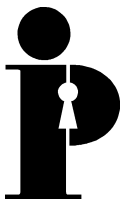
Intelligent Software Agents:
Turning a Privacy Threat into a Privacy Protector

Information and Privacy
Commissioner/Ontario
Canada

Registratiekamer
The Netherlands

April 1999



Information and Privacy Commissioner/Ontario

2 Bloor Street East, Suite 1400
Toronto, Ontario, Canada M4W 1A8
416-326-3333
1-800-387-0073
Fax: 416-325-9195
TTY (Teletypewriter): 416-325-7539

Registratiekamer

Sir Winston Churchilllaan 362
P.O. Box 3011
2280 GA Rijswijk, Netherlands
Tel. 011 (31) 70-3190190
Fax 011 (31) 70-3940460

J.J. Borking
B.M.A. van Eck
P. Siepel

With contributions from:

P.J.A. Verhaar, H.A.M. Luijf, M. Struik (TNO Physics and Electronics Laboratory – The Hague)

and

A. Cavoukian, G. Keeling, D. Duncan (Information and Privacy Commissioner/Ontario –Toronto.)

Achtergrondstudies en Verkenningen 13

Registratiekamer, The Hague. ISBN 90 74087 13 2

Intelligent Software Agents and Privacy
The Hague, 1999

This study was conducted in close co-operation with TNO-FEL, The Hague.

ISBN 90 74087 13 2

Druk: Sdu Grafisch Bedrijf

Preface

Currently, strenuous efforts are underway to develop software that will act as our 'agents' in the future. Increasing concerns about information overload and the pace of modern life have made these 'intelligent agents' an appealing concept. The notion of having an agent that would serve our needs and act on our behalf, to manage the day-to-day activities of our lives, much as a trusted personal servant would, is viewed not only as an advantage but a necessity in the warp-drive, networked world that we now live in. We wish to raise a note of caution, however, because such agents may also pose a serious threat to the privacy of their users – intelligent agents operate by accessing a detailed personal profile of the user, which enables them to execute their user's wishes. The potential loss of control over one's profile and the prospect of having the details of one's life accessed by unauthorized third parties looms like a black cloud over any potential benefits that may accrue. These issues are fully explored in the text of this report.

This is the second joint study ever undertaken by two organizations charged with the mandate of privacy protection in their respective jurisdictions, namely the Netherlands and Ontario, Canada. Our first report published in 1995, *Privacy-Enhancing Technologies: The Path to Anonymity*, marked a turning point towards seeking technological solutions to privacy (in addition to strengthening legislative efforts). Not only does the present study again demonstrate the benefits of international co-operation on a subject that touches the lives of all of us, it also clearly demonstrates that issues of privacy protection are indeed global in nature, no longer bound by national borders.

Our two organizations herald this as another opportunity to examine an emerging area of technology, one that holds the prospect of both promise and peril, and to do so from a privacy perspective. While technologies themselves may be privacy neutral, the manner in which they are used can easily affect privacy, either for the good or the bad – enhancing privacy or eroding it even further. Intelligent software agents have the potential to provide a valuable, much-needed service in the future. By reviewing the privacy aspects of this technology now and building privacy into the design criteria at the developmental stages, those responsible for creating these agents will maximize their ability to serve us all.

Peter Hustinx
President, Data Protection Authority
The Netherlands

Ann Cavoukian, Ph.D.
Information & Privacy Commissioner
Ontario, Canada

Abstract

At this moment, efforts are under way to develop software agents capable of handling 'information overload' and 'network overload.' These 'intelligent' agents, as they are known, 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 organization 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 (PETs), agents can protect the internally stored personal data against certain threats. Agents and PETs can also be used to help users search for locations where their personal data are stored, so that they can exercise their rights as laid down in international laws and treaties to control the collection of that personal data.

This report presents the results of the second joint study by the Dutch Data Protection Authority (in Dutch the 'Registratiekamer'), The Information and Privacy Commissioner/ Ontario (IPC), and the Netherlands Organization for Applied Scientific Research – Physics and Electronics Laboratory (TNO-FEL). With this study, the Registratiekamer, the IPC, and TNO-FEL have attempted to identify possible threats to the privacy of individuals resulting from the use of agent technology. Secondly, the study sought to identify ways of applying PETs to agent technology in such a way as to eliminate the impact of these threats.

The IPC and the Registratiekamer would like to recognize the efforts of Industry Canada for its on-going support, assistance, and dedication to promoting the development and use of PETs.

One key point to emerge from this report is that by applying clear procedural methodologies, PETs can be 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; and the Identity Protector.

These technologies can be wrapped around the agent or integrated in the agent. A combination of integration and wrapping is also possible. The technologies can also be used to build an infrastructure of trusted components. To find the right solution, designers, developers, suppliers, users, or providers will need to use a checklist of design criteria.

Executive summary

The hectic demands of modern lifestyles, combined with the growing power of information technology, is driving the development of products designed to help people get through their busy and information-laden days. Numerous services are currently available, ranging from simple push technologies such as 'PointCast,' which brings information to your doorstep by 'narrow-casting' or filtering information based on an individual's specified interests, to sophisticated systems that allow for the 'personalization' of network user sessions and the tracking of user activities. Collaborative filtering of a user's 'clickstream' (or history of Web-based activity), combined with neural networks, which look for detailed patterns in a user's behaviour, are just beginning to emerge as powerful tools used by organizations of all kinds.

While the majority of these technologies are, at the moment, essentially benign in design and utility, they are indicative of the types of products that are being developed. The end result culminates in the creation and development of Intelligent Software Agent Technologies (ISATs). Intelligent Software Agents (often referred to simply as agents or 'bots' [short for robot or knowbot]) are software programs, at times coupled with dedicated hardware, which are designed to complete tasks on behalf of their user without any direct input or supervision from the user.

At first glance, agent technologies appear to hold out great promise for automating routine duties and even conducting high level transactions. However, upon greater reflection, it becomes clear that ISATs could present a significant threat to privacy relating to the wealth of personal information in their possession and under their control. Accordingly, it is highly desirable that their development and use reflect international privacy standards in order to safeguard the personal information of their users.

Agent technology finds its roots in the study of Artificial Intelligence (AI), human computer user interface design, and software engineering. Currently available agents (which are typically 'smart' Internet search engines developed to support commercial Web sites) are beginning to display the characteristics envisioned by the visionaries, but do not yet have the full functionality that will lead to the great promise and threat identified in this paper.

While personalization programs and smart search engines may cause significant privacy concerns, we believe that, for the time being at least, privacy is not likely to become as major an issue with these services, due to their currently limited scope (recognizing the potential privacy concerns raised by data mining). Thus, this paper focuses on the emerging Intelligent Software Agent Technologies (ISATs) that extend these programs into new realms of automated activity and ungoverned behaviour.

User Profiling

The functionality and utility of user agents, lies in what they can do for the user. Remember, their whole raison-d'etre is to act on one's behalf and function as one's trusted personal servant, serving one's needs and managing one's day-to-day activities. Their powers are constrained by a number of factors: the degree of their software sophistication, the number of services with which they can interact, and, most importantly, the amount of personal information that they possess about the user.

It is this issue of 'user profiling' that is at the core of the privacy risk associated with the use of ISATs. Typically, an ISAT user profile would contain a user's name, contact numbers, and e-mail addresses. Beyond this very basic information, the profile could contain a great deal of additional information about a user's likes and dislikes, habits and personal preferences, frequently called telephone numbers, contact information about friends and colleagues, and even a history of Web sites visited and a list of electronic transactions performed.

Because agents could be requested to perform any number of tasks ranging from downloading the daily newspaper to purchasing concert tickets for a favourite singer, the agent is required to know a great deal of information about the user. In order to function properly, ISATs must also have the following characteristics:

- mobility, or a connection to a communications network;
- deliberative behaviour, or an ability to take an action based on a set of criteria;
- the following three abilities: to act autonomously, co-operatively, and to learn.

Depending upon the levels of security associated with the user profile, this information may be saved in a plain text file or encrypted by password, PIN, or biometric means. However, the security of the data residing within the agent is only one part of the concerns regarding privacy. The arguably more significant concern revolves around the dissemination of the information during transactions, and in the general conduct of the agent's activities on behalf of the user. Of even greater concern is the situation where the ISAT may not be owned directly by the user but is made available (rented, leased) to the user by an organization in order to assist in accessing one or more services.

This raises another risk, quite real, namely that the user's activities may be accessed, monitored, and disseminated to unauthorized third parties or otherwise subjected to 'data mining.' The user is required to place a certain degree of trust in his or her agent – that it will perform its functions correctly as requested. However, this trust could well come with a very high price tag, one that the user may have no knowledge or awareness of – the price of his or her privacy.

Privacy Threats Posed by Agents

There are 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).

As an agent collects, processes, learns, stores, and distributes data about its user and the user's activities, the agent will possess a wide variety of information which should not be divulged unless specifically required for a transaction. In the course of its activities, an agent could be required or be forced to divulge information about the user that he or she may not wish to be shared. The most important issue here is one of openness and transparency. As long as it is clear to the user exactly what information is being requested, what purpose it is needed for, and how it will be used (and stored), the user will be in a position to freely make decisions based on informed consent.

There are many possible scenarios whereby the agent may release information about the user that, while seemingly innocuous, could be potentially significant and adversely affect the user. An agent's visit to an online service to determine the cost of concert tickets would generate a wide variety of clickstream data that could ultimately jeopardize the user's privacy. For example, the online service could log an incoming request from the agent for Bob Smith (bsmith@open.net) looking for tickets to a particular concert on a particular night. In and of itself, this information seems relatively innocent. However, if the agent passes along the user's home address so that he can receive the tickets in the mail, then a more sensitive piece of information has been released into the wider cyber-ether about the user, linked to a particular interest.

There are numerous examples of how various types of information could be released, knowingly or unwittingly, which could result in significant repercussions for the user. If organizations interacting with the agent do not follow internationally accepted Fair Information Practices[1], then any accumulated information could be passed on to other groups, often without the knowledge or consent of the user. As a result, in no time at all (remember, practically everything is online these days), Mr. Smith is receiving offers from music companies, book clubs, magazines, and travel firms. While a great deal of this data mining already occurs today, the potential for even more significant data collection and exploitation about the most sensitive personal matters (one's finances, relationships, illnesses, insurance, employment, etc.) could result if this information was in the hands of one's agent.

Thus, if the use of agents could lead to so many potential privacy risks, one wonders if it could be possible for anyone to use ISATs safely. We believe this still remains within the realm of possibility, and that the answer lies with the use of privacy-enhancing technologies.

---

[1] For more information about Fair Information Practices, see the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data and the Canadian Standards Association (CSA) Model Code for the Protection of Personal Information.

The Identity Protector and Privacy-Enhancing Technologies (PETs)[2]

In and of themselves, ISATs do not necessarily pose a risk to privacy. However, the prevention of potential breaches of privacy depends on the implementation of a number of measures that can actively promote both the privacy and transparency of agent transactions.

The tracking and logging of a person's use of computer networks is a major source of potential privacy violation. Conventional information systems perform the following transactions: authorization, identification and authentication, access control, auditing, and accounting. At each phase, a user's identification is connected with the transaction. We will be suggesting that the adoption of a filter called the 'Identity Protector' (IP) will go a long way to protecting privacy. The introduction of an IP into an organization's information systems would improve the protection of the user's information by structuring the system in such a way as to remove all unnecessary linkages to the user's personally identifying information.

The Identity Protector filter can be placed between the user and the agent, preventing the ISAT from collecting any personal data about the user without the knowledge and prior consent of the user. Conversely, the IP can be located between the agent and the external environment, preventing the ISAT from divulging any personal information unless specifically required to do so in order to perform a particular task or conduct a specific transaction.

Additional technical means may also be integrated into the ISAT in order to bring even more transparency to the user in the operation of the agent, thus ensuring the user's knowledge, and if necessary, informed consent. The following provides just a brief indication of the types of processes that could be employed:

- registration, certification, and verification of the agent working methodology (to prevent any loss of control over the activities of the agent);
- clearly detailed audit logging and activity tracking of agent transactions so that the user can monitor and review the behaviour of the agent;
- the use of programs to render the user and/or the agent anonymous, or alternatively, the use of a 'pseudo-identity,' unless identification is specifically required for the performance of a transaction;
- the use of identification and authentication mechanisms such as digital signatures and digital certificates to prevent the 'spoofing' of a user or his or her agent by a malicious third party intent on committing fraud or agent theft;
- the exclusive use of data encryption technology to prevent unauthorized 'sniffing' or accessing of agent transaction details;

[2] For a more detailed discussion of these technologies, see: Hes, R. and Borking, J. (editors) e.a. (1998). Privacy-enhancing Technologies: The path to anonymity. Revised Edition. A&V-11. Den Haag: Registratiekamer.

- the exclusive use of trusted sources — the agent can be instructed to only visit sites that have been independently verified (through a variety of means such as trusted seals, audits, etc.) as having proper privacy provisions in place;
- placing limitations on an agent's autonomy so they only perform a certain range of activities. Limited activities will be permitted to be freely conducted without additional authorization; any requests for unauthorized transactions will be flagged for the user to scrutinize.

The integration of an Identity Protector and other PETs into the core of the ISAT, combined with a process that places similar technology between the agent and the external environment, would result in a system that enjoyed the maximum protection against threats to the user's privacy. Further, the development of a trusted source infrastructure would promote confidence (ensuring confidentiality) in the use of agents to conduct automated transactions of all types. The recognition and certification, by an independent certification and auditing council, of organizations that followed Fair Information Practices would greatly encourage the use of ISATs.

