Gossip-Based Aggregation of Trust in Decentralized Reputation Systems

Ariel D. Proccacia, Yoram Bachrach, and Jeffrey S. Rosenschein

School of Engineering and Computer Science
Hebrew University, Jerusalem, Israel
{arielpro,yori,jeff}@cs.huji.ac.il

Abstract. Decentralized reputation systems have of late emerged as the prominent method of establishing trust among selfish agents in today's online environments. A key issue is the efficient aggregation of data in the system; several approaches have hitherto been advanced, but are plagued by major shortcomings.

We put forward a novel decentralized data management scheme grounded in gossip-based algorithms. Rumor mongering is known to possess algorithmic advantages, and indeed, our framework inherits numerous salient features: scalability, robustness, globality, and simplicity. We also demonstrate that our scheme motivates agents to maintain a sparkling clean reputation, and is inherently impervious to certain attacks.

1 Introduction

In open multiagent environments, self-interested agents are often tempted to employ deceit as they interact with others. Fortunately, dishonest agents can expect their victims to retaliate in future encounters. This "shadow of the future" (e.g., as explored by Axelrod [5]) motivates cooperation and trustworthiness.

However, as the number of agents in the system grows, agents have an increasingly small chance to deal with an agent they already know; as a consequence, building trust in domains teeming with numerous agents becomes much harder. Reputation systems address this problem by collecting and spreading reports among agents, so that agents may learn from others' experience. To put it differently, agents are intimidated by the "shadow of the future" today, even though tomorrow they are most likely to meet total strangers.

Reputation systems can be decomposed into two major components: the trust model, which describes whether an agent is trustworthy, and the data management scheme. The latter component poses some interesting questions, since it is imperative to efficiently aggregate trust-related information in the system. A simple solution is maintaining a central database which contains the gathered feedback of past transactions. Unfortunately, this solution is inappropriate in distributed environments where scalability is a major concern, as the database soon becomes a bottleneck of the system. Moreover, this approach is not robust to failures. Previous work on decentralized reputation schemes (surveyed in Section 6) suffers from major problems: agents have to maintain complex data structures, evaluation of trust is based only on local information, or there are restrictive assumptions on the trust model.[1]

We try our hand at tackling this hornets' nest by designing a novel method of trust aggregation. The roots of our *gossip-based* approach can be traced to to a seminal paper by Frieze and Grimmett [13]: a rumor starts with one agent; at each stage, each agent that knows the rumor spreads it to another agent chosen uniformly at random. The authors show that the rumor reaches all agents very quickly (a result that coincides with real life). We directly rely on more recent results, surveyed in Section 2. It has been shown that aggregate information, such as averages and sums of agents' inputs, can be calculated using similar methods of uniform gossip in a way that scales gracefully as the number of agents increases. Furthermore, the approach is robust to failures, and the results hold even when one cannot assume a point-to-point connection between any two agents (as is the case in Peer-to-Peer networks).

In our setting, each agent merely keeps its private evaluation of the trustworthiness of other agents, based on its own interactions.[2] When an agent wishes to perform a transaction with another, it obtains the *average* evaluation of the other's reputation from all agents in the system, using a gossip-based technique. Some advantages are immediately self-evident: each agent stores very little information, which can be simply and efficiently organized, and evaluation of trust is based on global information. Additionally, this framework inherits the advantages of gossip-based algorithms: scalability, robustness to failure, decentralization (and in particular, applicability in Peer-to-Peer networks).

We show that our scheme has two other major advantages. An important desideratum one would like a reputation system to satisfy is motivating agents to maintain an untarnished reputation, i.e., to be absolutely trustworthy (as opposed to, say, being generally trustworthy but occasionally cheating). We show that our data management scheme, together with an extremely simple trust model, satisfies this property. We also demonstrate that our scheme is inherently resistant to some attacks (with no assumptions on the trust model). This is a positive side effect of the exponential convergence rates of the algorithms we use.

The paper proceeds as follows. In Section 2, we survey the gossip-based techniques that we utilize. In Section 3, we describe our framework. In Sections 4 and 5, we demonstrate that our framework has the two abovementioned features. In Section 6 we give an overview of related work, and finally, in Section 7, we give our conclusions and present directions for future research.

2 Gossip-Based Information Aggregation

In this section, we survey the relevant results of Kempe, Dobra and Gehrke [15].

We begin by describing a simple algorithm, Push-Sum, to compute the average of values at nodes in a network. There are n nodes in the system, and each node i holds an input $x_i \geq 0$. At time t, each node i maintains a *sum* $s_{t,i}$ and

[1] The "or" is not exclusive.

[2] The question of how agents set this valuation is outside the scope of the paper.

a *weight* $w_{t,i}$. The values are initialized as follows: $s_{0,i} = x_i$, $w_{0,i} = 1$. At time 0, each node i sends the pair $s_{0,i}, w_{0,i}$ to itself; at every time $t > 0$, the nodes follow the protocol given as Algorithm 1.

