

# An Incentives' Mechanism Promoting Truthful Feedback in Peer-to-Peer Systems\*

Thanasis G. Papaioannou and George D. Stamoulis

Department of Informatics, Athens University of Economics and Business (AUEB),  
76 Patision Str., 10434 Athens, Greece.  
{pathan, gstamoul}@aueb.gr

## Abstract

*We propose a mechanism for providing the incentives for reporting truthful feedback in a peer-to-peer system for exchanging services. This mechanism is to complement reputation mechanisms that employ ratings' feedback on the various transactions in order to provide incentives to peers for offering better services to others. Under our approach, both transacting peers (rather than just the client) submit ratings on performance of their mutual transaction. If these are in disagreement, then both transacting peers are punished, since such an occasion is a sign that one of them is lying. The severity of each peer's punishment is determined by his corresponding non-credibility metric; this is maintained by the mechanism and evolves according to the peer's record. When under punishment, a peer is not allowed to transact with others. We present the results of a multitude of experiments of dynamically evolving peer-to-peer systems. The results show clearly that our mechanism detects and isolates effectively liar peers, while rendering lying costly. Also, our mechanism diminishes the efficiency losses induced to sincere peers by the presence of large subsets of the population of peers that provide their ratings either falsely or according to various unfair strategies. Finally, we explain how our approach can be implemented in practical cases of peer-to-peer systems.*

## 1. Introduction

Peer-to-peer systems have recently become very popular as environments for exchanging services, i.e. files, storage capacity, etc. If there is no accounting of information about who is offering what to whom in such systems, then peers have the opportunity for free-riding, and for providing malicious services or services of unacceptably low quality. Due to this asymmetry in information among transacting peers, the risk for a peer of

placing much individual effort and receiving much less in return is high. Reputation on the basis of ratings can be a proper means for achieving accountability, since it reveals hidden information regarding the inherent quality and the behavior (i.e. *performance*) of peers [1], [2]. Reputation-based policies [1] determine the pairs of peers eligible to transact with respect to reputation. When such policies are employed the total value generated within the system is shared to peers according to their performance, thus, providing the right incentives to peers for offering services of high quality. However, reputation mechanisms are vulnerable to *false* or *strategic voting* (*rating*). For example, a particular peer may benefit by submitting unjustified positive ratings for his friends and/or negative ratings for his competitors. This problem is further augmented in case of pseudo-spoofing, i.e. use of multiple false identities, which may appear in a peer-to-peer system. In this paper, we deal with the issue of *credibility*. Many reputation systems deal with this issue together with performance [3], [4], [5]. Such an approach provides peers with the incentive for employing various malicious strategies; e.g. an adversary peer may obtain a high reputation by offering services of high performance and exploit it as a rater to demote his competitors or to promote his colleagues. Moreover, poor performance and lying are not necessarily related; e.g. poor performance may be inherent for a peer due to his limited resources.

In our approach, we deal with credibility *separately* from performance. In particular, we propose a proper mechanism for promoting *truthful reporting* of feedback information. This mechanism detects and penalizes peers that lie. A non-credibility value as well as a punishment state is maintained for each peer. We experimentally justify that our mechanism deals successfully with large fractions of liars in the peer-to-peer system (even if they are collaborated in order to gain unfair advantage) and practically expels them. Moreover, we show that the credibility mechanism can be combined very effectively

---

\* The present work has been carried out as a part of the IST project MMAPPS (IST-2001-34201) funded by the EU.

with a reputation system for performance, thus providing a complete and practically implementable solution for accountability in peer-to-peer environments. Our experiments reveal that this combination results in comparable efficiency for sincere peers, as when no liar peers exist in the system. The mechanism provides peers with the *right incentives* for truthful reporting of feedback information, as sincere peers receive *always* more benefit from the peer-to-peer system than liar peers, whose benefit is very low. Thus, the credibility mechanism is *strategyproof*.

