Computational Intelligence

DISTRIBUTED REPUTATION MANAGEMENT FOR ELECTRONIC COMMERCE

Bin Yu and Munindar P. Singh*

Department of Computer Science, North Carolina State University

One of the major challenges for electronic commerce is how to establish a relationship of trust between different parties. Establishing trust is nontrivial, because the traditional physical or social means of trust cannot apply directly in virtual settings. In many cases, the parties involved may not ever have interacted before. Reputation systems seek to address the development of trust by recording the reputations of different parties. However, most existing reputation systems are restricted to individual market web-sites. Further, relevant information about a party may come from several web-sites and from interaction that were not mediated by any web-site.

This paper considers the problem of automatically collecting ratings about a given party from others. Our approach involves a distributed agent architecture and adapts the mathematical theory of evidence to represent and propagate the ratings that participants give to each other. When evaluating the trustworthiness of a given party, a peer combines its local evidence (based on direct prior interactions with the party) with the testimonies of others regarding the same party. This approach satisfies certain important properties of distributed reputation management and is experimentally evaluated through simulations.

Key words: distributed reputation management, trust networks, electronic commerce

1. INTRODUCTION

It is important for participants (buyers, sellers, partners) to estimate each other's trustworthiness before initiating any commercial transactions. Not only do buyers need to trust sellers, but also sellers need to trust buyers. Let us consider a simple example. Before a buyer decides to provide his credit card information to a prospective seller, he must trust the seller sufficiently with regard to the quality of products the seller will send, the expected delivery time, and the seller's customer service. Likewise, the seller must trust the prospective buyer enough to realize that the buyer is seriously considering the purchase, will not attempt to cancel his payment, and will not try to cheat the seller in any way. Usually, a trust judgment of this sort cannot be made from information available from the seller's web-site. For example, the buyer may have to consult other buyers who have past experience with the seller. In other words, to estimate the trustworthiness of a given party, it is reasonable to try to find what reputation it has within some salient group.

Reputation systems are mechanisms that support such estimations. Current reputation systems are geared toward capturing information on sellers' past behavior as ratings given by buyers, e.g., [20]. In eBay, sellers receive feedback (+1, 0, −1) for their reliability in each auction. Their reputation is calculated as the sum of the ratings they received over the preceding six months. In Bizrate.com, each customer is asked to complete a survey evaluating the retailers after each purchase.

Explicit reputation systems are helpful in fostering trust among strangers. However, most existing reputation systems are completely centralized. Further, they require users to explicitly make and reveal their ratings of others. This would not be acceptable to many users. Further, such ratings would often be made strategically and may not reflect the trustworthiness of the rated parties. For example, Resnick *et al.* found that ratings on eBay

*This is a revised and extended version of a paper accepted by the First International Joint Conference on Autonomous Agents and Multiagent Systems, Bologna, Italy, 2002. Address correspondence to the authors at the Department of Computer Science, North Carolina State University, Raleigh, NC 27695-7535; email: byu@unity.ncsu.edu and singh@ncsu.edu.

are almost always positive [19]. Also they found there is a high correlation between ratings by buyers and sellers, suggesting that eBay users reciprocate and retaliate.

However, while current centralized approaches to reputation management suffer from the above limitations, the idea of reputation management is important and may often be the only viable source of estimations of trustworthiness. For this reason, we consider distributed reputation management, which involves aggregating ratings for a given party from others. No central authorities are assumed.

Briefly, this paper develops an evidential model of reputation management based on the Dempster-Shafer theory of evidence. In this approach, software agents assisting each participant in obtaining, evaluating, and combining the ratings. The agents cooperate by giving referrals to each other. An agent's value to another agent depends greatly upon the quality of the referrals it offers.

The rest of this paper is organized as follows. Section 2 introduces our technical approach, giving the key definitions for local rating and propagation through referrals. Section 3 presents our experimental results. Section 4 summarizes some related work in reputation management. Section 5 describes some interesting conclusions and directions for future research.

2. DISTRIBUTED REPUTATION MANAGEMENT

The proposed approach builds on our work on referral networks [25]. An agent-based referral network is a multiagent system whose member agents give referrals to one another (and are able to follow referrals received from other agents). To do so effectively presupposes certain representation and reasoning capabilities on the part of each agent. Each agent has a set of *acquaintances*, a subset of which are identified as its *neighbors*. The neighbors are the agents that the given agent would contact and the agents that it would point (refer) others to. An agent maintains a model of each acquaintance. This model includes the agent's abilities to act in a trustworthy manner and to refer to other trustworthy agents, respectively. The first ability we term *expertise* and the second ability we term *sociability*.

Each agent may modify its models of its acquaintances, potentially based on its direct interactions with the acquaintance, based on interactions with agents referred to by the acquaintance, and based on ratings of this acquaintances received from other agents. In practice, not all of these means are simultaneously useful. Importantly, in our approach, agents can adaptively choose their neighbors, which they do every so often from among their current acquaintances.

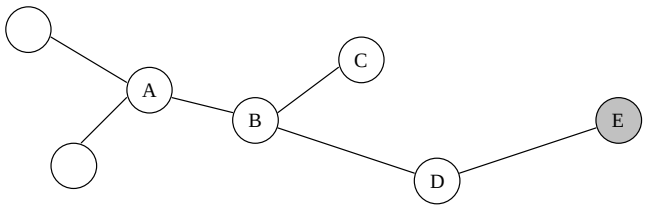

Figure 1. The process of finding witnesses, where agent A tries to estimate the trustworthiness of agent E, and D is a witness to E.

An agent may estimate the trustworthiness of a given party based on its own past interactions or may consult other trusted agents who have directly interacted with the same party. These agents are termed *witnesses*. An agent can find the right witnesses by seeking and following referrals from its neighbors. Figure 1 shows the process where agent A finds a witness D through the referrals $A \rightarrow B \rightarrow D$.