Contents of the Report

This paper is organized into six sections. The Introduction details the objectives of the study undertaken and initiates the discussion of Intelligent Agent technologies. Part two discusses agents in greater detail, including the general theory of intelligent agent technology and an exploration of the interactions between users and agents, including several concrete examples of agents in action. The following section discusses the legal issues surrounding the use of agents and introduces the concepts of Fair Information Practices, including anonymity, transparency, and control. The fourth section identifies various types of threats posed by agents acting on behalf of their users, and agents acting on behalf of others. Section five details various PETs, how they could be used when applied to agents, and the possible outcome of using these PETs. The final section outlines the conclusions and recommendations contained in the report. This is followed by a list of abbreviations, references, then appendices. The first appendix is a detailed description of Identity Protector technology, while the second addresses the process of reasoning and learning by intelligent agents.

Conclusion

While the development and use of ISATs is still in its infancy, it may well become a part of every-day-life within a few years. In order to safeguard the personal information of agent users, the privacy implications of such technologies must be addressed now, at the design stage — that is our primary recommendation: We encourage the developers of intelligent agents to ensure the proper means by which the privacy of users may be protected and

control maintained by data subjects over the uses of their personal information. As agents become more and more 'intelligent,' further consideration will have to be given to how individuals and organizations may react to the automation of tasks, especially higher-level functions that were once handled exclusively by humans.

The most important point to note is that the exchange of personally identifying information will not be necessary for all activities. Different tracks should be delineated at the early stages of development, with access to personally identifying information strictly limited to specific instances where it is clearly required. The reason for relinquishing personal information (unless clearly evident) must be demonstrated. Unprotected external agents will undoubtedly jeopardize the privacy of other users unless various forms of PETs are implemented to prevent the unauthorized collection and dissemination of personal information. Masking an agent's identity or having it act anonymously on one's behalf, where possible, would result in the greatest protection.

The increasing power and sophistication of computer applications offers both tremendous opportunities for individuals (beyond merely their capacity as consumers), but also significant threats to personal privacy. Provided that due care is taken to protect privacy in the creation of intelligent agent technology, and appropriate recourse is available in the event that agents erroneously or forcibly divulge confidential personal information (e.g., having it snatched away by external rogue agents), then ISATs will join the family of useful information technology products available to assist people in softening the mounting pressures of modern-day life.

Table of Contents

1 Introduction

Nowadays, computers are commonly used for an increasing range of everyday activities. Most of these activities are based on the acquisition of information. At present, users still interactively and directly initiate all actions needed for a computer to execute tasks.

Due to the enormous and fast-growing amount of data that is available, sometimes referred to as 'information overload,' it is impossible to sustain the current way users interact with their computers. Instead of direct user-initiated interaction, users and computers should be engaged in a co-operative process, a process in which both users and computers can initiate actions to accomplish tasks. In this way, a computer could continue its activities without waiting for the user to activate it. With the use of software agents[1], computer systems are capable of executing actions with minimal interference by their users. This gives the user more time to spend on other activities. The idea of software agents is not new. Over the years numerous researchers have been working on this issue.

For purposes of this study, a software agent is defined as a piece of software that acts on behalf of its user and tries to meet certain objectives or complete tasks without any direct input or direct supervision from its user. The lack of supervision however, could lead to undesirable actions, such as the violation of the privacy of individuals concerned. Besides acting independently on behalf of their users, agents may have a number of other properties, e.g., mobility, reasoning, learning, co-operation (negotiation) with other agents, and cloning.

It is still unclear what commercial direction this technology will take because the technology is still in the early stages of development. There are, however, two identifiable trends. The first trend concerns software agents that have been or are being developed to help people perform routine tasks; tasks that people could probably do themselves if they had the time. These software agents are far from 'intelligent'. The first wave of products is hitting the market now. The other trend is driven by researchers in the field of Artificial Intelligence (AI), who are trying to combine Artificial Intelligence with the agent philosophy to create an 'intelligent' agent. A great deal of research and development effort has and will continue to be devoted to the field of intelligent agents, but no products are commercially available yet. A good example of such research and development is the agent named Phil produced by Apple Computer. Phil appears in the promotional video called 'The knowledge navigator' made by John Sculley.

It is sometimes desirable to control the introduction of new technologies. There are several ways of doing so. One is by means of government regulation, where the application of new technologies has to meet current government rules. Due to the pace of present-day developments, the formulation of new government regulations governing new technolo-

[1] As will be explained later in the report, many different names are used for agents. Throughout this report 'software agent' or just 'agent' will be used as a generic term.

gies practically always lags behind. Most government regulations are therefore adopted or amended after these new technologies have been accepted by industry. Consequently, the responsible government organizations are responding reactively. This leads to a steadily widening gap between new technologies and adequate government regulation.

One of the independent organizations that executes and interprets government regulations designed to protect the privacy of all Dutch inhabitants is the Dutch Data Protection Authority (in Dutch: the 'Registratiekamer'). The Registratiekamer is a privacy protection agency that oversees compliance with the jurisdiction's privacy laws. It is the responsibility of the Registratiekamer to warn all Dutch consumers of, and protect them against, the possible consequences of technologies, especially new technologies, for their privacy. Its policy is to propose privacy regulations governing new technologies before these technologies hit the market. The Registratiekamer also looks for (new) technical measures, such as cryptographic tools like (blind) digital signatures, that could enforce these privacy regulations. Such technical measures to preserve the privacy of individuals are called Privacy-Enhancing Technologies (PETs). The Registratiekamer therefore needs to study new technologies, and the impact these technologies might have on the privacy of individuals. Hence, one of the roles of the Registratiekamer is to act as an adviser and partner in the development of these technologies.

The Information and Privacy Commissioner/Ontario (IPC) has a mandate under the Ontario Freedom of Information and Protection of Privacy Acts to research and comment upon matters relating to the protection of privacy with respect to personal information held by government organizations in Ontario. In the fulfilment of that mandate, the IPC is concerned that all information technologies, if not properly managed, could represent a threat to the privacy of the residents of Ontario.

TNO Physics and Electronics Laboratory (TNO-FEL) is one of the three institutes that form TNO Defence Research, part of TNO, the Netherlands Organization for Applied Scientific Research. With a long history in research and development, application, and integration of new defence technologies, TNO-FEL has traditionally devoted the majority of its resources to meeting the demands of the Netherlands Ministry of Defence and Armed Forces. Today however, TNO-FEL participates in international as well as national defence programmes and operates in close co-operation with technological institutes, industry and universities both inside and outside the Netherlands.

The Registratiekamer and the Information and Privacy Commissioner/Ontario, in association with TNO Physics and Electronics Laboratory (TNO-FEL), conducted an earlier study of technologies that could improve the privacy of individuals in 1995. The results of that study are published in (Hes, R. and Borking, J., editors, 1998, revised edition). A summary of the results of this study is included in Appendix A. Two of the technologies studied were blind digital signatures and Trusted Third Parties (TTP's).

The Registratiekamer believes that (intelligent) agent technologies could jeopardize the privacy of individuals. However, these technologies may also be used to protect the privacy of individuals. A special privacy software agent could be developed to exercise the rights of its user, and to enable this individual to protect him or herself against privacy intrusions with the aid of a PET. Therefore, the Registratiekamer decided to study the privacy aspects of these agent technologies pro-actively. Once again, this study was conducted in close co-operation with TNO-FEL.

1.1 Objective of this study

The objective of this study is to investigate the possibilities of protecting the privacy of a consumer in an agent-based environment. To accomplish this objective the following question needs to 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:

- Why do software agents exist?
- What are software agents?
- How do software agents work?
- Where are software agents applied?
- What is the future of software agents?
- What implementations of software agents already exist?
- What is privacy?
- What threats to privacy can be attributed to software agents?
- What (technical) measures are there to eliminate or reduce the impact of these threats?

1.2 A brief outline of the report

Chapter two provides answers to the following questions: Why do software agents exist? What are software agents? How do software agents work? Where are software agents applied? What is the future of software agents? What implementations of software agents already exist? Chapter three describes the rudiments of privacy. Chapter four answers the question: What threats can be attributed to software agents? Chapter five describes the PETs (technical measures) that could reduce or eliminate the impact of the identified threats. Chapter six provides a summary of the conclusions and recommendations arising from the study.

2 Agent technology

Software agents have their roots in work conducted in the fields of software engineering, human interface research and Artificial Intelligence (AI). Conceptually, they can be traced back to the late seventies when their predecessors, the 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 Multiple Agent Systems (MAS), one of the three branches of distributed AI research, the others being Distributed Problem Solving (DPS) and Parallel Artificial Intelligence (PAI) (Nwana, H.S. and Azarmi, N., 1997). Technically, they exhibit many of the properties and benefits common to distributed AI systems. These properties include:

- *Modularity*. A modular programming approach reduces the complexity of developing software systems.
- *Speed*. Parallelism, the concurrent execution of co-operating 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 the knowledge level*. Utilization of AI techniques allows high-level messaging.
- *Others*. These include maintainability, reusability and platform independence.

2.1 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. In 1982, the volume of publicly available scientific, corporate and technical information was doubling every five years. By 1988 it was doubling every 2.2 years; by 1992 every 1.6 years. With the rapid expansion of the Internet (the Net), one can expect this rate of increase to continue. It may now be doubling 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 should one organize 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 problem. 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 problems 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 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 the agent technology, designed to solve the problems.

2.2 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.' 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.'

For the rest of this study, the first trend mentioned in chapter one, the development of agents to help people perform routine tasks, will be ignored.

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, and 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), which is described in the following metaphor (Abdu, D. and Bar-Ner, O.), which describes the co-operative, mobile, and learning processes that are present in a PDA.

Metaphor:

> 'Bruce awoke instantaneously at 6:00 AM 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 AM or at 5:00 PM and that Hava personally preferred morning meetings. Doris confirmed 10:00 AM, 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 AM, 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 using Hava's VISA number. He returned with a big smile on his face and notified Hava of her evening plans. At about 1:00 PM, 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 to sell the tickets to someone else and then buy tickets to Sense and Sensibility (Hava just loves Emma Thompson). At 7:00 PM, 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, like:

- Hava's address, telephone numbers, and electronic mail address(es);
- Hava's relatives;
- other acquaintances of Hava;
- not only Hava's reading habits and preferences, but also all other habits and preferences.

It will be obvious that an infinite amount of personal data could be recorded in a user-profile. The only restrictions are the technical restrictions of the agent's memory (its capacity) and its processing capacity.

2.3 Agent ownership

Agents could be owned by individuals or organizations. These agent-owners can use their agents to carry out tasks to fulfil their owners' purposes; and to offer agent services to individuals or organizations that are not in a position to own an agent. In the metaphor provided above, the agent Bruce could be owned by its boss, Hava, but Hava could also have hired Bruce from a company or organization 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, let's say 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 will reduce the costs of the workstations and, therefore, increase the possibilities for financially less well-endowed individuals to use the Net. These workstations leave practically no room for the installation of additional (local) software, including user-owned agents. People who use these services will 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.

2.4 Interaction between users and agents

In activating an agent, a 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. By delegating tasks, responsibilities and competence, the user loses control over a considerable amount of the agent's activities. It is therefore imperative that the user can trust the agent that is used, 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 (Norman, D.A., 1994). If the user doesn't know what his or her agent is doing, or isn't content with the way the agent works, he might consider never using this agent again. There should be some kind of agreement between the agent and the user, as there is between secretary and boss where the agreement is often based on mutual understanding. The agreement will be tried out for a probationary period. During this period both parties can decide whether they accept the agreement. A user should have a description of the working method of the agent in order to learn more about it before using the agent. In this way, the user knows 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 will know about him or her. As in the relation between master and servant, where the servant knows practically everything about his or her master, it becomes very important that he handle this information with the utmost discretion. The servant will be engaged on account of this quality. It is essential that agents have the means to protect the privacy of their users. These means take the form of Privacy-Enhancing Technologies (PETs), which will be discussed in chapter 5.

Another way to increase trust is to provide assurances about the level of control individuals have over their agents. To give users the feeling that they are in control of their agents (Norman, D.A., 1994), the following items have to be taken into account: a description of the way the user and the agent interact, safeguards to prevent unwanted situations, and the setting of accurate expectations to minimize false hopes.

2.5 Classification of Agents

Agents can be classified according to the specific properties, or attributes, they exhibit (Nwana, H.S. e.a., 1997 and Abdu, D. e.a.). These include the following:

- *Mobility.* This refers to the extent to which an agent can move around a network. This leads to a distinction between static and mobile agents. Sometimes this includes cloning to distribute sub-tasks 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 the following three:
  - *Autonomy:* reflects the ability of agents to operate on their own, without immediate human guidance, although the latter is sometimes invaluable.
  - *Co-operation:* refers to the ability to exchange high-level information with other agents — an attribute which is inherent in multiple agent systems (MAS).
  - *Learning:* refers to the ability of agents to increase performance over time when interacting with the environment in which they are embedded. In Nwana, H.S. and Azarmi, N. 1997, agents combining several of the primary attributes are referred to by different names again: autonomous agents that co-operate 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 will be given. Agents may be classified, for example, by their pro-active versatility — the degree to which they pursue a single goal or engage in a variety of tasks. Furthermore, one might attribute social abilities to agents, such as truthfulness, benevolence, and emotions (anger, fear), although the last is certainly controversial. One may also consider mental attitudes of agents, such as beliefs, desires, and intentions (in short: BDI's).

By combining these properties and attributes (Caglayan, A.K. and Harrison, C.G., 1997), hybrid agents and heterogeneous agents can be constructed. 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 way that they interact via a particular communication language.

