A TRUST-BASED MODEL FOR MULTI-AGENT SYSTEMS

Jan Samek[1]

[1]Department of Intelligent Systems, Faculty of Information Technology
Brno University of Technology

samejan@fit.vutbr.cz (Jan Samek)

Abstract

Trust is very important aspect in our everyday interaction with people, groups and institutions in our society. We have trust in our environment, people and institutions as well. We are often rated and judged on the basis of our trustworthiness and this defines a different manner of the interactions in our social life. We behave more openly towards subjects on account of the strong confidence and this subjects can access different types of information which can be confidential. In the case of abuse of the information, the trust of the subject rapidly decrease and it is very hard to restore it. Some research in this area are aimed to use these trust principles from real-word and shift it into digital environment nowadays. In autonomous multi-agent systems, where agents are operated in a networked environment, it is particular possible to use these principles of trust to establish protocols for agents interaction. If we try to shift this real-word trust principles to multi-agent systems, we meet with some essential problems. This paper try to show a possible approaches to the basic problems with multi-agent systems based on trust. It presents the way to simulate the trust or the reputation from a viewpoint of application and safety in multi-agent systems.

Keywords: Trust, Reputation, Multi-agent system

Presenting Author's Biography

Jan Samek was born in 1982 in Opocno, Czech Republic. He attended on secondary secondary Technical School of Electrical Engineering, He graduated from Brno University of Technology, Czech Republic with MSc. degree in 2006 in Electrical Engineering and Computer Science. He is a student of PhD study Computer Science in Brno University of Technology since 2006. His MSc. theses was *Security of Information Systems* - this theses was about an analysis of information systems from the security point of view, especially analysis of risks and threats of current information systems approaches. He is also member of Brno University Security Laboratory (BUSLab). His research interests: IT security, information systems, intelligent systems – especially multi-agent systems and their security.

Acknowledgment: The research has been supported by the project MSM0021630528 (Ministry of Education, Youth and Sports, Czech Republic) "Security-Oriented Research in Information Technology".

ISBN 978-3-901608-32-2

1 Introduction

Latest researches shows, that system based on trust and reputation (for user rating) have great potentiality, mainly for e-commerce systems. This can be seen for example on world-famous auction server eBay, where the selection of seller (from buyer point of view) is based on his reputation. All users in the system can be participated on this reputation. Trustworthiness of seller so as buyer is represented by score value, which is update by system and depends on cumulating positive and non-positive ratings from other sellers or buyers. The reputation system in eBay is well studied, economics analysis shows that reputation has statistically significant effect on price and overall income for sellers [6]. This reputation system, from our point of view, can be accounted relatively simple a closely aimed.

The question is, how can we use trust and reputation principles – which is much more important in our real-life environment for behaviour and interaction – for other branches of information technologies? Towards this, some theoretical [11, 10, 7, 13, 8] also practical [1, 14, 3] studies were proposed, which were concentrate to different aspects of trust and reputations in different systems. Most of these studies describes only the practical aspects of using trust and reputation, typically without respect conceptual approaches.

Our objective is to provide elementary approaches in modelling multi-agent systems based on trust and reputation, allow to identify and understand some specific problems which brings these systems. After this, we can start to formally describe our model, construct elementary notation which will be used for model development in implementation and testing phase.

The remainder of this paper is organised as follows:

Section 2 presents some basic definitions which are basically adopted form [10, 7] and modified for our model requirements.

Main part is in section 3, where we describe elementary aspects of each trust model. At the beginning of this section, we define requirements to our model, describe the agent role in this model from different points of view. Next, we discuss trust representation, production, distribution and evaluation in the model. At the last part of this section, we focus on security questions about model, mainly to agent authentication and secure communication.

Section 4 describes the open issues and future work and concludes our paper.

2 Trust and Reputation

Before we start describe our model proposal, it is necessary to define some terms, such as *trust* and *reputation* and explain the difference between them and their relation. This definitions are primary adopted from other authors and at the end of this section, we refined them for our model purposes also for better understanding their meaning.