Intuitively, the agents' decisions depend upon the wishes of its user. This presupposes that the agents will act through user interfaces that support user feedback. That is, after each transaction, an agent can obtain its user's ratings of the other parties involved. This is effectively the same approach as in current commercial systems. The agent records these ratings. When a user has had direct interactions with another party, the user's own ratings are used to determine how much to trust the given party. When a user has not had sufficient interactions, the user's agent looks for agents of other users who are trustworthy and have had suitable interactions. In this sense, ours is a social approach for reputation. Each agent autonomously decides whether to entertain requests from other agents and whether to reveal its true ratings.

Finding if a party is trustworthy reduces to combining the rating given to it locally (based on direct interactions, if any) with the ratings assigned to that party by others. However, some important challenges must be addressed first.

- *Local ratings.* How may an agent rate another party based on direct interactions with it? Our approach does so by capturing the ratings over the last several interactions, which are recorded in the given agent's history, and then converting the ratings into belief functions.
- *Witnesses.* How may the agent find the right witnesses? Our approach applies a process of referrals, each agent being queried potentially offering referrals to other agents. This can lead to a focused search that does not send irrelevant messages to several agents and in which agents can help one another.
- *Testimonies.* How may the agent systematically incorporate the testimonies of the witnesses? Our approach includes the necessary representations and reasoning through which testimonies issued by witnesses can be combined in a principled manner.

The referrals-based approach developed here addresses these challenges directly. In particular, as discussed in Section 4, our approach has an advantage over other approaches in terms of the second and third challenges above. However, before we can describe the above three elements of our approach, we must consider a representational framework over which they are layered. There are three main choices in this regard.

- Simply average all the ratings over a period of time. This approach is followed in the reputation systems in eBay and Amazon. Everyone's ratings are centrally aggregated into one score. Whoever intends to base a decision on the aggregate score has no support for deciding if the constituent ratings are trustworthy. Sometimes the ratings are augmented with text notes by each rater. This adds the problem that prospective raters must reveal their ratings to everyone.
- Apply the Bayesian approach to combine evidence [17]. However, the Bayesian approach cannot distinguish between lack of belief and disbelief. That is, a low belief score for a proposition implies, a large belief score for its negation. Lack of belief must be modeled through the artificial construct of equiprobable prior probability distributions.
- Apply the Dempster-Shafer calculus, which handles the notion of supporting and refuting explicitly [13]. There is no causal relationship between a proposition and its negation, so the lack of belief does not imply disbelief. Rather, lack of belief in any particular

hypothesis implies belief in the set of all hypotheses, which is referred to as the state of uncertainty. This leads to the intuitive process of narrowing a hypothesis [10], in which the initial uncertainty is replaced with belief or disbelief as evidence is accumulated. Applied to the present setting, the proposition in question is whether a specified other party is trustworthy. Lack of belief would refer to the set $\{T, \neg T\}$, instead of $\{T\}$ or $\{\neg T\}$.

For the above reasons, this paper uses the Dempster-Shafer theory of evidence as its underlying computational framework.

2.1. Dempster-Shafer Theory

Let us now introduce the key concepts of the Dempster-Shafer approach. Let T mean that the given agent considers a given party to be trustworthy. A *frame of discernment* is the set of propositions under consideration.

Definition 1. Let $\Theta = \{T, \neg T\}$ be a frame of discernment. A *basic probability assignment (bpa)* is a function $m : 2^{\Theta} \mapsto [0, 1]$ where (1) $m(\phi) = 0$, and (2) $\sum_{\hat{A} \subset \Theta} m(\hat{A}) = 1$.

Thus $m(\{T\}) + m(\{\neg T\}) + m(\{T, \neg T\}) = 1$. A bpa is similar to a probability assignment except that its domain is the subsets and not the members of Θ. The sum of the bpa's of the singleton subsets of Θ may be less than 1. For example, given the assignment of $m(\{T\}) = 0.8$, $m(\{\neg T\}) = 0$, $m(\{T, \neg T\}) = 0.2$, we have $m(\{T\}) + m(\{\neg T\}) = 0.8$, which is less than 1.

For a subset $\hat{A}$ of Θ, the *belief function* $\mathsf{Bel}(\hat{A})$ is defined as the sum of the beliefs committed to the possibilities in $\hat{A}$. For example,

$$\mathsf{Bel}(\{T, \neg T\}) = m(\{T\}) + m(\{\neg T\}) + m(\{T, \neg T\}) = 1$$

For individual members of Θ (in this case, T and $\neg T$), Bel and m are equal. Thus, we have

$$\mathsf{Bel}(\{T\}) = m(\{T\}) = 0.8, \text{ and } \mathsf{Bel}(\{\neg T\}) = m(\{\neg T\}) = 0$$

2.2. Local Belief Ratings

When agent A_i is evaluating the trustworthiness of agent V_j, there are two components to the evidence. The first is an evaluation by A_i of the services offered by V_j. The second is testimonies from other agents in case A_i has had no transactions with V_j. Suppose A_i has the latest H services from V_j, $S_j = \{s_{j1}, s_{j2}, \ldots, s_{jH}\}$. We use the distinct values of $\{0.0, 0.1, \ldots, 1.0\}$ to denote the quality of service (QoS) $s_{jk}, where 1 \leq k \leq H$. Intuitively, these ratings are obtained from users.

Following Marsh [14], we define for each agent an upper and a lower threshold for trust. For each agent A_i, there are two thresholds Ω_i and ω_i, where $0 \leq \omega_i \leq \Omega_i \leq 1$. We use $f(x_k)$ to denote the probability that a particular value x_k of quality of services is obtained from V_j, where $x_k \in \{0.0, 0.1, \ldots, 1.0\}$. For example, given a specific value x_k, and that there are three services with that quality in the latest h responses, then $f(x_k) = 3/h$.

Definition 2. Given a series of responses from V_j, $S_j = \{s_{j1}, s_{j2}, \ldots, s_{jH}\}$, and the two thresholds Ω_i and ω_i of agent A_i, we can obtain A_i's bpa toward V_j: $m(\{T\}) = \sum_{x_k=\Omega_i}^{1} f(x_k)$, $m(\{\neg T\}) = \sum_{0}^{x_k=\omega_i} f(x_k)$, and $m(\{T, \neg T\}) = \sum_{x_k=\omega_i}^{x_k=\Omega_i} f(x_k)$.