Algorithm 1

```
procedure PUSH-SUM
    Let $\{(\hat{s}_l, \hat{w}_l)\}_l$ be all the pairs sent to $i$ at time $t-1$
    $s_{t,i} \leftarrow \sum_l \hat{s}_l$
    $w_{t,i} \leftarrow \sum_l \hat{w}_l$
    Choose a target $f_t(i)$ uniformly at random
    Send the pair $(\frac{1}{2}s_{t,i}, \frac{1}{2}w_{t,i})$ to $i$ (yourself) and to $f_t(i)$
    $\frac{s_{t,i}}{w_{t,i}}$ is the estimate of the average at time $t$
end procedure
```

Let $U(n, \delta, \epsilon)$ (the *diffusion speed* of uniform gossip) be an upper bound on the number of turns PUSH-SUM requires so that for all $t \geq U(n, \delta, \epsilon)$ and all nodes i,

$$\frac{1}{\sum_k x_k} \cdot \left| \frac{s_{t,i}}{w_{t,i}} - \frac{1}{n} \sum_k x_k \right| \leq \epsilon$$

(the relative error is at most ϵ) with probability at least $1 - \delta$.

Theorem 1. *[15]*

1. $U(n, \delta, \epsilon) = O(\log n + \log \frac{1}{\delta} + \log \frac{1}{\epsilon})$.
2. *The size of all messages sent at time t by* PUSH-SUM *is* $O(t + \max_i \textit{bits}(x_i))$, *where* $\textit{bits}(x_i)$ *is the number of bits in the binary representation of* x_i.

A major advantage of gossip-based algorithms is their robustness to failures: the aggregation persists in the face of failed nodes, permanent communication failures, and other unfortunate events. Further, no recovery action is required. The assumption is that nodes can detect whether their message has reached its destination; PUSH-SUM is modified so that if a node detects its target failed, it sends its message to itself. The following result is known to be true.

Theorem 2. *[15] Let $\mu < 1$ be an upper bound on the probability of message loss at each time step, and let U' be the diffusion speed of uniform gossip with faults. Then:*

$$U'(n, \delta, \epsilon) = \frac{2}{(1-\mu)^2} U(n, \delta, \epsilon).$$

In several types of decentralized networks, such as peer-to-peer networks, point-to-point communication may not be possible. In these networks, it is assumed that at each stage, nodes send messages to all their neighbors (*flooding*). When the underlying graph is an expander, or at least expected to have good expansion, results similar to the above can be obtained. Fortunately, it is known that several peer-to-peer topologies induce expander graphs [18, 16].

In the rest of the paper, we have $x_i \leq 1$, and in particular $\sum_i x_i \leq n$. Therefore, it is possible to redefine U to be an upper bound on the number of turns required so that for all $t \geq U$ and all nodes i, the *absolute error* $\left|\frac{s_{t,i}}{w_{t,i}} - \frac{1}{n}\sum_k x_k\right|$ is at most ϵ with confidence $1-\delta$, and it still holds that $U(n,\delta,\epsilon) = O(\log n + \log\frac{1}{\delta} + \log\frac{1}{\epsilon})$. Hereinafter, when we refer to U we have this definition in mind.

Remark 1. The protocol Push-Sum is presented in terms of a synchronized starting point, but this assumption is not necessary. A node which poses the query may use the underlying communication mechanism to inform all other nodes of the query; convergence times are asymptotically identical.

3 Our Framework

Let the set of agents be $N = \{1,\ldots,n\}$. Each agent $i \in N$ holds a number $r_i^j \in [0,1]$ for each agent $j \in N$ (including itself); this number represents j's reputation with respect to i, or to put it differently, the degree to which i is willing to trust j. As agents interact, these assessments are repeatedly updated. We do not concern ourselves with how agents set these values.[3] However, it should be assumed that different agents use identical (or at least similar) criteria for determining other agents' reputations.

When an agent i is deliberating whether to deal with another agent j, i wishes to make an informed evaluation of the other's reputation. Let $\bar{r}^j = \frac{\sum_k r_k^j}{n}$ be the average of j's reputation with respect to all agents. Knowledge of $\bar{r}^j$ gives i a good idea of how trustworthy j is.[4] The average is calculated via Push-Sum.

Algorithm 2

```
procedure Eval-Trust($i, j, \delta, \epsilon$)▷ $i$ evaluates $\bar{r}^j$ with accuracy $\epsilon$, confidence $1-\delta$
    for all $k \in N$ do
        $x_k \leftarrow r_k^j$          ▷ Inputs to Push-Sum are $j$'s reputation w.r.t. agents
    end for
    run Push-Sum for $U = U(n,\delta,\epsilon)$ stages
    return $\frac{s_{U,i}}{w_{U,i}}$
end procedure
```

The protocol Eval-Trust, given as Algorithm 2, is described somewhat informally, as the idea is very simple. Push-Sum is executed for $U = U(n,\delta,\epsilon)$ stages. At time U, it holds for all $k \in N$, and in particular for agent i, that

[3] See Sections 4 and 7, though, for some comments on this point.