## 2. The Credibility Mechanism

Consider a peer-to-peer system for exchanging services that employs a distributed reputation system for performance. The client peer, after a transaction, sends feedback rating his offered performance. For example, he may rate the transaction as “successful” (i.e. high offered performance) or as “unsuccessful” (i.e. low offered performance). Take that votes are aggregated into reputation values using the Beta aggregation rule [6]. That is, each peer’s reputation equals the fraction of the “weighted number” of his successful service provisions over the “total weighted number” of his service provisions, with the weight of each service provision being a negative exponential function of the elapsed time. The feedback messages are useful only if their content is *true*. Unfortunately, peers actually have the incentive of strategic rating of others’ performance, since they can thus hide their poor performance, improve their reputation, and possibly take advantage of others. Thus, a proper mechanism should make lying costly or at least unprofitable. “Punishing liars” is a known recipe [7], [8], but two questions arise: How can lying peers be discovered? How can they be punished in a peer-to-peer system, where there is no central control?

Under our approach peers submit ratings’ feedback according to the following rules: i) after a transaction, *both* peers involved have to send one feedback message each, and ii) besides voting the transaction as successful or not each feedback message *also* contains a quantifiable performance metric, e.g. the number of transferred bytes of useful content. We assume that the observed performance is with high probability the same with that actually offered. (The opposite may only occur due to unexpected events during a transaction like network congestion etc.) Thus, if feedback messages for a transaction *disagree* (either in their performance metric or in their vote), then, with high probability, at least one of the transacted peers is lying and has to be somehow “punished”, in order for the right incentives to be provided. However, the system cannot tell which of the peers does lie, and consequently whom to believe and

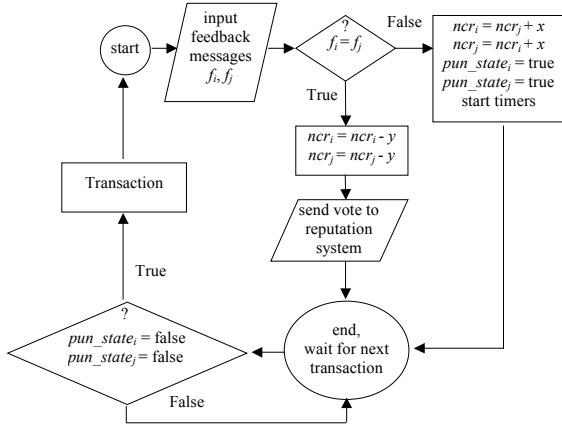
whom to punish. Thus, according to our approach, *both* peers are punished in this case. This idea was initially introduced in [7]. However, by simply applying it, a sincere peer is often punished unfairly.

Therefore, we need a complete mechanism specifying how to punish peers in such an uncontrolled system and how to limit potential unfairness. To this end, we introduce for each peer: i) the *non-credibility* metric  $ncr$ , which corresponds to reputation for non-credibility, and ii) a binary *punishment state* variable, declaring whether the peer is “under punishment” (if the variable is “true”) or not (if the variable is “false”). For each peer, both  $ncr$  and punishment state are public information, they are appropriately stored so that they are available to other peers (see Section 5 for practical implementation details). Upon entering the peer-to-peer system, each peer is assigned a moderately high initial non-credibility value  $ncr_0$ , while he is not under punishment. (Note that the lower  $ncr$  the better.) This choice of  $ncr_0$  is motivated later. The flowchart of the credibility mechanism is depicted in Figure 1. In particular, after a transaction between two not punished peers  $i, j$  their feedback messages  $f_i, f_j$  are sent as input to the mechanism: Upon *disagreement* (i.e. if  $f_i \neq f_j$ ), the non-credibility values of the transacted peers are both increased by  $x$  while both get punished. The duration of a peer’s punishment equals  $b^{ncr}$ , i.e. is exponential in his non-credibility, with a base  $b > 1$ . Upon *agreement* (i.e. if  $f_i = f_j$ ), the non-credibility values of the transacted peers are decreased (i.e. improved) by  $y$ , where  $0 < y < x$ , without ever dropping below 0. The common feedback is forwarded to the system computing reputation for performance.

Decrease of non-credibility in cases of agreement serves as a rehabilitation mechanism. This is crucial for the efficient operation of the credibility mechanism, because, as already mentioned, upon disagreement in reports, most probably one peer is unfairly punished. The ratio  $x:y$  determines the speed of restoring a non-credible reporting behavior. We employ additive increase/decrease of the non-credibility values for simplicity. Other approaches such as additive increase/multiplicative decrease are also possible.

Punishing peers is not an easy task to employ in the absence of any control mechanism, particularly if peers have full control over their part of peer-to-peer middleware. In our mechanism, a punishment amounts to loss of value offered by other peers. That is, a peer under punishment does *not* transact with others during his punishment period, while his ratings for such transactions are *not* taken into account. The latter measure provides incentives for peers to abide with the former one! Indeed, first, note that sincere peers under punishment are not expected to be willing to offer services as they would be subject to strategic voting without being able to disagree.