2.6 Intelligence and Agency

By varying the extent of the learning attribute, an agent's intelligence can range from more to less intelligent. By varying the extent of the attributes autonomy and co-operation, an agent's agency can vary from no inter-activity with the environment to total inter-activity with the environment.

In this case, intelligence relates to the way an agent interprets the information or knowledge to which it has access or which is presented to it (Caglayan, A.K. and Harrison, C.G. 1997). 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. With reasoning, preferences are combined with external events and external data in a decision-making process. The highest form of intelligence is called learning. Learning can be described as the modification of behaviour as a result of experience. Appendix B gives a more detailed description of reasoning and learning.

Agency relates to the way an agent can perceive its environment and act on it (Caglayan, A.K. and Harrison, C. G., 1997). 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 agency is user representation, where an agent has a model of the user's goals or agenda. In subsequent phases, the agent is able to perceive, access, act on, communicate, and interact with data, applications, services, and other agents. These phases are called: data inter-activity, application inter-activity, service inter-activity, and agent inter-activity.

By combining intelligence and agency, it becomes possible to indicate where 'intelligent' agents are positioned. Figure 2.1 illustrates this. Agents that are positioned in the shaded area are more or less 'intelligent' agents.

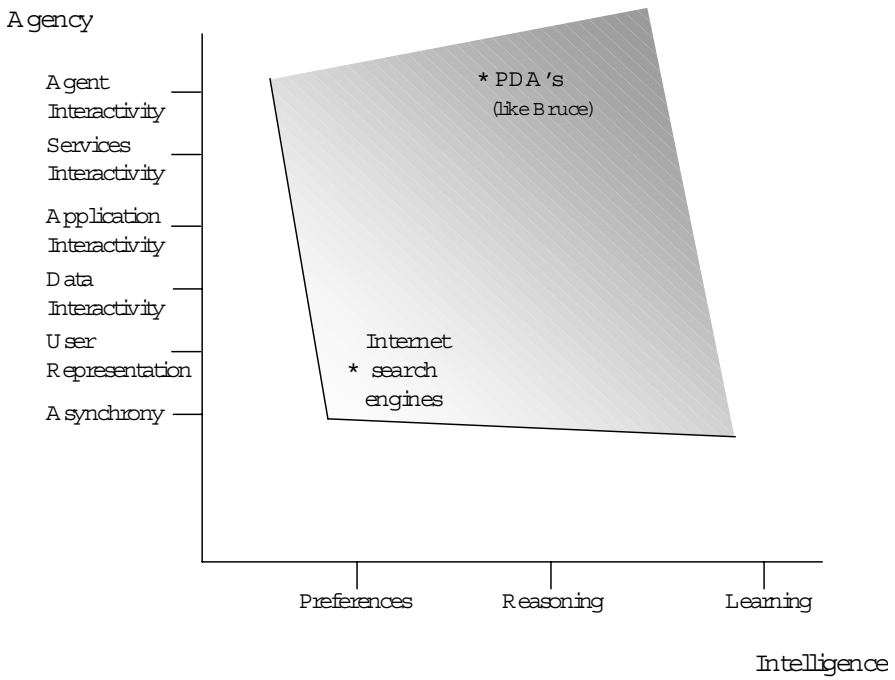


Figure 2.1: The position of intelligent agents in relation to intelligence and agency.

2.7 Examples of Agents

Agents can be classified according to the properties they exhibit. This section will 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, 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 for personalized 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 realize that interface agents can only be effective if the tasks they perform are inherently repetitive (otherwise, agents will not be able to

learn) and if the behaviour is potentially different for different users (otherwise, use a knowledge base).

> 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 — a technique similar to word-of-mouth recommendations.

Mobile agents. Mobile agents reduce communication costs and overcome limitations of local resources. Decentralization of the selection process prevents unwanted information being sent over networks, thus economizing on network utilization. As an example, imagine one has 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. 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 co-ordinating 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, or a wrapper which injects modified communications software into the existing software.

It should be noted that some of the examples given above refer to prototype systems. Moreover, the introduction and further development of agent systems usually involve having to overcome technical, as well as social, legal, and ethical hurdles.

2.8 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 tasks on behalf of their users. Moreover, there were some 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. In this section, the actual workings of agents will be addressed.

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.

From the conceptual description given above, it is clear that a general understanding of the working of agents requires an understanding of their internal workings, as well as an understanding of the mechanisms that underpin the communications behaviour among agents. It should 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 2.2. This figure describes the processing of received messages (input) via some function 'f' into performed actions and transmitted messages (output).

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

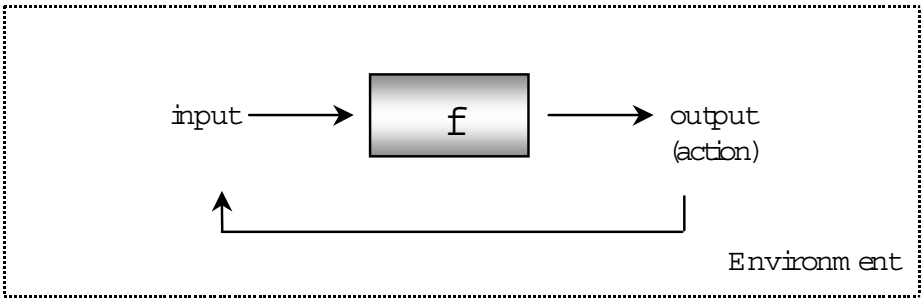


Figure 2.2: Black box model of an agent.

The mapping f is not directly controlled by an external authority: the agent is autonomous. The distinctions between agent models stem from differences in the nature of mapping f 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 2.2. Hence, there is a gap between the legitimacy of this model and its usefulness. In order to derive a more appropriate model, one that captures the essential characteristics of agents, it is useful first to discuss the parent disciplines of agent design in more detail. These disciplines are control theory, cognitive psychology, and classical AI planning theory.

2.8.1 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 should be estimated. A distinction is made between feed-forward and feedback control. With feedback control, actions are based on the monitoring of the behaviour of the environment and changes therein; with feed-forward control, the reaction of the process to be controlled can be predicted.

The analogy with agents is that agents recognize situations, derive goals, and engage in planning and scheduling in order to act according to the goals set. The interpretation of recognized situations employs (symbolic) reasoning using a specific perception of the actual model of the world, and hence 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 they derive information (truthfulness, benevolence, etc.).

2.8.2 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 sub-processes:

- 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 vs. instinct). 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. In making decisions, the agent needs to take into account instigating forces, consummatory forces, inhibitory forces, and forces of resistance. Implementing the DTA requires solving bottlenecks concerning how to represent and recognize situations and motives, and how to model the (inter)dependencies of the four forces at play so that smooth behaviour patterns result.

2.8.3 Classical Artificial Intelligence planning systems

The problem-solving behaviour of agents is viewed as a sense-plan-act (input, function (f), output: from Figure 2.2) 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 realize a goal of the agent in question. Most classical Artificial Intelligence (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 some of the constraints were relaxed, e.g., planning under resource limitations, interleaving of planning and execution. 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 realizing transitions in a discrete state space.

2.8.4 The agent model

With this insight into the parent disciplines of agent design, a detailed model of the internal workings of an agent can be given. First a layered model of an agent that is both deliberative, reactive, and co-operative will be given. Such an agent consists of three layers. These layers are: the behaviour-based layer, the local planning layer, and the co-operative 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 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 co-operative planning layer.

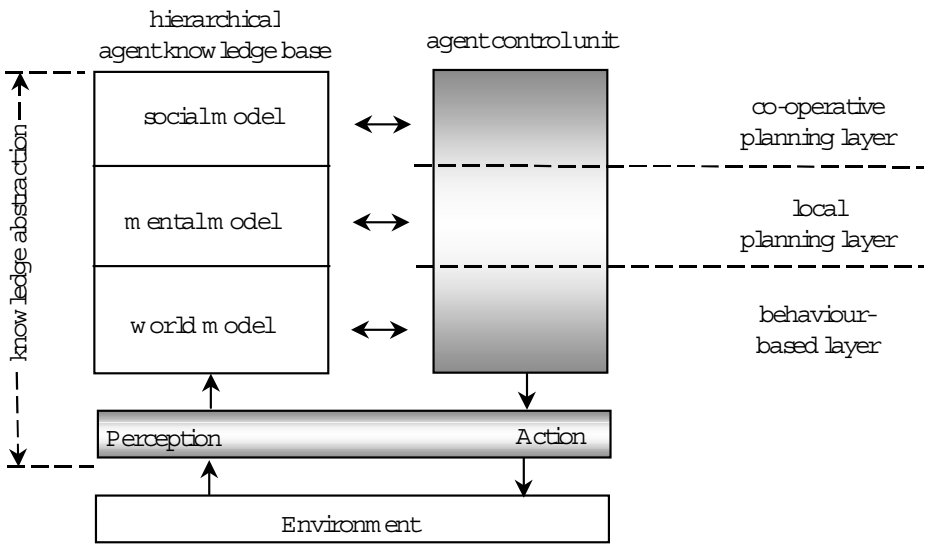


Figure 2.3: The layered design of an agent that is both deliberative, reactive, and co-operative (cf. (Müller, J. P., 1996)).

The models are hierarchical, see figure 2.3. The agent will 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 co-operate with these other agents and to avoid conflicts with them.

An example of an automated loading dock could illustrate this model. In this automated loading dock, autonomous agents are carrying out tasks, like moving goods from the storage facilities to the transportation area and vice versa. At first the agents need to know what their tasks are and what these tasks mean. The agents will use the world model to interpret their tasks and fix their behaviour. The mental model will 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. To avoid running

into each other and getting blocked by the other agents, the agents need to know what the other agents are doing. With the social model, agents could each decide to make a random move to get out of this situation. Another possibility for resolving the 'traffic jam' could be that the agents exchange their goals, which will lead to mutually agreed actions.

Another way to describe an agent that combines deliberative, reactive, and co-operative properties and attributes, is given in figure 2.4 (Müller, J.P., 1996).

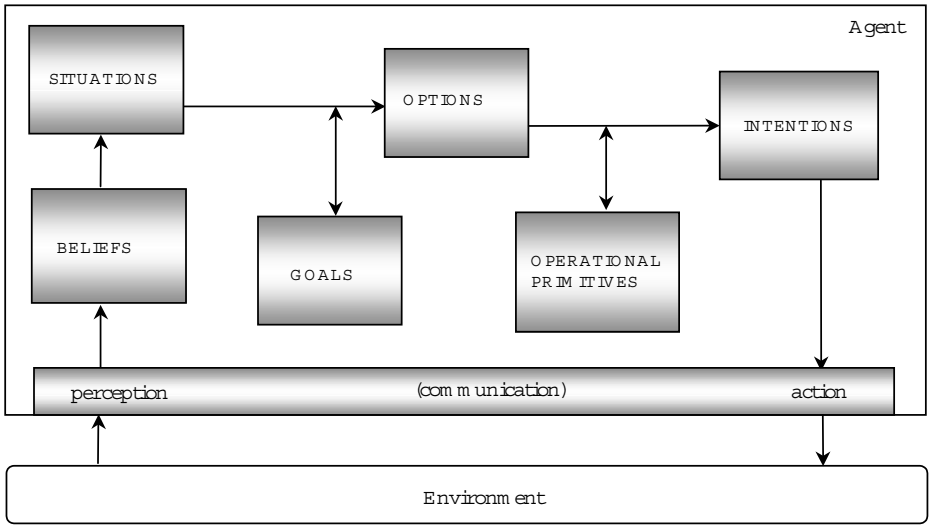


Figure 2.4: The agent model — a conceptual model of an agent that combines deliberative, reactive, and co-operative properties and attributes (cf. (Müller, J.P., 1996)).

Figure 2.4 depicts the agent control unit (see figure 2.3) 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.

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. According to the three models, there are three classes of situation. First, there are the behavioural situations, which are a subset of the agent's world model. Second, there are the situations describing local planning. These situations are based both on the world model and on the mental model. Last, there are the situations that describe co-operative 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 co-operative

goals. Reaction goals are goals that are triggered by external events. These goals require a fast reaction and are of short-term base. Local goals refer to the goals of the agent itself. Co-operative 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 or software techniques 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.

2.8.5 The communication model

To make communication possible, all agents need to be active in the same network infrastructure. 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 authenticate other 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 2.2 presents a black box model of the way agents communicate with their environment independently of their exact location in the environment.

2.9 The future of software agents

Some say 'intelligent' agents are the stuff of science fiction, but is this really so? No, we don't think so — the future is close at hand. Many current developments in R&D laboratories deal with the problems of intelligence, adaptive reasoning, and mobility. Nevertheless, people have 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. In practice, most agents available today are used to gather information from public networks, like the Net. 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. So, today's agents are simple in comparison to those that are being planned (Norman, D.A., 1994 and Minsky, M. e.a., 1994). However, already in 1990, philosophers (De Garis, H., 1990) warned that in the near future (50 years), it is likely that computer and communication technology will be capable of building brain-like computers containing billions of artificial neurons. This development will allow neuroengineers and neurophysiologists to combine forces to discover the principles of the functioning of the human brain. These principles will then be translated into more sophisticated computer architectures and intelligent software agents. This development might well become, in the 21st century, a global political issue. A new branch of applied computer ethics is needed to study the profound implications of the prospect of life in a world in which it is generally recognized to be only a question of time before our intelligent software agents and computers become smarter than we are.

3 Legal aspects

The activities of agents will lead to numerous ways of processing personal data, such as the personal data an agent provides to other agents during transactions, the personal data an agent collects for its user, and the data the agent-provider can extract from the use of his or her agent.

