

A Survey of Trust Models in Agent Applications

Barbara Pusey

College of Computer Science and Engineering

The Pennsylvania State University

pusey@cse.psu.edu

Dr. Carleen Maitland

Dr. Andrea Tapia

Dr. John Yen

College of Information Sciences and Technology

The Pennsylvania State University

cmaitland@ist.psu.edu, atapia@ist.psu.edu, jyen@its.psu.edu

Abstract

Trust is an important requirement for sustained interactions between agents. It is relevant to the research of many fields, such as Psychology, Sociology, Economics and a multitude of fields within Computer Science. This manuscript is intended to provide a comprehensive overview of trust models for those who wish to implement trust in an intelligent agent system, or for those who would like to familiarize themselves with the field. We will attempt to cover trust and its issues from the view points of the aforementioned fields, with regard to implementing trust in agent-based organization modeling. Due to the broad range of issues related to trust, we focus here on issues related to risk analysis, centralized versus decentralized management, situational context of the decision, boundaries of a trust value, the transient nature of trust and the concept of modeling reciprocity. We will then discuss the potential implications of these issues in designing agent-based models of organizations and their information sharing policy.

Contact:

Barbara Pusey

Department of Computer Science and Engineering

The Pennsylvania State University

University Park, PA 16802

Tel: (256) 585-0144

Email: pusey@cse.psu.edu

Key Words : Trust, Intelligent Agent, Trust Models

Acknowledgements : NSF, get grant # from dr Mailand

Support : National Science Foundation

A Survey of Trust Models in Agent Applications

Barbara Pusey, Dr. Carleen Maitland, Dr. Andrea Tapia, Dr. John Yen

1. Introduction

In this paper we will consider trust from the aspect of a multi-agent environment; these agents could represent corporations, organizations, or individuals. For the purpose of this paper the agents we are trying to model represent organizations. Ideally, all agents are competent, non-malevolent, and have no qualms with sharing potentially profitable information. This scenario is unrealistic when modeling real world environments which are potentially hostile and may include incompetent, malevolent, or rival agents. We will then assume these agents or organizations reside in uncertain and more risky environments.

With uncertainty in information exchanged between agents comes the need to model inter agent trust, as securing all information is not a viable option. Therefore, what is too costly to secure, or impossible to secure, must be trusted [2]. However modeling trust, a human trait, in intelligent agents can be a difficult undertaking. Modeling the trust one organization can have for another, and the evolution of that trust as events occur, can be even more challenging. One must sort through the many and diverse models that exist to find the one that is most suitable to the agent architecture and design, and the context in which these agents interact. Though difficult it can be necessary as it is vital in some environments that agents be able to identify other agents that can give them relevant advice [19, 10], it is also vital that they identify parties with whom they can share profitable or personal information without risk. The purpose of this manuscript is to give a comprehensive overview of the trust models that have been developed with an outline of how trust might be modeled in organizational models.

2. Literature Review

Luhmann defines trust as “a solution for specific problems of risk”[20]. Trust, is then inherently linked to risk [7, 8]. Marsh and many others define it as a variable with a threshold for action [23, 4, 2, 10, 17, 6]. In the Marsh model [23] the action defined is inter agent cooperation and in the Castelfranchi model [4] the action becomes task delegation such as passing a soccer ball in a game of RoboCup [27]. This implies the innate Boolean nature of trust, we either trust the agent or we do not, one organization either trusts the other or it does not. Jøsang and Pope go further to give two different types of trust, Reliability Trust and Decision Trust. It is this second definition of trust which applies to the scope of this paper. Using Mcknight and Chervany[24] as inspiration Jøsang *et al.* [15, 16] define decision trust as:

Definition (Trust) *Trust is the extent to which one party is willing to depend on somebody, or something, in a given situation with a feeling of relative security, even though negative consequences are possible.*

This definition may be vague but it includes all the relevant components of a trust interaction. It includes the situation or context, implies risk is involved and defines the two main actors of any trust based information exchange. These two actors are the trustor, or the agent who releases the information, and the trustee, the agent being trusted [2].