Trust is a subjective expectation an agent has about another's future behaviour based on history of their encounters. [10]

Reputation of agent A in our view is the average trust (whatever average means in that context) of all other agents towards A. [7]

Experience in this context (trust of agent A in an agent B) is therefore an observation of A about some behaviour of agent B in different aspects. At the base of this observation, is necessary to be able to judge whether an experience has been positive or negative in order to update A's trust in B.

Recommendation is a subjective information an agent A regarding some aspects (quality, reliability, ...) about the target (target of recommendation – another agent C) to agent B.

Terms *experience* and *recommendation* is partially adopted from [7] and redefined for our model purposes.

Differentiation of trust and reputation is either not made or the mechanism for inference between them is not explicit. Trust and reputation are different in way how they were developed, they are closely related. They are both used to evaluate a agent's trustworthiness.

Reputation can be centralised and decentralised. Centralise reputation is evaluated by a trusted third party. In the case, the reputation has a global aspect an trust has viewed as local and subjective value. This solution does not fit our model very well, because we are trying to find approaches to fit fully distributed model.

Toward this, for our model, decentralised case is, that reputation is evaluated independently by each agent after asking other agent for recommendations. Absolutely clearly, reputation is a agent's belief in another agent's capabilities based on recommendations received from other agents.

3 Trust model concept

3.1 Model requirements

In this part of work, we will concentrate on key parts of model, which are important and must be well studied for future formal specification. We will discuss available approaches for each problem toward to model requirements. This requirements can be summarised to the following items:

Distributivity – The most information systems of all kind are developed as centralised, this concept have some advantages and also disadvantages. The main disadvantage is, that in this systems exists only one central point, which ensure some critical mechanism in the system. If this central point is fails or is compromised, the whole system is typically quite functionless and its re-establish is very difficult, worse, it is often impossible. Also, multi-agent system are naturally decentralised. Toward

ISBN 978-3-901608-32-2

this, we pose requirement of distributivity at the top.

Heterogeneity – Agent and multi-agent system can be characterised by heterogeneous qualities and capabilities of each agent. Therefore our model must satisfy this requirements too. With this requirement, we also allow that each agent in the system can be implemented on different architecture, moreover we allow to the developers freedom of choice in using agent platform (on which system was build on), therefore each agents can connect to the system from various platforms.

Security – Our proposed model must be secure, it means that the system, based on our model, need to ensure basic security principles such as: *confidentiality*, *integrity* and *availability*. System must be resistant against different forms of attacks. System must ensure, that agent or group of agents collaborating together can not do intent or un-intent damage in the system or it's part. The system must be available for all authorised agents as well. High security requirements can relatively complicate the implementation phase of the system, but in the phase of the model proposals and formalisation do not play such important role. In this phase, we need to know security requirements but to describe elementary principles of trust-based models, define some terms and study specific characteristics of these models. We do not need to solve them in detail. Draft of some security approaches, mainly for authentication and secure communication between agents, are discussed at the last part of this section.

Flexibility – System based on our model must dynamically adapt for new states and changes. This adaptation must be considering to model functionality and purpose. First side of view to model flexibility is mainly from aspect of trust and reputation evolution algorithm toward to highly various properties of the system, for example: wide of system (count of agents), ration between agent capabilities (mainly *providers*/*advisors* – will be defined later), ability of system to react/response to different forms of attacks and similar actions. Second side of view is the self model proposal flexibility and implementation flexibility. That means the ability of model and system to flexible to add new functionality or change some functionality. This change can cause directly or indirectly, globally or partially changes in model/system. From both side of view, this is very strong requirements and it is not easy to achieve.

3.2 Agent

The main element of the system is an active autonomous entity, we call this entity as *agent*. Agent in our meaning is an autonomous active entity, which is able to decide independently, act and interact in its environment on its own behalf or on behalf of it's owner. Agent behaviour and decision in its environment is

- depending on his knowledge, capabilities and intentions;
- intelligent.

There are many formal representation of intelligent agents, for example *rational agent* and his typical representation in BDI (Beliefs, Desires and Intentions) logic [2], which can be implemented as architecture PRS (Procedural Reasoning System) [5]. Currently, its out of scope of this paper to describe approaches for agent representation and its will be described in future work.

