An Architectural Approach to Combining Trust and Reputation

Christopher J. Hazard and Munindar P. Singh

North Carolina State University, Raleigh NC 27695-8206, USA

Abstract. Though trust and reputation systems have been extensively studied, general architectural commonalities between the two have received little attention. In this paper, we present a life cycle model of reputation and trust systems, along with accompanying measures of how much effect signaling and sanctioning have on a given system. We map reputation attacks within our framework and apply our framework to an online auction model.

1 Introduction

Throughout the trust and reputation system literature, two techniques that stem from game theory are commonly applied for designing such systems. Signaling models are those in which agents attempt to assess private attributes about other agents, whereas sanctioning models are those in which agents behave strategically in an attempt to maximize their utility [2].

In real-world environments where agents must decide whether or not to trust one another, clean distinctions between signaling and sanctioning are rare. For example, an agent that allocates its own bandwidth and other resources may have little influence over the amount of resources it has available. Yet, it may be strategic and rational within those constraints. A manufacturer can acquire a good reputation for having tight quality controls, but new management may wish to see larger profit margins and may strategically slowly cut back on the quality controls as long as it remains ahead of its competitors.

Despite the complexity of the real world, few reputation systems are specifically designed to address both sanctioning and signaling. Typically, authors of reputation systems that involve signaling devise a variety of malicious behaviors to test their system against. Examples of the adversary agents include randomized acts of unfavorable behavior [9, 6], building up and spending of reputation [18, 10, 14], Sybil attacks where an agent creates multiple identities [10, 9, 17], and collusion with other agents [9, 17, 18]. Other systems are designed specifically around strategic agents to ensure good behavior, but do not attempt to measure attributes of the agents [8, 5]. A minority of reputation systems, such as that by Smith and DesJardins [16], examine both signaling and sanctioning explicitly.

Our primary contribution is a model that connects trust and reputation systems both architecturally and functionally. We examine the trust and reputation life cycle in an abstract form from which we can systematically determine how much influence signaling and sanctioning have on the particular system. We present a heuristic that indicates

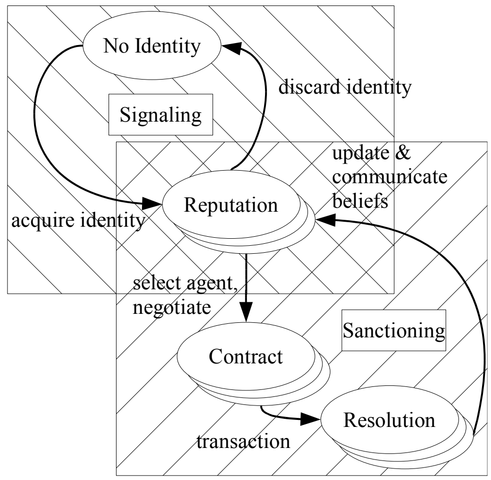

Fig. 1. Trust and reputation life cycle from an agent's perspective.

how a system is governed between the signaling and sanctioning, which is prescriptive in terms of what kind of a reputation or trust model should be used for a given situation. From this model, we identify and categorize different kinds of attacks against reputation systems. We use a running example of an online auction.

The general view of this paper is that trust is looking forward in time with respect to sanctioning and strategy, whereas reputation is looking backward in time with respect to signaling and determining agents' types. We discuss this dichotomy in detail. The focus of this work is on rational agents and e-commerce settings, rather than directly modeling human behavior. Emotional and cognitive factors of trust are outside of the scope of this paper.

The remainder of this paper is organized as follows. We first introduce the trust and reputation life cycle using a logic-based formalism and illustrate it via an online auction. We then discuss the signaling and sanctioning dichotomy, measuring the effect of each on a simple online auction model, and then discuss how attacks on reputation systems can occur at each of the points in our model. We conclude with a discussion of the benefits and limitations of our model.

2 Trust and Reputation Life Cycle

Although the specifics of particular trust and reputation systems can differ greatly, they all share some commonalities. In this section, we unify the systems to a common set of states and actions as outlined in Figure 1.