To protect the privacy of the persons involved, it is important that these personal data are used with care, that they are necessary for legitimate purposes, that the data will not be disclosed to the wrong persons, and that personal data are not processed without the knowledge of the persons concerned. Therefore, the use of agents and the processing of personal data have to meet certain conditions. These conditions derive from the principles of privacy, which are laid down in most laws and international treaties.

The privacy principles for the processing of personal data are laid down, for example, in the:

- European convention for the protection of human rights and fundamental freedoms;
- European convention for the protection of individuals with regard to the automatic processing of personal data;
- Organisation for Economic Co-operation and Development (OECD) guidelines;
- EU directive on the protection of individuals with regard of the processing of personal data;
- Canadian Standards Association (CSA) Model Code for the Protection of Personal Information.

From all these conventions, regulations, and directives, we can abstract the following privacy principles, which are strongly interrelated. Users, designers, developers, suppliers, or providers of agents have to take these principles into consideration when designing an agent, and must do so in the light of the fundamental right of an individual to decide when and under what circumstances personal data may be revealed. Violation of these principles may lead to severe liability and court actions, either by the data subject or by the data protection authorities involved.

Anonymity

Before data are processed, it has to be established whether it is necessary to process personally-identifying data instead of anonymous data. It is often possible to process data in such a way that they cannot be linked to specific persons. If processing can be conducted anonymously, there are no legitimate grounds for processing personally-identifying data. An important principle to keep in mind is if a user can choose to remain anonymous off-line, that same choice should also be available on-line. The user must have the possibility of remaining anonymous when the agent goes out on its search, or when it is possible for

the agent to fulfil its task without giving any personal information relating to its user. The wish to remain anonymous perhaps limits the agent's possibilities, but it is important that the user is aware of this and makes the decision about what information the agent is allowed to disclose, in what cases, and to whom. If it is possible for the agent to fulfil its task without giving any information relating to its user, that possibility has to be offered. Otherwise there will be no legitimate grounds for the processing of these personal data.

Purpose specification

The personal data must be collected for specific, explicit and legitimate purposes. Before collection, the purpose has to be described. This purpose also determines the further processing of personal data.

Legitimate grounds

There must be legitimate grounds for the processing of personal data. In some cases, this means that the data subject must give his or her consent. It is also possible that the personal data are necessary for the performance of a contract to which the data subject is party.

Compatible use

The personal data cannot be further processed in a way incompatible with the purposes as specified before collection.

Proportionality

Processing of personal data must be necessary and in proportion to the purposes identified. This means that, while there may be legitimate grounds for processing, the personal data must be necessary for the processing so that there is a balance. The interests that are to be served by the processing have to be in proportion to the invasion of the data subject's privacy. If it is possible to serve the same interest in a way that is less intrusive, then this way has to be followed.

Data quality

The personal data has to be adequate, relevant, and not excessive in relation to the purposes for which they are collected. The personal data must be accurate, and where necessary, kept up to date. Reasonable steps must be taken to ensure that data which are inaccurate or incomplete are erased or rectified.

Data subject's rights

Data subjects have the right of access and the right of correction if the personal data are incomplete or inaccurate. The right of objection means that in certain cases the data subject

can object to the processing of his or her personal data. Another right is the right not to be subject to a decision which produces legal effects concerning the subject or which significantly affects him or her and which is based solely on automated processing of data intended to evaluate certain personal aspects. Agents can also be used to execute the data subjects' rights. A data subject can send his or her agent to gain access to the personal data relating to him. After accessing the data, the subject, i.e., the user, has the opportunity of rectification, etc.

Transparency

During processing it has to be clear to the data subject what is going on with the data relating to him. He has to be informed not only about the fact that personal data are being processed, but also about the purposes for which the personal data are processed, and the identity of the person or organization processing the data, the controller. There must also be transparency regarding the organizations to which the personal data were disclosed, and about the way in which the data subject can exercise his or her rights (of access etc.). The obligation of notification can also contribute to greater openness towards the public and the data subjects.

Security

Appropriate technical and organizational measures must be taken to protect personal data against accidental or unlawful destruction or loss, alteration, unauthorized disclosure, or access, and against all unlawful forms of processing. Security is very important. If the controller wants to make it possible to disclose certain personal data to certain agents, this information must only be given to agents that are authorized. No other agents must be able to gain access, so there have to be technical and organizational measures to prevent this.

Accountability

Basic principles of law require that someone should be held accountable for what agents do. In the cases discussed in this report, the user and/or provider of the agent would be considered the controller, and therefore responsible for personal data under his or her control. Data subjects must have the opportunity to apply to the controller for damages. However designers, developers and suppliers of agents are responsible for the proper design and equipment of agents, and can be held liable if the use of agents would cause damage to data subjects, users, and/or providers of agents, or indeed to anyone else.

Supervision

An independent supervisory authority is necessary to verify whether or not these principles are observed. Where such an authority is lacking, there should be sufficient other means for independent verification of whether agents are reliable and are used in conformity with privacy principles.

4 Agent threats

An agent can collect, process, store, and distribute data. Some of these data could be personal data about individuals. It is important to note that some of these personal data may become privacy-sensitive, after combining or processing several pieces of personal data that may not, in themselves, be considered sensitive. In other words: part of these personal data may be privacy-sensitive, while the part that isn't may become privacy-sensitive when the agent processes it. As long as the agent doesn't 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 regulations to ensure privacy.

From the perspective of a specific user, an agent can have two roles. The first role is to act on behalf of its user. The second role is to act on behalf of others (i.e., organizations, other individuals, or their representatives). The two roles can generate different kinds of threats to privacy.

By describing an agent's task, as well as the steps that are necessary to fulfil that task, it will be possible to indicate when and how the privacy of the individuals concerned could be at risk. To illustrate this point, a description of an agent buying flowers for its user's mother-in-law will be given. First, the description will be restricted to a limited number of participants in the agent's environment. This description will be referred to as the first example. By adding more actors to the environment in the first example, the second example will become more complex. The third and last example will involve a user who uses an agent that is provided by an agent-provider. After the description of the examples, the privacy-threats will be enumerated.

4.1 First example

In this example the agent has to order flowers for the anniversary of the user's mother-in-law. The restrictions for the first example are:

- the actors participating in this example:
  - the user 'Mr. Jones;'
  - the mother-in-law 'Mrs. Smith;'
  - Mr. Jones' agent;
  - Mrs. Smith's agent;
  - a number of stores that sell flowers;
  - agents that represent these stores;

- no other actors are involved in this example;
- all communication will be by existing means. There will be no other actions by these means than to transport data that is sent by the actors, or to transport agents, if necessary.

In figure 4.1 the actors that are participating in the first example are sketched. Figure 4.1 also outlines the interactions between the actors. These interactions are described below.

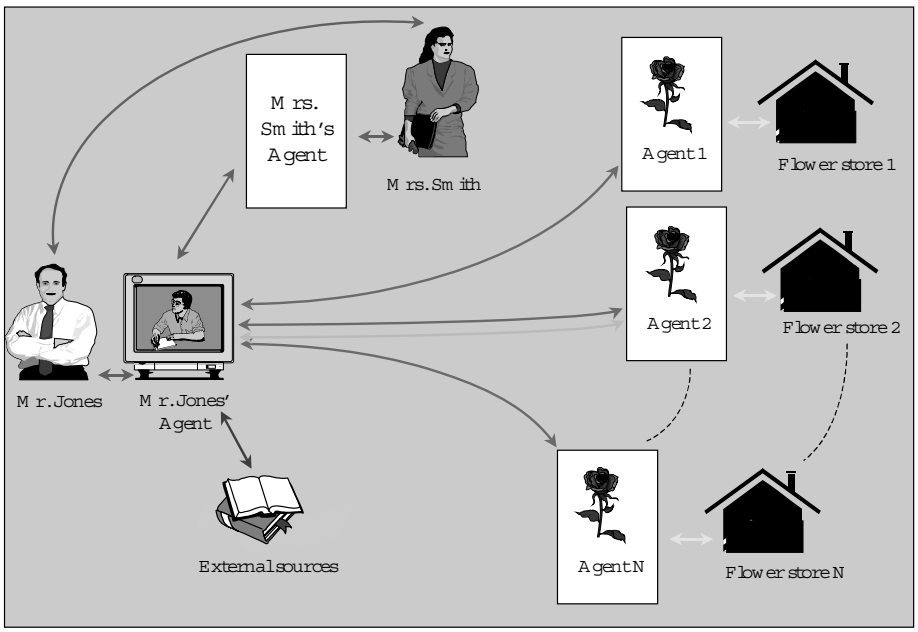


Figure 4.1: A visualisation of agents automating the buying process of flowers from Mr. Jones to Mrs. Smith in an environment with certain restrictions.

The agent should have an understanding of what the task entails. Hence, it must be able to determine the actual meaning of the individual words the task phrase is composed of and their mutual relationships. Obviously, this requires that the agent and its user have a common understanding of what the individual words mean. In this example, the agent has to know the meaning (perception) of the words flowers, mother-in-law, order, and so on. The beliefs (the way Mr. Jones thinks about Mrs. Smith), situations (which activities have to be carried out), goals (if it's an anniversary, then take care of it every year), and options (which stores are available; what types and prices of flowers can be selected) that are present in the agent will generate the necessary actions to accomplish the given task. With the operational primitives and the intentions (selecting one store) the agent will schedule the generated actions and execute them.

For this example the following actions can be generated:

- information gathering;
- the selection of a store where they sell flowers;

- the selection of the flowers;
- the actual buying of the flowers.

For the successful completion of this task, Mr. Jones' agent needs to have the following information relating to Mr. Jones and his mother-in-law (so this information would contain personal data):

- the nature of the relationship between Mr. Jones and the person for whom it has to buy flowers: In this example, that person is Mr. Jones' mother-in-law, Mrs. Smith. The agent also needs to know, for instance, whether Mr. Jones and Mrs. Smith are on friendly terms, dislike each other, or something in-between. This will allow the agent to decide the size of the bouquet and how much money to spend on the flowers;
- the occasion: why the agent has to buy the flowers. Different occasions call for different types of: flowers, bouquets, wrapping paper, and/or gift cards. Weddings, birthdays, and other celebrations require different bouquets of flowers than funerals. The date of the occasion is also needed so that the flowers are delivered at the right moment and not 3 months after the event, for example. In the case of a gift card, the agent needs to know what to write on it;
- the available budget and how it has to pay for the flowers. The agent needs to know what is the maximum amount that can be spent on accomplishing its task. The agent also needs to know how to settle the account;
- the delivery address. The agent needs to know the address of Mrs. Smith so that the flowers will be delivered to the right address;
- preferences concerning the flowers. The agent needs to know if Mrs. Smith is allergic to certain flowers, or if she likes or dislikes certain flowers.

The agent could obtain this information by consulting Mr. Jones, Mrs. Smith, and/or Mrs. Smith's agent. This interaction is visualised in figure 4.1 with blue arrows. In doing so, the agent needs to take into account that some information could be sensitive, e.g., that Mr. Jones dislikes his mother-in-law. What if Mr. Jones' agent asks Mrs. Smith, or her agent, what her preferences are for flowers, and at the same moment she, or her agent, can read Mr. Jones' profile, where it is written that he dislikes his mother-in-law.

Mr. Jones' agent also needs to know where to go to buy the flowers. This information can be obtained (blue and purple arrows in figure 4.1):

- from its existing knowledge base that has been programmed by the designer, Mr. Jones, or the agent itself, from tasks that have been carried out in the past;
- by consulting Mr. Jones;

- by consulting telephone books or yellow pages; therefore the agent must know what a telephone book is and what it can do with it;
- by consulting other sources of information, such as the Net.

In consulting external sources, like the Net, the agent has to be careful with the personal data it carries within the profile of its user or its knowledge base. There is no need for the agent to communicate these personal data, or parts of them, unless Mr. Jones is accountable for one of his agent's actions (Hes, R. and Borking, J. editors, 1998, revised edition). The agent has to protect the personal data stored in the user-profile or in the knowledge base against unauthorized access.

The choice of stores will be simple if Mr. Jones prefers a specific store. However, it is possible that there is no special reason to choose a specific store. The agent then needs to investigate which store will provide the best buy. The selection criteria the agent uses can vary widely. Some examples of these criteria might be: the location of the store, the layout of the ad in the yellow pages or on the Net, the selection of a store because of its membership of a trade organization, or because of a classification of these stores assigned by a particular organization or an interest group. The agent could also communicate with one or more stores, or their agents, before it makes the decision which store to buy from (red arrows in figure 4.1). If the agent does so, personal data about its user or other persons the user has connections with may be exchanged. According to (Hes, R. and Borking, J. editors, 1998, revised edition), there is no need to exchange personal data if the agent is only gathering information about stores that sell flowers.

The selection of the store could depend on the selection of the flowers, because some stores may be out of a particular flower that is preferred by the mother-in-law. It is therefore possible that the selection of the flowers takes place before the selection of the store. In this example, the selection of the flowers will take place after the selection of the store.

The choice of flowers or bouquet will depend on the available budget, the preferences of the mother-in-law, the feelings of Mr. Jones towards his mother-in-law (which will determine how nice the flowers will be), and the availability of certain flowers, wrapping paper, and gift cards. The agent has to exchange the preferences, and the occasion, so the store assistant can show the agent the flowers, the bouquets, the gift wrapping, and the gift cards to choose from. There is no need for the agent to exchange any other personal data than the preferences and the occasion.

When the agent has made a decision about the flowers and any other items, the actual purchase can be made. For that, it needs to exchange the delivery address, the date, and possibly the time of delivery. It also needs to exchange the text that has to be put on the gift card, and the way Mr. Jones wants to pay for the flowers. The actual exchange of personal data from Mr. Jones' agent to the agent that represents the chosen flower store is visualised in figure 4.1 with the green arrow. Any communication between the flower store agents and the flower stores is depicted with yellow arrows. Using blind digital signatures, Mr. Jones or his agent can also pay anonymously (see Hes, R. and Borking, J. editors, 1998, revised edition).

