserAgent to perform tasks for him. The UserAgent doesnot perform the tasks himself, but delegates the tasks to specialised task agents. In the demonstrator the JobSeekAgent is the only task agent that is implemented. The JobSeekAgent has the purpose of searching for jobs for its user.

The second user is the Company or its representative. The company uses a CompanyAgent to perform tasks for him. Again, the CompanyAgent delegates tasks to specialised task agents. The VacancyAgent is a task agent and can search for new employees to fulfil a vacancy for the company. VacancyAgents and JobSeekAgents find each other by using a JobMarketAgent as a broker. After a VacancyAgent and a JobSeekAgent have found each other, they will start to negotiate on the vacancy with a job interview between the user and the company as a potential outcome. If this is the case a match has been made and the UserAgent and CompanyAgent will exchange their contact information.

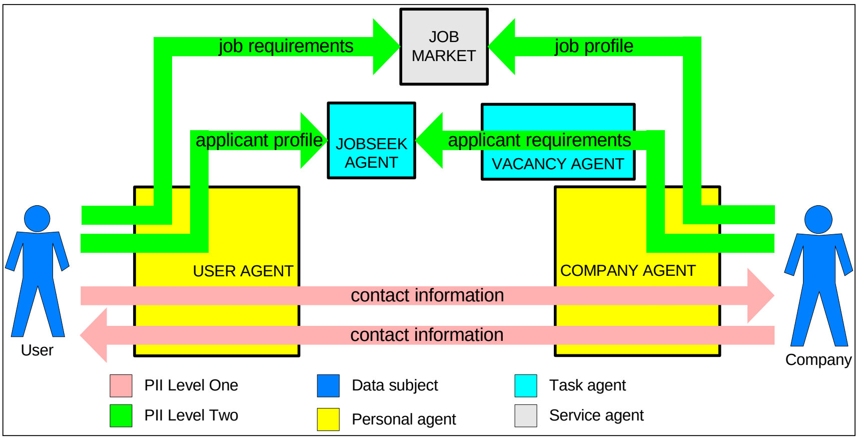


Figure B.1: Users of the PISA demonstrator, their agents and the data flow between the agents.

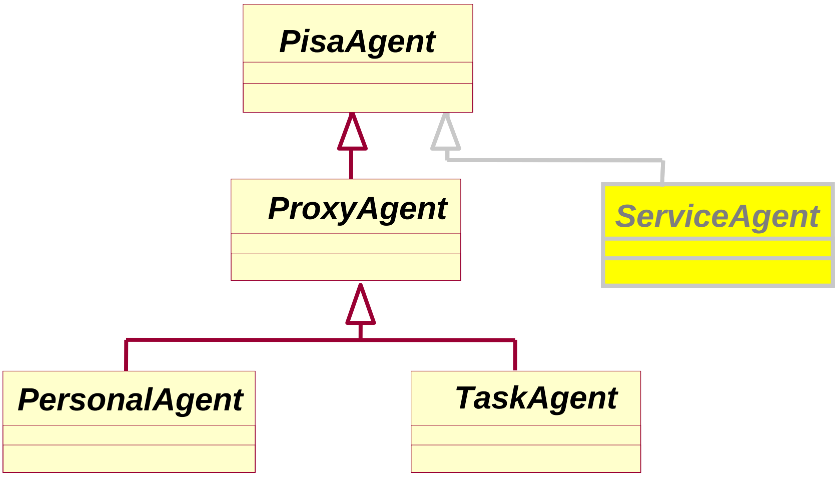


Figure B.2: PisaAgent class.

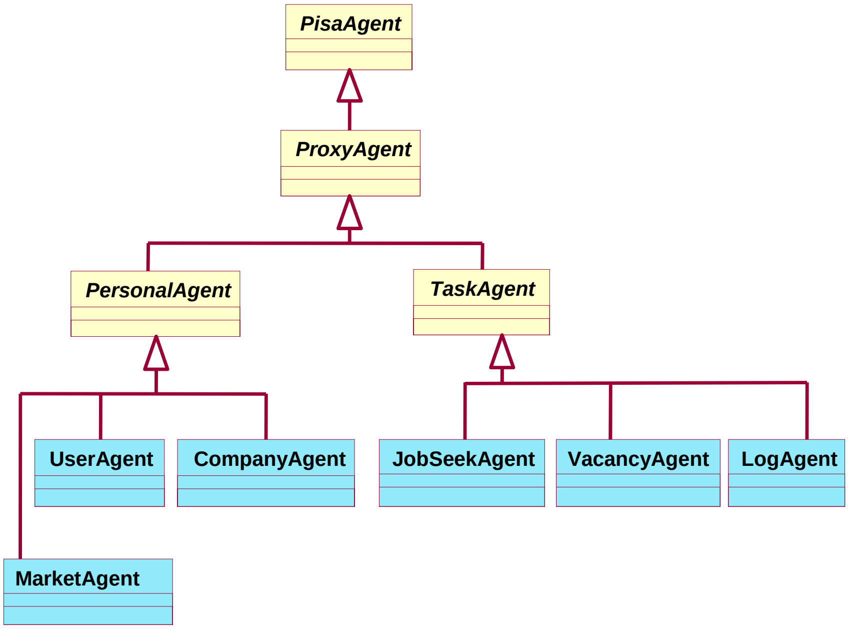


Figure B.3: Subclasses of the ProxyAgent in the demonstrator.

B.2 The PisaAgent, PersonalAgent and TaskAgent as base classes

Many of the agents in the demonstrator have a lot in common. For example, all agents in the demonstrator need to have knowledge on the privacy principles. To be able to reuse the common functionality and knowledge generic classes have been developed. The most generic agent class that was developed was the PisaAgent class, see Figure B.2. The ProxyAgent is a PisaAgent that works especially for a single user (in contrast to the ServiceAgent), is less generic and adds functionality to the PisaAgent, i.e. the ProxyAgent extends the PisaAgent.

The PersonalAgent and Task in turn extend the ProxyAgent. In the demonstrator, a PersonalAgent is an agent that communicates directly with its user and delegates tasks for the user to a specialised TaskAgent. A TaskAgent only works on a single task and is stopped after the task has been performed. In the demonstator, the PersonalAgent functions as a identity protector, i.e. it hides the identity of the user from other agents. The TaskAgents function as pseudo identities of the user.

B.3 The ProxyAgents and ServiceAgents in the demonstrator

Figure B.3 shows all subclasses of the ProxyAgent in the demonstrator. The picture shows that the UserAgent, CompanyAgent and MarketAgent are implemented as subclasses of PersonalAgent and that the JobsSeekAgent the VacancyAgent are (subclasses of) TaskAgent.

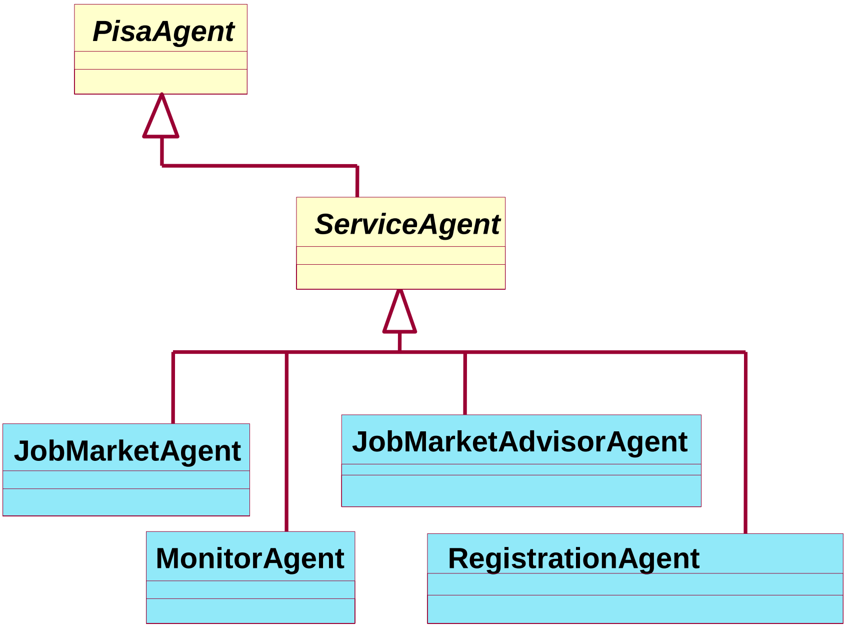


Figure B.4: All ServiceAgents.

Figure B.4 shows all ServiceAgents. As can be seen, the JobMarketAgent and some other agents are implemented as a subclass of the ServiceAgent.

B.4 The system components of the demonstrator and the PISA PKI

Figure B.5 shows the system components of the demonstrator. It shows that the agent application consists of several agent platforms that function as runtime environments for the agents in the demonstrator. The user uses a web browser to communicate with its personal agent and an application server is used to implement this interaction. Some agent use components, e.g. a database that are outside of the agent application. One of these components is the Certification Authority of the PISA PKI.

Figure B.6 shows the PISA PKI. As can be seen from the picture, a Certification Authority provides certificates to the users, application server and the agent platform. The RegistrationAgent is an agent that belongs to the PISA PKI and automatically provides certificates to agents when they are created. Each agent platform has its own RegistrationAgent.

B.5 The privacy ontology

Software agents use ontologies to give meaning to the messages they exchange. An ontology can be seen as a vocabulary of concepts and knowledge on the relations between those concepts. The concepts of an ontology can be used in the content of agent messages.