In real system, agent can be represented as stand-alone server or simple application running on some networked computer.

Agent characteristic

The agents can be characterised from two points of view. First, from the *function* (or *purpose*) the agent in system, second from agent *relation* to trust value. Agent functionalist in system means: agent quality to provide owns capabilities regarding to other agents is the system. According to this, we categorise agents by capability to *providing services* and reliability in *providing recommendations*.

- Service providers - agents, which provides services to another agent(s). Service rules are defined by providing agent.
- Recommenders (Advisors) - agents, which makes referrals or recommendations.

Each agent can be provider, recommender or both of them. This capability can be vary in time.

From aspect of agent relation to trust value, we categorise agents to trust *producers* and *consumer*.

- Trust producers - agents, which are producing relevant trust value regarding to another agents.
- Trust consumers - agents, which are using information produced by trust producers for our reasoning.

In most of the models, the agents are both, producers and consumers.

Agent strategies

Agent strategies in which the decision for an encounter with an agent is based on a few aspects. First of all is typically last interaction with that agent. If no interaction was made in the last, agent must compute some initial trust depend on expectations or recommendations. Based on this, agent must finally decide to *cooperate*, *defect* or *ask* other agent *for recommendations*.

Agents maintain interaction histories of other individual agents, and use these histories to ascribe reputations to individual agents. Agents use strategies in combination with observations and interaction histories for decision making.

ISBN 978-3-901608-32-2

3.3 Trust representation, producing and evaluating

Therefore the agent capability to decide to *cooperate* or *defect* base on trust value, to *produce* or *consume* trust value, we need to provide some metrics to represent trust value.

Representation

In some models [10, 15] is the trust value represented such as binary value, typically $t \in \{0,1\}$, it can means $t \in \{untrustworthy, trustworthy\}$ (also cooperate/defect, good/bad, high/low, etc). This representation is suitable for some case studies and simulations. However, this representation do not fit for our model purposes from some reason. Firstly, its not reflects real-word trust representation. Secondly, we need to express such kind of *partial trustworthy* or *partial untrustworthy* for modelling recommendations effects closely.

Toward this, we define trust value as real number on interval $t =< x, y >$, where x represent worst possible rating and y represent the best possible rating of agent trustworthy. It is not important if the $x = 0$ and $y = 1$ – so interval $t =< 0, 1 >$ or $x = -1$, $y = 1$ – so interval $t =< -1, 1 >$. Decision about this interval will keep on specific implementation. However, is important to ensure that trust value must change from x to y with difference Δt, which respect model requirements and trust evaluating manners of each agent.

Characteristics

Context specific – Trust and reputation both depend on some context [18]. For example, **Alice** trusts **Bob** as her doctor, because **Bob** is her doctor, but she does not trust **Bob** as a chief who can cook apple pie. So in the context of seeing a doctor, **Bob** is trustworthy, but in the context as chief, is untrustworthy. For our model, if agent provide more services, he can be different trustworthy in each of them.

Multi-faceted – Each agent evaluate trust from different aspects of capability of another agent (service) and each aspect has different value to overall trust. So overall trust is the combination of all these aspects by agent preference. For example, **Alice** evaluate restaurant from environment and kindly staff, while **Bob** from prices and speed with which is food served.

It is possible to say, that different agents evaluate and distribute different kind of trust and reputation depending on his particularity. These agents capability, evaluate and distribute recommendation for different aspects of one service, in case of large scale multi-agent system, implies big requirements for agent capabilities and therefore it will be the object of study in future work. Currently we define overall trust and reputation, which is some average value, evaluated from each aspect of service, by agent preference.

Direct and indirect trust and reputation

Trust and reputation can be derived *directly* and *indirectly*. *Direct trust* refers to trust, which is evaluated from direct agent interaction. *Indirect trust* refers to trust, which is observed in agent environment from other agent behaviour; or from beliefs about another agents capabilities.