[4] We address the question of how this knowledge is used in Section 4.

$\left|\frac{s_{t,i}}{w_{t,i}} - \bar{r}^j\right| \leq \epsilon$, with probability $1-\delta$. To put it differently, the algorithm returns a very good approximation of j's average reputation.

In practice, when two agents i and j interact, i may evaluate j's reputation (and vice versa) by calling Eval-Trust. The protocol quickly returns the approximation of $\bar{r}^j$, *based on the values* r_k^j *at the time* Eval-Trust *was called.* Each agent i keeps different values $s_{t,i}$ and $w_{t,i}$ for every different query that was issued by some other agent in the system, and updates these values repeatedly according to Push-Sum. In other words, at any stage every agent participates in many parallel executions of Push-Sum.

Remark 2. The size of messages depends on how the r_i^j are calculated, and as mentioned above, this issue is outside the scope of this paper. Nevertheless, there would usually be a constant number of reputation levels (say, for instance, $r_j^i \in \{0, 0.1, 0.2, \ldots, 1\}$), so the message size would normally be constant.

It must be stressed that as the above method of aggregating an agent's average reputation relies on the gossip-based algorithm Push-Sum, it inherits all the latter's benefits, in particular robustness to failure and applicability in Peer-to-Peer networks.

4 The Benefit of an Unstained Reputation

It is very desirable (indeed, crucial) that a reputation system be able to induce truthfulness in agents. Naturally, an agent with a stained reputation would be shunned by its peers, while an agent with a good reputation would easily solicit deals and transactions. A further step in this direction is motivating agents *never* to cheat. Indeed, an agent with a generally good reputation, which only occasionally cheats, would probably be able to win the confidence of peers; there is seemingly no reason why an agent should not play false now and again. Nevertheless, we study in this section an extremely simple and general trust model, and show that with the data management scheme that we have presented, there is a social benefit to having a very high reputation: the higher the agent's reputation, the shorter the time required to close deals.

We consider a model in which each agent i has a reputation threshold r_i^{thr} (similar to [20]) and a confidence level δ_i: agent i is willing to deal with an agent j iff i knows that j's average reputation is at least r_i^{thr}, with confidence $1-\delta_i$. Of course, i evaluates j's reputation as usual, using Eval-Trust. Recall that when the algorithm terminates, agent i only has an ϵ-close approximation of $\bar{r}^j$. If $\frac{s_{t,i}}{w_{t,i}}$ is very close to r_i^{thr}, i would have to increase the accuracy.

Remark 3. We still do not commit to the way the values r_j^i are determined and updated, so the above trust model is quite general.

The procedure Decide-Trust, given as Algorithm 3, is a straightforward method of determining whether $\bar{r}^j \geq r_i^{thr}$. Agent i increases the accuracy of the evaluation by repeatedly halving ϵ, until it is certain of the result. In this context, a stage of Eval-Trust identifies with a stage of Push-Sum.

Algorithm 3

```
procedure DECIDE-TRUST($(i, j)$)                ▷ $i$ decides if it wants to deal with $j$
    $\epsilon \leftarrow 1/2$                                       ▷ Initialization
    $k_1 \leftarrow 0$
    loop
        $k_2 \leftarrow U(n, \delta_i, \epsilon)$
        run EVAL-TRUST($(j)$) for another $k_2 - k_1$ stages     ▷ A total of $k_2$ stages
        if $s_{t,i}/w_{t,i} < r_i^{thr} - \epsilon$ then
            return false
        else if $s_{t,i}/w_{t,i} > r_i^{thr} + \epsilon$ then
            return true
        end if
        $k_1 \leftarrow k_2$
        $\epsilon \leftarrow \epsilon/2$
    end loop
end procedure
```

Proposition 1. *Let $i, j \in N$, and $\Delta_{ij} = |\bar{r}^j - r_i^{thr}|$. With probability at least $1 - \delta_i$,* DECIDE-TRUST *correctly decides whether agent j's reputation is at least r_i^{thr} after $O(\log n + \log \frac{1}{\delta_i} + \log \frac{1}{\Delta_{ij}})$ stages of* EVAL-TRUST.[5]

Proof. Assume w.l.o.g. that $r_i^{thr} < \bar{r}^j$, and that the algorithm reached a stage t_0 where $\epsilon < \Delta_{ij}/2$. At this stage, it holds that $|\frac{s_{t,i}}{w_{t,i}} - \bar{r}^j| \leq \epsilon$ (with probability $1 - \delta_i$), and therefore (see Figure 1):

$$\begin{aligned}\frac{s_{t,i}}{w_{t,i}} &\geq \bar{r}^j - \epsilon \\ &= r_i^{thr} + \Delta_{ij} - \epsilon \\ &> r_i^{thr} + \epsilon.\end{aligned}$$

Hence, the algorithm surely terminates when $\epsilon < \Delta_{ij}/2$. Now the proposition follows directly from the fact that $U(n, \delta_i, \Delta_{ij}) = O(\log n + \log \frac{1}{\delta_i} + \log \frac{1}{\Delta_{ij}})$. □

To conclude, Proposition 1 implies that there is a benefit for agent j in maintaining a high reputation: for any agent i with a reasonable threshold, Δ_{ij} is significant, and this directly affects the running time of DECIDE-TRUST.