2.1 Identity States

The following are the different states an agent can go through in a transaction in the presence of an open reputation or trust system. An agent is not limited to being in one state at a time, but can maintain multiple accounts and participate in multiple transactions simultaneously.

No Identity The agent begins without an identity or account in the system. This state is applicable for open systems where agents may enter or leave. From this state, an agent may acquire an identity and move to the reputation state. Acquiring an identity may be as trivial as using a nonvalidated screen name in an open text field where the agent simply claims to have some identity. Alternatively, the system may require extensive background checks, verifications from official organizations, or significant payments to create the account. An agent may asynchronously acquire multiple identities, and may acquire identities in different contexts or with different populations of agents.

Reputation Each identity that the agent has created will have its own reputation within the community. An agent may discard an identity, either actively by deleting an account or passively by simply no longer using an identity. When an agent decides to (or is forced to) interact with another agent, it must select an agent (or agents) with which to interact. It may communicate with this *target* agent, performing extensive negotiations and setting up a formal contract. Alternatively, the agent may simply rely on norms or not actively communicate with the target prior to the transaction.

Contract A contract expresses a promise or commitment to engage in some behavior. Contracts may be well-defined and policed by an external system or may be as ill-defined as the agents' a priori assumptions. From a contract, the agents involved undergo some transaction with the other agents involved. The transaction can involve active participation, such as exchanging money for an item, or a transaction can be passive, such as all agents timing out and not performing any task.

Resolution After a transaction has taken place, an agent will update its own beliefs about the agents involved in the interaction. The agent can evaluate, report, and communicate its new beliefs about another agent based on the results of the transaction, either directly to other agents or via a centralized reputation reporting mechanism. Concurrently, the agent may revisit the results and decide that further transactions are required. To set up future transactions, the agents may renegotiate to a new contract after having observed the other agents. A renegotiation can have positive connotations, such as providing additional services to supplement a previous transaction, or the renegotiation can have negative connotations, such as an agent demanding reparations from a transaction that did not fulfill the contract.

2.2 Agent Actions

To formalize our discussions about the life cycle of a reputation for further discussion and analysis, we use a logic-based framework. We formally describe abstractions of general interactions of a reputation system where comparisons between values are required to express agents' preferences. For example, we can represent an example of agent Alice's utility as

$$Util(Alice, 5) \wedge Util(Alice, 10) \rightarrow Util(Alice, 15) \tag{1}$$

given the quasilinearity utility rule

$$Util(agent, value1) \wedge Util(agent, value2) \rightarrow Util(agent, value1 + value2). \tag{2}$$

Similar such rules can be used to describe the values within reputation systems. We denote ground terms as those identifiers beginning with an upper case letter, and variables as lower case identifiers.

We can write each of the state transitions from Section 2.1 more formally as follows:

discard identity : $ID(agent, id) \wedge Util(agent, discardCost)$
acquire identity : $\neg ID(agent, id) \wedge Util(agent, acquireCost)$
select agent, negotiate : $Terms(agent, otheragent, contract)$
transaction : $Transaction(agent, otheragent, Terms(agent, otheragent, contract))$
update & communicate beliefs : $Transaction(agent, otheragent, Terms(agent, otheragent, contract)) \rightarrow Observation(agent, otheragent, terms)$

2.3 Example: Online Auction Representation

To illustrate the applicability of our life cycle and formalizations, we create an example Beta model reputation model with only positive and negative ratings resembling an online auction as follows.