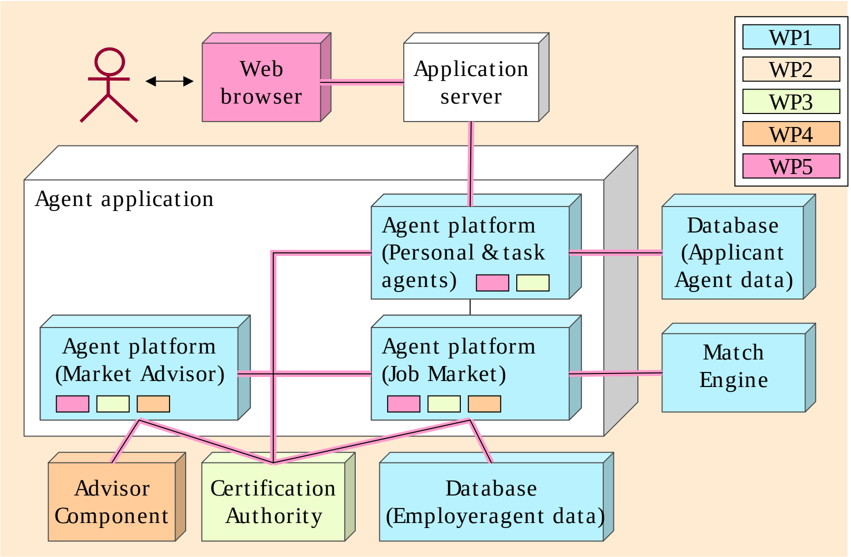

Web browser
Application server
WP1
WP2
WP3
WP4
WP5
Agent application
Agent platform (Personal & task agents)
Database (Applicant Agent data)
Agent platform (Market Advisor)
Agent platform (Job Market)
Match Engine
Advisor Component
Certification Authority
Database (Employeragent data)

Figure B.5: System components of the demonstrator.

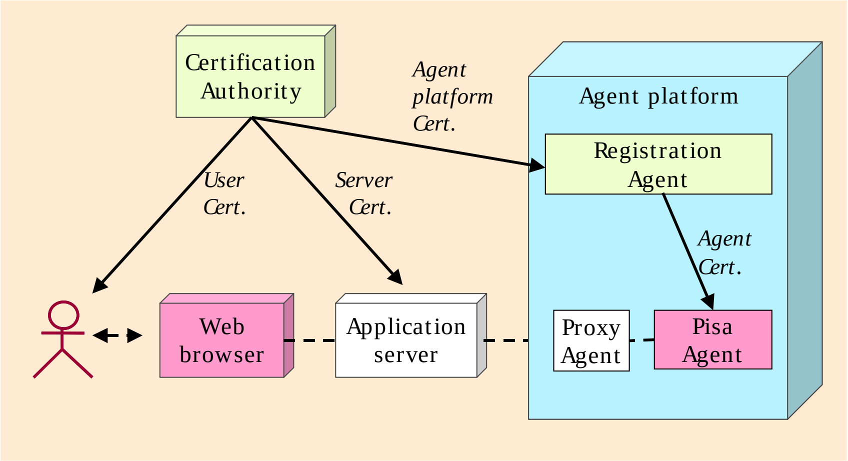


Figure B.6: The PISA PKI.

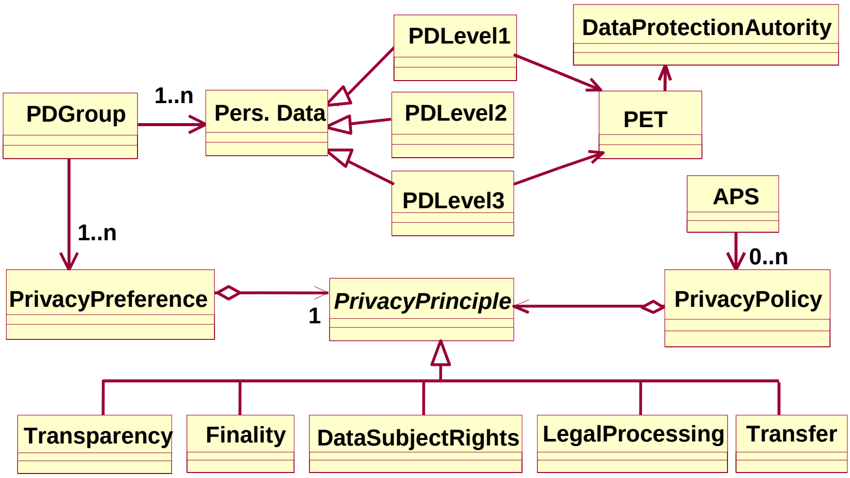


Figure B.7: The concepts of the privacy ontology.

One of the ontologies that was developed for the PISA demonstrator is the privacy ontology. A visualisation of the concepts of this ontology is presented in Figure B.7.

The privacy ontology is based on the privacy principles and contains concepts to deal with the preferences of data subjects for their personal data and with privacy policies.

Appendix C

Overview of threats per actor

This appendix gives in Table C.1 an overview of threats per actor.

Table C.1: Overview of threats per actor.

Actor	Threat (integrated with all previous ones)
Agent creator	Erroneous implementation of functionality Erroneous user documentation (loading of instructions, personal data and/or constraints)
Agent vendor	No notification to PPB of potentiality of personal data in commercial agent Erroneous commercial preparation of agent Unintentional disclosure of personal data
Agent user	Erroneous loading of instructions, resulting in erratic behaviour Erroneous loading of personal data, resulting in mutilated, superfluous or faulty data. Erroneous loading of constraints, leading to unnecessarily tight or unintended lenient or absent constraints No timely and accurate upgrade of changed personal data
Intelligent agent	No explicit approval of processing of personal data for reported (requested) purpose, or implicitly by membership. Erroneous application of instructions, resulting in the involuntary disclosure of software, instructions, personal data or constraints. Disclosure of information due to erroneous cloning Disclosure of information due to erroneous mobility and target searching Disclosure or unintended withholding of personal data due to erroneous interrogation procedures Loss or mutilation of personal data
Foreign actors: innocent variety	Unintentional disclosure of personal data Accidental loss or mutilation or personal data No timely and accurate update of reported data Unintentional processing of personal data
Foreign actors: dangerous variety	Forceful acquisition of personal data Unapproved or illegitimate processing of personal data or other actions in violation of the constraints
Institutional actors	No notification of eventual existence of collection of personal data No notification of processing of collection of personal data or parts thereof No deletion of personal data after expiration of authorisation No notification to private person of possession of personal data and/or purpose of processing
Privacy Protection Board (PPB)	Inadequate administration of notified personal data collections plus all pertaining information No notification to private person of possession of personal data and/or purpose of processing
Transmission medium	Wrong routing of agent over the transmission network Copying of agent (initiated by the transmission network, not the agent) Loss of agent

Appendix D

Overview of available PISA deliverables

This annex gives an overview of the deliverables that are currently (October 2003) available within the PISA project. The order is more or less chronological. This overview offers the reader a list of references with additional in-depth information on privacy and agents in the context of the PISA project. The deliverables can be found at the project web site: http://www.pet-pisa.nl.

D.1 WP1.1 D1

“Available and applicable Agent Technologies”

Abstract

An investigation of available and applicable Agent Technologies has been carried out. Goal of the investigation was to determine the usage of the available technologies. The technologies have been investigated and are described per ‘topic’. The following topics have been used: ‘Agent platforms’, ‘Agent Mobility’, ‘Standardisation initiatives’, ‘Legacy System Integration’, ‘Knowledge’, ‘Communication’ and ‘Implementation’. A list of most useful technologies is given in the conclusions.

Summary

One of the objectives of the PISA project is to demonstrate the feasibility of privacy enhanced technologies (PETs) to provide privacy protection in intelligent agents, possibly moving freely over the internet.
For this one or more cases of privacy sensitive applications using intelligent agents must be developed. This application must use advanced agent technology to mirror agent implementations that can be expected to appear on the market in the next couple of years.
So an advanced agent development environment must be selected, which meets a number of requirements. It must provide a wide range of facilities for agents, including the option to apply PETs.

The facilities encompass mobility, versatility and usability in an environment with legacy systems, communication with other agents, implementation support and, last but certainly not least, knowledge representation.
Knowledge representation is necessary to instruct the agent about its activities, what kind of information is needed, how to communicate with other agents, where to look and what to contact for information, and also: the representation of its own personal knowledge. The latter type of knowledge is the basis for interaction with other agents and the starting point for knowledge acquisition and exchange. In a real-life situation attempting to acquire another person's personal data without anything in exchange, would hardly be successful.

The result of the investigation is the selection of the most attractive candidate agent development environment, not just from the functional point of view but also from a standardisation point of view: how well does the intended environment comply with current standards, which are scarce up to now, and how well will it comply with the standards expected to emerge in the near future.

Ultimately, we chose the JADE package, with the remark that the Tryllian package might become a comparable candidate in the near future in this rapidly developing discipline.

D.2 WP2.1 D7

"Method for analysing privacy threats ISATs"

Abstract

An introduction to privacy and a number or privacy protection schemes is provided. Following this, a first overview of threats is provided. Also an overview of the differences between privacy, confidentiality and security is provided, to exemplify these terms.
An outline of the methodology for privacy consolidation is given, starting from the overall approach and zooming in on the part relevant for this WP.
Then the purpose of the system with intelligent agent is investigated, as an application of the method for privacy consolidation from the system purpose perspective.
Finally, the technical realisation is described and analysed for privacy threats that emanate from such realisation, to identify countermeasures to such threats.