Remark 4. The result is limited, though, when the number of agents n is large, as the time to evaluate an agent's reputation is also proportional to $\log n$.

5 Resistance to Attacks

In Section 3, we have seen that information about an agent's reputation can be efficiently propagated, as long as all agents consistently follow EVAL-TRUST.

[5] The probability is the chance that the algorithm will answer incorrectly; the bound on the number of stages is always true.

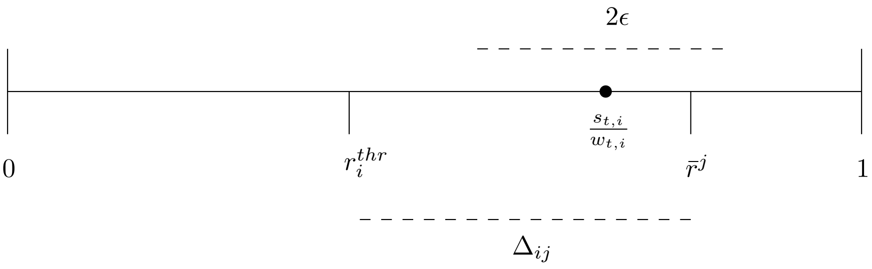

Fig. 1. Illustration for the proof of Proposition 1. Once $\epsilon < \Delta_{ij}/2$ and $s_{t,i}/w_{t,i} > r_i^{thr} + \epsilon$, it holds with probability $1 - \delta_i$ that $\bar{r}^j \geq r_i^{thr}$.

However, in multiagent systems we often deal with selfish, self-interested agents. In our context, a manipulative agent may artificially increase or decrease the overall evaluation of some agent's reputation by deviating from the protocol.

In the framework we have presented, trust is evaluated on the basis of global knowledge, i.e., the average of all reputation values in the system. Therefore, any small coalition cannot significantly change the average reputation of some agent j by setting their own valuations r_i^j to legal values in $[0, 1]$, and then following the protocol EVAL-TRUST.[6]

This is of course not the case when a manipulator is allowed to set its reputation value arbitrarily. As a very simple motivating example, consider a setting where agents propagate agent j's average reputation ($x_i = r_i^j$ for all i), and a manipulator i^m wants to ensure that for all i, $\frac{s_{t,i}}{w_{t,i}}$ converges to a high value as the time t increases. At some stage t_0, the manipulator updates s_{t_0,i^m} to be n, but except for this harsh deviation follows the protocol to the letter. In particular, the manipulator might initially set $r_{i^m}^j = x_{i^m} = n$. We refer to this strategy as *Strategy 1.* Clearly, for all i, $\frac{s_{t,i}}{w_{t,i}}$ eventually converges to a value that is at least 1.

Despite the apparent effectiveness of Strategy 1, it is very easily detected. Indeed, unless for all $i \neq i^m$ it holds that $s_{t_0,i} = 0$ at the time t_0 when the manipulator deviated by assigning $s_{t_0,i^m} = n$, the expressions $\frac{s_{t,i}}{w_{t,i}}$ would eventually converge to a value that is strictly greater than 1; this would clearly unmask the deceit. It is of course possible to update s_{t_0,i^m} to be less than n, but it is difficult to determine *a priori* which value to set without pushing the average reputation above 1.

We now consider a more subtle way to increase the values $\frac{s_{t,i}}{w_{t,i}}$, a deceit which is indeed difficult to detect; we call this strategy *Strategy 2.* For the first T stages of the algorithm, the manipulator i^m follows PUSH-SUM as usual, with the exception of the updates of s_{t,i^m}: after updating $w_{t,i^m} = \sum_l \hat{w}_l$ (as usual), i^m updates: $s_{t,i^m} = w_{t,i^m}$. In other words, the manipulator sets his personal

[6] In fact, this holds for every coalition that does not constitute a sizable portion of the entire set of agents.

evaluation of the average $\frac{s_{t,i^m}}{w_{t,i^m}}$ to be 1 at every stage $t = 1, \ldots, T$. For time $t > T$, the manipulator abides by the protocol. Using this strategy, it always holds that $\frac{s_{t,i}}{w_{t,i}} \leq 1$ for all i. In addition, for all t, it still holds that $\sum_i w_{t,i} = n$. Therefore, without augmenting the system with additional security measures, this manipulation is difficult to detect. We shall presently demonstrate formally that the manipulation is effective in the long run: $\frac{s_{t,i}}{w_{t,i}}$ converges to 1 for all i.