Because it costs only a small amount of time for an agent to set up an account, acquiring an identity becomes

$$ID(agent, id) \wedge Util(agent, acquireCost), \tag{3}$$

which could be, for example, $ID(agent, id) \wedge Util(agent, -\$0.5)$. Similarly, discarding an identity is easy, represented as

$$\neg ID(agent, id) \wedge Util(agent, discardCost). \tag{4}$$

The transaction itself becomes

$$\begin{aligned} &Transaction(agent, otheragent, Terms(agent, otheragent, contract)) \rightarrow \\ &\quad \Big(Terms(buyer, seller, contract) \leftrightarrow \big(Pay(buyer, seller, value) \\ &\quad \wedge Give(seller, buyer, good)\big) \vee (CurrentDate > SellDate + 30)\big)\Big), \end{aligned} \tag{5}$$

with

$$Pay(buyer, seller, value) \rightarrow Util(buyer, -value) \wedge Util(seller, value), \tag{6}$$

$$Give(a, b, good) \rightarrow \neg Has(a, good) \wedge Has(b, good), \quad \text{and} \tag{7}$$

$$Has(agent, good) \rightarrow Util(agent, Value(agent, good)). \tag{8}$$

We can represent the ratings mechanism, triggered by the observation of terms as

$$Terms(agent, otheragent, contract) \rightarrow Observation(agent, otheragent, contract). \tag{9}$$

The buyer can be rated positively using $AdditionalPositiveRating$ and $AdditionalRatings$ to each increment the respective value via

$$\begin{aligned} &Pay(buyer, seller, value) \wedge CurrentDate \leq SellDate + 30 \\ &\rightarrow AdditionalPositiveRating(buyer, 1) \wedge AdditionalRatings(buyer, 1) \end{aligned} \tag{10}$$

and negatively as

$$\neg Pay(buyer, seller, value) \wedge CurrentDate > SellDate + 30 \\ \rightarrow AdditionalRatings(buyer, 1). \quad (11)$$

The seller may be rated similarly as

$$Give(seller, buyer, good) \wedge CurrentDate \leq SellDate + 30 \\ \rightarrow AdditionalPositiveRating(seller, 1) \wedge AdditionalRating(seller, 1) \quad (12)$$

and

$$\neg Give(seller, buyer, good) \wedge CurrentDate > SellDate + 30 \\ \rightarrow AdditionalRating(seller, 1). \quad (13)$$

A simple buyer agent might just choose the highest rated seller as

$$\exists s(s \in Sellers) \wedge \forall t(t \in Sellers) \\ PositiveRating(s)/NumRatings(s) \\ \geq PositiveRating(t)/NumRatings(t) \\ \rightarrow Terms(buyer, s, contract). \quad (14)$$

However, this simplified rating system does not take into account strategy, which we discuss next.

3 Signaling Versus Sanctioning

The game-theoretic designations of signaling and sanctioning games are relevant to trust and reputation systems because they address the key mechanism of whether an agent must decide who to choose or how to act [2, 8]. In this section, we propose a way of determining the influence of signaling versus sanctioning and how these properties affect the design of a trust or reputation system, eventually connecting it back to our life cycle model.

In a signaling setting, agents have private information that they may use to their advantage. The asymmetric information can be used strategically to cause adverse selection, where agents perform transactions with agents they believe to be desirable but end up with an undesirable interaction. An example of a signalling situation is where agents are purchasing mass-produced products and deciding whether to buy the product from one manufacturer or another based on quality, price, and features. In this case, agents signal to each other what they believe about other agents (specifically, the manufacturers). Statistical and probabilistic measures are most effective at measuring agents' behaviors in the signaling setting.

Sanctioning mechanisms are useful in cases of moral hazard. Moral hazard occurs when agents' utilities are uncorrelated, meaning that one agent's gain may yield another's loss, and one agent can directly exercise control over another's utility. A purchase where a buyer pays the seller and then the seller has the option of not sending the product to the buyer is an example case of moral hazard. If the seller will not be sanctioned for its behavior and will have no future relations with the buyer, then it has no incentive to send the product. Sanctioning must be credible for the agents involved to be successful, and may be performed by the agent affected by refusing future transactions, or by other agents policing the system. Modeling behavior in a sanctioning environment with rational environments means employing game theory techniques to find Nash equilibria.

