Carbo J., Pedraza J., Lopez M., Molina J.M. (2014) Privacy Protection in Trust Models for Agent Societies. In: de la Puerta J. et al. (eds) International Joint Conference SOCO'14-CISIS'14-ICEUTE'14. (Advances in Intelligent Systems and Computing, 299), pp. 135-144. Springer.

Doi: https://doi.org/10.1007/978-3-319-07995-0_14

Privacy Protection in Trust Models for Agent Societies

Javier Carbo[1], Juanita Pedraza[2], Mar Lopez[1], and José Manuel Molina[1]

[1] Computer Science Dept., Univ. Carlos III of Madrid,
Campus de Colmenarejo, Madrid, Spain
[2] Public State Law Dept., Univ. Carlos III of Madrid,
Campus de Colmenarejo, Madrid, Spain

Abstract. In this paper we have motivated the use of privacy-protection measures in trust models, both in conscious exchanges of opinions and in an unconscious way when security attacks take place. Most of the privacy dimensions are concerned into trust communications. In particular we define the privacy rights that these trusting communications must legally be guaranteed. From them, we describe additional message exchanges that, acting as control mechanisms, would be required to exercise such rights. Furthermore, we also enumerated the corresponding privacy violations that would have taken place if these control mechanisms were ignored. From the possible existence of privacy violations, regulatory structures may establish what agents are allowed and forbidden to do according to the legal privacy rights. We have applied the control mechanisms as additional message exchanges to a particular application domain (the Agent Trust and Reputation testbed) implemented as JADE interaction protocols, and finally we plan to define an Electronic Institution that would rule the corresponding norms and violations to such control using the Islander specification tool.

Keywords: Privacy, Trust, Agents.

1 Introduction

The right to privacy or private life is enshrined in the Universal Declaration of Human Rights (Article 12), the European Convention of Human Rights (Article 8) and the European Charter of Fundamental Rights (Article 7). The Charter also contains an explicit right to the protection of personal data (Article8). This right has several dimensions and these have been defined for European and American judges. In [1] five dimensions of privacy were identified: Privacy of the person (bodily privacy), Privacy of personal behavior (media privacy), Privacy of personal communications (interception privacy), Privacy of personal data (data or information privacy) and Privacy of personal experience. Four out of these five dimensions apply, in some extent, to any (computer-based) Information System, but even more to decentralized trust models. While in closed systems, a central trusting entity ensures privacy through an exhaustive control of identities and

information exchanges, in open systems this trusting responsibility lies with participants. This is the case of distributed and open systems, often implemented as a collection of bio-inspired knowledge systems such as [2], also called agents. Interactions may then have the final intention of propagating reputation of agents in order to decide which agent to trust in.

Specifically, two central features of trusting agents pose the main challenge to privacy: the ability of them to collect large and detailed amounts of data about individuals' everyday activities over long periods of time; and the enhanced ability for classifying and integrating these large amounts of data [3]. These features demand reviewing trust models under light of the data protection law, particularly under principles of Directive of Data Protection 95/16 [4]. In this article we intend to exam these risks and we propose some solutions about the corresponding privacy protection measures that can be adopted. In order to enjoy the benefits of the assumed autonomy of agents, we must consider a approach to privacy and data protection,based on computer-based mechanisms of control rather than on law restriction and prohibition [5].

Although the associated risks to privacy of most recent technology advances have been addressed, such as in: cloud computing [6], profiling and data mining [7] and ambient intelligence [8], there is no publication specifically related to the issue of privacy protection in trust models with Agents and this paper intends to overcome this lack.

2 Trust Models and Privacy Decision

Trust is a very relevant issue in any social relationship, even when such relationship is distant and with electronic means. Therefore computer scientists have shown an increasing interest in the study in how trust is acquired and maintained. Specifically, when human users are represented by autonomous agents and they acted electronically on behalf of them, the interests of these users have to be considered in the decisions and relationships held by the corresponding agents that represent them. A trust model is then applied by autonomous agents in two ways: searching trustworthy partners and as an incentive/punishment mechanism to prevent dishonest behaviours. Every act of an autonomous agent may be then judged in order to compute the reputational image of such agent. This reputational image could be computed in a centralized way as a global property by a sole entity (as many actual commercial applications do [9]), but it implies a loss in personalization and privacy. Therefore we assume (as many researchers in Distributed AI) that each member of a society of agents is in charge of computing the reputation of all other agents that belong to this society. Many trust models been proposed and they are very different among them [10]. But most of them consider direct experiences and witness information as the main information sources. Between them we are interested in witness information because of its relevance to privacy issues. Additionally it is the most abundant source of reputation (but not the most reliable), and the way it is managed is the source of the most complexity involved in trust models. Witness information is often