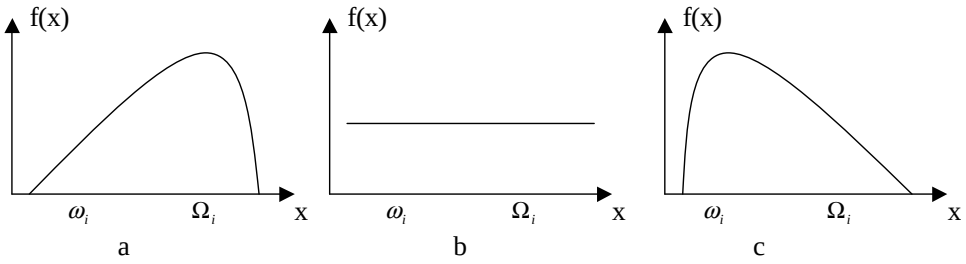

FIGURE 2. Example distributions of ratings for different agents

Figure 2 exhibits some distributions of ratings for different agents. Part (a) shows a distribution of ratings for high quality service providers, while Part (c) shows a distribution for low quality providers. Part (b) indicates that the quality of service could be random from a seller, although this usually does not happen in electronic commerce.

2.3. Combining Belief Functions

When an agent has not interacted often enough with a seller, it must seek the testimonies of other witnesses. Next we discuss how to combine such evidence.

A subset $\hat{A}$ of a frame Θ is called a *focal element* of a belief function Bel over Θ if $m(\hat{A}) > 0$. Given two belief functions over the same frame of discernment but based on distinct bodies of evidence, Dempster's rule of combination enables us to compute a new belief function based on the combined evidence. For every subset $\hat{A}$ of Θ, Dempster's rule defines $m_1 \oplus m_2(\hat{A})$ to be the sum of all products of the form $m_1(X)m_2(Y)$, where X and Y run over all subsets whose intersection is $\hat{A}$. The commutativity of multiplication ensures that the rule yields the same value regardless of the order in which the functions are combined.

Definition 3. Let Bel_1 and Bel_2 be belief functions over Θ, with basic probability assignments m_1 and m_2, and focal elements $\hat{A}_1, \ldots, \hat{A}_k$, and $\hat{B}_1, \ldots, \hat{B}_l$, respectively. Suppose

$$\sum_{i,j,\hat{A}_i \cap \hat{B}_j = \phi} m_1(\hat{A}_i) m_2(\hat{B}_j) < 1$$

Then the function $m : 2^{\Theta} \mapsto [0, 1]$ that is defined by

$$m(\phi) = 0, \text{ and}$$
$$m(\hat{A}) = \frac{\sum_{i,j,\hat{A}_i \cap \hat{B}_j = \hat{A}} m_1(\hat{A}_i) m_2(\hat{B}_j)}{1 - \sum_{i,j,\hat{A}_i \cap \hat{B}_j = \phi} m_1(\hat{A}_i) m_2(\hat{B}_j)}$$

for all non-empty $\hat{A} \subset \Theta$ is a basic probability assignment [23].

Bel, the belief function given by m, is called the *orthogonal sum* of Bel_1 and Bel_2. It is written $\mathsf{Bel} = \mathsf{Bel}_1 \oplus \mathsf{Bel}_2$. Let us now look at how beliefs obtained from two separate agents are combined. Suppose

$$m_1(\{T\}) = 0.8,\ m_1(\{\neg T\}) = 0,\ m_1(\{T, \neg T\}) = 0.2$$
$$m_2(\{T\}) = 0.9,\ m_2(\{\neg T\}) = 0,\ m_2(\{T, \neg T\}) = 0.1$$

Then m_{12} is obtained as follows:

$$m_{12}(\{T\}) = 0.72 + 0.18 + 0.08 = 0.98$$
$$m_{12}(\{\neg T\}) = 0$$
$$m_{12}(\{T, \neg T\}) = 0.02$$

Next suppose that one piece of the evidence confirms T, while the other disconfirms T. That is, we have the following situation:

$$m_1(\{T\}) = 0.8,\ m_1(\{\neg T\}) = 0,\ m_1(\{T, \neg T\}) = 0.2$$
$$m_2(\{T\}) = 0,\ m_2(\{\neg T\}) = 0.9,\ m_2(\{T, \neg T\}) = 0.1$$

Then m_{12} is obtained as follows:

$$m_{12}(\phi) = 0.72$$
$$m_{12}(\{T\}) = 0.08$$
$$m_{12}(\{\neg T\}) = 0.18$$
$$m_{12}(\{T, \neg T\}) = 0.02$$

In this case, 0.72 of our belief is committed to the empty set. Since there are no possibilities in this set, the belief in our other sets must be normalized to 1. This yields:

$$m_{12}(\{T\}) = 0.29$$
$$m_{12}(\{\neg T\}) = 0.64$$
$$m_{12}(\{T, \neg T\}) = 0.07$$

2.4. Deciding the Reputation

It helps to distinguish between two kinds of beliefs: *local belief* and *total belief*. An agent's local belief about another agent is from direct interactions with it and can be propagated to others upon request. An agent's total belief about another agent combines the local belief (if any) with testimonies received from all witnesses reached (if any). Total belief can be used for deciding whether the agent being considered is trustworthy. To prevent non-well-founded cycles, we restrict agents from propagating their total beliefs. However, in principle, the necessary information underlying a total belief can be obtained by the requesting agent from the original witnesses.