Then again this, *direct reputation* is obtain from recommendation another agent in question. *Indirect reputation* is reputation, which is obtained from recommendation received indirectly – recommendation obtained from communication with another agent, which is not based on direct interaction between recommender and target of recommendation. For example: **Alice** asks **Bob** for recommendation about **Catie**, **Bob** have no direct experience which **Catie**, but *he heard* that Catie is lovely. He response to **Alice**: "*I was heard* that **Catie** is lovely". This type reputation call Mui, *et al.* [11] as "word-of-mouth".

Individual vs Group reputation

Reputation can be used to describe an individual (agent *to* agent) or a group (agent *to* group of agents, where group are at least two agents) of individuals. Existing reputation systems such as those in eBay, Amazon or Slashdot concentrate on reputation of the individuals. Economists have studied group reputation from the perspective of the firm. A firm's (group) reputation can be modelled as the average of all its members individual reputation [11].

Halberstadt and Mui [9] have proposed a hierarchical group model and have studied group reputation based on simulations using the hierarchical model. Their group model allows agents to belong to multiple overlapping groups and permits reputation inferences across group memberships.

In this model, agents use interaction histories to ascribe reputations to individual agents and also to groups of agents for varying contexts. It is still very difficult problem to evaluate group reputation, because we need to take into account *context specific* and *multi-faceted* characteristics of reputation. For our model purposes and in next text we will study only individual reputation.

Evaluating

After each interaction, agents make an evaluation, on its base the trust value of their counterparts is updated. Agents for these evaluating have different criteria. For the same interaction may follow different kind of evaluation, therefore the overall trust value may be different for each interacting agent. So, the overall evaluation of an interaction is combination of evaluations of each aspect related to the interaction, such as quality of service, speeds, etc. How the agent combine each aspect of interaction to the overall trust value depends primary on agent preferences, capabilities and plans. The evaluating is also depend on the history of interaction and last experiences. The result of the overall evaluating, is used to update the agent trust value of his counterpart. Agent increase trust, if he evaluate inter-

ISBN 978-3-901608-32-2

action as "satisfying" or in "not satisfying" case, agent decrease the trust value.

In the reputation case, agent get directly trust value from another agent. This value can accept and appropriate as own trust value into recommendation target or just update own trust value recounted from recommended value. This recounting recommended value also depends on a few aspects, include aspect, how trustworthy is an recommender agent. In the reputation value based on recommendation, is necessary take into account that agent recommender purposely, toward to his strategy and purpose, provide incorrect recommendation value, which is not corresponding to his real trust value into target of recommendation. How the agent uses the reputation and its own trust to make decision with which agent to interact is an open question.

If the decision of interaction (cooperate/defect) is based on other agent recommendation, the agent will also update its trust in each agent that give recommendations.

3.4 Agent authentication

The agent authentication (unique identification) into system is very important for storing related trust value toward agent identity. Model have to allow agents to leave the system a join back with the same identity (same as when he leave the system). This requirement also allow to agents leave the system a join back with changed identity. This problem was discusses *Zacharia a Maes* [20] in our work. The result from this work is, we need to ensure, that agent joining into system with changed identity can not cause some damage, which can he use on his real identity.

Authentication can be centralised or decentralised. Advantage in centralised authentication is easy to implement and simple to manage, but from security point of view, is centralised authentication in our system unaccepted. If the central authentication point failure, whole system will be functionless.

Possibility approach for our model can be in using asymmetric cryptography, namely *public key encryption* and the *certificate infrastructure* – all together its typically called as PKI (Public Key Infrastructure) [19] standard. Principles to using public key for distributed authentication is described in a few studies [17, 16, 4, 12]. With using PKI standard, it will be possible to achieve not only authorisation mechanism but also mechanism to ensure secure communication between agents – confidentiality.

Disadvantage of this approach is relatively computationally costly and all agents have to need computational resources. Optionals this, advantages and disadvantages of using PKI standard for our model will be discussed in our future work.

4 Conclusion and future work

In this paper we present basic approaches to solve multi-agent distributed model based on trust and reputation principles. All interaction of agents is based on trust and reputation. While studying this problem, we found some open questions. How to effectively transform the reputation value to trust value?, How to evaluate all aspects of one service into one overall value of trust? Is possible to effectively evaluate group reputation with context and multi-faceted specifics? *etc*.