On the other hand, liar punished peers collaborated with other liar peers that strategically vote them (i.e. always positively) can raise their reputation anyway, thus having no incentives to perform well during their punishment. Thus, no peer has any incentives to *ask* for services from a punished peer except for strategic voting. Moreover, no peer has any incentive to perform well when offering services to a punished peer, because the corresponding feedback is not taken into account. Therefore, it is beneficial for the system to *prohibit* transaction with punished peers by rule. To this end, if a peer transacts with a punished one, then both of the transacting peers are punished as if they were involved in a new disagreement. Thus, the non-credibility value of a peer remains unchanged during his punishment period unless he transacts with other peers; in such a case it is further increased.



**Figure 1. The credibility mechanism.**

Peers should have the incentive to submit feedback, despite the risk of disagreement and subsequent punishment. Indeed, after a transaction that failed peers may not be willing to report the failure at all. Thus, to provide peers with the incentive to submit their feedback, our mechanism punishes both peers involved in a transaction if only one of them submits feedback. This also prevents unilateral submission of feedback messages for non-existing transactions. Note also that, since the proposed mechanism improves the long-term efficiency of the sincere peers, only liar peers are expected to have incentives to avoid submitting feedback. Yet applying the reasoning of [9] to our case, we expect that under certain circumstances, the existence of our mechanism will lead liar peers to give up their strategic behavior.

### 3. The Complete Reputation Mechanism

As explained in [1], regardless of the accuracy of reputation values' calculation, appropriate reputation-based policies should be employed in a peer-to-peer system. These determine the pairs of peers eligible to

transact, and provide the right incentives for peers regarding to perform well. Otherwise, the benefit obtained by high performing peers from the peer-to-peer system can be very low. A simple yet effective reputation-based policy, referred to as “Max-Max” [1], prescribes the following: i) peers select to be served by the highest reputed peer among the providing ones that offer the requested service and ii) peers select to serve the highest reputed peer among requesting ones that content for a particular service by them. Policies such as “Max-Max” either rely on the assumption of truthful reporting or can perform much more efficiently under such circumstances. Thus, it is highly beneficial to *combine* them with our proposed credibility mechanism for providing peers with incentives for truthful reporting.

The complete reputation mechanism operates as follows: After a transaction both peers send feedback about the offered performance to the reputation system, which aggregates it into reputation values, as described in Section 2. Ratings are taken into account in the calculation of the reputation values for performance only if the transacted peers agree on their evaluations for the performance of the service offered in their transaction. Upon disagreement in their evaluations, their credibility is diminished and they are punished, as specified in Section 2. Then, the reputation-based policy determines for each not punished peer the ones eligible for his next transaction. Experimental results, which are presented in Section 5, clearly show that combining our credibility mechanism with a reputation-based policy is capable of providing the right incentives *both* for truthful reporting and for high performance in providing services in peer-to-peer systems.

## 4. Experimental Results

### 4.1 The Model

We consider a peer-to-peer system where services of a certain kind are exchanged among peers. Similarly, with other articles [2], [10], [11], we assume that there are two types of peers with different performance in this system: *altruistic* and *egotistic*. Each peer exhibits (either inherently or intentionally) a mixed strategy regarding his performance in his service provisions; this strategy depends on the peer's type. In particular, each altruistic (resp. egotistic) peer provides a service successfully with a high probability  $\alpha=0.9$  (resp. with a low probability  $\beta=0.1$ ). Different service provisions by the same peer are taken as independent. At the same time, each peer exhibits a reporting strategy regarding the sincerity of his feedback: he is either (*always*) *sincere* or *liar*. The lying strategies considered are defined in Subsection 4.2. All liars follow the same such strategies. The performance

and the reporting types of each peer are *private* information, i.e. only the peer himself knows them. Furthermore, the population of peers is assumed to be *renewed* according to a Poisson process with mean rate  $\lambda=10$  peers/time slot, while the total size  $N$  of the population is kept constant, with  $N=1500$ . That is, each peer is assumed to live in the peer-to-peer system for a period determined according to the exponential distribution with mean  $N/\lambda$ . When a peer leaves the system, a new entrant of the same type takes his place. To make matters worse, the vast majority of peers (90%) are taken to be egotistic. The percentage of liar peers in each experiment varies. In fact, for each lying strategy, we present the results for the *maximum* such percentage that can be dealt with effectively by our mechanism.