Summary

In certain parts of the world the idea of privacy protection is well established, in other parts it is not. Although in the parts where privacy protection is well recognised and a couple of basic principles receive widespread recognition, the further elaboration is not uniform. To illustrate this, to prevent a biased interpretation and to promote widespread acceptance of the PISA work, a number of different, and reasonably valid elaborations is presented.
Each of such elaborations can be used as the starting point of a privacy consolidation analysis: the manner in which various facets of the intended system are analysed for threats to privacy and the way these threats can be annihilated or reduced in likelihood or consequences.

This situation of differing views also results in confusion about the differences between a number of pivotal terms here, notably privacy, confidentiality and security. For instance, privacy protection is by a number of people intuitively equated to confidentiality, which confounds the discussion about the subject of privacy protection. To enhance understanding of how these terms have to be interpreted an overview of their differences is presented.

For a repeatable and verifiable privacy consolidation approach a methodology is needed. This methodology should, if at all possible, be applicable to a wide array of privacy related, technological systems. To make this possible, a five-pronged approach is adopted.
The first one is the reasoning from general privacy principles, or any of the accepted sets of elaborated rules, to threats related to the players in the realm of privacy.
The second one is the line of reasoning from system purpose to threats that emanate as a result of such a system. This results in a more detailed list of threats.
The third one is related to the planned solution to be used for the system.
The fourth one is related to the technology of realisation: the line of reasoning from the technological approach applied to realise the system to related threats.
The last one is the consideration of the situation in which the system will be used; this situation may be different, or put more privacy stress on the system than envisaged originally.
The methodology is further clarified by an overview of the full analysis from regulations to implemented countermeasures and, to explicate the part addressed in the current workpackage, a more detailed presentation is provided of the WP's focus of attention.

To prove the workability of the methodology, as an example of its use, further steps are taken in the direction of an analysis of a system based on intelligent agents. Hence the purpose of that system is described and analysed with respect to potential privacy violations. This embodies a more detailed analysis of the way this system is intended to work, how the agents are generated, etc.

The last part of the deliverable addresses the technological side of the implementation of the system with intelligent agents, as far as this implementation is known, before actual system development starts. This encompasses an overview of the technical aspects of the agent as such and all other technical details that may endanger privacy.

The deliverable is finalised with a list of conclusions and remarks.

D.3 WP3.1 D10

“Privacy and security threats analysis”

Abstract

This document contains the privacy and security threats for all elements in an intelligent agent system together with the requirements for the different elements. A first outline of a security architecture is given, together with several solutions for the different functions.

Summary

PISA contributes at building a model of a software agent to demonstrate that it is possible to perform complicated actions on behalf of a person, without the personal data of that person being compromised. The emphasis of this project is to protect the user's privacy by use of technical solutions instead of only by law. This first deliverable of work package 3 gives a description of the possible threats with respect to privacy and security in an intelligent agent system. The main threats towards an agent are:

- Altering the agent's functionality
- Duplication of the agent
- Damaging the agent
- Fraud with agent and user's identity
- Unsecured data storage

Next to defining these threats in more detail, also a preliminary security architecture for the agent is described. This architecture divides the agent in smaller blocks such that each block can individually be looked at from a security point of view. Two logical blocks of the agent are the division of functionality and data. Functionality covers all the different functions an agent has to perform. This could be the task the user has given it, but could also be the ability to sign an electronic document or the ability to roam over a network. The data block is again split in two smaller pieces. First there is data that is stored at beforehand. Either the user or the agent developer stored this data. The second piece of data is the data, the agent receives while it is performing its task. This could be information it retrieved from a web site or from other agents. Having the agent split into data and functionality, some additional parts are necessary such that it is clear where to protect the privacy, how to react on the network environment, who to trust, what communication protocols to use and some legal aspects. Each of these aspects have a separate function block to overcome these problems.

Finally each work package has defined threats with respect to their own specialization, such as data mining or security or agent technology. These threats are defined in the appendix of the deliverable.

D.4 WP4.1 D16

"Impact of datamining on Privacy and ISAT"

Abstract

Data mining is a novel technology that aims to analyse data from different sources and compositions and integrate the results to arrive at as yet unknown information and knowledge. Obviously, this may involve privacy threats, if part of the integrated data refers to an identifiable person.

The relevance of privacy with data mining activities is articulated. Special attention is given to the data enrichment aspect of data mining, the use of data mining for privacy protection and special methods in defence of privacy that are more or less specific to data mining.

D.5 WP5.1 D21

"Network Privacy threats and opportunities"

Abstract

WP5 involves an investigation of issues associated with network, scalability and user interfaces for privacy enhancing technologies. The goal of this work is to produce an analysis of the design of PISA privacy enhancing technologies with a view to improving network security, scalability and usability of implementations. For deliverable 5.1 we outline the privacy threats and opportunities for some better known approaches for network privacy. We also examine the privacy needs for agent-based systems from a network perspective.

Summary

In this document we provide an overview of the better known network approaches for assuring anonymity and privacy over networks. The key approaches we discuss include: Dining Cryptographer Networks (DCNets), JANUS Proxy, MIX Network, Onion Routing, Crowds System, Freedom Network, and NymIP. While these approaches offer some possibilities for providing the anonymity and privacy, they do not answer all of the requirements for agent-based systems. For instance, many of these approaches are designed for World Wide Web access. Being protocol-specific these approaches may require further development to be used with agent-based systems, depending on the communication protocols used in those systems. For the most part the network-based privacy protocols are geared towards client-server configurations. This should not detract from using the protocols for agent-based systems, since the privacy approaches offer bi-directional exchanges. Core to many privacy developments is the use of multiple proxies. Many of the approaches may be compromised if the attacker gains access to multiple nodes of the privacy network. One approach to make this extremely difficult to do is to have the proxies located in different countries. To make these systems practical, approaches are needed especially to handle key distribution. Handling keys can be especially cumbersome in agent systems where there are many hundreds or thousands of agents.

While privacy and anonymity would be core requirements for agent operation, we discuss other requirements for the potential tracking privacy-related activities for the purpose of enforcement and management. In this case, the need for privacy and anonymity must be tempered by the requirement for accountability and provability to be compliant with potential privacy auditing requirements. This is an area we recommend PISA partners place considerable effort. We also recommend the adoption of ZKS freedom network for initial demonstrations of PET, while indicating there may be research opportunities in the development of alternate approaches for anonymity.

D.6 WP1.2 D2

“Agent Development Platform Evaluation”

Abstract

- The agent technologies for the demonstrator have been chosen in WP 1.1. The main choice was to use the JADE agent platform.

- Developing multi-agent systems requires additional modelling and development concepts, tools and technologies on top of ‘traditional’ object oriented systems. The goal of the current work package 1.2 is to determine the usefulness of available concepts, tools and technologies for developing the demonstrator.

- Division of demonstrator development work: Agent application, agent-user communication, employer simulation, privacy & security aspects.
- results and conclusions per part.

Summary

The goal of WP 1.2 is to investigate the usability of available tools, technologies and concepts for modelling and developing agent based applications in order to choose a development environment for the development of the demonstrator.
Based on the research that took place in WP1.2 we can draw the conclusion that enough information has been collected to build a preliminary demonstrator in WP1.3.
More detailed conclusions consider required and available modelling methods and tooling, access of legacy databases, method driven incorporation of privacy and security issues and a proposal for additional standardisation.
Research on these topics is reported in the following structure;
Chapter 2 will introduce a demonstrator set-up that is used throughout the rest of the report to illustrate the various topics.
Chapter 3 till 6 will focus on various aspects of designing and building an agent based system in general and will discuss various tools available to support this process.
Chapter 7 will discuss the possibilities for providing communication between users and agents using J2EE.
Chapter 8 will discuss privacy and security issues when developing an agent based system.

D.7 WP3.2 D11

“Cryptographic mechanisms to be applied”

Abstract

This deliverable gives a description of the state-of-the-art on privacy enhancing technologies with respect to cryptography. This description is then used to provide security and privacy mechanisms that can protect the agent against threats as were mentioned in deliverable D10, which results in an agent security and privacy architecture in combination with a public key infrastructure. Finally, a description is given about the identity protector, what it is and where it is located in the system.

Summary

This deliverable, D11, gives a description of the state of the art on Privacy Enhancing Technologies (PET) with respect to the cryptographic mechanisms. The other set of PETs, which are based on networking aspects are described by work package 5 in D21. It is seen that the common PETs can be used in the PISA-project such as pseudonyms and blind signatures, but also a new kind of cryptography called mobile cryptography. Mobile cryptography protects mobile code that is being executed at a foreign host and gives the possibility to provide privacy, such that functions can be executed without this host knowing what is executed or what the meaning of the parameters are. Unfortunately, not many of these mechanisms are developed yet and therefore a part of work package 3 will be to develop such mechanisms. The first one is developed, namely, that an agent is able to sign a message at a foreign host, without this host having access to its private key.

The second part of this deliverable describes mechanisms that solve the problems as they were stated in D10, written by work package 3. Because the PISA-project is based on agent technology, some conventional mechanisms must be adjusted to this environment, such as the agent signature. Also an overview is given of the agent security architecture. This security architecture exists out of several blocks and each has its own task to protect the agent against security and privacy attacks. For the overall security in the system, some possibilities for a public key infrastructure are given. A summary is given that shows the mechanisms that will be used in relation to the current status of these mechanisms. Some mechanisms still have to be developed, but most of them are available now.

The deliverable ends with a section on the identity protector. An identity protector can generate pseudo-identities as needed, convert pseudo-identities into actual identities or combat fraud and misuse of the system. It is shown where in the system the functionality of the identity protector is located. Because the identity protector is described by its functionality, it is not one element in the system, but the functionality can be found at various places.