For Mr. Jones, this decision is an automated process. In principle, Mr. Jones is unaware of the completion of this process because the agent is autonomous. The agent makes decisions to accomplish its task without telling Mr. Jones why or how the task is completed. This could lead to undesirable situations, for instance if the agent decides to exchange privacy-sensitive information about Mr. Jones, or Mr. Jones' mother-in-law. In addition, the agent should notify Mr. Jones of this activity so that he does not act surprised when Mrs. Smith thanks him for the lovely flowers she received from him.

4.2 Second example

As previously mentioned, by adding more actors to the environment, the first example will become more complex. The following actors will be added:

- agents that work on behalf of the communication-means suppliers (so-called network-agents);
- agents that work for other users that are using the communication-means;
- the users that delegated work to the above-mentioned agents.

Figure 4.2 illustrates the second example.

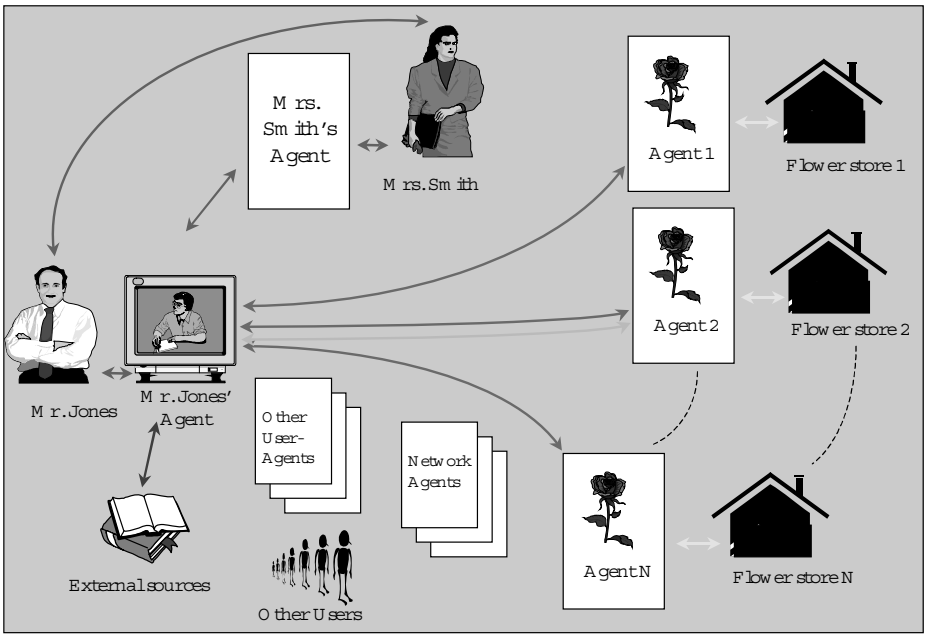


Figure 4.2: A visualisation of agents automating the buying process of flowers from Mr. Jones to Mrs. Smith in an environment without restrictions.

The network-agents could have more than one task to accomplish. One of these tasks could be to make sure the communication-means, i.e., communication networks, are working properly. Another task could be to monitor all traffic on the networks for management control. Without knowing it, these agents may collect information about individuals that could be seen as personal data, e.g., where the same telephone number is called frequently it is very possible that there is a relationship between the caller and the called person. From this relationship it could be possible to extract more personal data about both persons, e.g., the identities of these persons.

All agents involved in this example can communicate with all actors involved (including each other). Therefore, they can obtain personal data about every individual from every available source. These sources could be: the privacy domain of an individual, the user-profile or knowledge base present in an agent, or various databases where personal data are collected. All agents can also create new data, when combining and processing previously collected data, by means of automated decisions.

When Mr. Jones' agent is communicating with an agent there is no need to exchange personal data, unless Mr. Jones' agent has to account for some action it wants to execute or has executed (see also paragraph 4.1, appendix A, and (Hes, R. and Borking, J. editors, 1998, revised edition)). On the other hand, agents or organizations that collect personal data in user-profiles, knowledge bases, or databases need to protect these data against privacy-threatening situations (see chapter 3). So it should be difficult for anyone at all to obtain these personal data from agents or databases.

4.3 Agent-providers

Mr. Jones could use an agent that is provided by an agent-provider. This might be the case if Mr. Jones is not in a position (for financial or technical reasons) to own his own agent, or if it's of no use for Mr. Jones to own an agent (because he would only use the agent for a limited number of tasks). In using an agent provided by an agent-provider, Mr. Jones will run the risk of providing (some) personal data to the agent-provider through the agent, see figure 4.3.

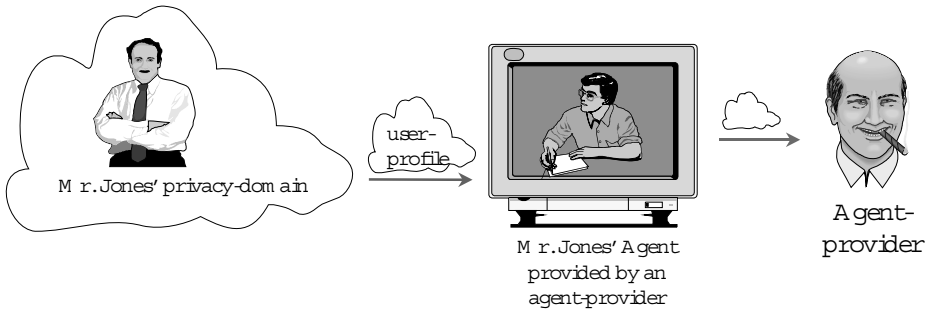


Figure 4.3: Possibility of leaking information to an agent-provider.

If the agent-provider wants to process traffic data and user-profiles, the question arises whether it is necessary to process that personal data. Furthermore, tracking someone's actions and searching for his or her digital traces is definitely a violation of his or her privacy. A person must have the freedom to give orders to his or her agent without the provider (or another organization) knowing precisely what the orders and results were. It is therefore very important to determine the interest the provider has in tracking the user's data in relation to the violation of privacy. Can one detect an interest more important than the violation of privacy? If not, the provider must consider another way to serve his or her interests.

4.4 Threats caused by agents acting on behalf of a user

As described earlier in chapter 2, people can delegate tasks, responsibilities, and competence to agents. Therefore, the agents need to have specific personal data about these individuals so that they can achieve the required results. These data are kept in a 'user-profile'. This user-profile contains part of the collection of personal data about a specific individual, which is called the privacy domain. With regard to the previously mentioned examples, if Mr. Jones is an individual who wants to delegate tasks to an agent, the agent will create a user-profile of Mr. Jones, see Figure 4.4.

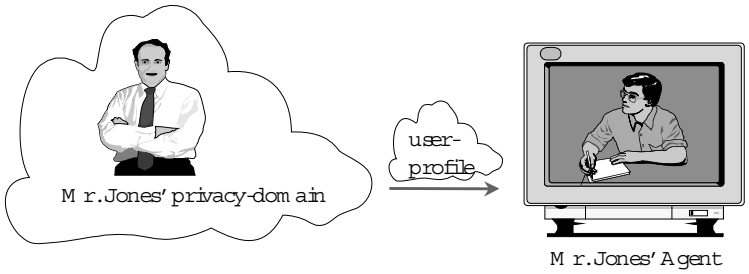


Figure 4.4: Personal data flow: the user-profile stored in the agent contains (part of) the personal data that is stored in the privacy domain.

4.4.1 Loss of control

In delegating tasks to his agent, Mr. Jones could lose control over the activities that are executed to get the right results. It is possible that some of the activities could be illegal, and therefore harm the user. The agent will also try to maintain and extend the user-profile so that its performance will improve. This could mean that the agent knows more about its user than it was supposed to know. Maintaining and extending the user-profile can take place by obtaining and processing (personal) data. Personal data that are processed or combined can become 'privacy-sensitive.'

The grounds on which (automated) decisions are made need to be clear to the user of the agent. If not, automated decisions could become threats to the user, because privacy-sensitive data could be exchanged, or illegal decisions or actions could be executed by the agent.

4.4.2 Agent-providers

Mr. Jones may endanger his privacy when giving (part of) his personal data to an agent that is provided by an agent-provider, see also figure 4.3.

4.4.3 The agent exchanges personal data with its environment

Mr. Jones' agent will communicate with its environment in order to secure the results Mr. Jones asked for. In doing so, the agent can exchange (personal) data with the other participants in the environment. Personal data only needs to be exchanged if the agent or Mr. Jones have to account for their actions. This is represented in figure 4.5.

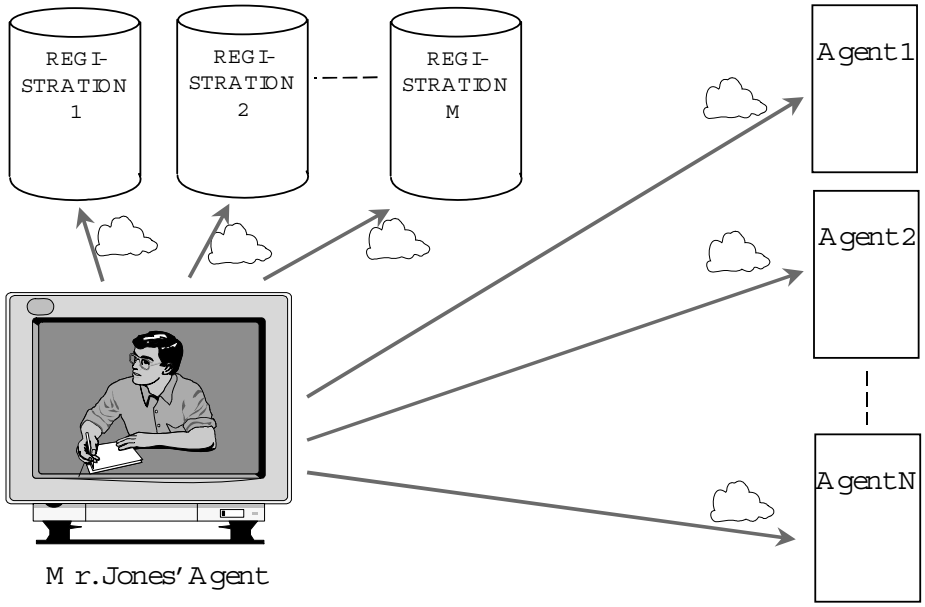


Figure 4.5: Possible ways for Mr. Jones' agent to leak personal data in its environment

The agent can also exchange (personal) data with its environment when it is delegating tasks to other agents or when it is cloning itself. When Mr. Jones' agent delegates tasks to other agents or clones itself, it doesn't know whether these other agents or clones are trustworthy. Nor does it know whether these other agents or clones are as respectful of privacy as it is itself. These other agents or clones could exchange Mr. Jones' personal data with their environment, either intentionally (e.g., Trojan horses or viruses), or unintentionally (e.g., because these agents are weaker than the agents they have to deal with).

An important privacy principle in this interaction is transparency. The user will be interested more in the results produced by his or her agent than in the procedures the agent has followed. But if the agent leaves traces and other personal data such as preferences, hobbies, profession, etc., the user will not be aware of this until he is later confronted with the fact. So the possibility that the agent will give away certain personal data must first of

all be made known to the user. The user must, in addition to the anonymity principle, have the opportunity to choose. If a deal cannot be made without the agent giving away personal information, it must be transparent to the user that certain information will be disclosed. There should preferably be the possibility to report to the user what information has to be disclosed, for what order, for what purposes, and to whom.

If the agent's actions required to fulfil its assignment cannot be performed anonymously, the user must be aware of this. It is up to the user to decide what he wants. Transparency means not only that the user knows that personal data are disclosed by the agent, but also that he is aware of the steps the agent takes during the performance of its task.

4.4.4 The agent runs into an agent that is in disguise

This possibility is closely related to the threats described in the next part of this chapter. It could be that Mr. Jones' agent runs into another agent that presents itself as someone else, and uses the rights of that person. This could mean that Mr. Jones' agent will exchange (personal) data with this agent-in-disguise, thinking that this agent is entitled to this information because it is using the rights of someone else.

4.4.5 The agent runs into an agent that is more powerful (e.g., has more processing power or capacity) than itself

This possibility is also closely related to threats that will be described in the next part of this chapter. When communicating with its environment, the chance exists that Mr. Jones' agent will meet an agent that is 'stronger' and 'smarter' than itself. This could mean that Mr. Jones' agent will release (personal) data without wanting to do so.

4.5 Threats caused by agents that act on behalf of others

There could be other agents searching the environment for information and making decisions with this information (automated decisions). This information may contain personal data about Mr. Jones. These data could be obtained by performing a traffic flow analysis or by communicating with Mr. Jones, the agent that works for Mr. Jones, or databases that contain personal data about Mr. Jones.

4.5.1 Agents can perform traffic flow analysis

When performing a traffic flow analysis, an agent can obtain personal data about Mr. Jones or others that are staying in the environment.

4.5.2 Agents can enter the privacy domain of a user and collect whatever they want

One way for agents to obtain personal data about a specific user is to enter the privacy domain of this user. This could be a threatening situation if this user can't do anything about it (see figure 4.6).