It is interesting to note the duality of this relationship, as the trustee, the agent receiving information also trusts the other agent’s information to be accurate. And in this manor the trustor is also a trustee in that they are being trusted to provide accurate information. Thus with information exchange we see two categories of trust, to trust another agent with valued information and to trust information provided by an agent. The latter essentially reduces to a competency evaluation. The former is a far more complicated and thus more interesting process to observe. So it is in this arena which we will focus our attention, on the inherent problems, the categories in which the models can be classified, and suggested solutions to the complicated issues of modeling inter agent trust.

2.1 Evaluating Risk

Some trust models, such as TidalTrust [11], choose not to add a risk evaluation scheme to the model. In fact, few models choose to explicitly add Risk to their trust model [13], as the notion is that risk is inherently represented in the cost [3]. However, Cahill *et al.* incorporate a Risk Evaluator into their trust model SECURE [3]. The risk evaluator determines the possible cost against the benefits the agent might obtain, and then after risk is determined trust is added into the equation. Manchala [22], approaches risk modeling in a slightly different way. He uses a fuzzy logic interface rule with a matrix of risk against trust

to determine whether the agent should take action. Jøsang and Presti [16] also see risk and trust as two separable tools for making decisions in potentially volatile environments. They extend the Manchala model further taking into account what is being invested or trusted. These models all assume that risk can be evaluated on a numerical level, such as stock market or other financial environments. There are cases however where risk cannot practically be calculated as there may be many factors in the transaction that need to be taken into account [1, 14], as in gambling where we might need to consider thrill a factor [16].

2.2 Centralized or Decentralized Management

There are two approaches to maintaining reputation records, either each agent could maintain the record themselves or have a central agent, such as the internet based agent ITIS [21], which keeps track of the activity of all the agents in the environment. Auction websites such as eBay[9], use a central agent which keeps track of the system's agent's reputations through a rating system, or reputation mechanism, where one agent rates the other after a transaction. Other approaches such as Yenta[31] or Khare and Rifkin [18] have the agents rate themselves and then have other verified agents on the system validate their rating. Such a centralized system takes into account the effect on the memory of a system [23]. Instead of all agents keeping records of their interactions, only one record for each agent exists. This makes the centralized model more applicable to environments with large numbers of agents.

For environments with fewer agents, where memory is not a consideration, it may be preferable to choose a decentralized model. Examples of such models are the SECURE project [3] and ReGreT [29, 20]. This negates the need of all agents to agree to trust a third party, and also allows agents to contain their own unique trust model to best suit their profile.

2.3 Initial Trust and limits

Many questions arise for what to do when one agent interacts with another for the first time. A new agent in an environment creates a dilemma, most especially for systems which base trust solely on reputation. It must assign an arbitrary level of trust, generally a median value, to this new agent [9, 32], so that other agents will interact with the new agent. It could not start off without any trust for there would be no way for a new agent on the system to ever gain trust. Another suggested method is "trust learning by observation", reducing the task of learning to trust to a statistical consideration [10], this can be achieved using methods such as Bayesian learning [28]. In such a model an agent observes a new agent of the environment to determine their competence and whether they have malevolent intentions or not.

We must also consider putting a limit on the level of trust one user can acquire. Zacharia [32] states that 'the reputation of users should not be allowed to increase at infinitum as in ebay'. Consider ebay, in these cases an agent could enact in many successful transactions and abruptly become malevolent. It would require a great deal of negative feedback before they could outweigh the large collection of positive input.

We should also consider the effect of lowering an agent's reputation. Models such as Sporas [32] do not allow the reputation rating of a user to fall below the reputation of a new user. If allowed to fall below their initial rating, we would find users simply creating a new identity or agent, a more profitable course of action than attempting to resurrect their original reputation.

2.4 Modeling with a Situational Context

Another aspect of a trust based situation that we may want to consider is context. Should the context of the situation influence the level of trust we are willing to accept before action? That is, is there such a situation in which one agent might divulge otherwise non-trusted information to another agent because the situation calls for more extreme measures? This inclusion of situational context on trust makes the threshold model of trust insufficient. Now we see varying levels of trust, from complete trust in which an agent would be willing to divulge any information given the right situation, to complete distrust in which an agent would never divulge some given information no matter the context. Marsh [23] includes situational trust into his model. In this model both a trust value and situational trust value are calculated, if the situational trust is greater than just the trust, then the agents will continue with their interaction regardless.