Proposition 2. *Under Strategy 2, for all $i \in N$, $\frac{s_{2T,i}}{w_{2T,i}} \overset{T\to\infty}{\longrightarrow} 1$ in probability.*

Proof. We first notice that $\sum_i s_{t,i}$ is monotonic increasing in the stage t. Moreover, as noted above, it holds that at every stage, $\sum_i w_{t,i} = n$, as for all $i \in N$: $\frac{s_{t,i}}{w_{t,i}} \leq 1$, and thus:

$$\sum_i s_{t,i} \leq \sum_i w_{t,i} = n.$$

Let $\epsilon, \delta > 0$. We must show that it is possible to choose T large enough such that for all $t \geq 2T$ and all $i \in N$, $\Pr[\frac{s_{t,i}}{w_{t,i}} \geq 1 - \epsilon] \geq 1 - \delta$.

Assume that at time t it holds that:

$$\frac{\sum_i s_{t,i}}{n} < 1 - \epsilon/2. \tag{1}$$

Let $I_t = \{i \in N : \frac{s_{t,i}}{w_{t,i}} \geq 1 - \epsilon/4\}$, $w(I_t) = \sum_{i \in I_t} w_{t,i_t}$. It holds that:

$$\begin{aligned} n(1 - \epsilon/2) &\geq \sum_{i \in N} s_{t,i} \\ &\geq \sum_{i \in I_t} s_{t,i} \\ &\geq \sum_{i \in I_t} w_{t,i} \cdot (1 - \epsilon/4) \\ &= w(I_t)(1 - \epsilon/4) \end{aligned}$$

It follows that $w(I_t) \leq n \cdot \frac{1-\epsilon/2}{1-\epsilon/4}$. The total weight of agents in $N \setminus I_t$ is at least $n - w(I_t)$. There must be an agent $i_t \in N \setminus I_t$ with at least a $1/n$-fraction of this weight:

$$w_{t,i_t} \geq \frac{n - w(I_t)}{n} \geq \frac{\epsilon}{4 - \epsilon}. \tag{2}$$

In order for the choice of i_t to be well-defined, assume i_t is the minimal index that satisfies Equation (2).

Now, let s'_{t,i^m} be the manipulator's sum had it updated it according to the protocol, i.e., $s'_{t,i^m} = \sum_l \hat{s}_l$ for all messages l sent to i^m. With probability $1/n$ (and independently of other stages), $f_t(i_t) = i^m$; if this happens, it holds that:

$$\begin{aligned} s'_{t+1,i^m} &\leq (w_{t+1,i^m} - 1/2 \cdot w_{t,i_t}) + 1/2 \cdot s_{t,i_t} \\ &\leq (w_{t+1,i^m} - 1/2 \cdot w_{t,i_t}) + 1/2 \cdot w_{t,i_t} \cdot (1 - \epsilon/4). \end{aligned} \tag{3}$$

For all stages t it holds that $\sum_i s_{t+1,i} - \sum_i s_{t,i} = s_{t+1,i^m} - s'_{t+1,i^m}$, as the manipulator is the only agent that might change $\sum_i s_{t,i}$. Therefore, in the conditions of Equation (3),

$$\begin{aligned}\sum_i s_{t+1,i} - \sum_i s_{t,i} &= s_{t+1,i^m} - s'_{t+1,i^m}\\ &= w_{t+1,i^m} - s_{t+1,i^m}\\ &\geq 1/2 \cdot w_{t,i_t} \cdot \frac{\epsilon}{4}\\ &\geq \frac{\epsilon^2}{32 - 8\epsilon}\\ &= \Delta(w)\end{aligned}$$

So far, we have shown that for each stage t where Equation (1) holds and $f_t(i_t) = i^m$, it is the case that $\sum_i s_{t+1,i} - \sum_i s_{t,i} \geq \Delta(w)$. This can happen at most $\frac{n(1-\epsilon/2)}{\Delta(w)}$ times before Equation (1) no longer holds, or to put it differently, before $\frac{\sum_i s_{t,i}}{n} \geq 1 - \epsilon/2$.

Let X_t be i.i.d. binary random variables, which are 1 iff $f_t(i_t) = i^m$. It holds that for all t where Equation (1) is true, $\mathbb{E}[X_t] = 1/n$. By Chernoff's inequality, it holds that:

$$\Pr[\frac{1}{T_1}\sum_{t=1}^{T_1} X_t \leq \frac{1}{2n}] \leq e^{-\frac{T_1}{2n^2}}.$$

It is possible to choose T_1 to be large enough such that this expression is at most $\delta/2$, and in addition $\frac{1}{2n} \cdot T_1 \geq \frac{n(1-\epsilon/2)}{\Delta(w)}$. Therefore, at time T_1, the average $\frac{\sum_i S_{T_1,i}}{n} \geq 1 - \epsilon/2$ with probability $1 - \delta/2$.