Concluding it can be said that the necessary mechanisms are provided to protect an intelligent agent. However, some of the mechanisms are in such a preliminary research phase, such as the mobile cryptography, that several algorithms will have to be developed by the PISA-project (work package 3).

D.8 WP4.2 D17

“Empirical Feasibility of using Data Mining for Privacy Invasion and Protection”

Abstract

In relation to privacy related data that is available over the internet, data mining can be used in two ways: defensively and offensively.
Defensive data mining means that the technology is used to identify leaks and loopholes in the privacy awareness and protection of internet sites and to identify how the privacy of an individual (‘data subject’) is endangered by lack of adequate protection. Offensive use means that data mining is used for privacy invasion: to collect and combine such information from different sources that privacy sensitive information about one or more persons emerges.
In this report the results are presented of research on both types of usage.

Summary

The internet environment is an excellent medium for communication; basically, that was the main reason this facility was developed. As a consequence, an enormous amount of information is available, among which, almost inevitably, privacy related information. A site containing such information may or may not display that it will adhere to privacy regulations and, if it does, this information may or may not be protected by adequate privacy protection measures.

Data mining is a technique to collect data from various web sites and, if possible, to integrate such information. The advantage of the integration is that a much clearer picture emerges than from any of the partial contributions. Obviously, data mining can be used in beneficial ways, for instance for the collection of scientific data on a certain subject. But the downside in relation to privacy related information is that data mining may be able to build a more or less complete picture of an individual, sure enough with damaging potentials. The extreme case of this is *Identity Theft*: another person exploits the social, financial and/or other identity of a victim. This is an example of the offensive and malafide use of the data mining technology.
Data mining (DM) is the best known approach of establishing LINKABILITY between what can factually (but not legally) be considered non identifiable information and identifiable information. DM will always have a place at the privacy table because of this.

On the other hand, in line with the beneficial use of data mining when privacy indifferent information is collected and integrated, it can also be used defensively and in two distinct manners. The first manner is to see what the privacy policy and associated protection of a certain web site means and to signal the findings. The second manner is to find out what can be collected about yourself, or someone you know, and to see how menacing this might be and whether protective measures should be taken. Both approaches are discussed in this report and the observations are that privacy awareness leaves a lot to be desired and that privacy protection is often rudimentary in the investigated segment of job broking sites.

Following that an investigation is started of offensive use of data mining, namely for privacy intrusion.
Data mining on the Internet is a viable approach for the collection and combination of the desired information for the individual pursuing such a goal, because of the array of information that can be found on the web, of which the data subjects are possibly not even aware of.

D.9 PISA System Architecture (PSA) and Datamining

On the role of datamining in PISA.

Summary

In this paper the role of data mining in the PISA System architecture and the PISA demonstrator will be explained. Two examples of the use of data mining for protecting the privacy of the private users of the platform will be investigated. The first example studies how data mining can be used in an agent-environment to train a model representative for the normal behaviour of agents, and how this model can be used to detect deviating behaviour of malicious agents. The aim is to be able to use such a model to predict malicious behaviour based on other properties than the behaviour of the agents. A data mining component, which trains a model of ‘normal behaviour’ based on the data about the interactions between the agents, will be implemented in the PISA demonstrator. The second case studies how data mining can be used to analyse and rank e-markets (job-markets in the demonstrator case), and use the analysis to inform Task Agents about the privacy risks of the communication.

D.10 WP1.3 D3

“System design in UML and additional documentation”

Abstract

One of the main goals of the PISA project is to prove that privacy and security requirements can be implemented in a realistic agent system. PISA proves this by building the PISA demonstrator that incorporates solutions for the privacy and security requirements that can be extracted from European privacy legislation.

This document describes the PISA demonstrator system design in UML and additional documentation. It first describes the job market as a demonstrator case, the privacy and other requirements that apply and the restrictions that were made for the demonstrator.

The description of the design starts with the design for the privacy and security solutions, followed by the design of the incorporation of the solutions in the agent demonstrator application. Finally, this report describes the design of the additional required components and the scenario’s to demonstrate the privacy solutions with the PISA demonstrator.

Summary

One of the main goals of the PISA project is to prove that privacy and security requirements can be implemented in a realistic agent system. PISA proves this by building the PISA demonstrator that incorporates solutions for the privacy and security requirements that can be extracted from European privacy legislation.

This document describes the PISA demonstrator system design in UML and additional documentation. It first describes the job market as a demonstrator case, the privacy requirements that apply and the restrictions that were made for the demonstrator.

The description of the design starts with the design for the privacy and security solutions, followed by the design of the incorporation of the solutions in the agent demonstrator application. Finally, this document describes the design of the additional required components and the scenario’s to demonstrate the privacy solutions with the PISA demonstrator.

The job market agent application that has been designed contains privacy protection mechanisms including a privacy preferences and policies mechanism that was inspired by P3P. Knowledge on privacy legislation was included into PisaAgents that deal with most generic aspects of privacy-awareness and can be used as a base agent for other agents. Some specialised agents have been identified to provide required functionality such as transparency. Each agent receives a certificate to provide secure communication. Task agents that need limited personal data to perform dedicated tasks act as pseudo-identities. Identifying personal data is called level 1 PII and is separated from non-identifying personal data. The identifying data is only disclosed when explicitly required and is always encrypted.

The privacy solutions are not specific for the demonstrator but can be reused in other applications.

In addition to the original idea of the PISA project to build an agent that can be used by a data subject to hide his personal data using PETs such as anonymity and using pseudo-identities two extra activities have been performed. First, PETs have been incorporated in the agent platform and other agents, instead of only in the PISA. and second Privacy legislation knowledge has been built into agents. Both activities have been performed because this gives much more possibilities for the agents to autonomously perform tasks for the data subject. The original idea can still be used however if the agents of the data subject encounter agents that do not implement the privacy preferences and policy approach.

Standardisation is needed for agents to be able to communicate with each other on privacy policies and privacy preferences and other aspects of privacy legislation. For this purpose PISA will continue to influence standardisation bodies like FIPA and W3C/P3P.

D.11 WP2.2 D8A

“Handbook Privacy and PET for ISAT’s”

Abstract

This pre-deliverable is the start of the Handbook “Privacy and PET”. It presents the results until mid 2002 of the joint study by the Dutch Data Protection Authority, TNO, University of Delft, Sentient Machine Research, FINSA Consulting, National Research Council Canada and GlobalSign, as partners in the PISA project. The Handbook will be used as a reference in the PISA-project and will identify and demonstrate ways of applying Privacy-Enhancing Technologies (PET) to agent technology.

Summary

At this moment, efforts are underway to develop software agents capable of handling ‘information overload’ and ‘network overload’. These ‘intelligent’ agents are able to act independently. Some can move through their environment, co-operate with other participants and learn from information provided by the environment and from every action they execute.

To delegate tasks to an agent, a user needs to provide the agent with a user-profile containing personal data about the user, e.g.: mail addresses, habits and preferences. On the other hand, the agent can also collect information about individuals on behalf of an organisation it works for. Because the agent possesses this personal data, the use of agents could pose a threat to the privacy of the individuals concerned. With the use of Privacy-Enhancing Technologies (PET), agents can protect the internally stored personal data against certain threats. Agents and PET can also be used to help users to search for locations where their personal data are stored, so that they can exercise their rights laid down in international laws and treaties to control the collection of that personal data.

This report presents the results of the joint study by the Dutch Data Protection Authority, TNO, Delft University of Technology, Sentient Machine Research, FINSA Consulting, National Research Centre Canada and GlobalSign, as partners in the PISA project. With this study have attempted to identify possible threats to the privacy of individuals resulting from the use of agent technology. Secondly, the study sought to identify and demonstrate ways of applying Privacy-Enhancing Technologies (PET) to agent technology in such a way as to eliminate the impact of these threats.

The following technologies are used to enhance privacy: certification of the agent’s working method; logging of all agent actions; identification and authentication of all agents; access control mechanisms; logging of all actions performed by other agents; audit mechanisms; integrity mechanisms for stored data, exchanged data, working methods or trusted components. The Identity Protector can be based on privacy preferences from the user, levels of personal data, an Agent Practises Statement (APS) that contains the privacy policy of the agent. An APS consists of two items: the privacy policy of the receiving agent as specified by its controller and the identification of this controller. This mechanisms will

be secured by an user/agent-PKI and new applications of cryptography, privacy protection in Agent-Networks and Human Computer Interfaces. Also a Privacy ontology and rules are developed to implement in the PISA-agent to act according the Directive.

To find the right solution, designers, developers, suppliers, users, or providers can use a this handbook as a checklist of design criteria.

The objective of this Handbook is to demonstrate the possibilities of protecting the privacy of a consumer in an agent-based environment. To accomplish this objective the following question will be addressed: What possibilities are there to protect the privacy of a consumer in an agent-based environment? In attempting to answer this question, the following, more specific, questions arise:

- What is privacy?
- What are software agents?
- What threats to privacy can be attributed to software agents?
- What (technical) measures are there to eliminate or reduce the impact of these threats?
- How to design a privacy incorporated system?
- How to audit and evaluate a privacy incorporated system?

D.12 WP3.3 D12

“Privacy protection software design”

Abstract

The objective of this report is to design mechanisms that provide privacy and security towards intelligent software agents that act on behalf of a user. A new mechanism is proposed for the storage of data in an agent, such that confidentiality is guaranteed. Also a new algorithm is presented that makes it possible for the agent, located at an agent platform, to sign a document without the agent platform having access to the agent’s private key. Two options for a public key infrastructure (PKI) are proposed to provide the possibility for the various elements in the system to obtain certificates, such that authentication between elements is possible.