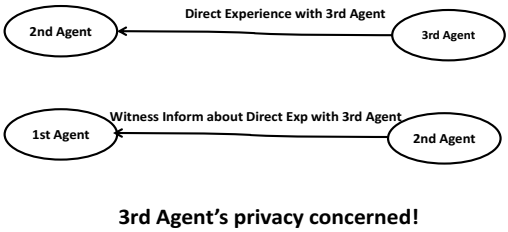

Fig. 1. Schema of communications hold in "classic" witness information

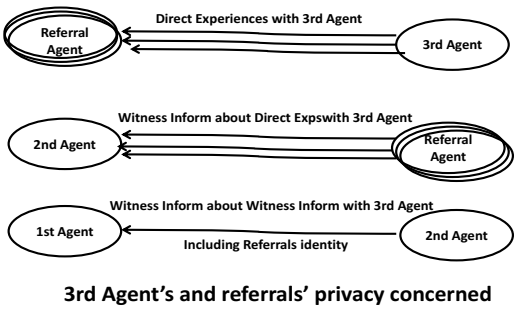

Fig. 2. Schema of communications hold when referrals ids were included in witness information

called indirect information or word-of-mouth, is the information that an agent (we call it in advance first agent) receives from a second agent about a third one. It can be based on the direct experiences of the second agent or it can be based on indirect information from other agents (the so called referrals). In this case, in many trust models, second agents just share the reputational image (a joint computation of several direct experiences and witness information) of the third agent, this is the "classic" way to do it. But some models include in the witness information about the third agents also the referrals of this indirect information, forming then a chain of trust [11] [12]. Therefore privacy of how third party agents behaved with second agents is involved in "classic" witness information (see figure 1). But in the case of trust models that include referrals identity, privacy problems become more extended (see figure 2). Both, third party agents and referrals are then concerned by the disclosure of their behavior (in direct experiences and witness information respectively) that could be violating the intention of those agents of exchanging its knowledge just to the receptors (second agents) and not to any other agent (first agents). In fact most of the trust models when they have to decide to share its information with other agents, they consider the reputation of such agents in order to reject or accept the information request. So they do not share it freely and publicly, and the knowledge of these opinions may have future consequences over its acts in the society as we will show in the

domain example of section 4. It has then full sense that some privacy limitations on the further exchanges of such shared information may take place.

3 Adapting Trust Models to Protect Privacy

With the objective to define levels or conditions of privacy protection for personal information in trust models, it is necessary to identify (according to the European Directive [4]) which legal conditions trust models have to satisfy related to privacy. In particular communications involved in the application of trust models must legally guarantee the exercise of the following rights:

1. Participating agents have to be informed that other agents will collect (trust opinions) personal data about them.
2. Participating agents have to know the name of other agents that will collect such personal data, what the processing is going to be used for, to whom your data may be transferred. They have to receive this information whether the data was obtained directly or indirectly.
3. Participating agents are entitled to ask other agents if these other agents are processing personal data about them;
4. Participating agents are entitled to receive a copy of this personal data in intelligible form;
5. Participating agents are entitled to ask for the deletion, blocking or erasing of the data.
6. Considering that decisions based on such personal data can significantly affect other agents, participating agents must adopt suitable safeguards, such as giving you the opportunity to discuss the thinking behind them, for instance contesting decisions based on inaccurate data.