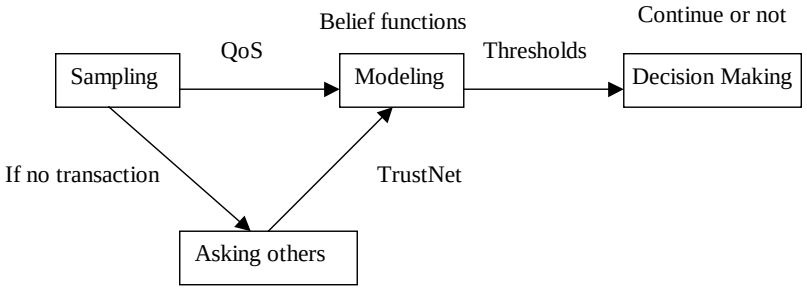

FIGURE 3. The process of deciding whether to cooperate with another agent

Agent A_r models all of the information he receives about agent V_g using belief functions, and then decides whether to trust V_g. Figure 3 shows the whole process. Total belief is

needed only if A_r has not previously had an interaction with V_g. Once A_r has had sufficient direct interactions with V_g, there is no reason for it to change its beliefs about V_g based on comments from others. Under our approach, A_r would have collected evidence from others regarding V_g prior to its first interaction with V_g and has no reason to seek it again after having interacted with V_g directly.

To evaluate the trustworthiness of V_g, A_r will check if V_g is one of its acquaintances, i.e., A_r has some local beliefs about its trustworthiness. If so, A_r will use its existing local belief to evaluate the trustworthiness of V_g. Otherwise, A_r will query its neighbors about V_g. When an agent receives a query about V_g's trustworthiness, it will check if V_g is one of its acquaintances. If yes, it will return the information about V_g; otherwise, it will return zero or more referrals to A_r. A_r, if it chooses, can then query some of the referred agents.

A referral r to agent A_j returned from agent A_i is written as $\langle A_i, A_j \rangle$. A series of referrals makes a referral chain. Observing that shorter referral chains are more likely to be fruitful and accurate [11] and to limit the effort expended in pursuing referrals, we define *depthLimit* as the bound on the length of any referral chain. The value of depthLimit and strategies to control the referring process are interesting questions for referral networks [28]. The referral process terminates in success when at least one rating is received and in failure when the depthLimit is reached or when it arrives at an agent neither gives an answer rating nor a referral.

To simplify the notation, we refer to the initial contact $\langle A_r, A_i \rangle$ as a referral as well. For simplicity, a chain is written as $\langle A_0, A_1, \ldots, A_k \rangle$, where A_0 is the querying agent and every agent A_i for $i < k$ gives a referral to agent A_{i+1}.

2.5. TrustNet

Now suppose A_r wishes to evaluate the trustworthiness of V_g. After a series of l referrals, a testimony about agent V_g is returned from agent A_j. Let the entire referral chain in this case be $\langle A_r, \ldots, A_j \rangle$, with length l. A TrustNet is a representation built from the referral chains produced from A_r's query. It is used to systematically incorporate the testimonies of the various witnesses regarding a particular party.

Definition 4. A TrustNet TN is a directed graph $TN(A_r, V_g, \mathbf{A}, R)$, where $\mathbf{A}$ is a finite set of agents $\{A_1, \ldots, A_N\}$, and R is a set of referrals $\{r_1, \ldots, r_n\}$.

Given a series of referrals $\{r_1, r_2, \ldots, r_n\}$, the requester A_r constructs a TrustNet TN by incorporating each referral $r_i = \langle A_i, A_j \rangle$ into TN. A_r adds r_i to R if and only if $A_j \notin \mathbf{A}$ and $depth(A_i) \leq depthLimit$.

Figure 4 shows how the testimonies propagate through a TrustNet. Suppose agent A_r wants to evaluate the trustworthiness of agent V_g, and $\{w_1, \ldots, w_L\}$ are a group of witnesses towards agent V_g. We now show how testimonies from witnesses can be incorporated into the rating of a given agent. Let τ_{A_i} and π_{A_i} be the belief functions corresponding to agent A_i's local and total beliefs, respectively.

Definition 5. Given a set of witnesses $\Delta = \{w_1, w_2, \ldots, w_L\}$, agent A_r will update its total belief value of agent V_g as follows

$$\pi_{A_r} = \tau_{w_1} \oplus \ldots \oplus \tau_{w_L}$$

Next we consider the situation where A_r needs to compute its total belief regarding V_g.

- **Case 1:** A_r has interacted with V_g. A_r will trust V_g if $\tau_{A_r}(\{T_{V_g}\}) - \tau_{A_r}(\{\neg T_{V_g}\}) \geq \rho$, where ρ is a threshold for trustworthiness and $0 < \rho < 1$. If we look at the belief function

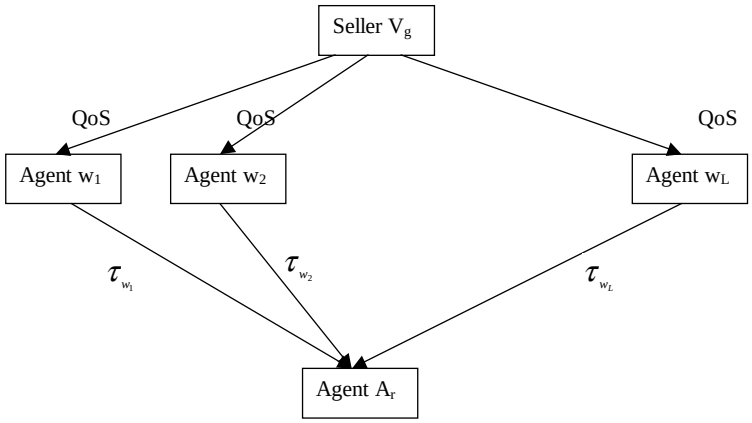

FIGURE 4. Testimony propagation through a TrustNet

more carefully, we see that the first transaction is important. If the first service is of high quality, then $\tau_{A_r}(\{T_{V_g}\}) - \tau_{A_r}(\{\neg T_{V_g}\}) = 1$; otherwise, the value would be -1.

- **Case 2:** A_r has not interacted with V_g. A_r computes its total belief about V_g. A_r will trust V_g if $\pi_{A_r}(\{T_{V_g}\}) - \pi_{A_r}(\{\neg T_{V_g}\}) \geq \rho$. This is the benefit of distributed reputation management in case that the rater and the rated agents have never interacted before.
- **Case 3:** V_g is totally new to the society. In this case, $\pi_{A_r}(\{T_{V_g}\}) = \pi_{A_r}(\{\neg T_{V_g}\} = 0$, and $\pi_{A_r}(\{T_{V_g}, \neg T_{V_g}\}) = 1$. V_g must initially establish its reputation in other ways, e.g., by advertising or obtaining endorsements from established agents [15].