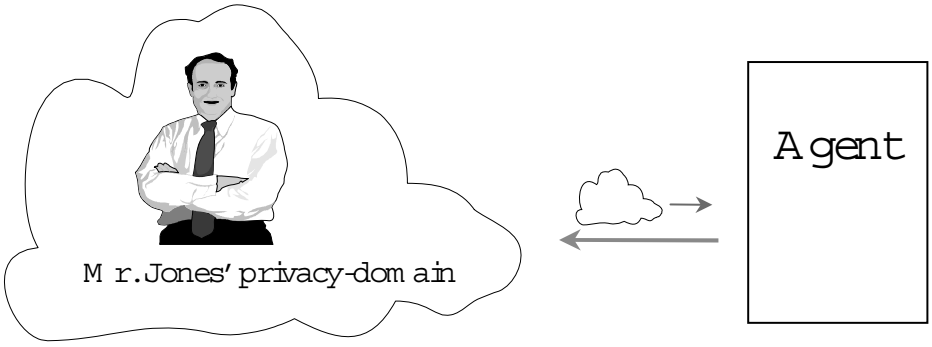


Figure 4.6: Flow of personal data when an agent enters the privacy domain.

4.5.3 Agents can enter databases and collect personal data

Another possible way of obtaining personal data is to enter databases where personal data are stored. 'Smart' (intelligent) agents, in particular, can try to break through the security that is built into the databases (see figure 4.7).

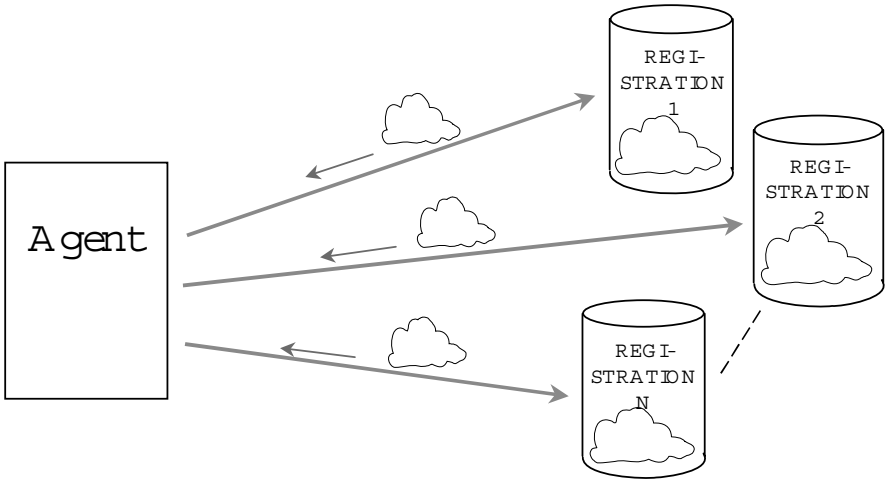


Figure 4.7: Flow of personal data when an agent enters databases.

4.5.4 Agents can steal personal data from Mr. Jones' agent

Another possibility is to obtain personal data by stealing this data from Mr. Jones' agent. This could be due to differences between the agents. For instance, some agents may be smarter than other agents and so mislead (or deceive) the inferior agents, see figure 4.8.

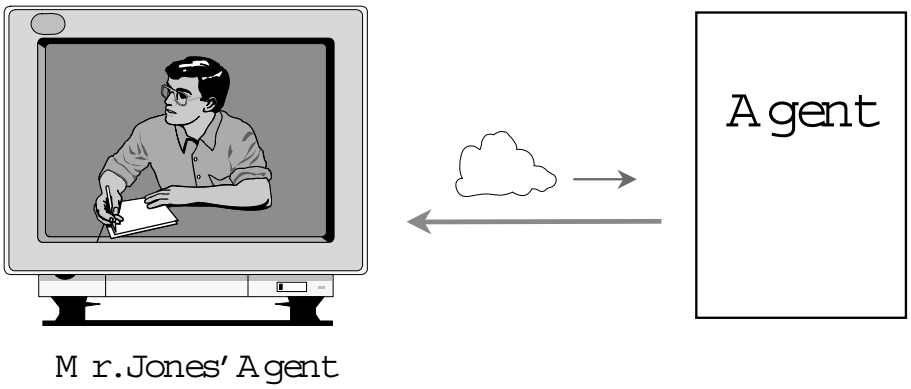


Figure 4.8: Flow of personal data when an agent misleads (deceives) the user-agent.

5 Privacy-Enhancing Technologies

As stated in chapter 3, privacy regulations and privacy guidelines have been drawn up by various governments and international governmental organizations. Tough measures are needed to enforce these regulations. Up to now, these have taken the form of inspections or audits to verify whether all organizations that collect personal data are complying with the privacy regulations. These inspections are time-consuming and therefore expensive. The Registratiekamer searches for technologies capable of replacing inspections for enforcing the privacy regulations. The IPC is also on the lookout for such privacy-enhancing technologies (PETs).

This chapter will describe 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 the Identity Protector (IP) described in *Privacy Enhancing Technologies: The Path to Anonymity* (Hes, R. and Borking, J. editors, 1998, revised edition). That edition also describes the technologies to implement the IP. These technologies are defined as PETs. The IP controls the exchange of the user's identity within an information system (for a more detailed description of the IP, see appendix A). In an agent-based environment the IP can be used in two ways:

- between the user and the agent, see figure 5.1(a);
- between the agent and the external environment, see figure 5.1(b).

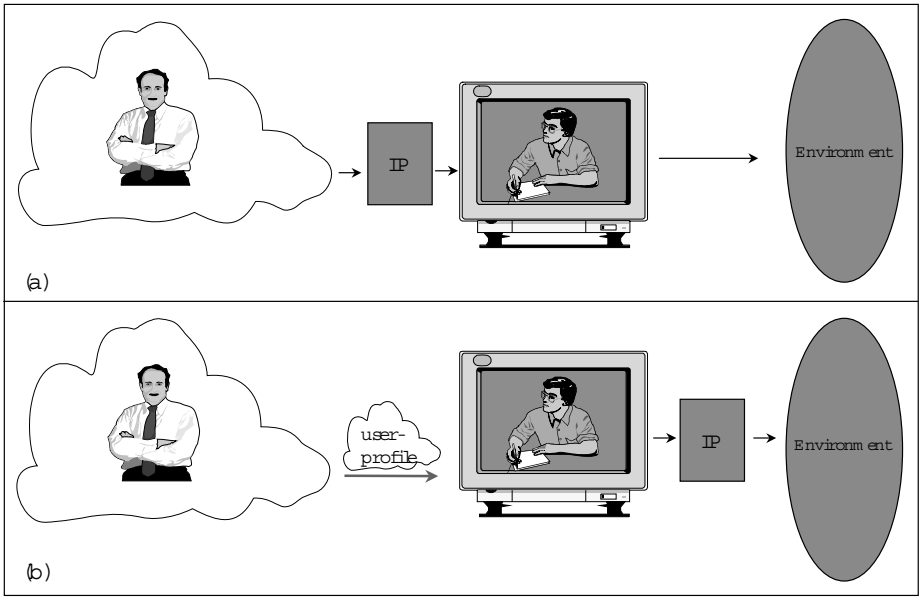


Figure 5.1: The Identity Protector (IP) placed in an agent-based environment: (a) the IP placed between the user and the agent; (b) the IP placed between the agent and the external environment.

When the IP is placed between the user and the agent, there will be no exchange of personal data from the user to the agent without the approval of the IP and the user. In this way, the user can 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 will help the agent to protect the personal data of its user against unwanted dispersion.

The PETs described in appendix A and *Privacy Enhancing Technologies: The Path to Anonymity* to implement an IP are only capable of managing some of the threats. To manage the remaining threats, existing security technologies that are not yet defined as PETs need to be applied in such a way that they can improve the privacy of individuals. Eventually, these technologies will also be called PETs.

In the previous chapter the threats were 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 will be studied for each group. The result of this will give PETs solutions for each group of threats. By combining the PETs solutions for both groups, 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.

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 to protect it or not. This leaves the individual with the consideration of whether or not to use PETs to secure his or her agent. If the individual chooses to protect his or her privacy, he or she 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 individual has with his or her environment. This relationship can consist of political, social, public, or other kinds of interactions. If the individual decides to take an agent with PETs with a high degree of privacy protection, this will have consequences for the performance of the agent.

5.1 PETs that manage the identified threats

The following threats were 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.

5.1.1 Loss of control

Loss of control can be prevented by increasing the user's trust towards his or her agent. This can be achieved by 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 organization is the first step towards increasing a user's trust in his or her agent. A certificate of this evaluation, in combination with the 'digital signature' of the agent itself, will provide users with a guarantee that the agent 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 *Privacy Enhancing Technologies: The Path to Anonymity*.

Besides certification, a user must have the possibility to verify the correctness of the agent's working method and thereby consolidate their relationship. So, besides certification of the working method, all actions that are taken by the agent need to be logged. The user of the agent should be able to audit all logged actions.

With certification of the working method and logging and auditing of the agent's actions, the threat of loss of control can be kept under control. 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 will show the user when, where, and what personal data was exchanged with the external environment. A well implemented logging will also help the user to backtrack a decision (automated decision) made by the agent.

5.1.2 Agent-providers

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

- 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 was already described in the previous paragraph. The agent's working method should 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. By inserting statements like: 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, the recording behaviour of the agent-provider can be controlled. If there is no confidence in the agent-provider at all, a user can place an IP between the user and the agent, as illustrated in figure 5.1(a).

5.1.3 Exchanging personal data

Personal data only needs to be exchanged if the agent or its user has to account for a specific action that has been executed (see the 'buying flowers' example in the previous chapter, and *Privacy Enhancing Technologies: The Path to Anonymity*). For all other actions, the agent should remain anonymous. The technologies that can be used to secure 'anonymity' can be found in *Privacy Enhancing Technologies: The Path to Anonymity*.

Other technologies can also be used to protect the privacy of a user of agent-technologies. Initially, if an agent needs to exchange personal data with its environment for the purposes of a specific action, these personal data must be accurate. 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. In other words: 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 will be calculated. This signature will accompany the (personal) data when it is exchanged. To verify the integrity of the (personal) data, the verifying party needs to execute the same

calculation. This will result in another digital signature. If this signature is identical to the signature received with the data, the data received are authentic. If they are not identical, the (personal) data has been modified. As long as the data are accompanied by the signature, authenticity can be verified. Data that are not accompanied by a signature can not 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. So, when using parity-bits or checksums to guarantee the integrity of personal data, the parties involved need to be trusted. With a digital signature it is still easy to change the personal data, but the signature can not be changed as easily, so 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 will 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 *Privacy Enhancing Technologies: The Path to Anonymity*, integrity mechanisms, or logging and auditing mechanisms. The strength of the technologies from *Privacy Enhancing Technologies: The Path to Anonymity*, the integrity mechanisms, and the logging and auditing mechanisms will depend on the level of privacy protection that is required.

5.1.4 Malicious agents

To avoid interaction with unwanted and malicious agents, agreements have to be made between friendly and trustworthy 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 can't be kept secret, it is necessary that agents present supplementary information to authenticate themselves. In this case, the identity of the agents 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. With identification and authentication, the threat 'the agent runs into an agent that is in disguise' can be reduced. 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.

5.1.5 More powerful agents

To reduce the occurrence and the impact of the threat 'the agent runs into an agent that is more powerful (e.g., has more processing power or capacity) than itself,' the use of identification and authentication alone is not enough. 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. By granting rights to other agents, the user-agent can control all activities inside itself. This can be done with an access control mechanism.

5.1.6 Traffic flow analysis

The impact and occurrence of the threat 'agents can perform traffic flow analysis' can be minimized 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. Unfortunately, this measure neutralizes 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 *Privacy Enhancing Technologies: The Path to Anonymity*.

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.

5.1.7 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 can't 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 will decrease the risks that are generated by the threat 'agents can enter the privacy domain of a user and collect whatever they want.'

5.1.8 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, 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. If there are no questionable intentions, all parties can draw up an agreement about how to handle the personal data. Besides this agreement, the agents need to have a logging and auditing mechanism to preserve the transparency of the personal data. Transparency means that the agents need to inform the persons concerned about what personal data they collected and when and where they collected it.

5.1.9 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 'the agent runs into an agent that is more powerful (e.g., has more processing power or capacity) than itself.'

5.2 PETs placed in the generic agent model

The evaluation and certification of the working method needs to be executed in conformity with an internationally agreed evaluation and certification scheme to receive an internationally accepted certification. The working methods can be certified by independent organizations, such as the Registratiekamer.

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. Between these two extremes every combination of integrating PETs and layering PETs is possible. The two extremes are represented in figures 5.2 and 5.3.

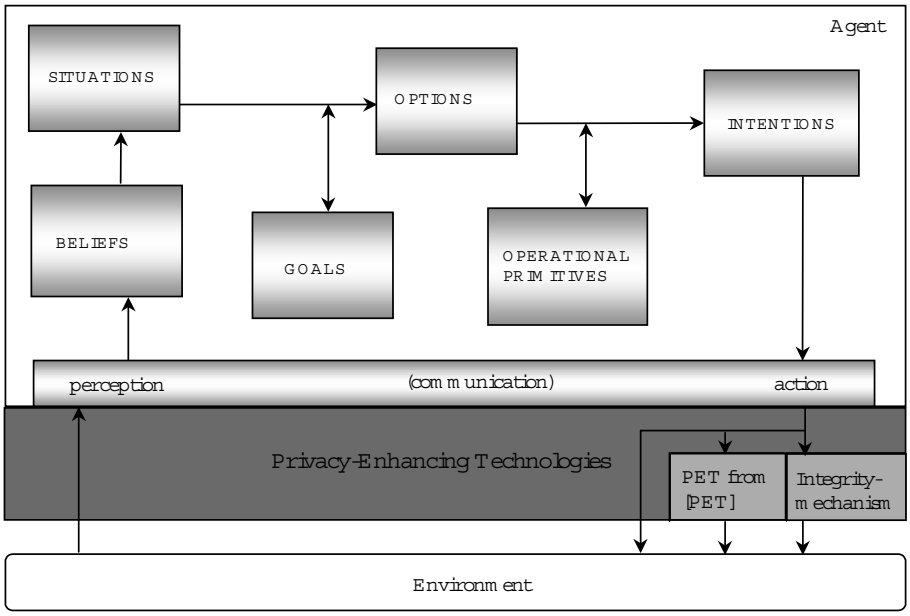


Figure 5.2: PETs wrapped around the agent.