Time is assumed to be slotted. The duration of the time slot is of the same order of magnitude as the average interval between two successive service requests. At each slot, every peer requests a service with a certain probability  $r=0.5$ . Service availability is Zipf-distributed. That is, assuming that services are ranked with respect to their popularity, a service with rank  $z$  is found at a certain peer with probability  $z^{-1}$ . A peer can serve only one peer per slot due to his limited resources.

After a transaction each of the peers involved sends feedback to the reputation system as explained in Section 2. (Votes are converted into reputation values using the Beta aggregation rule.) The reputation value for a peer is associated to his pseudonym, and expresses his probability of offering high performance given his past record. The Max-Max reputation-based policy described in Section 3 is employed. Employing other policies of [1] was seen to have similar effects. The peer-to-peer system is considered noiseless in the sense that the outcome of a transaction depends only on the performance of the providing peer in this transaction. A peer is assigned a low initial reputation  $h_0$  (i.e.,  $h_0=0.1$ ), in order to limit the incentive for name changes. That is, if  $h_0$  were high, then each peer would later have the incentive to drop his pseudonym and obtain a new one, thus clearing his past low-performance record.

The proposed credibility mechanism is employed too. Each peer is assigned an initial non-credibility value  $ncr_0$  that characterizes him as non-credible (i.e.  $ncr_0=6$ ), thus further limiting the incentive for name changes. The non-credibility values are increased upon disagreement with his transacted peer in their feedback by  $x=1$  and decreased upon agreement by  $y=0.5$ . The best possible non-credibility of a peer is 0. The duration of a peer's punishment equals  $2^{ncr}$ , where  $ncr$  is his non-credibility value upon punishment.

In the experiments conducted, we assess the *efficiency* attained in this peer-to-peer system when the credibility mechanism is employed, which is measured as the

number of successfully offered services per peer type. Particular emphasis is placed on the efficiency of sincere altruistic peers, as such peers offer most of the value to the peer-to-peer system. We also assess the *incentives* offered per type of peer for truthful reporting.

## 4.2 Lying Strategies

Depending on their objectives, liars follow various strategies for manipulating their ratings. We considered four possible lying strategies, some of which are similar with those in other related works [4], [12], [13]:

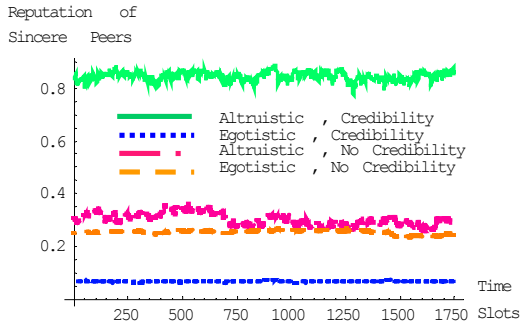
- *Destructive*, in which liar peers reverse the feedback on the outcome of their transactions.
- *Opportunistic*, in which liar peers claim that they always succeed in their transactions and that all other peers not collaborated with them fail.
- *Mixed*, in which a liar peer randomly selects which of the above lying strategies to employ. The selection probability may vary with time.
- *Discriminating*, in which a liar peer apart from being opportunistic, only serves peers collaborated with him, thus bypassing the Max-Max policy.

Note that *collaborated* liar peers always rate positively each other. In each experiment either all liar peers are collaborated or they all act autonomously. In the following subsection, we show that how the credibility mechanism deals with these lying strategies effectively.