3. EXPERIMENTAL RESULTS

Our experiments are based on an extension of a simulation testbed previously developed for information access [30]. The experiments involve between 100 and 500 agents. Some of the agents are identified as sellers and the rest as buyers. Each agent is modeled in terms of its *interest* (describing the services it is interested in purchasing) and its *expertise* (describing the services it is able to offer). Both interest and expertise are captures as terms vectors of dimension 5. The expertise vectors for each buyer are random, but for the sellers, each dimension of the expertise vectors is limited to 1.

Each agent keeps the latest 10 responses from another agent. The agents are limited in the number of neighbors they may have, here 4. The length of each referral chain is limited to 6. Moreover, we introduce a probability between 0 and 1 to model the cooperativeness of each agent A_i, denoted as C_{A_i}. The trustworthiness of a seller is viewed as the expectation of cooperative behavior from that seller. Agent A_i will generate an answer from its *expertise* vector upon a query with the probability C_{A_i} even when there is a good match between the query and its expertise vector.

In each simulation cycle, we randomly designate an agent to be the requester. The queries are generated as vectors by perturbing the interest vector of the requesting agent. An agent may query some of its neighbors. When an agent receives a query, it may answer it based on its expertise vector, or may give a referral to some of its neighbors. The originating agent

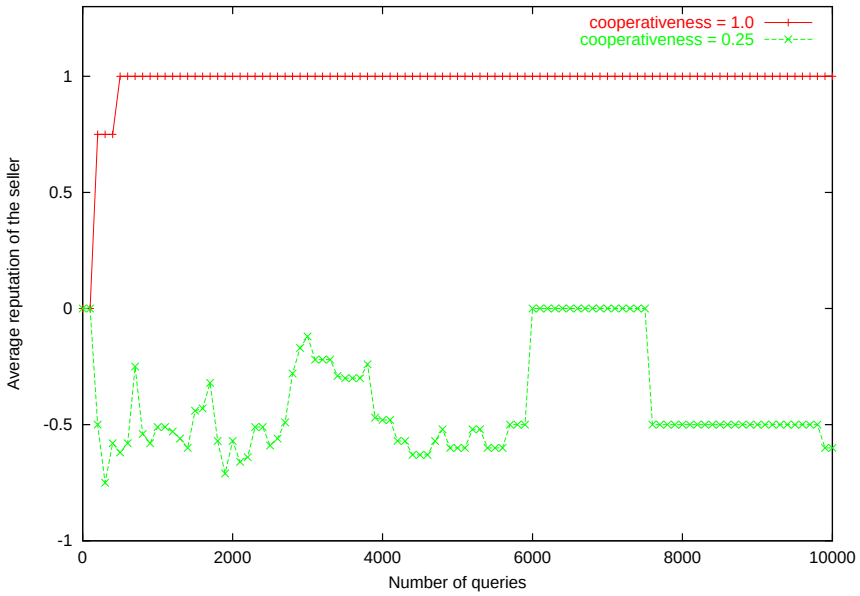

FIGURE 5. Reputations of the seller with different qualities of services in the bootstrapping stage (from a regular ring)

collects all possible referrals, and continues the process by following some of the suggested referrals. Each agent may keep track of certain acquaintances (a superset, usually a proper superset, of its neighbors). In our simulation, we allow 12 acquaintances. Periodically, each agent decides which of its acquaintances are dropped and which are promoted to neighbors.

3.1. Metrics

We now define some useful metrics in which to intuitively capture the results of our experiments.

Definition 6. Suppose $\{w_1, \ldots, w_L\}$ are exactly L agents whose neighbors include V_i. Then β_{V_i}, the cumulative belief regarding agent V_i is computed as

$$\beta_{V_i} = \tau_{w_1} \oplus \tau_{w_2}, \ldots, \oplus \tau_{w_L}$$

and the reputation of agent V_i is defined as

$$\Gamma(V_i) = \beta_{V_i}(\{T_{V_i}\}) - \beta_{V_i}(\{\neg T_{V_i}\}.$$

If $L = 0$ then $\Gamma(V_i) = 0$.

3.2. Bootstrapping

The neighborhood relation over the agents induces a graph wherein the agents are the vertices with edges to their neighbors. Our simulation must begin with a graph. Following Watts and Strogatz's study of small-world graphs [27], we begin our simulation from a ring (but with directed edges). The specific ring with which we begin is a regular ring with 100

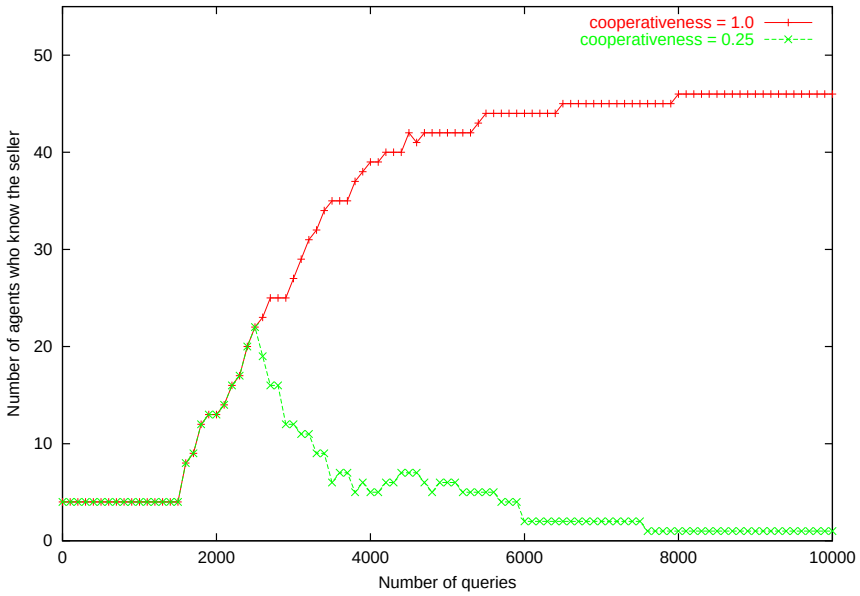

FIGURE 6. Number of the agents who know the seller in the bootstrapping stage (from a regular ring)

nodes and 4 edges per node. Initially, each node's out-edges point to its nearest neighbors in the ring. Since there are no real users in these simulation, the quality of service (QoS) is evaluated based on how close the answer it to the interest vector of the requesting agent.

