Agents, security and ethics: a framework for analysis

E Smith, MM Eloff, LM Venter, A Barnard, L Pretorius
Department of Computer Science and Information Systems
University of South Africa
P.O. Box 392, UNISA, 0003
South Africa
Tel: +27 12 429 6336
Fax: +27 12 429 6848
E-mail: smithe@unisa.ac.za

ABSTRACT
The domain of information security research is no longer exclusively of a technological nature as it has become permeated with aspects of human behaviour. Similarly the broad field of ethics is no longer only a human issue, as is reflected by the establishment of computing ethics as a separate research area. Advances in the past decade have led to the emergence of among others, new technologies, frameworks and methodologies in the field of computing. Examples include the Internet, global connectivity and agent technology – in particular intelligent agents. The attribute intelligent brings with it a concomitant human characteristic that is assigned to an inanimate technological object. It is even plausible to think of communities of intelligent agents, inhabiting cyberspace, interacting with other agents, human users and hosts, and in this way developing a social life. This raises issues concerning information security as well as the ethical and social behaviour of intelligent agents.

In this paper we thus briefly discuss agent computing and its impact on the environment in which it exists. In particular we focus on some relevant security and ethical issues associated with agent computing. The purpose of the paper is to present a framework within which the security and ethical behaviour of agents can be evaluated and analysed.

KEY WORDS
agents, intelligent agents, information security services, computing ethics
1. INTRODUCTION

The domain of information security research is not exclusively of a technological nature as it is permeated with aspects of human behaviour. Similarly the broad field of ethics is no longer only a human issue, as is reflected by the establishment of computing ethics as a separate research area. Advances in the past decade have led to the emergence of among others, new technologies, frameworks and methodologies in the field of computing. Examples include the Internet, global connectivity and agent technology – in particular intelligent agents. The attribute intelligent brings with it a concomitant human characteristic that is assigned to an inanimate technological object. It is even plausible to think of communities of intelligent agents, inhabiting cyberspace, interacting with other entities (agents, human users and hosts) and in this way developing a social life. This raises issues concerning information security as well as the ethical and social behaviour of intelligent agents.

Agent behaviour can be analysed from a multitude of perspectives, including the security and ethical concerns. Security analyses typically focus on evaluating the application of external measures to an entity to ensure the safety of the entire community. Alternatively an ethical analysis addresses the internal behaviour of an entity in order to highlight its possible performance of actions harmful to the community. These different perspectives complement one another and may lead to a simplification of the security system.

In this paper we thus briefly discuss agent computing and its impact on the environment in which it is applied. In particular we focus on security and ethical issues associated with agent computing. For this purpose we start off by explicating what we understand under the notion of an agent and we describe the typical environments in which these agents can operate, the so-called agent community. For illustrative purposes we also consider a well-known example of intelligent agent technology, namely the Microsoft Office Assistant, Clippy. This example is used to illustrate both the security and ethical analysis.

In section 3 we discuss a number of relevant security issues and ethical theories pertinent to agent computing and we present a framework within which the security and ethical behaviour of agents can be evaluated and analysed. We also briefly consider the significance of these analyses. Finally the framework for analysing the security and ethical considerations of agent computing is applied to the example.
We conclude by observing that agent computing presents certain ethical and security challenges that are worthwhile investigating and requires further research.

2. AGENT COMPUTING

In this section, we give a very brief overview of some aspects of agent computing. After presenting our definition of the concept, we explore some aspects pertaining to agent computing. We discuss some applications to show that agents are useful. Throughout this section, and indeed the entire paper, we refer to the Microsoft Office Assistant as an example of an intelligent agent.

2.1 What is an agent?

The literature contains a number of different descriptions and definitions of agents. According to the Principia Cybernetica (1992) an agent “... can be a rule-following interpreting, semantic system ....”. This definition encompasses many of the adjectives used to describe some agents types, including autonomous, intelligent, rational, software, intelligent software, mobile, social, heterogeneous, etc.

In (Franklin and Graesser, 1996) we find the following definition: “An autonomous agent is a system situated within a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.” In their taxonomy they subdivide this into biological, robotic and computational agents, with obvious definitions. Computational agents are subdivided into artificial life agents and software agents. This last category is the one we consider in this paper.

A more recent definition is found in (Fou, 2001): “An agent is a piece of software that has the capacity to autonomously conduct its work.” The same author states that an agent is considered to be mobile if we do not need to know where in the system it resides.