Recall that after T stages (where i^m deviated from the protocol), it still holds that $\sum_i w_{T,i} = n$. Assume that indeed $\frac{\sum_i S_{T_1,i}}{n} \geq 1 - \epsilon/2$. By modifying the proof of Theorem 3.1 from [15], it is possible to show that after another $T_2 = T_2(n, \delta, \epsilon)$ stages where all agents observe the protocol, it holds with probability $1 - \delta/2$ that for all i, $\left|\frac{s_{T_1+T_2,i}}{w_{T_1+T_2,i}} - \frac{\sum_i S_{T_1,i}}{n}\right| < \epsilon/2$, and thus for all i and $t \geq T_1 + T_2$: $\frac{s_{t,i}}{w_{t,i}} > 1 - \epsilon$ with probability $1 - \delta$.

The proof is completed by simply choosing $T = \max\{T_1, T_2\}$. □

Proposition 2 implies that Strategy 2 poses a provably acute problem, when Push-Sum is run a large number of turns. Fortunately, Push-Sum converges exponentially fast, and thus it is usually the case that the manipulator is not able to significantly affect the average reputation, as the following proposition demonstrates.

Proposition 3. *Let $T_1 \leq T$. Under Strategy 2,* $\mathbb{E}\left[\frac{\sum_i S_{T_1,i}}{n} - \bar{r}^j\right] \leq \frac{T_1}{2n}$.

Proof. Let $\{\hat{s}_l, \hat{w}_l\}$ be the messages which the manipulator received at time $t+1$. The manipulator sets $s_{t+1,i^m} = w_{t+1,i^m} = \sum_l \hat{w}_l$. Essentially, this is equivalent to setting for all l $\hat{s}_l = \hat{w}_l$, or in other words, raising each $\hat{s}_l$ by $\hat{w}_l - \hat{s}_l$. At turn t it was already true that $s_{t,i^m} = w_{t,i^m}$ (w.l.o.g. this is also true for $t = 0$), so it is enough to consider messages at time t from all $i \neq i^m$.

Therefore, for all stages t, it holds that:

$$\begin{aligned}
\mathbb{E}\left[\sum_i s_{t+1,i} - \sum_i s_{t,i}\right] &= \sum_{i \neq i^m} \left(\Pr[f_t(i) = i^m] \cdot (\frac{1}{2} w_{t,i} - \frac{1}{2} s_{t,i})\right) \\
&= \frac{1}{2n} \sum_{i \neq i^m} (w_{t,i} - s_{t,i}) \\
&\leq \frac{1}{2n} \sum_{i \neq i^m} w_{t,i} \\
&\leq \frac{1}{2n} \sum_{i \in N} w_{t,i} \\
&= \frac{1}{2}.
\end{aligned}$$

The last equality follows from the fact that for all t, $\sum_i w_{t,i} = n$.

As $\bar{r}^j = \frac{\sum_i s_{0,i}}{n}$, and from the linearity of expectation, we obtain that

$$\begin{aligned}
\mathbb{E}\left[\frac{\sum_i s_{T_1,i}}{n} - \bar{r}^j\right] &= \frac{1}{n}\mathbb{E}\left[\sum_{t=0}^{T_1-1}\left(\sum_i s_{t+1,i} - \sum_i s_{t,i}\right)\right] \\
&= \frac{1}{n}\sum_{t=0}^{T_1-1}\mathbb{E}\left[\sum_i s_{t+1,i} - \sum_i s_{t,i}\right] \\
&\leq \frac{1}{n} T_1 \cdot \frac{1}{2}.
\end{aligned}$$

□

In particular, since $U(n, \delta, \epsilon) = O(\log n + \log \frac{1}{\delta} + \log \frac{1}{\epsilon})$, Push-Sum is executed $O(\log n)$ stages, and thus the difference in the average is at most $O(\frac{\log n}{n})$, which is quite insubstantial.

Remark 5. It is not guaranteed at time T_1 that each $\frac{s_{t,i}}{w_{t,i}}$ is close to $\bar{r}^j$, because the inputs were dynamically changed during the execution of Push-Sum.

Remark 6. The above discussion focused on a setting where the manipulator attempts to increase the average reputation of an agent. It is likewise possible for a manipulator to decrease an agent's average reputation, or indeed set it eventually to any value it wants.

Remark 7. Jelasity, Montreso and Babaoglu [14] propose a general method to completely prevent malicious agents from deviating in gossip-based algorithms, by augmenting the protocol with exchange of certificates. However, the authors describe their approach in a very general manner, and so this approach was not implemented here.