At this time, our model makes explicit the difference between trust and reputation. We defines reputation as a quantity inferred from recommendation, which can bee highly relatively toward to evaluating agent (recommender) mental state and interaction history. Trust we define as a quantity between two agents – the trustor and the trustee – which can be inferred from trustor interaction with trustee or inferred from reputation data about the trustee.

In our future work, we will concentrate on decision discussed problems and fully formally describe our model. Next phase will be implementing some of our proposals and simulate it, to achieve some practical results. Simulations will also show, if proposal is realisable and applicable in real information system. At the last, we need to bring some new approaches to solve complex security mechanism for this model.

The formalisation of our model can be used as base notation for describing new system constructed on trust and reputation principles in future. On grounds of our studies, some implemented systems such as access control mechanism or protocols can be updated toward our results.

5 Reference

[1] A. Abdul-Rahman and S. Hailes. Using recommendations for managing trust in distributed systems. In *IEEE Malaysia International Conference on Communication '97 (MICC'97)*, 1997.

[2] M. E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.

[3] D. Cvrcek. Alternative security for wifi networks. *Brno University of Technology*, 2006.

[4] A. Das and D. S. Wallach. Distributed authentication using web of trust. In *South Central Information Security Symposium (SCISS '04)*, 2004.

[5] M. P. Georgeff F. F. Ingrand and A. S. Rao. An architecture for real-time reasoning and system control. 7(6):34–44, 1992.

[6] D. Houser and J. Wooders. Reputation in internet auctions: Theory and evidence from ebay. 2000.

[7] M. Kinateder and K. Rothermel. Architecture and algorithms for a distributed reputation system. In *Proceedings of the First International Conference on Trust Management*, volume 2692, pages 1–16, 2003.

[8] Y. H. Lam, Z. Zhang, and K. L. Ong. A reputation-based trust model for agent societies. In *8th Pacific Rim International Conference on Artificial Intelligence, Auckland, New Zealand*, volume 3157, 2004.

[9] L. Mui and A. Halberstadt. Group and reputation modeling in multi-agent systems. In *NASA Goddard/JPL Workshop on Radical Agent Concepts*, 2001.

[10] L. Mui, M. Mohtashemi, and A. Halberstadt. A computation model of trust and reputation. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)*, volume 7, page 188, 2002.

[11] L. Mui, M. Mohtashemi, and A. Halberstadt. *Evaluating Reputation in Multi-agents Systems*, volume 2631/2003. Springer Berlin / Heidelberg, 2002.

[12] V. Pathak and L. Iftode. Byzantine fault tolerant public key authentication in peer-to-peer systems. *Computer Networks, Special issue on Management in Peer-to-Peer Systems: Trust, Reputation and Security*, 50/4, 2006.

[13] S. D. Ramchurn, D. Hunyh, and N. R. Jennings. Trust in multi-agent systems. *Knowledge Engineering Review*, 19:1:1–25, 2004.

[14] P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. In *NBER Workshop on Emperical Studies of Electronic Commerce*, 2000.

[15] S. Sen and N. Sajja. Robustness of reputation-based trust: boolean case. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 288–293, 2002.

[16] M. Sirbu and J. Chuang. Distributed authentication in kerberos using public key cryptography. In *Symposium on Network and Distributed System Security*, 1997.

[17] Joseph J. Tardo and Kannan Alagappan. Spx: Global authentication using public key certificates. 00:232, 1991.

[18] Y. Wang and J. Vassileva. Trust and reputation model in peer-to-peer networks. In *P2P '03: Proceedings of the 3rd International Conference on Peer-to-Peer Computing*, page 150, 2003.

[19] Wikipedia. Public key infrastructure — wikipedia, the free encyclopedia, 2007. http://en.wikipedia.org/wiki/Public_key_infrastructure, [Online; accessed 17-May-2007].

[20] G. Zacharia and P. Maes. Trust management through reputation mechanisms. *Applied Artificial Intelligence*, pages 881–907, 2000.

ISBN 978-3-901608-32-2