If we restrict ourselves to software agents there are some common features that emerge. The agent is autonomous, it can act, its actions are specified beforehand, it operates within some environment, and finally its position within the environment is not fixed. For the purposes of this paper, we consider that a combination of these properties distinguishes an agent from any other computer program.
2.2 An example

As noted in the introduction we consider the Microsoft Office Assistant (Microsoft named it Clippy because of its paper clip persona) as a representative example of intelligent agent technology. Clippy is the little animated figure that appears on the user’s screen and presents tips about using Microsoft programs. When first released, critics dismissed Clippy as the equivalent of training wheels for computer novices. Yet the friendliness of Clippy conceals a great deal of computing potential. “In fact, it’s essentially a back door for Microsoft to allow macros that can take control of a PC and help out users” (Lemos, 2003).

Clippy obviously exhibits all of the features of an agent as described in the previous section. Clippy’s settings are global for all programs in the Microsoft Office suite. When Clippy is set to provide a set of help options for one program in the suite, it will do the same for all the others within that suite.

2.3 Are agents useful?

The single most important ability of an agent, which makes it extremely useful, is that it takes the computation to the data rather than to bring the data to the computation. In many cases it is unknown where the data resides. As an example, the BargainFinder agent will search through a number of databases to find the availability and pricing for a specified music CD (Wilder, 1995). In other cases, the volume of the data is such that it is impractical to move the data. A fairly typical example of this situation occurs where agent technology is utilized in Intrusion Detection Systems (Kruegel and Toth, 2001).

2.4 Agent communities

Wagner (2000) argues that agents form an integral part of society, and as such they interact with one another, as well as with the environment in which they operate. This environment is referred to as an agent community and includes other agents, network elements as well as humans. Within such an environment, different examples of agent interaction are possible, such as (Wagner, 2000):

- Time-planning-agent: Arranges dates with other agents and/or hosts in the community;
- News-filtering-agent: Selects news and information that are of interest to the user;
- Matchmaking-agent: Brings sellers and buyers of products together; and,
- Shopping-agent: Searches for products, negotiates and buys on behalf of the user.

According to Eichmann (1994) there are three general categories of agent functionality, viz:

- user-agent interaction,
• agent-host interaction, and
• agent-agent interaction.

It is clear that the actions performed by the agent within its community must be specified such that it does not jeopardise the integrity of the community. Hence there is a need to analyse the security aspects and ethical behaviour of agents. In our security analysis (section 4.1) we focus on agent-host interaction, while our ethical analysis (section 4.2) will focus on agent-user interaction. These perspectives are dictated by the choice of Clippy as our example. Similar analyses of the other types of interaction can be performed, but falls outside the scope of this paper.

3. A FRAMEWORK FOR ANALYSIS

In this section we discuss the necessity to analyse security issues and ethical aspects of intelligent agents. This then leads to the formulation of a suitable framework that can be used for analysis of agent behaviour.

3.1 Relevance of the analysis

Computer users may be classified as either aware or unaware of security aspects. The former group mistrusts unfamiliar agents while the latter group are not at all aware of potential security risks associated with agent computing. A framework to analyse the security risks of agent computing will create and raise awareness of how secure agents are.

Intuitive assessment of agent behaviour may be misleading and it can be argued that a systematic ethical analysis will provide a more reliable basis for assessment. For example the actions of Clippy may be considered as unethical by an expert user due to Clippy's obtrusive character – however the systematic ethical analysis of Clippy’s actions in section 4.2, reveals that Clippy’s actions can at most be considered irritating, but certainly not unethical.

An a posteriori systematic analysis of the behaviour of an agent can assist developers of said agent to improve the modelling of the secure and ethical behaviour of future versions of the agent. Once the behaviour of a number of agents have been analysed in this systematic fashion, norms and criteria for the design of new agents that will exhibit acceptable secure and ethical behaviour can be formulated and continually refined. This may lead to a simplification of the security measures imposed on the agent.
3.2 Relevant security theory

As part of the evaluation phase of the security risks posed by agent computing, we consider the five services of Information Security (IS), viz. identification and authentication, authorisation, confidentiality, integrity, and non-repudiation as defined according to the ISO 7498-2 standard, produced by the International Standards Organisation (ISO, 1999). These services are required to ensure that information is protected and secured during its storage, transmission and usage by an agent (Schneier, 2000). A definition for each of these services (Pfleeger and Pfleeger, 2003) as appropriate for agent computing is presented below. Note that this discussion focuses on the interaction between a possibly malicious agent and possibly trusted host. This is in accordance with the approach of Chess, Harrison and Kershenbaum (1995) relating to security aspects of mobile agent computing.