Summary

The objective of this deliverable is to translate the mechanisms that should be applied for privacy protection into concrete cryptographic algorithms, protocols and other tools. To reach this objective the document is divided in three main parts: agent security and privacy, agent communication and a Public Key Infrastructure (PKI). Each of these parts deals with the general case of security and privacy for agent technology. This means that the agent is mobile, intelligent and does not know beforehand at which platforms it will be located during its lifecycle. A consequence of the latter is that the agent does not know beforehand whether it can trust the platform. This consequence is the reason that not in

all cases standard security measures can be used to protect the agent. New algorithms and measures must be designed in order to fulfill the requirements.

Agent security and privacy

Securing an intelligent software agent that has the property of mobility is different than securing conventional software systems, because initially it is not known whether the execution environment can be trusted or not. Also storing data in an intelligent software agent is not equal to the conventional problem of secure storage of data on one's personal computer, although many similarities do exist. Equivalent to the requirements of conventional data storage, stored data in the software agents must be able to provide properties such as confidentiality and integrity. The difference in data storage between conventional software systems and agent technology is not the storage itself, but the access to the secured data, especially the location of access.

An often-used technique to provide confidentiality in conventional data storage, is to encrypt the data using the data-owner's encryption key and at the moment this data should be exchanged it is decrypted with the owner's decryption key and he again encrypts the data, but now using the receiver's encryption key. This technique is not completely sufficient to provide data confidentiality for agents, because if this method would be applied there will always be a period in time that the data is readable to the agent platform (between decryption with the owner's key and encryption with the receiver's key). In some cases this is not a wanted situation. Hence, a new mechanism must be designed to overcome this problem and this new mechanism is proposed in this document. Summarizing this mechanism, the data is first encrypted with the owner's encryption key, followed by an encryption of the receiver's encryption key. At this moment the data is encrypted twice. This is followed by a decryption operation using the owner's decryption key. Finally, at time of receiving the data, the receiver can use his decryption key to obtain the readable data. Using this technique, the data is never readable to anyone but the entities that have access to the used keys. This technique works as long as entities are located at different computers or if the data is only collected for storage and not for further processing at that particular location. If the agents are located at the same platform and the confidential data must be further processed on that platform by one of the agent's, it is not possible using this technique to provide full confidentiality, because at time of processing the data must be in the clear in order to execute the correct function, unless mechanisms like "computing with encrypted data" can be used.

Providing integrity in data storage is easier, because at the moment of data exchange no extra operations must occur. If the data owner signs the data, it provides integrity and it can be done beforehand.

Agent communication

During communication, it must be possible for the agent to have access to mechanisms to secure the communication, hence provide integrity and confidentiality. Similar to the case of confidential data storage, it is different to communicate securely between agents from the conventional secure communication, where the communicating partners are located at trusted computers.

Providing communication confidentiality means that the above-described mechanism for secure data storage should be implemented and that the order of the operations necessary for that technique should be guaranteed. This technique is useful in case the communicating partners are located at different platforms or if the data is exchanged only for collection purposes. When the agents are located at the same host, the mechanism must be extended with a new one, computing with encrypted data. In the literature some of these mechanisms have been developed, but they do not satisfy yet, either because they are inefficient or for performance reasons. Hence, this is an interesting research problem for the next stages in the PISA project.

Providing integrity in data communication is not as easy as for data storage. At some point the agent will receive information from some party and needs to sign the document to provide integrity. However, in this case the agent must use private information, that if other parties (including the agent platform) would have access to this private information, they could pretend to be that agent. A mechanism has been developed that prevents the access to this private information by the agent platform, called an agent digital signature. The way the agent digital signature is designed, provides also extra privacy to the agent and can therefore be called a privacy enhancing technology.

Public Key Infrastructure (PKI)

In an agent environment different elements are present, such as agent platforms, software agents and various users. In order for each element to authenticate each other, all elements must be in possession of a certificate that binds the element's identity to a key pair. Using the combination of the key pair and the certificate a person can proof his identity in the electronic world and has the possibility to provide confidentiality and integrity during communication. A PKI provides the infrastructure and issues these certificates. Two solutions are proposed. The first solution is an external PKI. This means the PKI platform is located outside the agent platform and this PKI platform issues the certificates to the various agents, users and platforms. A problem with this approach is the management of the certificates. This is difficult because some entity must revoke the certificate when the agent is killed and using an external PKI platform this is not easy. A second disadvantage is the amount of communication outside the platform. Every time an agent is created or killed, communication between the PKI platform must take place. This can put heavy requirements on the network.

The second proposed solution is a PKI where the PKI platform is embedded in the agent platform. This has the advantage that managing and requesting certificates is much more efficient, because everything can occur locally. Also, most of the communication with the PKI platform can occur on the platform, such that the amount of traffic on the network is much smaller than in the previous case. However, a consequence is that a new module must be implemented in a secure way, that can support this PKI on an agent platform.

Conclusion

Summarizing, it can be said that several solutions to provide security and privacy to agents have been proposed in this document. However, many open problems remain and it is necessary to perform research on these problems in the remaining of the PISA project in combination with the design of the PISA demonstrator.

The core of the document describes the security and privacy of the general case of agent technology. IN addition to this work has also been done on the PISA demonstrator, which is described in the appendix. A first design is given to provide a PKI and secure agent creation.

D.13 WP4.3 D18

“Design & technical description of customise online matching environment PISA”

Abstract

In the PISA demonstrator a virtual job market is used as a test site for the PISA agents. This job market presents a database of vacancies to the jobseeker agent. Resumes of the jobseeker are processed by the matching engine of the job market and suitable vacancies are returned. In this deliverable the matching engine workings are described and the

testbed content of vacancies.
The interactions of the agents with the job market also provide the data for data mining analyses. A Jobmarketadvisor uses the resulting rules to provide directions to the jobseeker.

D.14 WP5.2 D22

"PISA system scalability"

Abstract

WP5 involves an investigation of issues associated with network, scalability and user interfaces for privacy enhancing technologies. The goal of this work is to produce an analysis of the design of PISA privacy enhancing technologies with a view to improving network security, scalability and usability of implementations. For deliverable 5.2 we outline approaches for scalability for privacy provisions within PISA Systems, define test cases for PISA system scalability, test PISA system scalability for these cases and analyse PISA scalability.

Summary

In this document we examine two areas: modeling aspects of PISA system design, and testing PISA system scalability. Since at this stage the PISA prototype is under development, the objective of this work is to provide an overview of approaches that may be applied to modeling and testing. More specifically, we develop early models of the PISA prototype system design, define test cases for PISA system scalability, test the design for PISA system scalability and indicate the direction we will take in WP5.

D.15 WP1.4 D4

"Experiments with implementation concepts"

Abstract

This document describes experiments with various implementation approaches for aspects of the PISA system design. The goal of the implementation experiments is to prove feasibility of the concepts and to make a well-substantiated choice for the implementation of the PISA demonstrator as a next step. The experiments are grouped in the three themes security, privacy and job market case. Implementation solutions have been found for the implementation concepts for security, privacy and the job market case, as specified in deliverable 3 of the PISA project. The results of the implementation concepts make us believe that it is possible to implement the PISA prototype.

Summary

One of the main goals of the PISA project is to prove that privacy and security requirements can be implemented in a realistic agent system. PISA proves this by building the PISA demonstrator that incorporates solutions for the privacy and security requirements that can be extracted from European privacy legislation.

This document describes the PISA demonstrator system design in UML and additional documentation. It first describes the job market as a demonstrator case, the privacy requirements that apply and the restrictions that were made for the demonstrator.

The description of the design starts with the design for the privacy and security solutions, followed by the design of the incorporation of the solutions in the agent demonstrator application. Finally, this document describes the design of the additional required components and the scenario's to demonstrate the privacy solutions with the PISA demonstrator.

The job market agent application that has been designed contains privacy protection mechanisms including a privacy preferences and policies mechanism that was inspired by P3P. Knowledge on privacy legislation was included into Pisa Agents that deal with most generic aspects of privacy-awareness and can be used as a base agent for other agents. Some specialised agents have been identified to provide required functionality such as transparency.

Each agent receives a certificate to provide secure communication. Task agents that need limited personal data to perform dedicated tasks act as pseudo-identities. Identifying personal data is called level 1 PII and is separated from non-identifying personal data. The identifying data is only disclosed when explicitly required and is always encrypted.

The privacy solutions are not specific for the demonstrator but can be reused in other applications.

In addition to the original idea of the PISA project to build an agent that can be used by a data subject to hide his personal data using PETs such as anonymity and using pseudo-identities two extra activities have been performed. First, PETs have been incorporated in the agent platform and other agents, instead of only in the PISA and second Privacy legislation knowledge has been built into agents. Both activities have been performed because this gives much more possibilities for the agents to autonomously perform tasks for the data subject. The original idea can still be used however if the agents of the data subject encounter agents that do not implement the privacy preferences and policy approach.

Standardisation is needed for agents to be able to communicate with each other on privacy policies and privacy preferences and other aspects of privacy legislation. For this purpose PISA will continue to influence standardisation bodies like FIPA and W3C/P3P.

D.16 WP3.4 D13a

“Privacy in agents: theory”

Abstract

This report describes the theory of privacy in an agent environment from the cryptographic perspective. It is analysed until what extent privacy can be offered based on a trust model, where only the agent's owner can be completely trusted. Three approaches are followed. The first two approaches conclude that the success whether privacy can be protected in an agent environment depend on whether it is possible to make decisions on encrypted data. The third approach results in a cryptographic dilemma based on information theory.

Summary

This report is the fourth deliverable about security and cryptography in the PISA-project. The objective of this report is to draw conclusions about the possibility of providing privacy in an agent environment.
Autonomy is a typical property of an intelligent software agent and it is the question if this can be combined with strong privacy requirements.