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 5.1(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 its advantages. One of them is that a user can buy, separately, a relatively cheap agent and PETs-modules (PETs-tools) containing only the specific protection functionality that is required by the user. This is in contrast with the PETs-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 PETs-integrated agent will 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 with the right proportions could provide an agent with the advantages of both wrapping and integration.

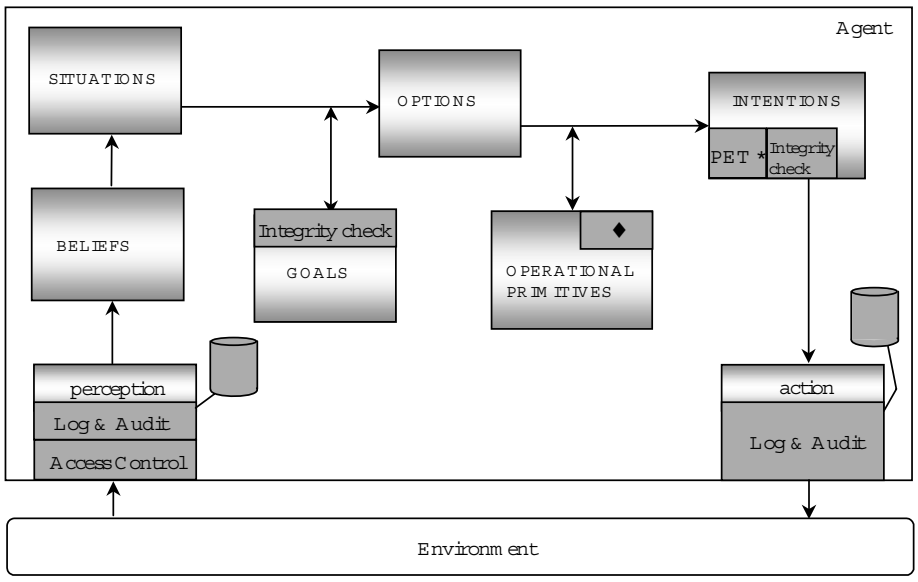


Figure 5.3 PETs integrated in the agent.

5.3 Alternative use of PETs

There could still be individuals who can't afford to wrap or integrate the above-mentioned PETs around or into their agents. These individuals should 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 should handle privacy-sensitive data (personal data) in a manner respectful of privacy. 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 will lead to the certification of these components with an indication of the guaranteed level of trust. With this alternative, individuals who can't protect their agents with PETs can still interact with other participants without giving up their privacy.

How does this alternative work? An agent can interact directly with other agents which may be unknown or not to be trusted, and risk revealing personal data of its owner. An agent can also interact indirectly with other agents by using a trusted component, and protect the personal data of its owner. To do so, the agent will move to the nearest trusted component, identify and authenticate itself to the trusted component, and authenticate the trusted component. The movement of the agent is visualised in figure 5.4 as the agent putting itself in an envelope and going to the trusted component. After the mutual authentication, the agent will ask the trusted component to execute the intended activities with the other agents (which could be unknown or not trusted). The trusted component will execute the desired activities in a privacy-secure way, and will come up with the results.

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

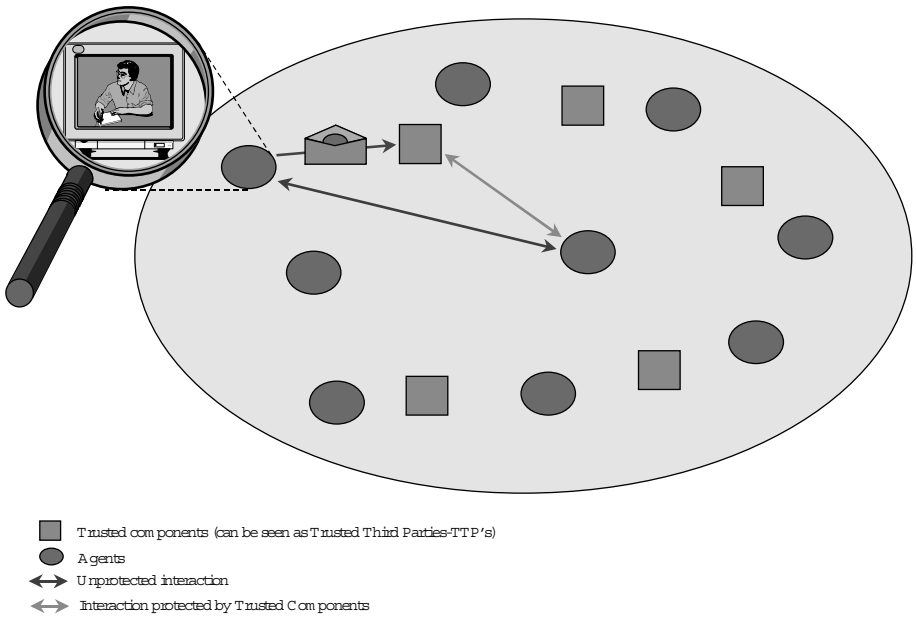


Figure 5.4: PETs used to create trusted components.

The functionality of the trusted component infrastructure isn't enough to secure the privacy of its users. The functionality needs to be implemented in a safe way. This is also applicable for the use of PETs in general.

5.4 Consequences of using PETs

When there is no need for privacy, there is no need for measures to guarantee privacy.

When there is a need for privacy, at least all the agents that a user-agent wants to communicate with need to have a unique identity. In addition, they need to authenticate themselves. Therefore, the agents need to 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 will add the new identity to an authorization table. Another way to register could be that a new agent needs to apply to a mutually recognized party that will add the new agent to a participant list. This list will be distributed to all attending participants.

If this is not enough, because more privacy is wanted, every participant in the environment will have to be known. This means that the environment will be limited to the group of participants listed in the participants list.

If a user sets high requirements for the protection of his or her privacy, the measures taken need to be very strong. It also means that the measures need to be enforced by the agent software. In other words: it must 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. So, if a user needs strong protection of his or her 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 will be reduced, because properties such as mobility and cloning are not allowed;
- the costs of an agent that satisfy the need for the strong protection of privacy will be high.

Privacy-consciousness will 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.

5.5 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 PETs-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 needs. These agents will be relatively expensive.

For individuals who use agents with no privacy protection, PETs-tools or trusted components could offer a better solution. Organizations that develop security tools or agents should actively start developing PETs-tools. These tools can be used as an IP to help individuals protect their privacy. When these tools are available, everybody can compose privacy protection agents according to their own specific privacy requirements.

The development of trusted components will also help individuals to protect their privacy, but this is a bit more difficult than the development of PETs-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 will have to be spent on setting up mutual agreements between all participating parties. Although this is a time-consuming and costly process, it will be a good alternative for individuals who want to protect their privacy in an agent-based environment. A subscription to a trusted component infrastructure will be less expensive than the procurement and maintenance of a PETs-integrated agent or PETs-tools.

5.6 PETs 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.

A user, designer, developer, supplier, or provider of an agent can ask himself how the privacy of the user and all other individuals involved can be protected. To help them, there is a checklist of considerations during the different phases of the design process.

During the analysis phase of the development of an agent, it must become clear whether the agent will collect and handle personal data of both the future user and other individuals. The personal data of the future user (user-profile) should be protected with proper PETs. In addition, collection of personal data of other individuals, particularly identifying data, should be minimized in accordance with the privacy regulations described in chapter 3.

During the design phase, the way PETs will 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 will 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 to the environment without its permission.

Figure 5.5 indicates how the designer can take the privacy of everyone involved into account during the different phases of the design process.

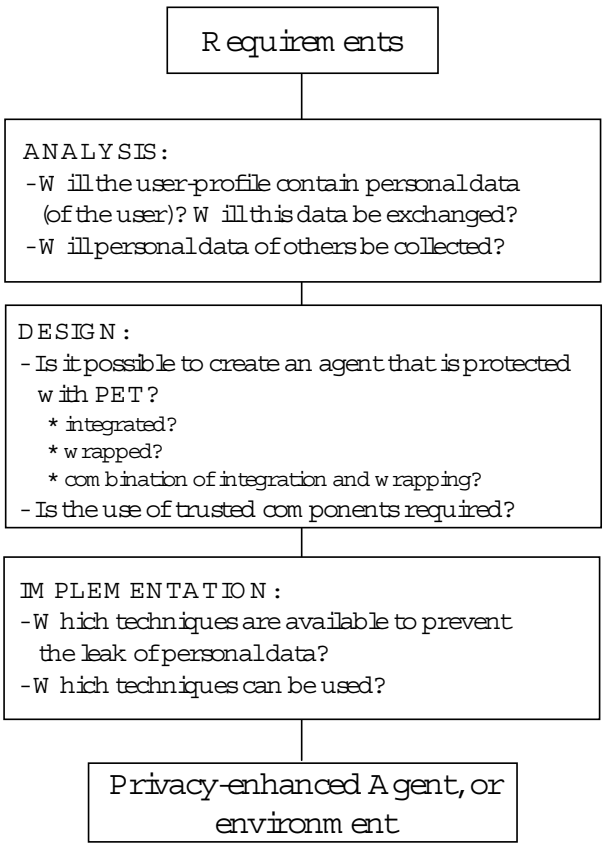


Figure 5.5: Aspects to take into account during the different phases of the design process of a privacy-protecting agent.

In appendix A the criteria to design a privacy-enhanced information system are given.

6 Conclusions and Recommendations

This chapter contains the conclusions and recommendations that emerged from this study.

Conclusion: 'Intelligent' agents are the near future. They are being studied and developed in quite a number of research & development laboratories. Nevertheless, the agents that are available today still require a lot of user-initiated actions to produce the right results.

Recommendation: In spite of the fact that agents are not yet as sophisticated as researchers claim, the implications of the use of (intelligent) agents for the privacy of individuals already need to be taken into account. This is necessary to control both today's consequences and consequences that may arise in the (near) future.

Conclusion: Future intelligent agents might have advanced computing powers, which enable them to take over human tasks, and to interact with people in human-like ways. 'Some agents have the potential to form their own goals and intentions, to initiate actions on their own without explicit instruction or guidance, and to offer suggestions to people' (Norman, D.A., 1994). This could lead to certain privacy threats.

Conclusion: To ensure a smooth introduction of agent technologies, two aspects are relevant. The first aspect deals with the way people feel about agents. The second aspect deals with the comfort and acceptance of the agent's automatic, autonomous actions (Norman, D.A., 1994).

Recommendation: Developers of agents need to make sure that people do not loose control over their computational systems and information contained therein. Adding control and feedback mechanisms and safeguards to prevent runaway computation will help agent-users to increase trust in using agent technologies.

Conclusion: Privacy and confidentiality of actions will be amongst the major issues confronting the use of intelligent agents in the future, when the society will be fully automated and interconnected.

Conclusion: The exchange of personal data is only necessary in some cases, for example for the authorization or accounting of the individuals who want to access a system, environment or service. In all other cases, the exchange of personal data is not necessary.

Conclusion: Unprotected agents will jeopardise the privacy of individuals. Agents can exchange personal data of their owners with others, but it is also possible that agents collect personal data of individuals in the interest of their owners. This could lead to the following potential threats to privacy:

- loss of control;
- agent-providers;
- the exchange of personal data with the environment:
- agents that are in disguise;
- agents that are more powerful;
- traffic flow analysis performed by agents;
- the collection of personal data of individuals, by:
  - entering the privacy domain of the individual;
  - entering databases that contain information about the individual;
  - entering the user-profile of an individual's agent.

Conclusion: Measures have to be taken to reduce the impact of the privacy threats. These measures are:

- certification of the agent's working method;
- logging of all internal and external actions of the agent itself;
- identification and authentication of all agents;
- access control mechanisms;
- logging of all actions performed by other agents that collect personal data;
- mechanisms to audit the logged activities;
- integrity mechanisms to control the integrity of stored or exchanged data and to control the integrity of working methods of agents or trusted components, like digital signatures;
- the Identity Protector: implemented with existing Privacy-Enhancing Technologies (PETs) such as: digital pseudonyms, blind digital signatures, and Trusted Third Parties (TTP's).

Recommendation: These measures can be wrapped around the agent or they can be integrated in the agent. A combination of integrating and wrapping is also possible. The measures can also be used to build an infrastructure of trusted components.

Conclusion: The consequence of using identification, authentication and access control mechanisms is that all agents that want to co-operate in the environment need to have a unique identity. To obtain a unique identity the agents all need to be registered.

Conclusion: When a high degree of protection is required the measures that are implemented in the agent need to be enforced, and must, therefore, be impossible to bypass. Enforcing the measures will have an impact on some of the agent's properties and attributes, like mobility and cloning. The agent will not be allowed to be mobile and to clone itself. This will lead to reduced performance of the agent.

Recommendation: Due to the fact that the research is in the early stages, the results of this research may change, because of new developments or new views on the use of agents. The results need to be discussed with developers of agents, agent-technologies, and privacy-enhancing technologies.

