A Social-Network Defence against Whitewashing *

(Extended Abstract)

Adrian Perreau de Pinninck
IIIA - CSIC
Campus de la UAB,
Barcelona, Spain
adrianp@iiia.csic.es

Marco Schorlemmer
IIIA - CSIC
Campus de la UAB,
Barcelona, Spain
marco@iiia.csic.es

Carles Sierra
IIIA - CSIC
Campus de la UAB,
Barcelona, Spain
sierra@iiia.csic.es

Stephen Cranefield
Dept. of Information Science
University of Otago
Dunedin, New Zealand
scranefield@infoscience.otago.ac.nz

ABSTRACT

We provide a defence against whitewashing for trust assessment mechanisms (TAM) by using an underlying social network in MAS and P2P. Since interaction requests are routed through the social network, routers can block requests from portions of the network known for whitewashing. Furthermore, by limiting feedback spread to the interaction routers, the trust assessment can be done without querying for feedback with a small loss in efficiency.

Categories and Subject Descriptors

I.2.11 [**Distributed A.I.**]: Multiagent systems

General Terms

Security

Keywords

Whitewashing, Social Network, Reliability

1. INTRODUCTION

The aim of a TAM is to assess the likelihood of another agent delivering the expected service. Malicious agents will try to subvert such mechanisms through different types of attacks. Whitewashing is an attack in which the malicious agent changes its identifier in order to avoid negative assessments from previous feedback. State-of-the-art TAMs have poor response to whitewashing attacks. The defence we present in this paper make TAMs robust to whitewashing.

*This work is supported by the Generalitat de Catalunya grant *2009-SGR-1434* and the Agreement Technologies Project *CONSOLIDER CSD2007-0022, INGENIO 2010*

Cite as: A Social-Network Defence against Whitewashing (Extended Abstract), A. Perreau de Pinninck, M. Schorlemmer, C. Sierra and S. Cranefield, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. . 1563-1564

We assume that all interactions among agents follow a basic protocol. i) An *initiator* agent selects the *potential partners*. ii) The initiator assesses the trust on each potential partner, selects an agent with which to interact, and sends an interaction *request* to the selected *target*. iii) Upon receiving the request, if the target accepts the request, an interaction starts between the initiator and target (the *partners*). iv) After the interaction is over the partners may send *feedback* about the interaction.

We use a social network structure in which each agent is connected to a set of *contact* agents. Our experience in [1] showed that structuring a MAS as a social network aids in achieving norm compliance. Interaction requests are routed through a path of contacts towards the target. Therefore, in order to interact an agent needs to know at least one other agent already in the network. This alone makes simple whitewashing fruitless, since if an agent changes its identity, its contacts no longer recognise it and its requests will not get routed. Nonetheless, having many temporary identities that whitewash by hiding behind a permanent identity is still possible.

In the proposed defence, partners only send feedback to agents that routed the request, and agents only take into account feedback about interactions that were triggered by requests they routed. This brings about smaller message overhead and less feedback available for assessment. However, since trust assessments are integrated into the routing mechanism and an agent's contacts are bound to have more feedback regarding it, the routers can realise the best assessments. A router must decide whether to forward, re-route, block, or discard the request depending on the trust assessment (see Figure 1 for the full protocol).

The focus of the defence against whitewashing is to use feedback of interactions routed by the current routers to estimate the trust on an agent for which there is no feedback. The following equation describes the social-network defence applied to the PeerTrust [2] metric[1]. In the equation u is the agent assessing the trust, R is the set of routers that

[1]We have chosen PeerTrust because it is robust to most known attacks without relying on centralisation or pre-trusted entities.

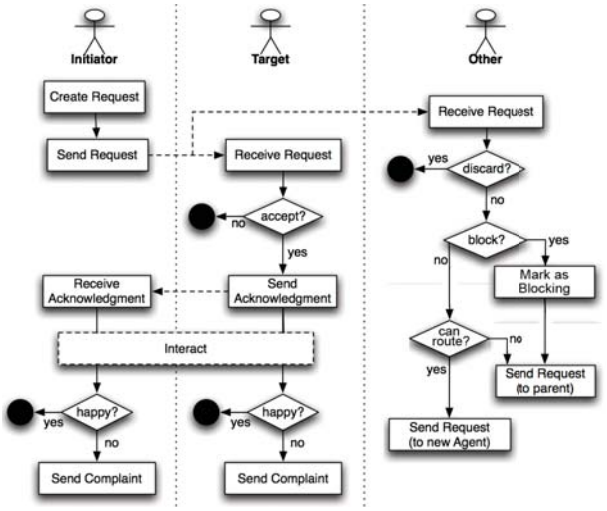

Figure 1: Interaction protocol

have trusted the assessed agent (the request forwarders), $RP(r)$ is the set of routed partners, $RS(r)$ is the routed satisfaction, and $Sim(W, p)$ is the similarity between two feedback vectors, one for the peer being assessed and the other for the feedback of interactions routed by the given set of routers.