2.5 Reciprocity

In choosing to more realistically model human agents, reciprocity, or the mutual exchange of deeds like favor or revenge [25], should be considered in the trust model. This proves even more difficult than trust to model as it is a less predictable human trait. Marsh [23] states that reciprocity can be an

extension of the decision making process. For example, if Agent 1 does a favor for Agent 2, then Agent 2 would be expected to reciprocate in some point in the future. This could also overlap with situational context, in this example the situation is a favored owed. However, for marketplace agents this would be ill-advised as it opens the door for one agent to deceitfully gain trust over another.

2.6 Trust Transitivity

Some trust models consider a case in which we could choose to trust someone who we trust trusts. Meaning, if Agent 1 trusts Agent 2, and Agent 2 trust Agent 3, we might infer that Agent 1 to some degree trusts Agent 3. This is an inductive model of trust, i.e. the friend of my friend is my friend, and its inverse. Zacharia *et al.* builds upon this concept describing a model Histos [32] which finds a path of connected PGP [12] signed web pages between agents using a search algorithm very similar to a breadth first search. Marsh contradicts this logic, noting that over long trains of trust this chain logic will result in conflicts with regards to distrust [23, 26], the paths of trust may still have application in short logic trains, and Christianson and Robinson [5] prove that this transitivity does not hold in real world contexts. It also would not hold in a multi-agent environment where each agent has its own unique trust mechanism. With independent trust models, an agent could not be sure that its 'friend' would come to the same conclusions regarding trust or reputation.

3. Discussion

In this part we will examine the issues discussed from the view of agent-based organizational modeling. We ask the important question: Does any one model, or combination of models, adequately cover the trust realm? The answer quite simplistically is that any implementation could model trust but not to the degree we would like to simulate in its nuances of human nature. An organization is run by humans and therefore has human traits such as reciprocity, and situational contexts, both of which should weigh in more heavily than the models reviewed in this paper.

There are many intangibles which the risk models lack, such as good will or malevolence. This again leads back to reciprocity, or the granting of a present benefit in anticipation of a future return, whether it is profit or simple good will that enables a organization to proceed. For example if company A lends money to a politician in order to gain access to future land rights, the company would consider this a benefit, i.e. the reciprocation of investments, yet it is difficult to place a simple monetary gain on it. This is a simple idea of enabling gain instead of solely creating monetary gain.

Another issue, which may never be resolved, is the removal of situational judgment in the malevolent case. A computer agent lacks the necessary skills to interpret actions in a situational context. For example the con man; a human may recognize the con through prior observational experience in regards to the con man's body language, something which a computer has yet to be able to do with any reliance. However a computer has an ability to maintain an absolute threshold, which implies that there could always be information that it would never be persuaded to release.

Closely related to situational judgment is the concept of situational context. In the example of Non-Government Organizations we see little inter-organizational information sharing, however the organization may become more willing to share vital information in the context of a disaster or threat. The only model discussed would release all information if the situational trust became higher then the threshold value. The flaw being that no information was held back if the situation was judged to be dire enough, that is, surpassed a threshold value. In the context of real organizations there would always be information which could not or should never be shared.

4. Summary

Trust is a very human trait, which can often be irrational. A number of efforts have been made to implement a logical basis for trust, having a variety of value bases. If one were to attempt to build an agent-based model of organizations, the concepts of reciprocity and situation context would have to be expanded on and developed into more comprehensive, human emulating algorithms. A good starting point in this effort would be the works reviewed here in this manuscript. It is important that a proper modeling of organizations take into account the sometimes predatory, or incompetent nature of interactions in the real world, else the model is at best incomplete, and could lead to false reactions within a potentially hostile simulation.