As we remarked above, many real-world situations do not fall cleanly into either signaling or sanctioning situations. An agent may have some control over the quality of its products, but it is rarely impossible for an agent to make any changes to quality (pure adverse selection) or for an agent to have perfect control over quality (pure moral hazard). This distinction is blurred further by agents having differing levels of patience that influence their strategic behavior [5, 16] and also by the blurred distinction of whether an observation was intentionally communicated [**?**]. The amount of sanctioning comes down to how much explicit control an agent has over its communications, and also intent, which may be subtle.

In broad terms, we can distinguish two varieties of trust that apply in many computational settings with intelligent agents. We abstract the terms *Competence* and *Integrity*, as described by Smith and DesJardins [16], into

Capabilities, which are what an agent *can* do, and
Preferences, which are what an agent *will* do.

From these definitions, it is clear to see that when agents want to determine which other agents have capabilities, they need a signaling system which looks into what the agents have done before. When agents want to determine another agent's preferences and ensure that the agent will perform a desirable behavior in the future when it has the choice, then they need a sanctioning system. This is consistent with the notions of **reactive** and **anticipatory** coordination [**?**].

To examine the role of signaling versus sanctioning on reputation systems, it is instructive to consider three interrelated terms—trust, trustworthiness, and reputation—that are used in nonstandardized ways in the literature. We begin from basic definitions in order to capture the general intuitions about them.

Trust is an agent's assessment of another party along some dimension of goodness leading to expected outcomes.
Trustworthiness is how good a party is in objective terms. In other words, this is a measure of how worthy it is to be trusted.
Reputation is the general belief (among the agents in a society or community) about an agent.

Specifically, Alice may or may not trust Bob for possessing desirable attributes (these could be capabilities, resources, bandwidth, and such). Alternatively, Alice may or may not trust Bob for having his preferences aligned with hers or rather for having his preferences aligned with hers under a particular incentive mechanism. Bob may or may not be worthy of any trust Alice may place in him. Bob may or may not have a reputation for being trustworthy in the specified ways. And such a reputation may or may not be well earned.

Reputation and trust therefore can be fit into our dual categorization. Reputation involves what an agent is, as measured from its past; an agent has a reputation of having some attribute or capability, and so a reputation system in this sense is a signaling system. Trust is concerned with what an agent will do in a future situation, which concerns the agent's preferences and must be handled by a sanctioning system. However, as trust and reputation have other connotations in specific domains, such as emotion, we will maintain the distinction using the terms signaling and sanctioning.

3.1 Measuring Influence of Signaling and Sanctioning

Consider agents A and B that have strongly typed behavior, meaning that they will always behave almost the same way regardless of the situation (e.g., by offering products of some specific quality). An example of such an agent is one that controls a high-volume web service with specific offerings and finite bandwidth with little autonomy and business logic. Consider an agent C that is deciding which agent to interact with between A and B. If C chooses A, then C will receive some benefit (or loss) of utility, b_A. If C chooses B, then C would receive a change in utility of b_B. Since the agents are strongly typed, C's behavior other than choosing A or B will not make much difference. To maximize utility, C should use statistics to measure A and B's attributes.

Conversely, consider that agents A and B are rational, have full and precise control over each of their actions, and may change their behavior without any switching costs. An example of these agents would be low-volume reseller agents that have sufficient supply of substitutable products. In this case, whether C chooses A and B matters little to C's utility. Instead, C's choices in negotiation and behavior with respect to A or B dominates C's change in utility. Finding an optimal interaction strategy is how C can maximize its utility.