6 Related Work

The main focus of research on trust and reputation systems has been on the semantic aspects of these systems, and their effect on social welfare. Previous work has highlighted the advantages of reputation systems in overcoming social pitfalls in several domains. Akerlof [4], for instance, has considered markets where information asymmetry exists between buyers and sellers, in the sense that buyers can only guess the quality of goods; in such a setting, a reputation system can improve social welfare.

Several works have analyzed manipulations of general reputation mechanisms. Friedman and Resnick [12] have discussed the effects of *cheap pseudonyms.* When agents can enter the system using pseudonyms, and the cost of recreating an identity is cheap, agents who have a stained reputation may easily shed it. The authors have considered several solutions to this problem: disallowing anonymity, entry fees (which make pseudonyms more expensive), and using a central authority for irreplaceable ("once-in-a-lifetime") pseudonyms. However, each approach has major drawbacks. Dellarocas [9] has studied a setting where agents manipulate a reputation system by providing unfair ratings to some of their peers, and suggests several solutions.

The next few paragraphs survey previous work on distributed reputation systems. An early work is that of Abdul-Rahman and Hailes [1], which relied on the results of Marsh [17] to design a model of trust in online environments. In this framework, each agent must maintain and update large data structures, which contain knowledge about the entire system. Updating this data may be inefficient, and in particular it is not certain that the scheme scales well when the number of agents grows.

P2PRep [6] and Xrep [8] are P2P reputation systems that can be piggybacked on existing P2P protocols (such as Gnutella). P2PRep allows peers to estimate trustworthiness of other peers by polling; XRep takes another step forward: each peer keeps trust evaluations both of other peers and of resources. No guarantees are given with respect to computational efficiency and scalability.

Aberer and Despotovic [3] introduce a reputation system that consists of both a semantic model and a data management scheme. The latter relies on P-Grid [2], and uses distributed data structures for storing trust information; the associated algorithms scale gracefully as the number of agents increases. In addition, a limited resistance to manipulation and failure is achieved through replication of data. This approach suffers from several shortcomings compared to ours. Agents in this scheme assess others' reputation only on the basis of complaints filed in the past; the framework is generally limited to such binary trust information. In addition, trust is evaluated only according to referrals from neighbors, whereas in our scheme the evaluation is based on all the information in the system.

Xiong and Liu [20] introduced a sophisticated framework specifically applicable in peer-to-peer networks, where the decision whether to trust a peer is based on five metrics: satisfaction, number of transactions, credibility of feedback, transaction context, and community context. This work was extended in [19]. Both papers concentrate on the trust model, and generally do not elaborate on the data management scheme. Specifically, in [20] a P-Grid [2] is used. Therefore, this work is in a sense orthogonal but complementary to ours. Dewan and Dasgupta [11] propose self-certification and IP-Based safeguards as ways of inducing trust; this work also complements ours.

Finally, gossip-based algorithms[7] have many applications in other domains, for instance replicated database maintenance [10].

7 Conclusions and Future Research

We have presented a data management scheme which is based on gossip-based algorithms, and have demonstrated that it possesses the following features:

- Decentralization: no central database, and further, applicability in networks where point-to-point communication cannot be assumed.
- Scalability: the time to evaluate an agent's average reputation with confidence $1 - \delta$ and accuracy ϵ is $O(\log n + \log \frac{1}{\delta} + \log \frac{1}{\epsilon})$.
- Robustness to failure.
- Globality: evaluation of trust is based on all relevant information in the system, rather than local information.
- Extremely simple data structures: each agent merely keeps an assessment of the agents with which it personally interacted.
- Motivates absolute truthfulness, as the time to close deals may decrease as reputation increases.
- Resistance to some attacks, such as carefully tampering with the updates performed by Push-Sum.

We have focused on the data management scheme, and have largely ignored the trust model (with the exception of Section 4). However, we believe that many existing trust models can be integrated with our framework. A very simple example is the binary trust model of [3], where agents can file complaints against other agents. In our framework, each agent i sets its value r_i^j to be 0 if it wishes to file a complaint against j; otherwise, the value is 1. More sophisticated models may require tweaks in the framework. Consider the trust model presented in [20], where five factors are taken into account. Three of the factors mentioned simply determine the way an agent updates its own values r_j^i, and our framework of course supports any update formula. The "number of transactions" factor is already taken into account, as we compute the average reputation. The "credibility of feedback" factor requires a small change: given credibility ratings c_i for agents, a weighted average can be computed; the initial inputs of agents are $x_i = s_{0,i} = c_i \cdot r_i^j$, and their weights are $w_{0,i} = c_i$.

An interesting direction for future research is augmenting the framework with an option to efficiently choose among *service providers*:[8] agent i requires a specific service, and there are m other agents that offer to service i's request. Agents

[7] Also called *epidemic algorithms.*

[8] In a sense similar to [6].

can be matched to service providers using a *matchmaking* service, but this problem has been dealt with [7]. We are concerned with the following question: once an agent is given a list M of m-service providers, which one should it choose? The obvious answer is, the one with the highest reputation: $\text{argmax}_{j \in M} \bar{r}^j$. However, as the size of M may approach n, it is difficult to estimate the reputation of all agents in M.