Recommendation: By using a checklist of design criteria during the design process, the user, designer, developer, supplier, or provider of an agent have a tool to help them develop an agent or an agent-environment with proper privacy-enhancing technologies.

Recommendation: Privacy Commissioners and Data Protection Authorities should ask designers and developers of agents if they used the design criteria during the development of their agents.

List of abbreviations

ACL	Agent Communication Language
AI	Artificial Intelligence
BDI	Beliefs, Desires and Intentions
DPS	Distributed Problem Solving
DTA	Dynamic Theory of Action
EU	European Union
IP	Identity Protector
MAS	Multiple Agent Systems
OECD	Organization for Economic Co-operation and Development
PAI	Parallel Artificial Intelligence
PDA	Personal Digital Assistant
PETs	Privacy-Enhancing Technologies
TTP	Trusted Third Party
WWW	World Wide Web

References

Abdu, D., and Bar-Ner, O., 'Software Agents: A general overview,' http://t2.technion.ac.il/~s3180501/agent.html.

Caglayan, A.K., and Harrison, C.G., 'Agent sourcebook: a complete guide to desktop, internet, and intranet agents,' in: Wiley Computer Publishing, 1997.

Council of Europe, 'Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data and the Explanatory report on this convention,' 1981.

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.

Helmers, S., Hoffmann, U. and Stamos-Kaschke, J., '(How) can software agents become good Net citizens?', http://www.december.com/cmc/mag/1997/feb/helend.html.

Garis, de Hugo, (1990). The 21st century artilect moral dilemmas concerning the ultra intelligent machine, in: Revue Internationale de Philosophie. Vol. 44, no. 172. p 131–139.

Hes, R. and Borking, J. (editors) e.a. (1998). Privacy-enhancing Technologies: The path to anonymity. Revised Edition. A&V-11. Den Haag: Registratiekamer.

Minsky, M., and Riecken, D., 'A conversation with Marvin Minsky About Agents,' in: Communications of the ACM, July 1994-volume 37, Number 7, p. 22–29.

Müller, J.P., 'The Design of Intelligent Agents: A layered approach,' Springer, 1996.

Norman, D.A., 'How might people interact with agents,' in: communications of the ACM, July 1994-volume 37, Number 7, p. 68–71.

Nouwt, J. and Vorselaars, H.A.C.M., 'Privacy in Cyberspace,' from: Emerging Electronic Highways, Bekkers, V. et al (eds), Kluwer Law International, 1996.

Nwana, H.S., and Azarmi, N., 'Software Agents and Soft Computing: Towards enhancing machine intelligence,' Springer, 1997.

Recommendation C (80) 58 (Final) of the OECD concerning and guidelines governing the protection of privacy and transborder flows of personal data (Organisation for Economic Co-operation and Development 1980).

Appendix A: The Identity Protector

Conventional information systems generally record a large amount of information. This information is often easily linked to a private individual. Sometimes these information systems contain information that could be privacy-sensitive to some private individuals. To prevent information systems from recording too much information, the information systems need to be adjusted.

There are a number of options to prevent the recording of data that can be easily linked to a private individual. The first is not to generate or record data at all. The second option is not to record data that is unique to an individual. This data is called identifying data. The absence of such data makes it almost impossible to link existing data to a private individual. These two options can be combined into a third one. With this third option, only strictly necessary identifying data will be recorded, together with the non-identifying data.

The conventional information system contains the following processes: authorization, identification and authentication, access control, auditing, and accounting. In the conventional information system, the user's identity is often needed to perform these processes. The identity is used within the authorization process, for instance, to identify and record the user's privileges and duties. The user's identity is thus introduced into the information system. Because in a conventional information system all processes are related, the identity travels throughout the information system.

The question one must ask is: is identity necessary for each of the processes of the conventional information system? For authentication, in most cases, it is not necessary to know the user's identity in order to grant privileges. However, there are some situations in which the user must reveal his or her identity to allow verification of certain required characteristics.

For identification and authentication, access control, and auditing, the identity is not necessary.

For accounting, the identity could be needed in some cases. It is possible that a user needs to be called upon to account for the use of certain services, e.g., when the user misuses or improperly uses the information system.

The introduction of an Identity Protector (IP), as a part of the conventional information system, will structure the information system in order to better protect the privacy of the user. The IP can be seen as a part of the system that controls the exchange of the user's identity within the information system. The IP offers the following functions:

- reports and controls instances when identity is revealed;
- generates pseudo-identities;

- translates pseudo-identities into identities and vice versa;
- converts pseudo-identities into other pseudo-identities;
- combats misuse.

An important functionality of the IP is conversion of a user's identity into a pseudo-identity. The pseudo-identity is an alternate (digital) identity that the user may adopt when consulting an information system (see figure A.1).

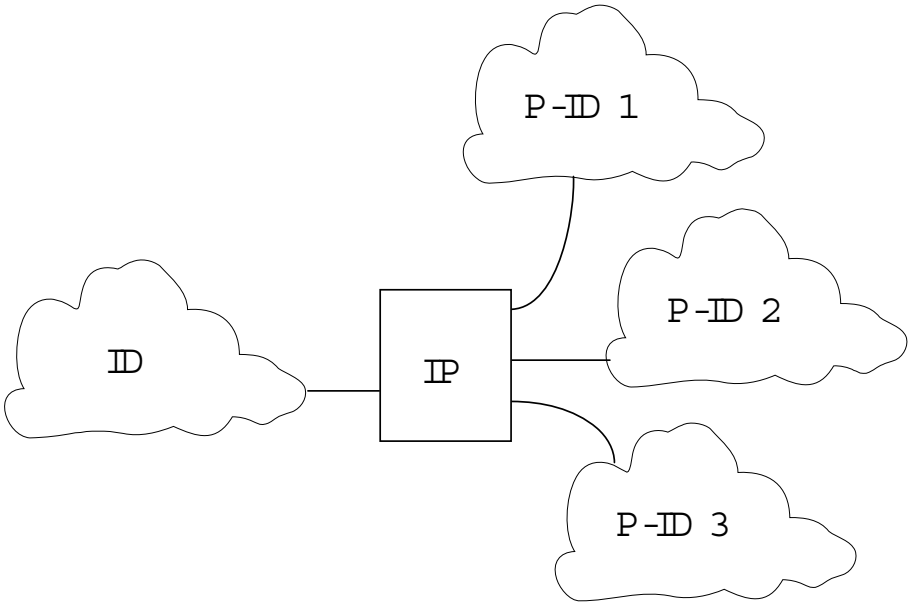


Figure A.1: The identity protector separates the identity and pseudo domains.

The user must be able to trust the way his or her personal data is handled in the domain where his or her identity is known. The IP can be placed anywhere in the system where personal data is exchanged. This offers a couple of solutions for an information system that handles the privacy of an individual.

Techniques that can be used to implement an IP are: digital signatures, blind digital signatures, digital pseudonyms, and trusted third parties.

To design an information system that protects the privacy of the user, the design criteria showed in figure A.2 need to be considered.

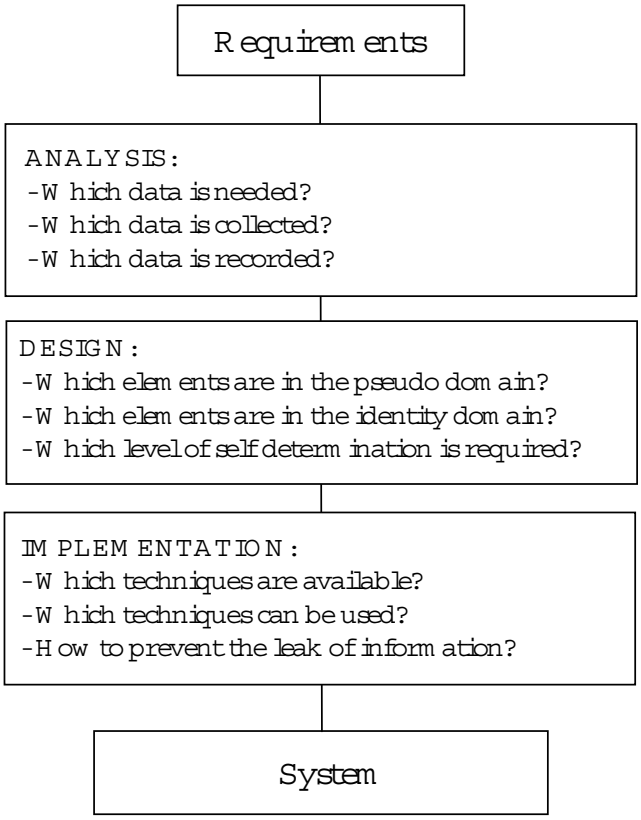


Figure A.2: Aspects to take into account during the different phases of the design process of a privacy information system.

Appendix B: Reasoning and Learning

An 'intelligent' agent needs to be able to interpret the events that take place in the environment to make the appropriate decisions for operating autonomously. Therefore, agents rely on the knowledge they possess.

The first possibility for an agent to act on events that take place in the environment is with the use of specified preferences. Preferences are statements of desired behaviours that describe a style or policy an agent needs to follow. The agent doesn't reason any external events or data, it just acts on them.

The next possibility is to interpret external events or data by means of reasoning. With reasoning, preferences are combined with external events and external data in a decision-making process. The logic or behaviour of the mechanism is called the 'rule base.' Depending on both external events and external data, the reasoning mechanism will create results that are adjusted to the environment. Therefore, the reasoning mechanism needs to contain the following elements:

- short-term facts: short-term facts describe the state of the mechanism. There are new facts, which have been produced by the analysis of events or the acquisition of information, and derived facts, which are the results of the decision-making process. The short-term facts are present during operation of the reasoning mechanism, but will be gone when the reasoning mechanism is restarted. These short-term facts constitute the agent's real-time perception of the environment (world), i.e., the entities and events that take place in the environment.
- long-term knowledge: long-term knowledge is the knowledge that is provided by the designer of the reasoning mechanism. Most of the time, this knowledge is included in the rule base.
- control structure: the control structure describes the sequence of decisions that have to be made (procedural structure) by the reasoning mechanism to come to a specific action.
- event/action interfaces: these are interfaces the reasoning mechanism needs to receive data from its environment (external events or external data) or to perform actions in the environment.

Most reasoning mechanisms deal with events one at a time and do not look for correlations among sequences of events. A reasoning mechanism can't change the rule base by itself, but it is possible to program the mechanism to detect new knowledge. To add this new knowledge to the reasoning mechanism, the user or designer needs to modify the rule base. Reasoning mechanisms, where the mechanism itself can modify the rule base, are called adaptive reasoning mechanisms. In adaptive reasoning mechanisms, the rule base may be modified either by the mechanism itself, or by some external process. It will not be possible

for an adaptive reasoning mechanism to delete, or change drastically, the present policies that are included in the rule base. It can only add new rules (knowledge) to the rule base.

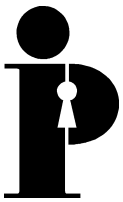
The last possibility is to learn from external events and data. Learning can be described as the modification of behaviour as a result of experience. Learning can be achieved in three ways:

- by adding new rules or modifying existing rules: if an agent recognizes a new behaviour, it could create a new rule from this behaviour or it could modify an existing rule. According to the agent sourcebook (Caglayan, A.K. and Harrison, C. G., 1997), it is still very difficult at this moment to automate the creation or modification of rules, 'not because writing rules is difficult, but because maintaining a logically consistent rule is very difficult' (Caglayan, A.K. and Harrison, C.G., 1997).
- adding new facts or modifying existing facts: automatically identifying new facts is still a big problem. It is possible that an agent recognizes a new fact, but it would be very difficult for the agent to recognize the role of this new fact. Recognizing the role of the new fact means that the agent must already know about the role of the fact and how to identify it.
- modifying the level of confidence of a belief: as a result of experience, an agent can change the level of confidence that it has of one of the beliefs it has been given.

There is still the possibility that an agent ends up in a totally new situation with more than one alternative action with unknown results to take, where its existing experience cannot help it. Depending on which algorithm the designer implemented, the agent will try to cope with this situation. There are a couple of possible ways to create such algorithms. One of the solutions is a trial and error method, where the agent tries the actions until it comes up with a result that satisfies the user. The agent will add this action to its knowledge[1]. Within another solution, the agent still tries to connect beliefs or (other) existing experience to one of the possible actions and tries to solve it. When the action gives the user a satisfying result, the agent knows that it took the right action. The agent will then add this action to its knowledge. If it wasn't the right action then the agent will execute the action that is next in line, and so on, until it finds the action that will satisfy the user.

[1] According to the agent sourcebook (Caglayan, A.K. and Harrison, C. G., 1997) 'Knowledge can be acquired by employing any of the following techniques:

- Developer specified: Models for the application domain and the intended user serve as the basis for the inference mechanism in a knowledge base. Formal knowledge is organized in rules and/or frames. The disadvantage of knowledge-based learning agents is their lack of customizability after deployment;
- User specified: user-programmable rules for carrying out user-defined tasks are typical of rule-based systems. Some systems allow users to record their actions. The main disadvantage of rule-based agents is their imposition of the agent's programming onto the user;
- Derived from other knowledge sources: common agent languages enable knowledge acquisition among agent communities;
- Learned by the system: agents that derive knowledge from the user and environment.'



Information and Privacy Commissioner/Ontario

2 Bloor Street East, Suite 1400
Toronto, Ontario, Canada M4W 1A8
416-326-3333
1-800-387-0073
Fax: 416-325-9195
TTY (Teletypewriter): 416-325-7539

Registratiekamer

Sir Winston Churchilllaan 362
P.O. Box 3011
2280 GA Rijswijk, Netherlands
Tel. 011 (31) 70-3190190
Fax 011 (31) 70-3940460