If we write the benefit C will gain with behavior x when choosing agent A as $b_{A,x}$, then magnitude difference of utility change between these choosing A and B while C maintains consistent behavior is $|b_{A,x} - b_{B,x}|$. Using C's ideal behavior, this can be written as $\max_x |b_{A,x} - b_{B,x}|$. When evaluated against all agents available for interaction, S, agent C's value of the utility difference between two agents, $d_{\text{selection}}(C)$, can be written in terms of the expected rate of interaction between C and another agent A as $r_{A,C}$, as

$$d_{\text{selection}}(C) = \frac{1}{\sum_{A \in S} r_{A,C} + r_{C,A}} \cdot \sum_{A \in S} \sum_{B \in S} \max_{x \in H} |r_{A,C} \cdot b_{A,x} - r_{B,C} \cdot b_{B,x}|. \quad (15)$$

The normalizing term $\frac{1}{\sum_{A \in S} r_{A,C} + r_{C,A}}$ represents the reciprocal of the total interaction rate. Similarly, we may write the expected value of the utility difference between any two behaviors, $d_{\text{strategy}}(C)$, of the set of all behaviors in H across all agents, as

$$d_{\text{strategy}}(C) = \frac{1}{\sum_{A \in S} r_{A,C}} \cdot \sum_{A \in S} \max_{x \in H, y \in H} |r_{A,C} \cdot b_{A,x} - r_{A,C} \cdot b_{A,y}|. \quad (16)$$

Seller Agent	Refurb. Value	Refurb. Market Price	Unrefurb. Price	Refurb. Cost
A	\$500	\$400	\$200	\$150
B	\$490	\$350	\$250	\$80

Table 1. Online auction refurbished laptop example data.

As $d_{\text{selection}}$ measures the impact of an agent's type and d_{strategy} measures the impact of an agent's strategy, we can use these values to determine the impact of signaling and sanctioning on a multiagent interaction mechanism. In aggregation, we express the expected value of each of the values across all agents as $E(d_{\text{selection}})$ and $E(d_{\text{strategy}})$ respectively. The fraction of agents' total utility in a system that is governed by signaling, $i_{\text{signaling}}$ can be represented as

$$i_{\text{signaling}} = \frac{E(d_{\text{selection}})}{E(d_{\text{selection}}) + E(d_{\text{strategy}})}. \tag{17}$$

The fraction of utility governed by sanctioning, $i_{\text{sanctioning}}$, can be represented as

$$i_{\text{sanctioning}} = \frac{E(d_{\text{strategy}})}{E(d_{\text{selection}}) + E(d_{\text{strategy}})}, \tag{18}$$

with $i_{\text{signaling}} + i_{\text{sanctioning}} = 1$.

3.2 Example: Online Auction Representation

We reuse our general interaction model from Section 2.3 to show an example of applying our signaling versus sanctioning measure. Suppose agents are participating in online market for refurbished laptops outlined in Table 1.

A buyer agent, C, values its own utility of the refurbished laptop from A at \$500 and the refurbished laptop from B at \$490. It needs to decide whether to buy from A or B for the market price of \$400 or \$350 respectively. It costs A \$150 to refurbish its laptop that it bought unrefurbished at \$200, and costs B \$80 to refurbish the laptop it purchased at \$250. Both A and B are claiming that the laptop on sale is refurbished, but C does not know for sure.

First, we investigate the case of selection. Agent C can select to buy from A or B, but A and B have no choice in the matter because of the online auction format. The rates of interaction from A's perspective are $r_{A,A} = 0$, $r_{A,B} = 0$, $r_{A,C} = 1$, and B is analogous. The rates from C's perspective are $r_{C,A} = 1$, $r_{C,B} = 1$, $r_{C,C} = 0$.

First we evaluate $d_{selection}(C)$. Agent A only can interact with C, and the maximum profit C could make while still providing a laptop is \$200. Therefore, $d_{selection}(A) = \frac{1}{2} \cdot (|(\$400 - \$200) - \$0| + |\$0 - \$400 - \$200)|) = \$200$. Similarly, $d_{selection}(B) = \$100$. To compute this value for agent C, we must first evaluate which strategy yields the greatest difference between choosing A or B. When the seller performs the refurbishment, C's difference in utility between choosing seller A and B is $|(\$500 - \$400) - (\$490 - \$350)| = \$40$. When the seller does not perform the refurbishment, the difference becomes $|(\$200 - \$400) - (\$250 - \$350)| = \$100$. As the rates of interaction are symmetric, the larger of these two yields $d_{selection}(C) = \$100$. The combined expected value of the difference of selection across all three agents is $E(d_{selection}) = (\$200 + \$100 + \$100) \approx \$133$.