To answer this question it is first necessary to define the trust model of the system. In this report only the agent's owner (originator) is said to be completely trusted. All the other agent platforms are trusted in the sense that the agent is executed correctly, but they are interested in the agent's content and goals. These platforms can execute the agent as many times as it wants to acquire some information about the agent.

Three different approaches are followed to answer the main question. The first approach is to achieve privacy by using cryptographic mechanisms. This means that the task is encrypted. It is possible to encrypt a certain type of functions, however in order to achieve full privacy also a decision must be made on encrypted data. It is still an open research question to find a solution to make decisions efficiently on encrypted data.

In the second approach it is assumed that encryption of the entire task is possible. However, now a look is taken from a broader perspective. The agent's task can be modelled as a black box with some input and output parameters. The host is capable of performing this task multiple times with different input parameters. In this way the host can determine it's optimal input-output pair and in this way it can gain some knowledge about the agent's task. The task must be protected such that the most efficient way to find the optimal input-output pair is doing an exhaustive search. This can only be achieved if the output space is large enough and when knowledge about some input-output pairs does not give certainty about another input-output pair.

The problem is also looked at from an information theoretic perspective. Shannon's secrecy model is extended such that it is also applicable to an agent environment. Two extensions are necessary. First, the execution environment (agent platform) cannot be trusted while in conventional secrecy systems this is one of the assumptions. The second extension is that the agent platform can execute the agent's task many times and therefore can obtain knowledge about the agent's task. Following this approach a similar dilemma to Shannon's cryptographic dilemma is derived. A maximum uncertainty about the key is required, but that implies that there is certainty about the task given the protected task.

Following these three approaches a conclusion can be drawn that providing privacy in an agent environment depends heavily on the assumptions one is willing to take. It has also a great influence on the autonomy of an agent. If full privacy cannot provided, this means that the agent must perform privacy-sensitive operations at its originator and therefore is not completely autonomous anymore, because it cannot even contain that privacy-sensitive task.

D.17 WP3.4 D13b

"Software Description Privacy Protection - PKI-part"

Abstract

This report describes the requirements and the specification of the PISA-PKI platform that has been introduced in D.11. The need for the introduction of such platform is to derive a secure communication mechanism between different agents, within an agent platform.
The architecture and the functional description, give the specifications of all services of the PISA-PKI platform. Evaluation and Testing Procedures gives a list of evaluative criteria that will be provided in the testing procedure of the PISA-PKI platform. The list will be based on ISO Testing Standard (ISO 9646) terminology and contains three classes of tests: basic tests, capability tests and behaviour resolution tests.

The implementation plan gives a guideline on the implementation phase and the set up of such PISA-PKI platform.

Summary

PISA targets the creation and implementation of agent based privacy protective software for next generation electronic business applications. The framework privacy being developed links to the European standard privacy legislation. Specifically the European Directive 95/46/EC, 97/66/EC and, indirectly the OECD privacy guidelines from 1980.

The PISA project aims to integrate its work into FIPA (Foundation for Intelligent Physical Agent) in the area of agent privacy, by developing a FIPA based agent framework for privacy including ontologies, agent technologies and communication patterns.

An other goal of the PISA project is to demonstrate that PET (Privacy Enhanced Technology) can be used as secure technical solution to protect privacy of citizens when using intelligent agents. However, PET introduces a set of requirements, which together with requirements coming from an agent platform constitute inputs for requirement for secure agent platform. We believe that most of those requirements can be implemented by using customised PKI. For this purpose, PISA demonstrator, integrating a PKI module will be set up in order to illustrate the feasibility of such an approach and to show the limit of performances if there are any.

The overall objective of the present deliverable is to describe the requirement and the specification of the PISA-PKI platform that has been introduced in D.11. We will show that the introduction of two registration authorities to manage PISA certificates can fulfil the overall requirement.

The deliverable is organised as follows:
In Section 1 (“Introduction”), an introduction is provided with an overall view of the work in WP3. In Section 2 (“Requirement”), the overall requirement of a PISA-PKI platform is described based on the requirements posed by the agent platform and the PET. Section 3 (“Overview of architecture”), describes the characteristics of such architecture. Section 4 (“Functional description”), gives the specifications of all services of the PISA-PKI platform. In section 5 (“Evaluation and Testing Procedures”), a list of evaluative criteria will be provided in the testing procedure of the PISA-PKI platform. The list will be based on ISO Testing Standard (ISO 9646) terminology and contains three classes of tests: basic tests, capability tests and behaviour resolution tests. Section 6 (“Implementation plan”) gives a guideline on the implementation phase and the set up of such PISA-PKI platform and gives some recommendations for future research in this area.

D.18 WP4.4 D19

“Prototyping datamining/matching components demonstrator”

Abstract

The functional design of the agent communication, simulation model of the labour market and the log activity of the agents for the data mining purposes about the agent behaviour is completed in this work package. Requirements of the agent platform and of the simulation for the data mining module are also analysed in this deliverable. At the moment we are finishing the technical design and the implementation of the planned data mining functionality of the agent platform. The matching engine is integrated through a web service to the agent platform and the performance is tested. The data mining module is designed and can be integrated after the implementation of the agent negotiations and log activity is ready.

Summary

In the PISA demonstrator a virtual job market is used as a test site for the PISA agents. This job market presents a database of vacancies to the jobseeker agent. Resumes of the jobseeker are processed by the matching engine of the job market and suitable vacancies are returned. The integration of the matching engine to the agent platform through a web service is completed and tested. The simulation of the online labour market, which includes the generation of an appropriate number of job seekers and companies with vacancies from the vacancy database, is designed in this work package.

Also the interactions between the agents in the PISA demonstrator are designed. The interactions of the agents in the PISA platform provide the data for data mining analyses. Data mining is used to control the correct behaviour of the agents by presenting a model of the normal behaviour and by detecting the deviation of the behaviour of the agents from this model. The data mining module finds rules based on the properties of the agents, which can be used at run time to detect new malicious agents. A job market advisor uses the resulting rules to provide directions to the jobseeker. The data mining module will be integrated in the agent platform using the implementation of the agent interactions and the log data of those interactions.

D.19 WP5.3 D23

“Agent User Interfaces and Documentation”

Abstract

This is the first report of the Human-Computer Interaction (HCI) research and development activities being conducted in the PISA project. The goals was to build an agent-based service that people will trust with sensitive, personal information and one that will operate according to privacy-protection requirements coming from legislation and best practices. Two different research activities were conducted. The first was to carefully examine the concept of “trust” and develop requirements and guidelines for building trustworthy agent systems. The second was to examine the privacy legislation and principles to determine the human-factors implications and consequences. The result was a process that begins with

privacy legislation, works through derived privacy principles, examines the HCI requirements, and ends with specific interface design solutions. Following this general research, specific system design requirements for the PISA Demonstrator were developed. Further, in order to demonstrate the HCI design concepts and kick-start the interface design portion of the project, a stand-alone interface demonstration was developed and demonstrated. Finally, planning has begun for usability evaluations to be conducted in 2003 that will test the PISA interface and the completed system.

Summary

This is the first report of the Human-Computer Interaction (HCI) research and development activities being conducted in the PISA project. The goal of the HCI activities in the PISA project is “to understand and develop technologies that will improve human interaction with the distributed virtual marketplace that is emerging in the information and communications industries” (from the original PISA Description of Work). More specifically, the task is to build an agent-based service that people will trust with sensitive, personal information and one that will operate according to privacy-protection requirements coming from legislation and best practices.

To meet these goals, two different research activities have been conducted. The first was to carefully examine the concept of “trust” and review what is known about building trustworthy systems. It was found that intelligent, autonomous agents have the potential to facilitate complex, distributed tasks and protect users’ privacy. However, building agents users will trust with personal and sensitive information is a difficult design challenge. Agent designers must pay attention to human factors issues that are known to facilitate feelings of trust. These include providing transparency of function, details of operation, feedback, and predictability. They must also consider factors that lead to feelings of risk taking. This means reducing uncertainty, collecting the minimal amount of information, and carefully considering the amount of autonomy an agent will have. These guidelines are being used in the development of the PISA Demonstrator, and they are also of interest to other system designers.

The second research activity was to examine the privacy legislation and principles to determine the human-factors implications and consequences. The goal of this work was to document a process that begins with privacy legislation, works through derived privacy principles, examines the HCI requirements, and ends with specific interface design solutions. This research involved a phrase-by-phrase analysis of the European Privacy Directive (95/46/EC) to determine the human behaviour requirements that were implied by the legal constructs. Interface design techniques were then outlined for each of the requirements, and specific design solutions were developed for the PISA Demonstrator. The result is a set of recommendations for implementing “usable compliance” with privacy legislation and principles. For the first time, this work specified what must be included in human-computer interfaces to satisfy the spirit of European privacy legislation and principles, and satisfy the privacy needs of the users (“usable compliance”).

These research activities have led to the publication of a journal article on building trustworthy agents and a workshop submission on deriving HCI requirements and usable compliance from privacy principles and best practices.

Following this general research, specific system design requirements for the PISA Demonstrator were developed. Some of these requirements are related to building trustworthy interfaces, and they include are features that should be present in the interface, supporting information that is required (i.e., help, documentation), overall system characteristics and capabilities, and system performance issues. Some of the requirements come from the

analysis of privacy legislation and principles. These requirements are organised around user tasks, and include registering and agreeing to service terms, creating a task, tracking a task, modifying a task, and dealing with system messages.