The cooperativeness for each agent is set to 1 if not specified. For any two agents A_i and A_j, $\tau_{A_i}(\{T_{A_j}\}) = \tau_{A_i}(\{\neg T_{A_j}\}) = 0$, $\tau_{A_i}(\{T_{A_j}, \neg T_{A_j}\}) = 1$ in the beginning. Only one agent is identified as a seller, called V_g. We use a fairly low value for the lower threshold ω_i to facilitate modeling the cooperativeness of seller agents. For each agent A_i, we have $\Omega_i = 0.5$ and $\omega_i = 0.1$. Whether a given neighbor will remain a neighbor depends on how close the neighbor's expertise is to the agent's interest. After every 100 rounds, we compute the reputation for the single seller agent using the metrics defined above. The computation is not counted in the simulation cycle.

In the first simulation we evaluate the convergence of our approach. Consider the following two settings for the cooperativeness of the seller: (1) 1 or good quality service and (2) 0.25 or bad quality service. Figure 5 shows the reputation of the two kinds of sellers. The reputation of seller with $C_{V_g} = 1$ climbs quickly to 1 and then stabilize at 1. The reputation for a seller with cooperativeness $C_{V_g} = 0.25$ is low.

Figure 6 shows the number of agents who know the seller. More and more buyers have the good seller (with $C_{V_g} = 1$) as a neighbor (about 46 out of 99 agents). Fewer and fewer agents have the bad seller as a neighbor. Several agents have it as a neighbor early on in the simulation, but eventually few if any agents do. In the figure, it is only one agent's neighbor after 10000 queries.

This is the reason why, in Figure 5, the reputation is negative for a while and finally approaches zero. The convergence to zero indicates that the other agents have forgotten about this agent.

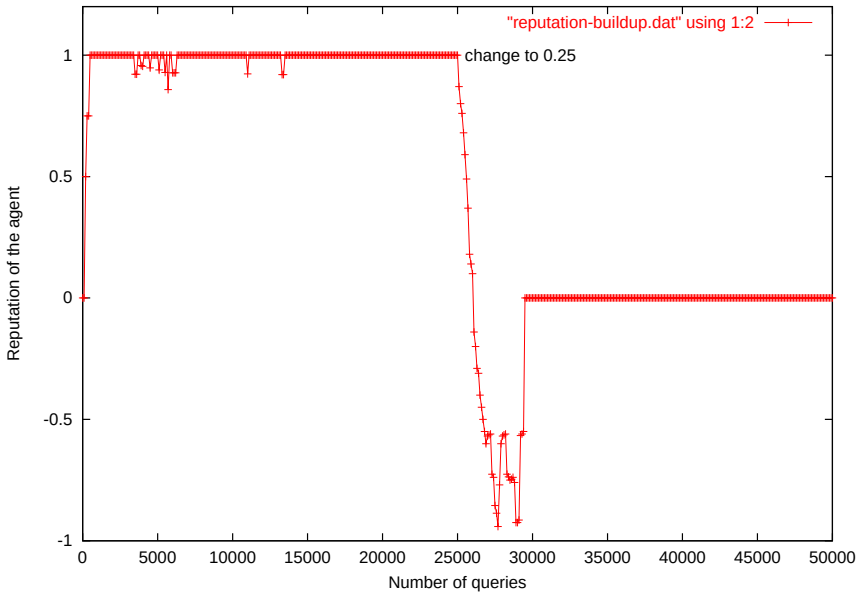

FIGURE 7. Reputation buildup and crash of seller V_g

3.3. Reputation Buildup

Sometime, the reputation of a seller will not remain stable for long. They might be good in the very beginning, but later they may lower their quality of service. In the second simulation, we show that a *very good* seller V_g who accumulates a high reputation during the first simulation cycle of $25,000$, behaves cooperatively with a cooperativeness factor 1 until it reaches a high reputation value, and then starts abusing its reputation by decreasing its responsiveness factor to 0.25. Thus its average reputation begins to drop, ultimately settling at a reputation of 0. Figure 7 illustrates this case. A reputation of 0 indicates that V_g is no longer a neighbor of any agent. That is, it ends up isolated from the other agents. However, in order to distinguish the agent with *zero* reputation (new agent) from the agent with -1 reputation, we might introduce a *blacklist* to remember all V_g with bad reputation. We defer this enhancement to future work.

3.4. Community Size

Usually there is a better chance to select a partner in a large (virtual) city of $300,000$ people than in a small town of $3,000$ people. Conversely, it is much easier to collect "bad" testimonies in a small town. We conjecture that the average reputation of an agent in a smaller group should change faster than that in a larger community.

Given two groups of agents, with the number of agents 20 and 100, respectively. Suppose sellers V_{g_1} and V_{g_2} are two cooperative agents in the beginning with cooperativeness factors of 1. After several simulation cycles, both V_{g_1} and V_{g_2} decrease their cooperativeness factor to 0.25. Thus, their average reputation starts dropping because of their non-cooperative behavior. Figure 8 shows that the reputation of agent V_{g_1} drops faster (measured by number of queries sent by each agent) since it is in a smaller community.