References

- [1] Bachman R (2001), 'Trust, Power and Control in Trans-Organizational Relations' Organization Studies, vol 2 pages 341-369
- [2] Braynov S, (2005) 'Trust Learning Based on Past Experience' KIMAS pages 197-201
- [3] Cahill V., et al., (2003), 'Using Trust for Secure Collaboration in Uncertain Environments', IEEE Pervasive Computing, vol. 2(3), pages 52-61
- [4] Castelfranchi
- [5] Christianson B, and Harbinson W.S., (1996), 'Why Isn't Trust Transitive?', In Proceedings of the Security Protocols International Workshop. University of Cambridge
- [6] Coleman J., (1990), 'Foundations of Social Theory', Harvard University Press
- [7] Deutsch, M (1958), 'Trust and Suspicion', The Journal of Conflict Resolution 2, 265-79
- [8] Deutsch, M (1962), 'Cooperation and trust: some theoretical notes', Nebraska symposium on motivation, 275-319
- [9] eBay: <http://www.ebay.com>
- [10] Esfandiari B, and Chandrasekharan S (2001), 'On How Agents Make Friends: Mechanisms for Trust Acquisition', In Proceedings of the Fifth International Conference on Autonomous Agents Workshop on Deception, Fraud and Trust in Agent Societies, pages 27-34
- [11] Golbeck J., (2005), 'Computing and Applying Trust in Web-Based Social Networks', PhD Thesis, University of Maryland, College Park
- [12] Gardfinkel S., (1994), 'PGP: Pretty Good Privacy', O'Reilly and Associates
- [13] Grandison T, and Sloman M, (2000) 'A Survey of Trust in Internet Applications' IEEE Communications Surveys Fourth Quarter
- [14] Grandison T, (2003) 'Trust Management for Internet Application' PhD Thesis, University of London
- [15] Jøsang A, and Pope S, (2005) 'Semantic Constraints for Trust Transitivity', 2nd Asia-Pacific Conference on Conceptual Modelling (APCCM2005), New Castle, Australia
- [16] Jøsang A, and Presti S, (2004), 'Analyzing the Relationship between Risk and Trust', 2nd International Conference on Trust Management
- [17] Kaplan S., Garrik B., (1981), 'On the Quantitative Definition of Risk', Risk Analysis, vol. 28, pages 11-27
- [18] Khare R and Rifkin A, (1997), 'Weaving a Web of Trust', Summer Luhmann97 Issue of the World Wide Web Journal, vol. 2(3) pages 77-112
- [19] Lashkari Y., Metral M., and Maes P., (1994), 'Collaborative Interface Agents' In Proceeding of the 12th National Conference of Artificial Intelligence, AAAI-Press
- [20] Luhmann, Niklas (2000), 'Familiarity, Confidence, Trust: Problems and Alternatives', in Gambetta Diego (ed.) Trust: Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford, chapter 6 pp 94-107
- [21] Ma M., and Meinel C., (2002), 'A Proposal For Trust Model: Independent Trust Intermediary Service (IT IS)', Trier Germany
- [22] Manchala D.W, (1998) 'Trust Metrics, Models and Protocols for Electronic Commerce Transactions' In Proceedings of the 18th International Conference on Distributed Computing Systems, pages 312-321, IEEE Computer Society
- [23] Marsh S, (1994) 'Formalizing Trust as a Computational Concept', PhD Thesis, University of Sterling

- [24] McKnight d.H., and Chervaney N.L., (1996), '*The Meanings of Trust*', Technical Report MISRC Working Paper Series 96-04, University of Minnesota, Management Information Systems Research Center
- [25] Mui, L., Mohtashemi, M., Halberstadt, A., (2002) '*A computational model of trust and reputation for e-businesses*', In: Proc. of the 35th Annual HICSS, Vol. 7, Washington, DC, USA, IEEE Computer Society
- [26] Ries S., Kangasharju J., and Muhlhauser M., (2006), '*A Classification of Trust Systems*', OTM Workshops, pages 894-903
- [27] RoboCup : <http://www.robocup.org/>
- [28] Russel S., and Norvig P., (1995), '*Artificial Intelligence: A Moddern Approach*', Prentice Hall, New Jersey
- [29] Sabter J., Sierra C., (2005), '*Review on Computational Trust and Reputation Models*' Artificial Intelligence Review' vol. 24(1), pages 33-60
- [30] Sabter J., Sierra C., (2002), '*Reputation and Social Network Anaalysis on Multi-Agent systems*', In Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems, New York, NY, ACM Press, pages 475-482
- [31] Foner L, (1997) '*Yenta: A Multi-Agent, Referral Based Matchmaking System*', First International Conference on Autonomous Agents, Marina del Rey, California
- [32] Zacharia G, and Maes P (2000), '*Trust Management through Reputation Mechanisims*', Applied Artificial intelligence, 14(9) pages 881-908