So we have seen the ways data and media privacy has to legally be protected in real life. So in advance we try to integrate a protection of the 6 privacy rights enumerated before into the corresponding trust communications between agents. Therefore we propose to include additional message exchanges in the protocols of trusting relationships, that would act as control mechanisms that allow trust models to satisfy the 6 privacy rights derived from the European privacy directive. These message exchanges that we propose are:

1. An one way communication: A single message informing to each third agent about the future collection of opinions about them, what the opinions are to be used.
2. Two pairs of additional messages: corresponding to a negotiation protocol (a proposal followed by a counterproposal) on whom these opinions may be propagated (possible first agents in our notation). Although agents collecting opinions (the role of second agents) send an initial proposal (to everyone, to a list of possible first agents, or to none) about the two types of possible opinion transmission (direct or indirect) the final decision has to correspond to the third agents, either considering or ignoring the proposal of second

agents. That decision has to take into account several criteria: whether such third agents are interested in propagating their behaviour, or whether the cooperation with the second agent is of special interest, or whether some possible first agents are possible competitors or the opposite case, when they are potentially interesting cooperative partners. We additionally define an additional privacy constraint to each possible first agent according to the similarity of security policies applied in the communication, in order to limit the possibility of unconscious disclosure of opinions. The corresponding final decision takes then the form of a privacy statement.

3. Agents acting as third agents will request any other collecting agent (first agents) if they are already collecting information about them and what is this information. It involves a pair of messages: one requesting the information and the corresponding response.
4. A one way communication: A single message ordering the deletion or blocking of the already collected opinions.
5. An argumentative dialog between second and third agents about the reasons behind the collected and propagated opinions arguing about the inaccuracy of such decisions. Such argumentation may involve several message exchanges discussing the different factors or criteria involved in such opinions. This sequence of messages may conclude into a final agreement (one of the agents acknowledging the reasons of the other one) or with a disagreement. Such disagreement may then lead to a third agent deciding to order a blocking/deletion communication.

Since all this additional message exchanges may take place or not, it is necessary to define the possible privacy violations in order to effectively control/verify the satisfaction of the legally required privacy rights. Such violations take the next forms:

1. A first/second agent is collecting opinions about another one without its knowledge (no previous informative message was sent)
2. A second agent is propagating opinions about a third agent to first agents that were not included in the corresponding privacy statement of the third agent.
3. A second agent informed about an incomplete or inaccurate collected opinions about a third agent (in other words, it is sending a different or more extended opinions to first agents).
4. A first/second agent ignored the order of blocking or deletion of already collected opinions (in other words, it is still propagating them).
5. A first/second agent do not explain/justify/motivate the opinions about a third agent that is propagating (in other words, it does not respond to the request of justification from a third agent).
6. We also include an additional privacy violation corresponding to the possibility of agents breaking the security of communications where opinions were propagated: A first agent propagating opinions about a third that were never collected (no previous direct or indirect reception of such opinion from any second or third agent).

Therefore six regulatory structures should establish what agents are allowed and forbidden to do according to the required message exchanges and their corresponding violations that were described before. They will be automatically deployed using descriptive semantic elements called norms by the so called Electronic Institutions as we will see in next section.

4 Application of Privacy Protection to ART Testbed Domain with an Islander-built Electronic Institution

As an illustrative example we can use Agent Reputation and Trust (ART) testbed domain [13]. In this application domain, agents act as painting appraisers. Each agent has high expertise appraising paintings on some given eras but not in the others. Additionally each agent receives a set of paintings of any era to appraise from a central entity that simulate the painting owners. So each agent requires the cooperation of other agents to appraise paintings belonging to eras with own low expertise. But since the expertise of each agent is unknown, knowing them and obtaining the cooperation of just the complementary agents (those who has high expertise in the eras with own low expertise) become the real goal of the ART game.

In this ART domain, each agent decides which agents are interesting partners according to several criteria: honest and cooperative attitude, valuable knowledge about others and a complementary expertise in the eras. While being honest and cooperative with the requests from any other agent is always a good strategy to improve our own reputational image and as a general incentive mechanism for providing truthful opinions [14], sharing own opinion about third agents with any other agent is not because it can propagate our information advantage (the already known expertise of others) to our natural competitors (those who has high expertise in the same eras than us). Therefore, a gaining strategy in this ART game would be to limit the propagation of the knowledge about others to those agents that are of our interests (complementary expertise) while avoiding such knowledge to reach agents who are our natural competitors. We obtain the illocutions, roles and relationships corresponding to ART interactions from our previous work moving the adhoc ART testbed platform to JADE environment [15]. Such protocols correspond to those protocols involved in ART testbed which were formalized in a FIPA-compliant way in our previous work [16].

Therefore, we can use ART application domain to define the additional messages (including the corresponding concepts, predicates and actions used to define the message contents) required to attend the corresponding privacy requirements of the five types defined before in the previous section. Next we show the five privacy-preserving (FIPA-compliant) protocols applied to ART-testbed that we have implemented in JADE[17]:

1. a message with INFORM as FIPA performative, and with a *IsCollecting* predicate as content. This predicate has the next properties with the corresponding concepts as values: Who: Appraiser Agent, On: Era, Value: Reputation.