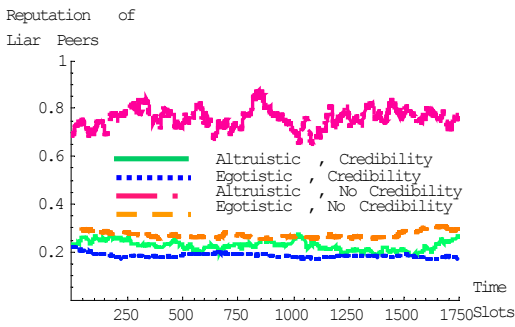
## 4.3 Effectiveness of the Mechanism

Initially, liar peers are assumed to be *collaborated*, to follow the *destructive* lying strategy, and to constitute 45% of the population of the peer-to-peer system. In all of the experimental results to follow, we omit an initial "bootstrapping" period of operation of the peer-to-peer system in the beginning of which all peers are newcomers. (This period lasts for 250 slots; in general its duration depends on various parameters, but mainly on the service request probability.) We assess the efficiency of peers during the normal operation of the peer-to-peer system, of course with dynamically renewed population. In Figure 2a, depicted are the mean reputation values of sincere peers, which are very accurate when the credibility mechanism is employed. Indeed, the values for altruistic (resp. egotistic) peers are very close to the corresponding a priori probability for successful service provision  $\alpha=0.9$  (resp.  $\beta=0.1$ ). On the contrary, if the mechanism is not employed, then the two performance types cannot be distinguished by means of their reputations. On the other hand, as depicted in Figure 2b, the mean reputation values of liar peers are very low when the credibility mechanism is employed; liar peers are mostly under punishment during their lifetime (see

below discussion on Figure 4) and as a result they receive too few votes. Also note that altruistic liar peers benefit from the absence of the credibility mechanism as opposed to altruistic sincere ones! Therefore, peers have wrong incentives if our mechanism is absent.



(a)



(b)

**Figure 2. (a) Reputation values for sincere peers: their accuracy is greatly improved by employing the credibility mechanism. (b) The reputation values of liar peers are also affected.**

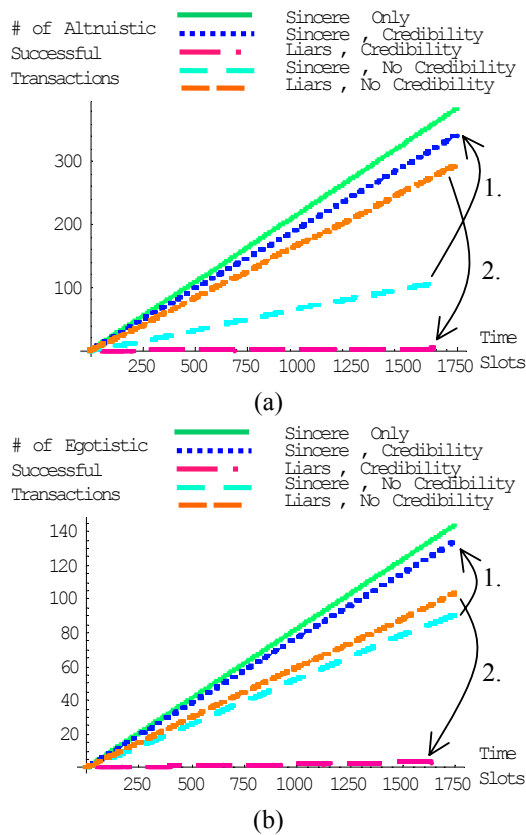
Next, we deal with the efficiency issues for the same set of experiments. The number of total successful transactions per peer (i.e. efficiency) increases for both altruistic and egotistic sincere peers when the credibility mechanism is employed, as depicted by arrow 1 in Figure 3(a) and (b) respectively. On the contrary, when the credibility mechanism is employed, the efficiency of liar peers (which was greater than that of sincere ones) becomes now almost zero as depicted by arrow 2. Also, when the credibility mechanism is employed, the efficiency achieved by sincere peers in the presence of liars is very close to that achieved in the *ideal case* where no liar peers are present in the peer-to-peer system. In fact, after time slot 250, the relative difference between the two topmost lines in Figure 3(a) is always very close to 10%. The same conclusion also applies to egotistic sincere peers, whose efficiency is of course considerably lower than that of altruistic sincere ones. That is,

credibility mechanism enables the right operation of reputation for performance. Therefore, when our credibility mechanism is employed, the disturbance of sincere peers by liars is minimal. Introduction of the mechanism is very beneficial for sincere peers and very harmful for liar ones, who receive a much lower efficiency than sincere peers. Therefore, the strategy of collaborative destructive lying strategy is *dominated* by the “always be sincere” strategy. Our mechanism provides peers with the right incentives for truthful reporting and is incentive-compatible for sincere peers. On the other hand, liars spend most of their lifetimes being under punishment. This is confirmed by the results depicted in Figure 4. The mean non-credibility values of liars under our credibility mechanism are slightly above the initial value of 6. The mechanism discovers liar peers early and punishes them for a large proportion of their lifetimes, thus suffering only a few punishments. On the other hand, mean non-credibility values of sincere peers are very low. This implies that they recover soon enough from both the initially high non-credibility value as well as from credibility losses due to unfair punishments by the mechanism.