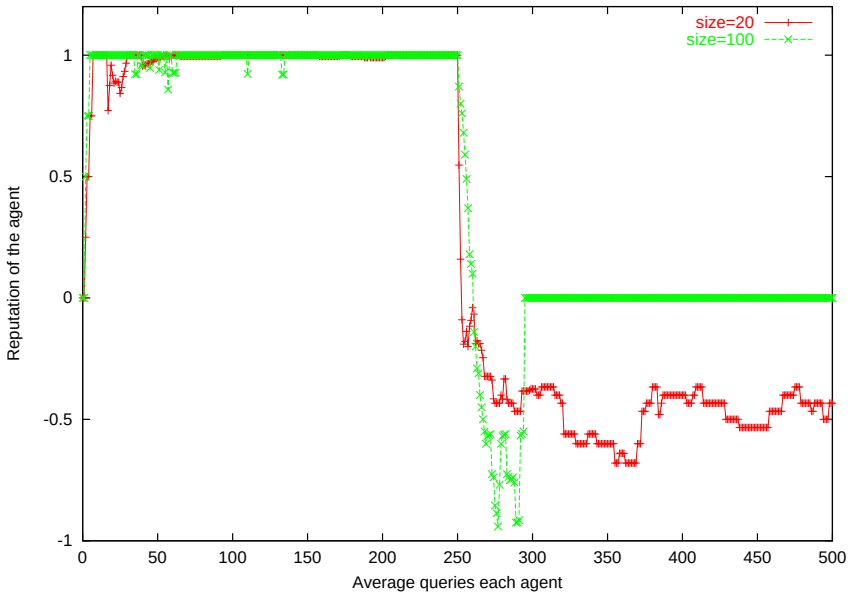

FIGURE 8. Low quality sellers in different community sizes

Another interesting phenomenon is that the reputation of agent V_{g_1} oscillates around -0.5, a low reputation level, but agent V_{g_2}'s reputation changes back to 0. In conjunction with the above-mentioned experiments, this indicates that a bad agent is more easily forgotten in a big community than in a small group.

4. RELATED WORK

Reputation management in electronic commerce has drawn much interest lately. Yenta [8] and Weaving a Web of Trust [12] are two well-known prototype approaches. Yenta clusters people with common interests according to referrals of users who know each other and verify the assertions they make about themselves, while Weaving a Web of Trust relies on the existence of a connected path between two users. These systems require preexisting social relationships among the users of their electronic community. It is not clear how to establish such relationships and how the ratings propagate through this community.

A social mechanism of reputation management was implemented in Kasbah [7, 31]. This mechanism requires that users give a rating for themselves and either have a central agency (direct ratings) or other trusted users (collaborative ratings). A central system keeps track of the users' explicit ratings of each other, and uses these ratings to compute a person's overall reputation or reputation with respect to a specific user in a directed graph. However, it is not clear how the agents collect the ratings in an open environment where the number of agents grows to very large.

Trusted Third Parties (TTP) [18] are often employed to facilitate trust in commercial transactions. Typical TTP services for electronic commerce include certification, time-stamping, and notarization. TTPs act as a bridge between buyers and sellers in electronic

marketplaces. However, they are most appropriate for closed marketplaces. In loosely federated, open systems a TTP may either not be available or have limited power to enforce good behavior.

One of the earliest works that tried to give a formal treatment of trust was that of Marsh [14]. Marsh's model attempted to integrate all the aspects of trust taken from sociology and psychology. Since Marsh's model has strong sociological foundations, the model is rather complex and cannot be easily used in today's electronic communities. Moreover the model only considers an agent's own experiences and doesn't involve any social mechanisms. Hence, a group of agents cannot collaborate to rate others.

Tan and Thoen [26] discuss the trust that is needed to engage in a transaction. In their model, a party engages in a transaction only if its level of trust exceeds its personal threshold. The threshold depends on the type of the transaction and the other parties involved in the transaction. The trust in a transaction includes the trust in the other party and the trust in the control mechanisms. Tan and Thoen use these design principles to make people trust electronic commerce. By contrast, we focus on the computational model of distributed reputation management for electronic commerce.

Another computational method is the *Social Interaction Framework* (SIF) [21]. In SIF, an agent evaluates the reputation of another agent based on direct observations as well through other *witnesses*. Moreover, Schillo *et al.* tested the performance of two groups of agents with different settings for honesty versus dishonesty for altruism versus egotism [22]. This work motivates some of our experiments for reputation management. However, SIF does not describe how to find such witnesses, whereas in the electronic communities, deals are brokered among people who may have never interacted before.

We previously developed an approach for social reputation management [29], in which they used a scalar value to represent an agent's belief ratings about another and combine them with testimonies using combination schemes similar to the certainty factor model. The drawbacks of the certainty factor models, discussed in Section 2, led us to consider alternate approaches.

Abdul-Rahman and Hailes [1] proposed an approach in virtual communities. This work adapts Marsh's work. It too uses concepts such as situation or contexts and simplifies others such as trust, which is limited to have only four possible values. Abdul-Rahman and Hailes require each agent to keep complex data structures that represent a kind of global knowledge about the whole network. This is a serious limitation, because usually maintaining and updating these data structures can be laborious and time-consuming. Also it is not clear how the agents get the needed information and how well the model will scale when the number of agents grows.

Aberer and Despotovic [2] simplified our model and use that to manage trust in a peer-to-peer network where no central database is available. Their model is based on binary trust, i.e., an agent is either trustworthy or not. When a dishonest transaction happens, the agents can forward their complaints to other agents. Aberer and Despotovic use a special data structure, namely P-Grid, to store the complaints in a peer-to-peer network. In order to evaluate the trustworthiness of another agent B, an agent A searches the leaf level of the P-Grid for complaints on agent B.

Breban and Vassileva [4] present a coalition-formation mechanism based on trust relationships. Their approach extends existing transaction-oriented coalitions, and might be an interesting direction for distributed reputation management for electronic commerce.

There has been much work on social abstractions for agents, e.g., [5, 9]. The initial work on this theme studied various of relationships among agents. More recent work on these themes has begun to look at the problems of deception and fraud [6]. However, our proposed approach goes beyond their approach in its representations of trust, propagation algorithms,

and formal analysis.