3.2.1 Identification and Authentication

The identification and authentication of any agent (and hence its owner) who wants to access a host is the first step towards enforcing IS. An agent requesting access needs to present a user-id that uniquely identifies it. On presentation of such a user-id, the user-id should be verified to ensure that it does, in fact, belong to the agent who presented it.

3.2.2 Authorisation

The next step towards enforcing IS is to determine whether the authenticated agent has the right to access the host in question. Therefore, in terms of the authorisation process, control is exerted over the access rights of all authenticated agents.

3.2.3 Confidentiality

All information must be strictly accessible to authorised agents only. Protecting the confidentiality of information provides assurance that only authorised agents will have access to the information in question.

3.2.4 Integrity

Information should not only be kept confidential, but its integrity should also be guaranteed. Only authorised agents should be able to change the content of protected information. In other words, unauthorised changes to information must be prevented thus ensuring that the information can be deemed accurate and complete.
3.2.5 Non-repudiation

The last step towards enforcing IS, namely non-repudiation, is to ensure that no action performed by an agent (and hence its owner) to affect IS could be denied at a later stage.

3.3 Relevant ethics theory

This section is concerned with a discussion of some of the better-known ethics theories that may be applied in the analysis of the ethical behaviour of agents. In this respect we review the basic principles of two deontological theories, viz. duty-based and rights-based ethics, the teleological theory of utilitarianism (Spinello, 1997), and the theory of just consequentialism (Moor, 2001). Note that these theories are discussed in the context of agent computing.

3.3.1 Duty-based ethics theory

The duty-based ethics of Kant may be summarized as “the absolute principle of respect for other” entities (i.e. users, agents and hosts) “who deserve respect because of their rationality and freedom” (Spinello, 1997: 34). Ranau, Davies and Rogerson (Maner, 2002) suggest that answers to the following (relevant) questions should be considered with regards to the action of the user, agent and host:

- **Fidelity**: Is there a promise that should be kept in contemplating or performing some action?
- **Reparation**: Is there a wrong that should be righted due to the contemplation or performance of said action?
- **Justice**: Should the outcome of the action be fair?
- **Beneficence**: Can the lot of others be improved as a result of the contemplation or performance of the action?
- **Gratitude**: Is an expression of gratitude due to the performance of an action appropriate?
- **Non-injury**: Can others be protected from injury or harm due to the contemplation or performance of said action?

3.3.2 Rights-based ethics theory

This approach focuses on individual rights and respect for these rights which are equal. According to Spinello (1997: 39) everyone (i.e. the user, the agent and the host), “for example, equally shares in the rights to life and liberty regardless of their nationality or status in society”. Ranau, Davies and Rogerson (Maner, 2002) suggest that answers to the following questions should be considered, i.e. is the right of the user, the agent and the host:

- to know respected?
- to privacy respected?
3.3.3 Consequence-based ethics theory

Utilitarianism is a widely used form of consequentialism (Spinello, 1997: 27). For the purposes of this paper, we concur with Spinello (1997: 28) that “utilitarianism is the moral doctrine that an action is morally right if it produces the greatest happiness for the greatest number of” entities (i.e. users, agents and hosts) “affected by it”. One thus needs to determine who (i.e. the user, the agent and/or the host) would be affected by the contemplation or performance of an action, and to what degree.

3.3.4 Just consequentialism

Moor (2001) summarises the theory of just consequentialism to imply that the ends, however good, “do not justify using unjust means”. Regarding the contemplation, and in particular the performance of some action, one would thus need to determine whether unjust means would be required to facilitate performance of the action by the user, the agent or the host. Therefore, if it is not possible to achieve the envisaged end (performance of the action) without utilizing unjust means, the requirement of just consequentialism is not satisfied.

3.4 Framework for analysis

To determine whether the actions of an agent are secure, the following questions need to be answered:

- Has the agent been properly identified and authenticated?
- Does the agent have the appropriate access rights?
- Can the agent be trusted not to reveal confidential information?
- Can the agent’s actions be trusted not to compromise information integrity?
- Can the actions of the agent be audited?

The answers to the above questions would highlight the information security services that might be compromised due to the actions executed by an agent.