The next step was to determine how these requirements could be met in the PISA Demonstrator. A new technique called “Privacy Interface Analysis” was developed to describe how UML models can be created from a HCI point of view, and their role in the developing trustworthy, privacy-protecting interfaces. This work involved creating and annotating a UML sequence diagram for each use case that was expected in the application. For the PISA project, this technique was quite successful in making assumptions concrete and stimulating discussions within the development team.

In order to demonstrate the HCI design concepts developed for PISA, and kick-start the interface design portion of the project, a stand-alone interface demonstration was developed at NRC. The goal was to develop a set of WWW pages and back-end applications that would demonstrate a “look and feel” for the PISA demonstrator before the actual agent platform was available. This prototype as designed to be both stand-alone, so it could be demonstrated and tested, and also modular and well structured, so it could be integrated into the final PISA demonstrator. These design goals have been met. A stand-alone interface demonstration is available and the interface concepts and page designs have been integrated into the main PISA Demonstrator.

Planning has begun for usability evaluations of the PISA interface. Following the original project plan, there will be two rounds of usability evaluations. The first round, to take place early in 2003, will test the interface concepts and implementation in the interface prototype, and the second, to take place in the fall of 2003, will be an evaluation of the complete PISA Demonstrator.

D.20 WP1.5 D5

“PISA demonstrator documentation”

Abstract

The PISA demonstrator has the aim to proof that privacy protection of the personal data of a data subject and privacy legislation can be incorporated in a working system. This deliverable describes and documents its implementation. Its functionality and features include providing the conditions for consent and the usage of pseudo-identities. Other implemented aspects are a privacy ontology, the PISA-PKI and a logging and auditing system. The implemented demonstrator has some limitations due to decisions that were made beforehand and some that were unforeseen. Functionality tests for the demonstrator have been described and further evaluation and fine tuning will take place in the next work package.

Summary

The PISA demonstrator is a proof of concept implementation of the PISA system with the aim to proof that privacy protection of the personal data of a data subject and privacy legislation can be incorporated in a working agent based system for e-commerce and m-commerce. This deliverable describes and documents the implementation of the PISA demonstrator. It demonstrates the functionality and features in a job market application

like providing the conditions for consent by using privacy preferences and policies and the usage of 'transfer rules' to determine whether the preferences and the policies match. In the demonstrator, pseudo-identities are implemented as task agents that only perform a single task. These task agents only disclose the personal data if both the preferences match the policies and the disclosure is beneficial for the task the agent is executing for the data subject.

A privacy ontology is used by the agents in the demonstrator and a PISA-PKI has been implemented to provide the basis for trust in users, application servers, agent platforms and agents. A Logging and auditing system provides the required information required for both transparency and data subject rights.

The implemented demonstrator has some limitations due to decisions that were decided beforehand including the trusted agent platform assumption, the limitation of privacy principles and a limitation in the implementation of pseudo-identities. The demonstrator also has some unforeseen limitations including an unfinished PISA-PKI, limited usage of APS matching and secure conversations and a partial implementation of user rights.

Functionality tests for the demonstrator have been described and further evaluation and fine tuning will take place in the next work package.

D.21 WP3.5 D14

“Mechanisms for secure data exchange and storage with PKI”

Abstract

This fifth deliverable written by work package 3 gives two descriptions. First a description is given about the implementation of the PISAPKI as this is done by TNO-FEL. In addition to conventional PKI, the PISA-PKI has a special agent (RAAgent) on the agent platform dedicated to receive certificate requests. The second part describes the research done at Delft University of Technology to define theoretical boundaries of protecting privacy in an agent environment. A concrete result of this work is that three dilemmas are derived which define the boundaries of providing privacy to an agent's actions.

Summary

The work of milestone 5 for work package 3 was twofold. On one hand PISA security was implemented (work done by TNO-FEL). On the other hand more research was done on the long term approach (Delft University of Technology). These two topics can be seen as separate parts as the underlying assumptions are different. In the implementation the assumption is made that the agent platform can be completely trusted. This assumption is reasonable, because otherwise security cannot be guaranteed, therefore it would not be possible to make any claims (with respect to the legal implementation) for the demonstrator. However, for the research this assumption cannot be made, implementation and research can be seen as two different parts of the PISA-project. In this deliverable, the first part gives a description of the implementation of the PISA-PKI. Originally GlobalSign was responsible for delivering this part of the demonstrator software, however due to circumstances they were not able to implement this and this part has been taken over by TNO-FEL. TNO-FEL built a PKI simulation such that the PISA demonstrator also has this functionality. The PISA-PKI looks similar to a conventional PKI, with the main difference that the an RegistrationAuthorityAgent is added to the agent platform. This RAAgent is responsible for certificate request from the various agents. The second described part in this deliverable is the research done during milestone 5. This research was mainly focussed on defining boundaries of what is possible in protecting agent's actions. The assumption is that the agent platform cannot be considered trustworthy in the sense that it is interested in the actions, content and strategy of the agent, however it does not change the agent and it executes it correctly. The platform is capable of executing the agent as many times as possible. In this deliverable a model is presented that takes this assumption into account. Using information theory it is then possible to determine some boundaries of the possibilities of task protection. The result is that three dilemmas exist that restrict the possibilities to protect a task. In the original model, made by Shannon, one dilemma existed, but because of the assumption made in the agent environment, this dilemma is extended with two new dilemmas. Informally, these dilemmas can be described as follows. The first dilemma states that maximizing the uncertainty of the key, given input to the task, task and encrypted task, will result in a minimum of the uncertainty of the task, given the encrypted task and input. The second dilemma shows that the uncertainty of the input, given the encrypted task, shall be small when the uncertainty of the task, given the input and encrypted task, is maximized. Finally the third dilemma combines the uncertainty of the key, given the input and encrypted task, and the uncertainty of the input, given the encrypted task. When the latter is maximized, the former will be minimize. Knowing these dilemma's, it is now clear what in theory can be achieved. It has also been shown that task encryption is only possible using public key cryptography.

D.22 WP5.4 D24-1

"Testing of User Interfaces for the Privacy Agent"

Abstract

This is the second report on the human-computer interface (HCI) aspects of the PISA project and it contains the results of a preliminary usability test of portions of the interface design prototype. The purpose of the testing was to determine if the design concepts in the prototype were successful in constructing an agent technology that (1) users can use, (2) users can understand, and (3) users can trust with secure and private information. This testing involved laboratory sessions where subjects interacted with the software system on a computer and answered questions about the features and performance that they experienced. The results indicated that the prototype worked fairly well and was reasonably easy to navigate, but it had poor visual appeal. Users generally understood the concept of a personal assistant who could provide services, and most of them understood the major functions (create, modify, track, get results). The most notable problem was that users had trouble understanding what information was being protected when they completed the privacy preferences portion of the data entry screens. The users also had to rely heavily on rollover help information to understand the privacy preference terms, such as ”retention period” and ”require tracking”, and they were frustrated by having to set their preferences three times.

Summary

This is the second report on the human-computer interface (HCI) aspects of the PISA project. This report contains the results of a preliminary usability test of portions of the

interface design prototype. The purpose of the testing was to determine if the design concepts in the prototype were successful in constructing an agent technology that (1) users can use, (2) users can understand, and (3) users can trust with secure and private information.

Fifty people were recruited for the study from the Carleton University community through posters, the Introductory Psychology notice board, and also by approaching individuals in common areas. The usability testing involved sessions where subjects interacted with the software system on a computer and answered questions about the features and performance that they experienced. After completing the usability test, participants were given a 16-item questionnaire to assess the overall usability of the prototype. In addition, this questionnaire enquired about their attitudes towards the trustability of the Internet in general, Internet services, and the PISA prototype that they had just tested.

It should be noted that the participants were not representative of the general population, or of the likely user group for a real PISA service. Instead, the participants were young, Canadian college students, who may have very different attitudes towards WWW site design, Internet services, job searching, or privacy preferences than older adults or European citizens. This important limitation should be kept in mind when interpreting the results.

The usability testing of the prototype PISA interface revealed a number of strengths and weaknesses. Notable findings were:

Result 1. users were able to complete or understand an average of 17.5 out of a possible 27 (65%) major concepts or functions in the interface

Result 2. the average ease-of-use rating was 4.7 on a 7-point scale, corresponding to a rating of ”slightly easy”

Result 3. the average ease-of-navigation ratings was 3.64 on a 5-point scale, corresponding to a slightly positive rating

Result 4. the average rating of proper operation of the prototype was 3.62 on a 5-point scale, corresponding to a slightly positive rating

Result 5. 30% of the users attempted to login without registering first

Result 6. 38% of the users did not like the graphics, describing them as copied from another site. Others thought the graphics looked ”professional” and ”cute”

Result 7. 42% of the users did not like the colours in the interface, describing them as ”bland” and ”boring”. Others described them as ”soothing” or ”relaxing”

Result 8. 50% of the users thought that the visual appearance was unlike other sites that they think highly of

Result 9. 88% of the users liked the fonts in the interface because they were familiar

Result 10. 92% of the users understood the concept of a personal assistant created for them

Result 11. some users questioned the need to login as a second step after a successful registration

Result 12. users were generally able to understand the major functions presented in the prototype, such as ”create task”, ”modify task”, etc.