$$T_{WW}(u, R) = \frac{\sum_{r \in R} RS(r) \times Sim(u, RP(r))}{\sum_{r \in R} Sim'(u, RP(r))}$$

A theoretical analysis shows that the cost of PeerTrust when enhanced with the social-network defence is $O(\ln n)$ for scale-free and small-world networks, whereas the cost of the original PeerTrust is $O(n^2 \cdot \log_2 n)$. Empirical tests comparing the original PeerTrust and the enhanced version to a system without a TAM support the claim that the cost is reduced by orders of magnitude, and that the defence is more robust against collusion and whitewashing attacks. Experiments have shown that the defence reduces the robustness against simple attacks. However, the results are always better than not having trust assessment.

2. EXPERIMENTS

There are three main experimental scenarios. i) Single attack, in which malicious agents try to cheat as much as they can without using any complex attack schemes. ii) Collusive attack, in which malicious agents form a collective that tries to boost the members' reputations in order to cheat more often with the non-malicious agents. iii) Whitewashing attack, in which malicious agents help each other in attempting whitewashing attacks.

The single attack experiment revealed that PeerTrust reduced the non-satisfactory interactions by 96.8% and the enhanced version achieved a 68.7% reduction. PeerTrust does better because agents have more information. The collusive attack experiment showed that PeerTrust reduced the non-satisfactory interactions by 58.7%, and the enhanced version achieved a reduction of 78.2%. The fact that the enhanced version was more robust for colluding scenarios came as a surprise. We believe this happened because the colluding feedback is scattered throughout the network, whereas the feedback important for the blocking decision is concentrated in the routers close to the partners. Finally, the experiment on whitewashing cheaters showed that PeerTrust achieved no statistically significant reduction in non-satisfaction, on the other hand the social-network defence reduced the non-satisfactory interactions by 27.1%. (see Table 1).

Attack	PeerTrust	RBR	Error rate
Simple Cheating	96.8%	68.7%	0.8%
Collusion	58.7%	78.2%	0.22%
Whitewashing	0.002%	27.1%	0.13%

Table 1: Reduction in non-satisfaction

Table 2 shows the mean number of messages per agent and round for all experiments. The data shows that the cost was much higher for PeerTrust. Although the number looks exceedingly large for PeerTrust, they conform to the expected analytical value.

Attack	None	PeerTrust	RBR
Lone	5.0	1.7×10^7	8.6
Colluding	3.7	3.3×10^7	6.7
Whitewashing	4.0	2.1×10^7	7.1

Table 2: Mean number of messages per interaction

3. CONCLUSIONS AND FUTURE WORK

We have presented an innovative approach to defend a trust assessment mechanism against whitewashing attacks by using an underlying social network. Trust assessment techniques should be designed in order to be robust against different kinds of attacks: badmouthing, ballot-stuffing, dynamic personality, collusion, and whitewashing. Most systems treat the former four to a good degree, but whitewashing seems to be an attack that is hard to counteract when identifiers are freely available. The proposed defence has the following benefits: it is totally distributed, relatively easy to implement, it reduces the overhead compared to other approaches, and it is robust against whitewashers and colluders.

In future work we plan to test whether our defence can be made robust to free riding by using MANET trust based routing techniques. We also want to develop middle-ware that implements the social-network defence, and to embed it into existing distributed systems. Furthermore, we plan to test mechanisms for dynamically changing agent's contacts so that the routers achieving most satisfaction become hubs, thus making the system more robust.

4. REFERENCES

[1] Adrián Perreau de Pinninck Bas, Carles Sierra, and Marco Schorlemmer. A multiagent network for peer norm enforcement. *Autonomous Agents and Multi Agent Systems*, In Press.

[2] Li Xiong and Ling Liu. PeerTrust: supporting reputation-based trust for peer-to-peer electronic communities. *IEEE Transactions on Knowledge and Data Engineering*, 16:843–857, 2004.