One possible solution is to hold an election: the voters are the n agents, and the candidates are the m service providers. It is possible, for instance, to determine the winner using the simple *plurality* rule: each agent votes for one candidate; the winner is the candidate that secured the largest number of votes. It is also possible to resolve this election in our framework, using Push-Sum. Denote $M = \{j_1, j_2, \ldots, j_m\}$. The input x_i of each agent for Push-Sum is (conceptually) a base $n+1$ number with m coordinates; the l'th coordinate of x_i is n if i votes for j_l, and 0 otherwise. The average is calculated by using Push-Sum with an *absolute* error $\epsilon = 1/3$, and some confidence $1 - \delta > 0$ (the x_i are translated to base 10). After the average $\frac{s_{t,i}}{w_{t,i}}$ is calculated, the result is rounded to the nearest integer and again translated to base $n+1$; the agent j_l such that the $l'th$ coordinate is largest wins the election.

Unfortunately, since in this case the inputs x_i are large, obtaining such a small absolute error requires a large number of iterations of Push-Sum, and furthermore, the message size is large. Is there a way of holding an election (using some other voting rule, perhaps) in our framework in a way that scales well with respect to both the running time and message size?

A different direction is using our framework to prevent attacks based on cheap pseudonyms [12]. This is made possible due to the fast aggregation of information and its globality. If an agent cheats, *all* other agents will soon know. It is possible to restrict newcomers, with an unestablished reputation, to only one transaction in every period of length $O(\log n)$. This way, each identity would be good for only a single deceitful transaction, since after the period is over, the information could have already been obtained by any agent. Granted, it would still be possible to shed stained identities, but the flow of transactions a cheater would be able to complete would be severely diminished.

8 Acknowledgment

This work was partially supported by grant #039-7582 from the Israel Science Foundation.

References

1. A. Abdul-Rahman and S. Hailes. Supporting trust in virtual communities. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*, 2000.
2. K. Aberer. P-grid: A self-organizing access structure for P2P information systems. In *Proceedings of the 9th International Conference on Cooperative Information Systems*, pages 179–194, 2001.

3. K. Aberer and Z. Despotovic. Managing trust in a peer-2-peer information system. In *Proceedings of the Tenth International Conference on Information and Knowledge Management*, pages 310–317, 2001.
4. G. A. Akerlof. The market for lemons: qualitative uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84:488–500, 1970.
5. R. Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.
6. F. Cornelli, E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, and P. Samarati. Choosing reputable servants in a P2P network. In *Proceedings of the 11th International World Wide Web Conference*, pages 376–386, 2002.
7. G. Cybenko and G. Jiang. Matching conflicts: Functional validation of agents. In *Proceedings of the AAAI Workshop on Agent Conflicts*, pages 14–19, 1999.
8. E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, P. Samarati, and F. Violante. A reputation-based approach for choosing reliable resources in peer-to-peer networks. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, pages 207–216, 2002.
9. C. Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *Proceedings of the 2nd ACM Conference on Electronic Commerce*, pages 150–157, 2000.
10. A. J. Demers, D. H. Greene, C. Hauser, Wes Irish, and John Larson. Epidemic algorithms for replicated database maintenance. In *Proceedings of the Sixth Annual ACM Symposium on Principles of Distributed Computing*, pages 1–12, 1987.
11. Prashant Dewan. Peer-to-peer reputations. In *Proceedings of the 18th International Parallel and Distributed Processing Symposium*, 2004.
12. E. J. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(2):173–199, 2001.
13. A. M. Frieze and G. R. Grimmett. The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, 10:57–77, 1985.
14. M. Jelasity, A. Montresor, and O. Babaoglu. Towards secure epidemics: Detection and removal of malicious peers in epidemic-style protocols. Technical Report UBLCS-2003-14, Department of Computer Science, University of Bologna, 2003.
15. D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 482–491, 2003.
16. P. Keyani, B. Larson, and M. Senthil. Peer pressure: Distributed recovery from attacks in peer-to-peer systems. In *Proceedings of the International Workshop on Peer-to-Peer Computing*, pages 306–320, 2002.
17. S. P. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling, Apr 1994.
18. G. Pandurangan, P. Raghavan, and E. Upfal. Building low diameter P2P networks. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 492–499, 2001.
19. M. Srivatsa, L. Xiong, and L. Liu. Trustguard: Countering vulnerabilities in reputation management for decentralized overlay networks. In *Proceedings of the 14th International World Wide Web Conference*, pages 422–431, 2005.
20. L. Xiong and L. Liu. A reputation-based trust model for peer-to-peer ecommerce communities. In *Proceedings of the 4th ACM Conference on Electronic Commerce*, pages 228–229, 2003.