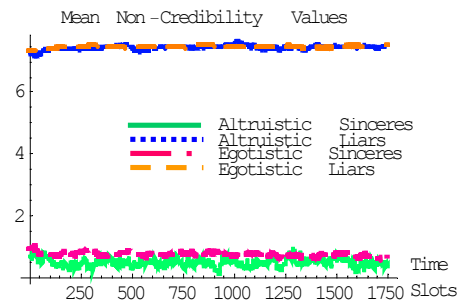
Similar conclusions on efficiency and incentives also apply when liars are collaborated but follow the *opportunistic* lying strategy. The credibility mechanism can deal effectively with as many liars as 41% of the entire population of peers. The effectiveness of our mechanism is also preserved in the case of collaborated liars following the *discriminating* strategy. The fraction of such liar peers that can be dealt with successfully by the mechanism is now lower (namely, 12%), due to the fact that sincere peers cannot transact as clients with liar peers. Thus, a significant number of disagreements in feedback messages is avoided for liar peers. Fortunately, the achievable efficiency of sincere peers is not affected significantly by this lying strategy; only that of liar peers is increased compared to other lying strategies.

Finally, we consider the effectiveness of the credibility mechanism in case of liar peers that follow the *mixed* strategy. In particular, liar peers are taken to constitute 33% of the entire population and to employ the destructive strategy with certain lying probability per transaction, rather than constantly. As depicted in Figure 5, when the credibility mechanism is employed, altruistic sincere peers are provided with more successful services by time slot 1750 than altruistic liars for *any* lying probability. For small values thereof the differences are small, although increasing with the lying probability. Hence, to confirm that the advantage of sincere peers is preserved even in such cases, the experiments were conducted multiple times and average values and confidence intervals were calculated; see Figure 5. Thus, truthful reporting dominates this lying strategy too.

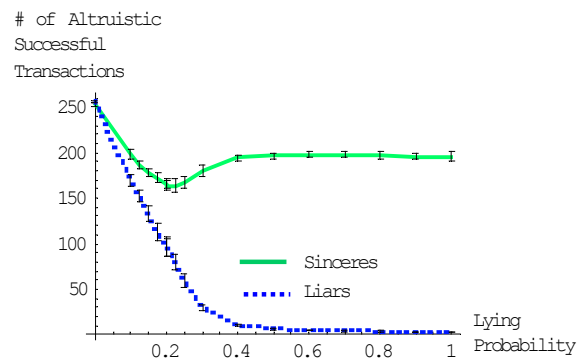
Employing the credibility mechanism is still incentive-compatible for sincere peers. Higher lying probabilities are dealt with more successfully by the credibility mechanism, as in this case liar peers are discovered faster. However, it is fair for peers that lie with a very small probability to receive a benefit by the peer-to-peer system that is close to that of sincere peers. Also notice that, as the lying probability increases, the efficiency of altruistic peers initially decreases, and reaches a minimum when this probability is approximately 25%. Then, the efficiency of altruistic peers increases again until it stabilizes for lying probabilities higher than 40%, for which the efficiency of liars is almost zero.



**Figure 3. The credibility mechanism clearly rewards both altruistic (a) and egotistic (b) peers for their sincerity resulting in almost the same efficiency as in the absence of liar peers.**



**Figure 4. The non-credibility values of sincere and liar peers under the credibility mechanism**



**Figure 5. The credibility mechanism effectively deals with collaborated liar peers that constitute 33% of the population, for all lying probabilities.**