Regarding an ethical analysis of an agent’s behaviour we use the Five-step Process of Ethical Analysis of Rananu, Davies and Rogerson (Maner, 2002) as basis. Other similar procedures for ethical analysis may be found in Maner (2002). The analysis procedure of Rananu, Davies and Rogerson, originally designed primarily for the analysis of human behaviour and ethical decision-making, was chosen because it can be readily applied to the ethical analysis of agent behaviour. For the purposes of this paper we modify this process to be applicable to agent computing:
Step 1: analysis of the scenario
In analysing the behaviour of an agent, the following must be considered:

- **What are the facts?**
- **Who are the stakeholders?**
- **Identify relevant ethical and social issues.**

Step 2: application of appropriate formal guidelines
- **Does the behaviour of the agent conform to or violate the Golden Rule which states “do unto others as you would have them do unto you”** (Spinello, 1997: 37)
- **Who benefits from or is harmed by the agent’s actions?**

Step 3: application of ethics theories
We use the four ethics theories presented in section 3.3 to analyse the agent’s actions.

Step 4: application of relevant law
For the purposes of this paper, we exclude this step.

Step 5: application of informal guidelines
Rananu, Davies and Rogerson (Maner, 2002) suggest that answers to the following appropriate informal questions should be considered:

- The TV test: Would an agent inform the entire agent community of its actions?
- The Other Person’s Shoe test: What if the roles were reversed?

An ethical conclusion regarding the agent’s actions and behaviour can be made based on the above five steps.

4. **APPLICATION OF THE FRAMEWORK**

In this section we apply the framework suggested in section 3.4 to analyse the security risks and ethical considerations posed by Clippy.

4.1 **Security considerations**
Clippy as such does not pose any harm to the user. However, a security hole was discovered that subverts the powerful functions of Clippy (Lemos, 2000). In particular, an ActiveX control supplied with Office 2000 is incorrectly marked as *safe for scripting*. This control is used by Clippy and allows Office functions to be scripted. A malicious web site operator could use this control to carry out Office functions on the machine of a user who visited his site (Microsoft Security Bulletin, 2000). The security researcher, who discovered the hole, claims: “*because its abilities are marked ‘safe for scripting’, anything is possible*” (Lemos, 2000).
A patch is available to fix the problem, but users still need to download the patch and update their Office program. There is no agent that will do it automatically for users! The Security firm @Stake Inc. made the following comment: “The fact that this control exists and is installed in the particular fashion would permit the construction of a worm of unparalleled devastation.” This control, however, can be manipulated to harm users as well - a test program created by @Stake can set the system security to “low” and copy a text document to the hard drive (Lemos, 2000).

The proposed framework yields the following results regarding Clippy’s security aspects:

- Has the agent been properly identified and authenticated?
  
  As Clippy is an integral part of MS Office there exists no explicit need for identification and authentication.

- Does the agent have the appropriate access rights?
  
  When providing useful hints and tips Clippy primarily accesses the appropriate help files to guide the user through a series of steps. Clippy has appropriate access rights for these actions. However, the security hole as described earlier, poses a serious threat of unauthorised access if the patch file has not been installed.

- Can the agent be trusted not to reveal confidential information?
  
  In assisting the user Clippy accesses the appropriate help files, which are not confidential. However, the security hole allows an intruder to write a script that could for example, access other confidential information on the host.

- Can the agent’s actions be trusted not to compromise information integrity?
  
  Clippy itself does not pose an integrity threat, but the security hole can allow malicious script to add or delete files which can compromise the integrity of information.

- Can the actions of the agent be audited?
  
  Clippy as such does not pose any harm to the user and therefore does not require non-repudiation. The actions of an attacker exploiting the security hole can, however, cause serious damage.

4.1 Ethical considerations

It is instructive to perform a systematic *a posteriori ethical analysis* of the actions of Clippy.

Step 1: analysis of the scenario

In analysing Clippy’s behaviour, we take note of the following:
Facts: The agent Clippy is a little animated figure that appears on the user’s screen and provides tips about using Microsoft Office programs. It also opens a dialogue box that allows the user to bypass the Help menu and enter a simple question in natural language.

Stakeholders: The human user, the agent Clippy, and the host on which the Microsoft Office package is installed.

Ethical and social issue: Does Clippy exhibit any unacceptable or unethical behaviour by being present on the user’s screen and employing continual intrusive animation in order to offer unsolicited assistance?

Step 2: application of appropriate formal guidelines

- Clippy’s conformance/violation of the Golden Rule which states “do unto others as you would have them do unto you” (Spinello, 1997: 37): One can argue that Clippy’s continual intrusive animation in order to offer unsolicited assistance, can be viewed by the user as distracting him/her from the task at hand. More fundamentally, Clippy’s continued presence and monitoring of the user’s actions and keystrokes can be viewed as an invasion of the privacy of the user. The fact that Clippy sometimes also goes to sleep when a period of inaction on the part of the user is detected, can be viewed in a negative light, and even experienced as intimidating behaviour on the part of Clippy towards the user. On a certain level thus it may seem as if Clippy violates the golden rule. One however needs to bear in mind that the user has the option to control or de-activate Clippy’s presence and one can hence argue that if Clippy is in violation of the golden rule, it is with the consent of the user. As an independent agent thus Clippy does not violate the golden rule.