Next we investigate the case of sanctioning. Beginning with A, we find $d_{\text{strategy}}(A) = 1 \cdot |(\$400 - \$200 - \$150) - (\$400 - \$200)| = \$150$, which is the cost of refurbishing the laptop, and accordingly $d_{\text{strategy}}(B) = \$80$. To find $d_{\text{strategy}}(C)$, we also examine the sellers' behavior. If A does not refurbish the before shipping laptop, but instead delivers a broken laptop, then C regains only \$200 from selling the laptop at the unrefurbished price and loses its \$400 payment. Applying this evaluation with both A and B, $d_{\text{strategy}}(C) = \frac{1}{2} \cdot \big(|(\$500 - \$400) - (\$200 - \$400)| + |(\$490 - \$350) - (\$250 - \$350)| \big) = \$270$. Putting the three of these agents' results together, we obtain
$E(d_{\text{strategy}}) = (\$150 + \$80 + \$270)\,/3 \approx \$167$.

The system has $i_{\text{signaling}} = \frac{\$133}{\$133+\$166} \approx .44$ and $i_{\text{sanctioning}} = \frac{\$166}{\$133+\$166} \approx .66$. An effective reputation system for this system should emphasize sanctioning mechanisms slightly over signaling mechanisms.

4 Attacks on Trust and Reputation Systems

Given our logic model to represent a trust or reputation system, we can use these formalisms to give systematic treatment of types of attack or exploits for each type of action.

discard identity: An agent may discard its identity to remove a bad reputation and potentially acquire a new one [4], which can be expressed as $\neg ID(agent, id1) \wedge Util(agent, discardCost) \wedge ID(agent, id2) \wedge Util(agent, acquireCost)$. At a higher level, an agent could discard an identity which is not lost in anonymity, but rather used to frame another agent as a threat, for blackmail, to remove a competitor, or for other forms of sanctioning.

acquire identity: Sybil attacks occur when one agent creates multiple identities in order to manipulate a reputation or other aggregation system [11], expressed as $ID(agent, id2) \wedge Util(agent, acquireCost)$. Such attacks may directly influence reputation or flood out other behavior.

select agent, negotiate: An agent may select another agent that is known to be in a weak position or that is easy to manipulate or exploit, offer terms in negotiation without intent to fulfill them or with intent to deceive or harm the other agent, or demand a commitment by threatening to harm the other agent if not fulfilled. In subsequent negotiations, an agent may mislead another agent that a previous transaction was problematic and that reparations are needed to continue the relationship. Agents can also be selected against for sanctioning purposes.

transaction: An agent may not fulfill a commitment at all, fulfill it only part way or of lesser quality than expected, or provide something unexpected [15].

update & communicate beliefs: This transition is particularly rich in terms of the numbers and types of attacks, and can range from coordination with other transitions (e.g., acquiring identities in a Sybil attack) to simply communicating [3]. An agent can lie about another agent's performance for positive or negative reciprocity.

Revisiting Figure 1, the different types of attacks can be categorized by whether or not they can be primarily addressed by signaling or sanctioning matter. Selecting agents and updating and communicating beliefs are two of the interactions in the life cycle that can be exploited in the most complex strategic manners, but also map directly to signalling systems.

5 Discussion

In this paper, we have examine an architectural view of the trust and reputation life cycle. The motivation of this paper was not to introduce any new mechanisms of trust and reputation, but rather to take a step toward formalizing the taxonomy and structure of trust and reputation. Our model aids in the discussion of systems' mechanisms, as it presents a broadly applicable view that can be used in conjunction with other taxonomies of reputation systems [12, 1, 13, 7] and taxonomies of attacks [10]. Further, we offer an indication as to the extent that a given interaction system is affected by both signalling and sanctioning, which is useful in determining what kind of a reputation or trust system should be deployed. Our work unifies reputation systems, trust systems, and related game theory under a common architectural framework.