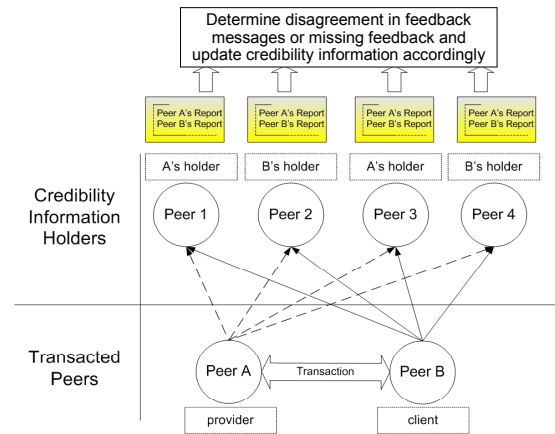
In general, the effectiveness of the mechanism in expelling liars improves for lower population-renewal rates, for all lying strategies considered. This was expected, as at any time the proportion of peers for which the mechanism has already converged to their true non-credibility and reputation values is increased in this case. Also, the effectiveness of the mechanism improves for lower fractions of collaborated liar peers in the system, as the unfair punishments for sincere peers are fewer in this case. Such large fractions of collaborated liar peers, as those considered in the simulation experiments and dealt with successfully by the credibility mechanism, are not expected to emerge in real peer-to-peer systems with large populations. Note also that *no* mechanism can deal effectively with collaborated liars when they are more than 50% of the total population; in such cases the ratings of liars essentially reverse reality. Other experiments that we conducted have revealed that if liar peers are *not* collaborated, then our mechanism can effectively deal with even higher fractions of liars (70% or more of the entire population for the case of the destructive strategy), as liars are punished for disagreement in the feedback messages even when they transact with each other.

## 5. Implementation Issues

We have already demonstrated the effectiveness of our proposed mechanism for promoting credible reporting of feedback in a peer-to-peer system, as well as the right incentives provided thereby. Next, we discuss how this mechanism can be implemented in a completely insecure, anonymous and distributed peer-to-peer environment. The credibility information for each peer has to be efficiently stored and traceable. Authentication, integrity and non-repudiation of the credibility information and the feedback messages are also required. The security issues can be dealt with by means of the public-key infrastructure (PKI). Upon registering in the peer-to-peer system, each peer chooses a public-private key pair and creates his own certificate, which is signed by the system; that is, it is signed by a certain number of peers, as in Pretty Good Privacy (PGP) [14].

Throughout the paper we have assumed that no peers are pre-trusted. Thus, we propose an implementation that does not rely on such a requirement. Peers are assumed to be organized in a hash-indexed structure enabling search of data. Such a structure is already available in systems such as Chord, P-Grid (see [4] and references therein). Peers are required to submit their feedback messages to other peers (referred to as credibility holders) based on their node identifier in the hash-indexed structure and on a number of hash functions employed for this purpose. Each peer is responsible for storing non-credibility values and punishment states of multiple other peers. Thus, multiple peers are responsible for holding credibility information of each fixed peer. After a transaction, each peer sends his feedback message (provider identifier, client identifier, rating and performance metric) and its digest signed by his private key to *all* peers that store credibility information of *both* transacted peers, as depicted in Figure 6. Peers that receive feedback messages verify the sender and the integrity of messages. Then, they detect agreement or disagreement of the feedback messages and compute non-credibility values and update the punishment states of the transacted peers as necessary. If only one feedback message is received, then this is also regarded as a disagreement and both transacted peers are punished. The credibility information is vulnerable to strategic modification by malicious peers. To avoid this, the credibility information provided by the majority of holders can be taken as valid. If there is enough redundancy in storing credibility information, then any malicious modification thereof can be observed by the peer himself. Indeed, the peer can monitor the credibility information about him periodically, by asking the corresponding information holders and comparing their responses. Thus, if a peer detects significant inconsistency in these responses, then the minority of

holders should be punished for misreporting. The credibility holders of the misreporting peers should be informed for this inconsistency, which should be observable by these holders too. If there are less collaborated liars in the peer-to-peer system than sincere peers, then the inconsistency will be revealed and corrected, and the corresponding credibility information will be updated accordingly.



**Figure 6. Determining disagreement in feedback messages in a peer-to-peer environment.**

## 6. Comparison with Related Work

Below, we overview a variety of articles dealing either explicitly or implicitly with the consequences of lying, and, in certain cases, with how to alleviate them. We emphasize on the differences of these works with our assumptions as well as with our credibility mechanism and its effectiveness, in order to clarify our contribution.