5. CONCLUSION

This paper examines the problem of distributed reputation management for electronic commerce. We directly consider how agents may place trust in each other and refine the ways in which an agent may convey its ratings of the trustworthiness of an agent to another agent. Explicit distributed reputation management can potentially help the agents detect selfish, antisocial, or unreliable sellers in electronic commerce and thus lead to more reliable relationships among buyers and sellers.

The iterated, multi-player prisoners' dilemma is intimately related to the evolution of trust [3]. On the one hand, if the players trust each other, they can both cooperate and avert a mutual defection where both suffer. On the other hand, such trust can only build up in a setting where the players must repeatedly interact with each other. Our observation is that a reputation mechanism sustains rational cooperation, because good players are rewarded by society whereas bad players are penalized. Both the rewards and penalties can be greater from a society than from an individual [16, 24].

Our present approach does not fully protect against spurious ratings generated by malicious agents. It relies upon there being a large number of agents who offer honest ratings to override the effect of the ratings provided by the malicious agents. In future work, we plan to study the special problems of lying and rumors in extensions of the present evidential framework. We also plan to study evolutionary situations where groups of agents consider rating schemes for other agents. The purpose is not only to study alternative approaches for achieving more efficient communities, but also to test if our mechanism is robust against invasion and, hence, is more stable.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under grant IIS-9624425 (Career Award) and ITR-0081742. We are indebted to the anonymous reviewers for their helpful comments.

References

[1] Alfarez Abdul-Rahman and Stephen Hailes. Supporting trust in virtual communities. In *Proceedings of the 33rd Hawaii International Conference on Systems Science*, 2000.

[2] Karl Aberer and Zoran Despotovic. Managing trust in a peer-2-peer information system. In *Proceedings of the 10th International Conference on Information and Knowledge Management (CIKM)*, pages 310–317, 2001.

[3] Robert Boyd and Jeffrey P. Borderbaum. No pure strategy is evolutionary stable in the repeated prisoner's dilemma game. *Nature*, 327:58–59, 1987.

[4] Silvia Breban and Julita Vassileva. Long-term coalitions for the electronic marketplace. In *Proceedings of Canadian AI Workshop on Novel E-Commerce Applications of Agents*, pages 6–12, 2001.

[5] Cristiano Castelfranchi. Modelling social action for AI agents. *Artificial Intelligence*, 103:157–182, 1998.

[6] Cristiano Castelfranchi and Rino Falcone. Principle of trust for MAS: cognitive anatomy, social importance, and quantification. In *Proceedings of 3rd International Conference on MultiAgent Systems*, pages 72–79, 1998.

[7] Anthony Chavez and Pattie Maes. Kasbah: An agent marketplace for buying and selling goods. In *Proceedings of the 1st International Conference on the Practical Application of Intelligent Agents and Multiagent Technology (PAAM)*, pages 75–90, 1996.

[8] Lenny Foner. Yenta: A multi-agent, referral-based matchmaking system. In *Proceedings of the 1st International Conference on Autonomous Agents*, pages 301–307, 1997.

[9] Les Gasser. Social conceptions of knowledge and action: DAI foundations and open systems semantics. *Artificial Intelligence*, 47:107–138, 1991.

[10] Jean Gordon and Edward H. Shortliffe. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, 26:323–357, 1985.

[11] Henry Kautz, Bart Selman, and Al Milewski. Agent amplified communication. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 3–9, 1996.

[12] Rohit Khare and Adam Rifkin. Weaving a web of trust. *World Wide Web*, 2(3):77–112, 1997.

[13] Henry E. Kyburg. Bayesian and non-bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[14] Steven P. Marsh. *Formalising Trust as a Computational Concept.* PhD thesis, Department of Computing Science and Mathematics, University of Stirling, April 1994.

[15] E. Michael Maximilien and Munindar P. Singh. Reputation and endorsement for Web services. *ACM SIGEcom Exchanges*, 3(1):24–31, 2002.

[16] Martin A. Nowak and Karl Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393:573–577, 1998.

[17] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference.* Morgan Kaufmann Publishers Inc., San Mateo, California, 1988.

[18] Tim Rea and Peter Skevington. Engendering trust in electronic commerce. *British Telecommunications Engineering*, 17(3):150–157, 1998.

[19] Paul Resnick and Richard Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. In *Working paper for the NBER Workshop on Empirical Studies of Electronic Commerce*, 2000.

[20] Paul Resnick, Richard Zeckhauser, Eric Friedman, and Ko Kuwabara. Reputation systems: Facilitating trust in Internet interactions. *Communications of the ACM*, 43(12):45–48, 2000.

[21] Michael Schillo and Petra Funk. Who can you trust: Dealing with deception. In *Proceedings of the Autonomous Agents Workshop on Deception, Fraud and Trust in Agent Societies*, pages 95–106, 1999.

[22] Michael Schillo, Petra Funk, and Michael Rovatsos. Using trust for detecting deceitful agents in artificial societies. *Applied Artificial Intelligence*, 14:825–848, 2000.

[23] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

[24] Susan P. Shapiro. The social control of impersonal trust. *The American Journal of Sociology*, 93(3):623–658, 1987.

[25] Munindar P. Singh, Bin Yu, and Mahadevan Venkatraman. Community-based service location. *Communications of the ACM*, 44(4):49–54, April 2001.

[26] Yao-Hua Tan and Walter Thoen. An outline of a trust model for electronic commerce. *Applied Artificial Intelligence*, 14:849–862, 2000.

[27] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.

[28] Bin Yu. *Emergence and Evolution of Agent-based Referral Networks*. PhD thesis, Department of Computer Science, North Carolina State University, 2001.

[29] Bin Yu and Munindar P. Singh. A social mechanism of reputation management in electronic communities. In *Proceedings of the 4th International Workshop on Cooperative Information Agents*, pages 154–165, 2000.

[30] Bin Yu, Mahadevan Venkatraman, and Munindar P. Singh. An adaptive social network for information access: Theoretical and experimental results. *Applied Artificial Intelligence*, 2002. In press.

[31] Giorgos Zacharia and Pattie Maes. Trust management through reputation mechanisms. *Applied Artificial Intelligence*, 14:881–908, 2000.