Towards a Human Trust Model for Mobile Ad-hoc Networks

Licia Capra
Dept. of Computer Science, University College London
Gower Street, London WC1E 6BT, UK
L.Capra@cs.ucl.ac.uk

1 Background and Motivation

Rapid advances in wireless networking technologies have enabled mobile devices to be connected anywhere and anytime. While roaming, applications on these devices dynamically discover hosts and services with whom interactions can be started. However, the fear of exposure to risky transactions with unknown entities may seriously hinder collaboration. In order to advance the goal of anywhere-anytime computing, the exposure to risky transactions has to be reduced as much as possible. This requires the existence of a *trust management framework* that enables devices to form, maintain and exchange trust opinions. These opinions can then be used to customise the way interactions take place: for example, to decide from where to download a file, what service provider to contact, what access rights to grant, and so on. Trust is obviously not the only aspect that must be taken into account when making these decisions: the perceived risk inherent to a transaction, and the quality of service (QoS) requirements will all contribute to the final configuration decisions. However, feelings of trust, risk and QoS can be formed independently of each other, and thus dealt with separately, before being combined. At present, we are concerned with trust management issues only.

A trust decision framework for mobile ad-hoc networks must be fully *decentralised*, as we cannot assume the existence of a trusted third party that can be contacted on demand to acquire reputation information about an entity. Approaches such as [4] cannot therefore be applied to the mobile setting, as they assume the existence of a central specification server where trust information is stored and used. The framework must be highly *customisable*, in order to capture the varying and complex natural disposition of an individual to trust into computer models; this should be achieved without causing disruption to the device computation and communication resources. Approaches such as [2] work well at the routing level, where trust decisions are automatic and homogeneous, but suffer severe limitations at the application level, where subjectivity in the decision making process becomes fundamental. Other approaches (e.g., [5]) deal with trust in a less automatic fashion, but they still fail to capture a variety of aspects peculiar to human trust (e.g., ways to recover from a bad reputation, natural disposition to trust unknown entities, etc.). Finally, a trust decision framework for mobile ad-hoc networks must be *selfish*: in a resource constrained environment, selfishness is likely to prevail over cooperation, for example, to save battery power. A trust management framework cannot therefore completely rely on the assumption that entities have a social conscience that will make them exchange reputation information whenever asked. This limits the applicability of approaches such as [1].

We are currently designing and formalising a trust management framework that meets these requirements. Our approach is completely decentralised: each entity acts as a self-contained unit, carrying along a *portfolio of credentials* derived from the past interactions of the entity, and that the entity uses to prove its trustworthiness to others. This portfolio is created and maintained during peer interactions, and can be used as the unique source of reputation information when having to make a trust decision, in case the social context is populated by a majority of selfish agents that are not willing to propagate reputation information. Finally, our model makes intensive use of customisable functions to adapt the behaviour of the trust management framework according to the agent's disposition, thus capturing human models of trust in computer models. Altogether, these functions enable the model to semi-automatically derive new trust relationships from previously formed ones. In the following section, we provide a more detailed description of the major characteristics of this trust management framework.

2 Trust Management Model

In our model, we define *trust* as the degree of belief about the behaviour of other entities, also called *agents*, upon which we depend (for example, to have a service delivered). Trust is dynamic and it tends to be reduced if entities are misbehaving; viceversa, it increases if agents are doing well. Our trust management framework thus characterises as a self-adjusting system, or infrastructure, used to form and exchange trust opinions about the agents that populate the network.

Closely related to trust management is the issue of identification: we must be able to bind a trust opinion to an identity; however, creation and deletion of identities is very quick and easy in mobile settings, and malicious agents could exploit it to repeatedly misbehave, without being isolated. We do not address the issue of identification in our trust management framework. We assume each agent has got a pair of public/private keys (perhaps more than one), that is managed via an independent public-key management system (e.g., [3]).

Based on these assumptions, key issues our trust management model deals with are the following.

