Comprehensive Trust Management

(Extended Abstract)

Sandip Sen
University of Tulsa
Department of Computer Science
Tulsa, Oklahoma, USA
sandip@utulsa.edu

Kuheli Chakraborty
University of Tulsa
Department of Computer Science
Tulsa, Oklahoma, USA
kuhi85@gmail.com

ABSTRACT

Trust has been viewed as an integral component of agent decision making in the context of multiagent systems (MASs). Various formal and semi-formal trust schemes, motivated by diverse considerations and influenced by various fields of study, have been proposed, implemented, and evaluated. We believe that there still exists a pressing need for developing a comprehensive trust management scheme that addresses most, if not all, issues underlying trust development, maintenance, and use. To facilitate the discussion of a general and comprehensive trust management scheme, we provide our own operational definition of trust motivated by uncertainty management and utility optimization. We identify the various components required of a comprehensive trust management scheme and their roles in determining agent performance in a competitive, open MAS.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

General Terms

Management, Performance, Reliability

Keywords

trust management, engagement, trust establishment, trust evaluation

1. INTRODUCTION

A core component of human reasoning in societal settings is the use of *trust*. Trust is truly a multi-dimensional and multi-faceted, even somewhat nebulous, concept and is used to capture a somewhat loosely related set of influences on human decision-making. While some researchers have attempted to formalize the role of trust in multiagent interactions, others have proposed trust models that allow agents to represent, update, and use their trust in other agents and services in their environment [1, 2, 5]. Though notable advances have been made, we believe it would be productive to discuss the aspects of trust that have received fair treatment from researchers and those that have been relatively unexplored. This reflective evaluation of the requirements of a trust model and the existence of matching trust models can identify the critical needs that can drive future trust research. Our goal in this research is then to analyze and recommend the necessary components of a comprehensive trust management scheme (CTMS) that can be used by researchers to both evaluate existing trust models and develop the next-generation trust management schemes that will be more robust and effective in handling a rich set of decision-making contexts. We also present a definition of computational trust which captures the fundamental need of an agent to effectively handle uncertainty and optimize performance.

Cite as: Comprehensive Trust Management (Extended Abstract), Sandip Sen and Kuheli Chakraborty, *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. 1573-1574

2. INTEGRATED TRUST BASED REASONING

To effectively and consistently maximize local utility, individual agents will need to coordinate, collaborate, and work with other agents in their environment. This often means that agents have to rely on other agents' decisions, e.g., that they fulfill negotiated commitments. Without any centralized authority or enforcement mechanisms in most of these open environments, commitments are non-binding. In addition, the likelihood of external offers and opportunities may provide short-term incentives for agents to deviate from commitments. Hence, agents in open environments need to rely on distributed reputation and trust mechanisms that encourage agents to fulfill their commitments. Distributed trust schemes produce and maintain agent reputations reflecting their performance and trustworthiness and can support and sustain mutually beneficial medium to long-term relationships between self-interested agents.

Various definitions of trust exist in literature focusing on either the philosophical or pragmatic aspect of the concept [1, 2]. We propose the following operational characterization that captures what it means for an agent to trust another agent (also see Figure 1):

> Trust in another agent reduces the uncertainty over that agent's independent actions which positively correlates with the truster's utility.

According to this interpretation of trust, trust in another agent can both reduce uncertainty about outcomes and improve performance. From a decision theoretic perspective, given a set of outcomes influenced by another agent, the actions of a trusted agent results in higher utility outcomes

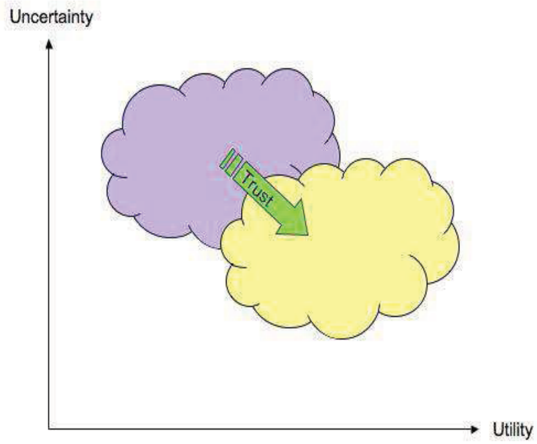

Figure 1: The set of outcomes for an agent changes from the dark region to the light region when interacting with a trusted agent, thereby reducing uncertainty and increasing utility.