Applying our measure of the effect of signaling versus sanctioning to a live system requires some judgement. For example, if the strategy which offers the worst utility for a given situation is something that no rational agent would do, then it is best left out. In general, only those actions that lie on the Pareto frontier of utility or on a Nash equilibrium (or approximate Nash equilibrium) should be considered. The actual attributes of the agents, including preferences, utilities, and capabilities may also be unknown to a system designer, so estimates may often need to be substituted.

A key theme in our model is the relative autonomy of the agents involved. If the agents behave in a fixed manner that is largely independent of the other agents' strategies, then are best measured by a signaling system. An agent's autonomy is further reflected by the level of bounded rationality and information available to an agent in the system.

Future work involves investigating time preferences of agents and combinations of actions. Though our model is high level, some strategies require involvement of several distinct pieces of the model. Further future work involves deepening the connection between our logic framework with the signaling and sanctioning measures to give more prescriptive results.

Acknowledgments

Research was partly sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

1. D. Artz and Y. Gil. A survey of trust in computer science and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5:58–71, 2007.
2. C. Dellarocas. Reputation mechanisms. In T. Hendershott, editor, *Economics and Information Systems*, volume 1 of *Handbooks in Information Systems*, chapter 13, pages 629–660. Elsevier Science, 2006.
3. J. Farrell and M. Rabin. Cheap talk. *The Journal of Economic Perspectives*, 10(3):103–118, 1996.
4. M. Feldman, C. Papadimitriou, J. Chuang, and I. Stoica. Free-riding and whitewashing in peer-to-peer systems. *IEEE Journal on Selected Areas in Communications*, 24(5):1010–1019, 2006.
5. C. J. Hazard. ¿Por favor? Favor reciprocation when agents have private discounting. In *AAAI Workshop on Coordination, Organizations, Institutions and Norms (COIN)*, pages 9–16, Chicago, Illinois, July 2008.
6. T. D. Huynh, N. R. Jennings, and N. R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Journal of Autonomous Agents and MultiAgent Systems*, 13(2):119–154, 2006.
7. A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, March 2007.
8. R. Jurca and B. Faltings. Obtaining reliable feedback for sanctioning reputation mechanisms. *Journal of Artificial Intelligence Research*, 29:391–419, August 2007.
9. S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th International Conference on World Wide Web*, pages 640–651, Budapest, Hungary, 2003.
10. R. Kerr and R. Cohen. Smart cheaters do prosper: Defeating trust and reputation systems. In *Proceedings of the eighth international conference on autonomous agents and multiagent systems*, pages 993–1000, 2009.
11. J. Newsome, E. Shi, D. Song, and A. Perrig. The sybil attack in sensor networks: analysis & defenses. In *Proceedings of the 3rd international symposium on Information processing in sensor networks*, pages 259–268, New York, NY, USA, 2004. ACM.
12. S. D. Ramchurn, D. Huynh, and N. R. Jennings. Trust in multi-agent systems. *The Knowledge Engineering Review*, 19:1–25, 2004.
13. J. Sabater and C. Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, September 2005.
14. A. Salehi-Abari and T. White. Towards con-resistant trust models for distributed agent systems. In *International Joint Conference on Artificial Intelligence*, pages 272–277, 2009.
15. M. P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7(1):97–113, March 1999.
16. M. J. Smith and M. desJardins. Learning to trust in the competence and commitment of agents. *Autonomous Agents and Multi-Agent Systems*, 18(1):36–82, February 2009.
17. J. D. Sonnek and J. B. Weissman. A quantitative comparison of reputation systems in the grid. In *Proceedings of the 6th IEEE/ACM International Workshop on Grid Computing*, pages 242–249, 2005.
18. M. Srivatsa, L. Xiong, and L. Liu. Trustguard: Countering vulnerabilities in reputation management for decentralized overlay networks. In *Proceedings of the 14th International Conference on World Wide Web*, pages 422–431, 2005.