Exchange of Credentials. Whenever agent a has to make a trust decision about b, she first asks b for her portfolio of credentials. Each credential has the following format: $[x, b, l, t, c, time, n]_{SK_x}$, meaning that agent x trusts b at level l to carry on task t in context c. This trust is dated *time*, after n transactions have taken place between x and b. This letter of presentation is authentic, as it has been signed with x's private key. At this point, agent a may use the information contained in the portfolio to make a trust judgment about b, or she may query the neighbor agents to obtain further information (for example, because the portfolio provided by b is outdated, or because it contains recommendations signed by agents unknown to a, or that a does not trust). Once the neighbor agents have replied, a finally forms her trust judgment. A customisable *trust decision function* τ is used to drive the behaviour of the credential exchange protocol, and to output a trust judgment, based on both the portfolio and the recommendations received. Each agent builds her portfolio during direct interactions: our protocol demands that, whenever interaction between a and b takes place, they exchange at the end a letter of presentation in the format described above. This is a far less demanding assumption than requiring an agent to answer all reputation information requests from neighbor agents. Malicious agents that continuously create new public/private keys to conceal past misbehaviors will have no portfolio to introduce themselves, making it hard to interact in the future.

Isolation of Malicious Agents. A malicious agent may trick other agents by spreading: *fake bad reputation*, in order to compromise an agent, and *fake good reputation*, in order to build a false portfolio. The mechanism we provide to detect and isolate malicious agents is basically a *conflict detection* mechanism. A conflict is detected whenever there is a high discrepancy among the letters of presentation regarding an agent. A customisable *inconsistency resolution function* $\mathcal{I}$ defines the behaviour the agent is willing to adhere to whenever a conflict is detected: mark the agent as ‘suspect’ and be alert in the future, mark the agent as ‘cheating’ and do not interact with her anymore, or do nothing, in case the conflict is considered a non-malicious divergence of opinions. We assume the existence of a higher number of benevolent agents in the network than of malicious agents; in this case, the more cohesive and dense is the network, the quicker misbehaviors are detected.

Tacit Information Extraction. Both the trust decision function and the inconsistency resolution function should capture the varying disposition to trust of the human entity involved in the trust decision process (i.e., the user of the device) into computer models. In order to do so, they rely on tacit information that is extracted from each interaction, and that assesses the trustworthiness of agents both as service providers and as recommenders, as perceived by the user. The inconsistency resolution function then uses this information, for example, to decide in favour of whom to solve a conflict of trust opinions, while the trust decision function may use it to weigh letters of presentation coming from agents that share our opinions (i.e., trusted recommenders) more than those coming from unknown agents.

3 On-going and Future Work

The long-term goal of this research project is to design a trust management framework for mobile ad-hoc networks, that enables mobile devices to form, maintain and exchange trust opinions over the network. Trust information should then be combined with risk information and QoS requirements to customise interactions in this context.

We are currently working on the design and formalisation of a trust management framework that meets the requirements of distribution, customisation and selfishness previously discussed. This will be followed by a simulation phase where we plan to evaluate our model with respect to: speed at which reputation information is spread, ability to detect malicious agents (and consequently accuracy of trust judgments), adaptability of the model to the user's disposition.

This work is part of the “Trusted and QoS-Aware Provision of Application Services” project (IST-2001-34069).

References

[1] A. Abdul-Rahman and S. Hailes. Using Recommendations for Managing Trust in Distributed Systems. In *Proc. of IEEE Malaysia International Conference on Communication (MICC'97)*, Kuala Lumpur, Malaysia, November 1997.

[2] S. Buchegger and I.Y. Le Boudec. The Effect of Rumor Spreading in Reputation Systems for Mobile Ad-hoc Networks. In *Proc. of WiOpt 2003: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, Sophia-Antipolis, France, March 2003.

[3] S. Capkun, L. Buttyán, and J.P. Hubaux. Self-Organized Public-Key Management for Mobile Ad Hoc Networks. *IEEE Transactions on Mobile Computing*, 2(1):52–64, 2003.

[4] T. Grandison and M. Sloman. Specifying and Analysing Trust for Internet Applications. In *Proc. of 2^{nd} IFIP Conference on e-Commerce, e-Business, e-Government*, Lisbon, Portugal, October 2002.

[5] L. Rasmusson and S. Janson. Simulated Social Control for Secure Internet Commerce. In *New Security Paradigms Workshop*, pages 18–26, Lake Arrowhead, CA, September 1996. ACM Press.