becoming more probable (correspondingly lower utility outcomes becoming less likely) and hence results in higher expected utility. Risk neutral agents can then choose actions based on the Maximum Expected Utility (MEU) principle [4]:

$$\arg\max_{a \in A} \sum_{o \in O} Pr(o|a, M)U(o),$$

where A is the set of actions available to the agent, O is the set of outcomes possible, M is the world model of the agent and U is its utility function over outcomes. In the context of trust management, we consider the set of outcomes to be also dependent on other agents in the environment. We concentrate on bilateral interactions, and hence, outcomes are determined by the current agent and another agent and influences the trust between them. Assuming prior knowledge of the set of actions A, the set of possible outcomes O, and the utility function, a trust model in another agent will then estimate the outcome probabilities, $Pr(o|a, M)$. Either a frequency based approach can be used to estimate these probabilities or one can use Bayesian priors and associated update rules for model updating. Often these outcomes will depend on unobservable parameters and may involve time dynamics. Efficient approximate inference schemes may be used to estimate these probabilities in such cases.

An agent usually develops trust estimates of another agent both from direct interactions with that agent and from trust values reported by other agents (also called *reputation*). In particular, for various reasons often cited in favor of multiagent systems, including flexibility of use, low infrastructural overhead, robustness, etc. we are interested in reputation frameworks that are distributed and peer-level rather than centralized and monolithic.

The need for distributed trust schemes also arises in distributed systems susceptible to security threats. Malicious sources can compromise the nodes of a distributed system, e.g., a sensor network, to undermine the performance of the entire system. The problem is compounded when multiple nodes are compromised and collude to adversely affect the system. Distributed trust schemes can be used to screen and identify irregular activities in distributed systems exposed to intrusion threats and take responsive measures to limit damage to the system from malicious intruders [3].

Trust is also a resource that can be leveraged to gain influence [6]. When agent interactions are based on trust, trustworthy agents will have a larger influence on negotiated outcomes. For example, agents who are trusted to provide higher quality service may demand larger fees for their services. Trust often has to be earned at a cost. For example, a manufacturing agent may have to spend extra time and resources to meet stringent delivery deadlines when upstream suppliers delay delivery of raw materials. If, however, improved trustworthiness is rewarded with additional profitable contracts, the cost expended can be recouped many times over. In such scenarios, establishing a high reputation may be a priority for rational agents.

We believe that trust is a complex, multifaceted concept and involves more than merely evaluating others' trustworthiness. An integrated approach is necessary and should address *engagement* of other agents, creating situations to *evaluate* trust, investing resources and time to *establish* your own trustworthiness, *strategic use* of trust information, etc. Strategic reasoning involving trust considerations will trade-off the cost of establishing and maintaining trust in the community with the future expected profits from leveraging the trust earned. Though prior research have proposed and evaluated various trust and reputation approaches that evaluate the trustworthiness of other agents, little attention has been paid to a comprehensive trust management scheme. Our proposed CTMS scheme addresses trust modeling, exploration, learning, as well as both tactical and strategic reasoning to achieve the desired properties of reducing uncertainty and increasing utility. Additionally, we believe that any holistic, robust, and flexible agent architecture must include a trust module that is integrated with and directly influence agent decision making. Our continuing research is focused on elaborating the details of such an agent architecture and evaluating the effects of various trust sub-modules on agent performance in competitive, open multiagent scenarios.

3. REFERENCES

[1] C. Castelfranchi and R. Falcone. Principles of trust for MAS: Cognitive autonomy, social importance, and quantification. In *Proceedings of the Third International Conference on Multiagent Systems*, pages 72–79, Los Alamitos, CA, 1998. IEEE Computer Society.

[2] D. Gambetta. *Trust*. Basil Blackwell, Oxford, 1990.

[3] O. Mistry, A. Gursel, and S. Sen. Comparing trust mechanisms for monitoring aggregator nodes in sensor networks. In *Proc. Eighth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 985–992, 2009.

[4] J. V. Neumann and O. Morgenstern. *Theory of games and economic behavoir*. Princeton University Press, New Jersey, 1944.

[5] S. Sen. Believing others: Pros and cons. *Artificial Intelligence*, 142(2):179–203, 2002.

[6] S. Sen and D. Banerjee. Monopolising markets by exploiting trust. In *Proc. 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2006)*, pages 1249–1256. ACM, 2006.