Result 13. users were sometimes confused by the term ”task” when used to describe a set of instructions given to the assistant

Result 14. 50% of the users could not understand why they needed to name a task being created

Result 15. users frequently did not understand the terms ”job sector” or ”company type”

Result 16. users sometimes did not know what units to use for input fields like ”desired salary” or ”desired location”

Result 17. users sometimes did not understand why they were asked to provide their ”current employer”

Result 18. 64% of users failed to associated the privacy preferences with the job search parameters

Result 19. users often did not feel that job search parameters required privacy protections

Result 20. users often had difficulty understanding the terms used in the privacy preference interface. Rollover help did provide assistance, but it should not have been necessary

Result 21. many of the users failed to notice or use the preset buttons available for setting privacy preferences

Result 22. the function of the check boxes in the privacy preferences was often unclear, such as whether checking ”other allowed purposes” had the effect of allowing the purpose or not

Result 23. users were often unclear about how to complete the ”retention period” field

Result 24. understanding of the privacy preference screens increased when contact information and resume information were entered

Result 25. a Just-In-Time Click-Through Agreement (JITCTA) to seek explicit confirmation when processing ”union membership” information failed to appear reliably, but when it did reaction was mixed. Some users appreciated the warning about the sensitive information, while others ignored the message completely.

Result 26. a JITCTA to seek final confirmation before the task agent is launched also had mixed reactions, with some users finding the pop-up box redundant and annoying

Result 27. results from the trust questionnaire revealed that, whereas only 54% of participants were willing to send personal information on the Internet at large, 84% would provide their resume to the prototype, 80% would provide their desired salary, and 70% would provide name, address, and phone number

Result 28. whereas only 34% thought that Internet services at large acted in their best interest, 64% felt that the prototype service would act in their best interest.

Result 29. 66% of the users agreed that they would ”feel comfortable depending on the privacy provided by [the prototype]”

Result 30. only 48% of the people characterized the prototype service as ”trustworthy”, with 44% of the users being ”undecided”.

These findings have led to a number of recommendations for further development of the interface:

Recommendation 1. improve on the terms used throughout the interface. Small usability tests can be conducted on the terms to ensure that potential users share the meaning that was intended by the developers.

Recommendation 2. consider using more polished and professional looking graphical images

Recommendation 3. consider a brighter, more attractive color scheme

Recommendation 4. keep to standard fonts, such as Times Roman

Recommendation 5. improve the registration process so that it is clear that users must register before they can use the service

Recommendation 6. integrate registration with login so that users are entered into the system automatically after they register

Recommendation 7. continue to highlight when information is transferred securely

Recommendation 8. the mental model of a personal assistant being created for the user is working, and users understand that the assistant can be given tasks

Recommendation 9. look for a word other than ”task” to label the process of creating a job for a personal assistant to do

Recommendation 10. move the step of naming a task to the end of the creation process, and make it clear why users must name their task (so they can differentiate them later)

Recommendation 11. remove or replace the terms ”job sector” and ”company types”

Recommendation 12. use selection menus instead of text boxes whenever a list of items is appropriate

Recommendation 13. make units of measurement clear, such as for ”desired salary” or ”desire location”

Recommendation 14. make the reason of entering the current employer clear at the time that it is entered, and make sharing the resume with the current employer a separate and explicit step

Recommendation 15. integrate the privacy preferences screen into a single screen that is completed once, after all the personal information is collected

Recommendation 16. collect the contact information first, since the need for privacy protection is clearest with this information

Recommendation 17. continue to use rollover help, but make it more prominent and easier to discover

Recommendation 18. change preference parameters so they are all yes/no choices, not check boxes whose function is unclear

Recommendation 19. make the privacy presets more prominent and a clear alternative to customized choices for each parameter

Recommendation 20. make it clear that using a privacy preset will erase any custom parameter settings

Recommendation 21. fix the implementation of the presets in the interface code

Recommendation 22. fix ”retention period” so it is either a duration for keeping the information or a date

Recommendation 23. change the JITCTAs to be integrated into the interface screens instead of being distant, separate interface windows

Recommendation 24. fix the implementation of the JITCTAs for the union membership field

Recommendation 25. improve the wording for the sensitive information JITCTA

Recommendation 26. fix the ”start over” feature so that previously entered information is saved and can be edited

Overall, the results indicate that users can use the major features of the interface, such as creating a job-searching agent. However, some of the specific features, such as controlling specific privacy preference parameters, are in need of more attention. Concerning understanding, the results clearly indicate that users have difficulty understanding the privacy preference terms used in the interface, and this is the most important characteristic to improve. Finally, users did find the service to be trustable, although it is clear that, with the problems in understanding the interface, the maximum possible trust was not created.

D.23 WP5.4 D24-2

“Testing of PISA Scalability”

Abstract

WP5 involves an investigation of issues associated with network, scalability and user interfaces for privacy enhancing technologies. The goal of this work is to produce an analysis of the design of PISA privacy enhancing technologies with a view to improving network security, scalability and usability of implementations. For part of deliverable 5.4 we overview the PISA system architecture, refine the PISA scalability models, and test the PISA scalability according to the refinement models.

Summary

In this document we examine two areas: refining PISA system scalability models, and testing PISA system scalability. The objective of this work is to provide the testing results of PISA system scalability. More specifically, we define the testing cases for PISA system scalability, and test the PISA system scalability on which the different privacy protection technologies, like PET, Onion Routing Network, etc., are employed.

Appendix E

PISA project information

This appendix gives a brief overview of the PISA project (Privacy Incorporated Software Agent). Detailed information and project publications can be found at the PISA website: http://www.pet-pisa.nl.

E.1 Contribution

PISA contributes to key action lines of the IST-programme of the EC:

> II4.1: "To develop and validate novel, scalable and interoperable technologies, mechanisms and architectures for trust and security in distributed organisations, services and underlying infrastructures".
>
> II4.2: "To scale-up, integrate, validate and demonstrate trust and confidence technologies and architectures in the context of advanced large-scale scenarios for business and everyday life. This work will largely be carried out through trials, integrated test-beds and combined RTD and demonstrations".

E.2 Goal

The objectives of the PISA-project are:

- Demonstration of PET as a secure technical solution to protect the privacy of the citizen when he/she is using Intelligent Agents (called shopbots, buybots, pricebots or just "bots", a short for robot[1]) in E-commerce or M-commerce applications, according to EC-Directives on Privacy.
- Interaction with industry and government to launch new privacy protected services. The PISA-project will produce a handbook on Privacy and PET for ISAT and a PISA-agent as shareware. Also a plan for the dissemination of the results of PISA will be produced.
- Propose a standard for Privacy Protected Agent Transactions to Standardisation Bodies.

[1] In E-commerce, "Bots" will slug It Out for Us; International Herald Tribune, 21 August 2000

E.3 Results

PISA contributes at building a model of a software agent within a network environment, to demonstrate that it is possible to perform complicated actions on behalf of a person, without the personal data of that person being compromised. In the design of the agent an effective selection of the presented Privacy-Enhancing Technologies (PET) will be implemented. We label this product as a Privacy Incorporated Software Agent (PISA).

The PISA demonstration model is planned to be a novel piece of software that incorporates several advanced technologies in one product:

- Agent technology, for intelligent search and matching;
- Data mining or comparable techniques to construct profiles and make predictions;
- Cryptography for the protection of personal data, as well as the confidentiality of transactions.

Additionally the project involves:

- Legal expertise to implement the European privacy legislation and the needed development of new rules and norms;
- System design knowledge to turn legal boundary condition into technical specifications;
- Advanced software-programming skills to implement the privacy boundary conditions.

In order to prove the capability of the PISA-model, we propose to test it in a model environment in two cases in e-commerce that closely resembles a real-life situation.

Appendix F

PISA Project consortium

The PISA consortium consisted of the following partners:

- TNO-FEL Physics and Electronics Laboratory
  Oude Waalsdorperweg 63
  P.O. Box 96864, 2509 JG The Hague, The Netherlands
  www.fel.tno.nl
  Project coordination, Privacy-Enhancing Technologies

  TNO-TPD Institute of Applied Physics
  Stieltjesweg 1
  P.O. Box 155, 2600 AD Delft, The Netherlands
  www.tpd.tno.nl
  Intelligent Software Agents Platform and PISA-demonstrator

- College bescherming persoonsgegevens (CBP)
  Netherlands Data Protection Authority (NDPA)
  Prins Clauslaan 20
  P.O. Box 93374, 2509 AJ The Hague, The Netherlands
  www.cbpweb.nl
  Privacy Protection and Legal Issues

- Delft University of Technology, Faculty of Information Technology and Systems, Information Theory Group
  Mekelweg 4
  P.O. Box 5031, 2600 GA Delft, The Netherlands
  www.its.tudelft.nl
  Cryptography

- Sentient Machine Research
  Baarsjesweg 224
  1058 AA Amsterdam, The Netherlands
  www.smr.nl
  Data Mining, Data Matching and Cases

- FINSA Consulting, Italsoft
  52, Corso del Rinascimento, 00186 Rome, Italy
  www.finsa.it
  Intelligent Software Agents and Multimedia Development

- National Research Council Canada
  Institute for Information Technology
  Montreal Road, Building M-50
  Ottawa, Ontario Canada K1A 0R6
  www.nrc.ca
  Network, Scalability and User Interfaces

- GlobalSign
  Ubicenter
  Philipssite 5
  3001 Leuven, Belgium
  www.globalsign.net
  Public Key Infrastructure

PISA Project consortium

COLLEGE BESCHERMING PERSOONSGEGEVENS

zerøknowledge

National Research Council Canada

TU Delft
Delft University of Technology

GlobalSign
TRUST ON THE NET

SENTIENT MACHINE RESEARCH

FINSA consulting

TNO

College bescherming persoonsgegevens
Prins Clauslaan 20
Postbus 93374
2509AJ Den Haag
Phone: +31 (0)70 3811300
Fax: +31 (0)70 3811301
Email: info@cbpweb.nl
Internet: www.cbpweb.nl

ISBN 90-74087-33-7