- Who benefits from or is harmed by Clippy’s actions: By design Clippy is intended to assist the user - a novice user may find the continued assistance helpful, whereas a more advanced user can customise Clippy’s level of assistance and presence (and in the extreme even de-activate Clippy). Therefore the user can benefit from Clippy.

One can thus conclude that Clippy does not intentionally violate these formal guidelines.

Step 3: application of ethical theories

As was explicated earlier, we apply the following ethical theories (Rananu, Davies and Rogerson as in (Maner, 2002)):

Duty-based ethical theory

- Fidelity: Clippy does offer the user relevant assistance, and thus lives up to the promise of user support.

- Reparation: Not applicable.
- **Justice**: Clippy’s assistance is available to all Office users.
- **Beneficence**: Clippy’s design implies that assistance is freely available to all users irrespective of competency levels. Thus this agent may improve the lot of the user in general. The expert user may find Clippy’s presence distracting but still has the option to either customise or de-activate Clippy.
- **Gratitude**: Not applicable.
- **Non-injury**: Not applicable.

In terms of the duty-based theory thus, Clippy’s actions towards the user are not regarded as unethical.

**Rights-based ethical theory**
- Clippy’s visual presence or not is a true reflection of the agent’s activity, and thus the user is always fully aware of its presence. Therefore the *user’s right to know is respected.*
- The *default design* of the agent is that it is always present and active. The deactivation ability is only an option. Thus we contend that the user’s right to privacy is not respected.
- Clippy has no autonomous intervention capabilities, and thus the *user’s right to property*, i.e. his/her control and possession of electronic data and the concomitant integrity thereof, *is respected.*

Clippy poses a minor threat to the user’s right to privacy (which can be counteracted by the user), while respecting the user’s right to know and right to property. In terms of rights-based ethics thus, Clippy’s actions towards the user are not regarded as unethical.

**Consequence-based ethical theory**
The user has final control regarding the agent’s activities and existence and is thus subject to the user’s discretion. In this respect the agent does not influence the user, whereas the user determines the lifespan of the agent. We can thus conclude that the impact of the agent on the (single) user is limited, and as the agent interacts only with the Office applications of the (single) user, general impact is also limited. Therefore Clippy’s actions are not in conflict with utilitarian principles.

**Just consequentialism**
We are reminded that just consequentialism implies that the end, however good, “*do not justify using unjust means*” (Moor, 2001). We again note that the default design of the agent is that it is always present and active in an attempt to provide the user with assistance. This action of the agent compromises the user’s right to privacy and is an instance of using unjust means towards a good end. Clippy’s actions can thus be viewed as a violation of just consequentialism.

We conclude that the majority of ethical theories applied in this step, suggest that Clippy is a relatively benign agent that does not pose malicious (autonomous) intentions *towards the user.*
Step 4: application of relevant law
Not applicable.

Step 5: application of informal guidelines
Rananu, Davies and Rogerson (Maner, 2002) suggest that answers to the following appropriate informal questions should be considered:

- The TV test: Not applicable.
- The Other Person’s Shoe test: Clippy’s obtrusive and even intimidating character may be demonstrated by its continual intrusive animation in order to offer unsolicited assistance, its continued presence and monitoring of the user’s actions and keystrokes, and the fact that Clippy sometimes also *goes to sleep* when a period of inaction on the part of the user is detected. These inherent character flaws imply that Clippy would have difficulty in passing the Other Person’s Shoe test.

From the above it is apparent that the unethical aspects of Clippy’s behaviour can be counteracted or managed by the (expert) user. Although some may view Clippy’s actions as irritating or distracting, the above ethical analysis clearly demonstrates that on the whole, Clippy’s actions towards the user cannot be regarded as unethical.

5. CONCLUSION

In this paper we presented a framework for the analysis of security aspects and ethical behaviour of an agent within its community. In particular we discussed the agent’s behaviour towards the host and the user. Through considering our practical example we have demonstrated that the framework is indeed useful. The proposed framework can furthermore also be used to analyse other relationships in an agent community, for example the ethical issues concerning Clippy’s agent-host interaction which were not explored in this paper.

6. REFERENCES