2. a pair of messages: The first one with a PROPOSE performative and a *StatesPrivacy* action as content. This predicate has the next properties with the corresponding concepts as values: Who: Appraiser Agent, On: Era, Whom: None/All/Appraiser Agent, Type: Indirection Level, How: Security Policy. Where Indirection Level concept may have two values: direct (direct experiences) and indirect (witness information) and Security Policy would describe the rules to be applied into cryptographic algorithms of communications. The second message, the corresponding response to this PROPOSE message may be an ACCEPT PROPOSAL or an REJECT PROPOSAL. In case of a rejection, the message will include a *StatesPrivacy* action as content in order to be considered a counterproposal.
3. a pair of messages: The first one with QUERY-REF as FIPA performative with a *Is Collecting* predicate as a content, where property Value has a void Concept associated. The corresponding response message is a INFORM-REF performative with the value property of *Is Collecting* predicate fulfilled with the actual Reputation collected.
4. a message with REQUEST as FIPA performative, and with *Blocking* or *Deleting* action as content. Such predicates have the next properties: Who: Appraiser Agent, On: Era, Value: Reputation. Next, the other agent has to answer with an AGREE performative in the response message.
5. a sequence of messages: The first one from an iniciator agent with QUERY-REF as FIPA performative with a *Justification* predicate as a content, where this predicate has the next properties with the corresponding concepts as values: Who: Appraiser Agent, On: Era, Value: Reputation, From: Appraiser Agent, Type: Indirection Level, Initial Value. Where the properties From, Type and Initial Value have a void Concept associated that the corresponding response INFORM-REF message would fulfill with the Agent source of such argued reputation value, the way this reputation value was collected (direct vs. indirect) and the value originally sent by this source agent. After this second message an additional REQUEST message might take place from the initiatior agent to suggest the other agent to rectify the reputation value collected from the source agent. In order to motivate such rectification, the initiator agent would include the details of the direct interaction with such source agent (if that interaction really took place). This REQUEST message includes a *Rectifying* action that includes the real and appraised value of the painting corresponding to such interaction. Finally, the other agent could answer with a REFUSE or either an AGREE performative in the response message.

To illustrate the defined protocols to protect privacy in ART testbed we include the figure 3.

Once the additional message exchanges to be applied in the ART application domain to protect legally privacy were defined, we now have to explicitly formalize a set of norms that constraints the behaviour of agents to the right use of such message exchanges. The specification of these norms corresponds to an electronic institution. We have chosen to do it with Islander [18] because

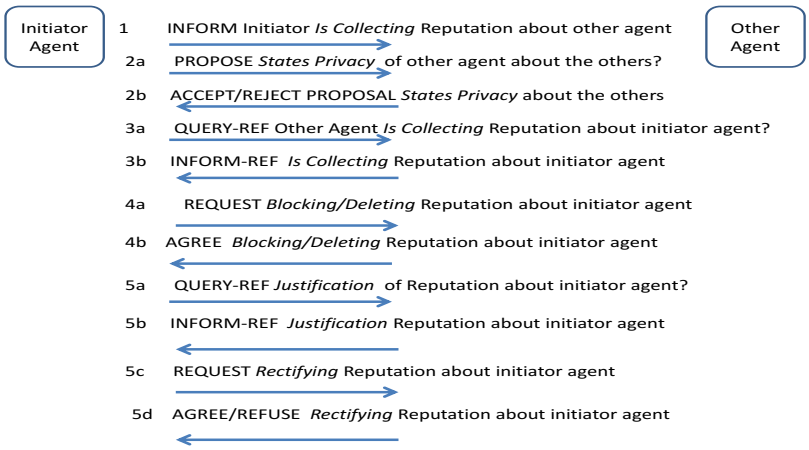

Fig. 3. Privacy Preserving JADE Protocols in ART Domain

it designs such social constraints with a combination of textual and graphical elements. It is also remarkable that Islander does not assumes any particular agent architecture or language for the participating agents. This specification tool is complemented with other tools that simulates and tests the execution of electronic institutions. Therefore we intend to define the six privacy violations described in the end of section 3 with the BNF syntax format of Islander as an extension of this work.

5 Conclusions

In this paper we have motivated the use of privacy-protection measures in trust models, both in conscious exchanges of opinions and in an unconscious way when security attacks take place. We have enumerated five privacy protection requirements to be applied in trust models according to the current European Directives. Such five requirements have been implemented as interaction protocols with JADE in the ART testbed domain. This protocols involve the definition of additional concepts, predicates and actions to be included to the ART ontology defined for JADE in our previous works. Furthermore, six possible privacy violations that might take place have been formalized as norms of an Electronic Institution designed with Islander tool. This is the first serious effort of formalizing and implementing privacy protection (JADE protocols) on trust models in agent societies. As future works we propose to implement privacy preserving norms with Islander and to evaluate the influence that such privacy protection causes in the trusting reasoning and decisions, for instance, has the fact that an agent can ask about the information that is collected and can ask for deletion or blocking such information, or the effects caused by the agent interaction

constraints that control privacy violations. Using the privacy protection provided, we intend to design some experiments in order to analyze the effects and performance of this privacy protection.

Acknowledgements. This work was supported in part by Projects MINECO TEC2012-37832-C02-01, CICYT TEC2011-28626-C02-02, CAM CONTEXTS (S2009/TIC-1485).

References

1. Clarke, R.: Information technology and dataveillance. Commun. ACM 31(5), 498–512 (1988)
2. Calvo-Rolle, J.L., Corchado, E.: A bio-inspired knowledge system for improving combined cycle plant control tuning. Neurocomputing 126, 95–105 (2014)
3. Wozniak, M., Graña, M., Corchado, E.: A survey of multiple classifier systems as hybrid systems. Information Fusion 16, 3–17 (2014)
4. Parliament, E.: Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (October 1995), `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML`
5. Pedraza, J., Patricio, M.A., de Asks, A., Molina, J.: Privacy-by-design rules in face recognition system. Neurocomputing 109 (2013)
6. Pearson, S.: Privacy, security and trust in cloud computing. In: Pearson, S., Yee, G. (eds.) Privacy and Security for Cloud Computing. Computer Communications and Networks, pp. 3–42. Springer London (2013)
7. Sattar, A.H.M.S., Li, J., Ding, X., Liu, J., Vincent, M.W..: A general framework for privacy preserving data publishing. Knowl.-Based Syst. 54, 276–287 (2013)
8. Pallapa, G., Francescoy, M.D., Das, S.K.: Adaptive and context-aware privacy preservation schemes exploiting user interactions in pervasive environments. In: 2013 IEEE 14th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 1–6 (2012)
9. Dellarocas, C.: The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management Science 49, 1407–1424 (2003)
10. Sabater-Mir, J., Sierra, C.: Review on computational trust and reputation models. Artificial Intelligence Review 24, 33–60 (2005)
11. Yu, B., Singh, M.P.: An evidential model of distributed reputation management. In: In Proceedings of First International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 294–301. ACM Press (2002)
12. Esfandiari, B., Chandrasekharan, S.: On how agents make friends: Mechanisms for trust acquisition. In: In Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies, pp. 27–34 (2001)
13. Fullam, K., Klos, T., Muller, G., Sabater, J., Schlosser, A., Topol, Z., Barber, K.S., Rosenschein, J., Vercouter, L., Voss, M.: A specification of the agent reputation and trust (art) testbed: Experimentation and competition for trust in agent societies. In: The Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2005), pp. 512–518 (2005)

14. Gómez, M., Carbo, J., Benac-Earle, C.: Honesty and trust revisited: the advantages of being neutral about other's cognitive models. Autonomous Agents and Multi-Agent Systems 15(3), 313–335 (2007)
15. Moya, J., Carbo, J.: Distributing art agents with jade. In: 10th European Workshop on Multi-Agent Systems, EUMAS (2012)
16. Carbo, J., Molina, J.M.: A jade-based art-inspired ontology and protocols for handling trust and reputation. In: Ninth International Conference on Intelligent Systems Design and Applications, ISDA, pp. 300–305 (2009)
17. Bellifemine, F.L., Caire, G., Greenwood, D.: Developing Multi-Agent Systems with JADE. Wiley (2007)
18. Esteva, M., Cruz, D.d.l., Sierra, C.: Islander: an electronic institutions editor. In: The First International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS, pp. 1045–1052. ACM (2002)