Dellarocas deals in [12] with the problem of unfair ratings and discriminatory behavior in on-line trading communities where collaborated liars constitute at most 10% of the entire population of buyers. He classifies false reporting as follows: a) unfair high or low ratings by clients to sellers (“ballot stuffing” or “bad-mouthing”), and b) negative or positive discrimination by sellers to clients that offer low or high quality services to a few specific other clients and thus indirectly affect the efficiency of other buyers. Only ballot stuffing and positive discrimination are dealt with, by clustering of ratings of buyers with similar tastes based on commonly rated sellers. Moreover, this approach is not directly amenable to peer-to-peer environments where consumers are also producers of services, and bad-mouthing and negative discrimination can also arise due to peers’ personal interest. Also, finding buyers with common taste requires a global view of the transaction history and raises privacy issues. Chen *et al.* [15] deal with the credibility of raters based on the quality and the quantity of the ratings

they provide. However, the method assigns high confidence to ratings that agree with a majority opinion. Therefore, lying adversaries can still improve their credibility by submitting a large amount of feedback and thus forming the majority opinion.

Schillo *et al.* [10] deal separately with behavior and credibility using the so-called disclosed prisoners' dilemma game with partner selection. Credibility and performance (due to strategic behavior) of other agents are updated by an agent's own observations. Testimonies of witness agents are used for partner selection. It is assumed in [10] that witnesses may hide positive feedback but not tell lies in order not to be discovered. The approach approximates hidden feedback of witnesses and calculates a transitive credibility metric over a path to an agent using Bayes' rule. However, an adversary may still strategically gain high credibility by being truthful in his claims about his high offered performance and then manipulate as a witness the partner selection of other agents. Furthermore, collaboration among lying agents is not considered in [10]. The need for discovering witnesses for an agent is also a drawback of applying this approach in large electronic communities where the same agents meet very rarely. Damiani *et al.*, in [16], extend Gnutella protocol to calculate performance and credibility of other peers based on a peer's own experience and votes from witnesses. This approach (referred to as P2Prep) is similar to the one of [10] in many aspects and hence it has the same limitations.

Credibility and performance (due to strategic behavior) are addressed by Yu and Singh [3]. However, this approach has no explicit mechanism for assessing the credibility of the witnesses; this issue is dealt together with a trust metric regarding behavior, which is determined by direct observations or by asking witnesses. Therefore, it is possible for an adversary peer to maintain a good reputation by performing high quality services and send false feedback for its competitors or his colleagues. This argument also applies to the approach of Kamvar *et al.* [5] where, a global reputation metric regarding performance of each peer is calculated. To this end, each peer's local beliefs (based on observations) on the performance of other peers are weighted by the others' beliefs on his own performance. Aberer *et al.* [4] present an approach to evaluate trustworthiness (i.e. the combination of credibility and performance) of peers based on the complaints posed for them by other peers following transactions. The approach also aims to provide incentives for truthful submission of complaints. The main idea is that a peer is considered less trustworthy the more complaints he receives or files. An agent trusts another if the latter is at least as trustworthy as the former. The experiments conducted showed that the approach does not succeed in identifying a significant part of liar

peers if they constitute 25% of the population. Note that the effectiveness of the approach in the case of collaborated liars was not examined and the approach is not robust against various types of peers' misbehavior. Feldman *et al.* [11] address the problems of free-riding (i.e. poor performance) and misreporting of feedback on contributions (i.e. low credibility) by an indirect reciprocity scheme. Their objective is for each peer to offer to any other peer roughly equal benefit as indirectly offered by the latter to the former. However, their approach provides opportunity for peers to lie about the contribution of other peers in order the latter to be unfairly exploited or for another liar collaborated with the former to prevail in competition. Ngan *et al.* [13] have proposed another indirect reciprocative approach for avoiding free-riding and false claims in a peer-to-peer system for sharing storage capacity. This approach requires peers to publish auditable records of their capacity and their locally and remotely stored files. However, collaborated adversaries can exploit this mechanism by claiming to have stored huge files of one to another.

A side payment approach for eliciting honest feedback in electronic markets has been proposed by Miller *et al.* in [18]. In particular, a payment charged to a buyer is paid to a second buyer according to a scoring rule for his prediction of the rating of a later buyer for their common seller. In the environment considered, honest reporting proved to be Nash equilibrium. However, strategic voting was considered to generate no value for buyers, which is not the case in general, particularly in cases of strategic collaborations. This approach does not deal with collaborated liars, while it is not appropriate for peer-to-peer systems, as it involves the employment of a central bank that distributes payments to peers. Jurca and Faltings [19] have proposed a similar approach that also has similar limitations.

An approach for providing incentives for truthful reporting of feedback in e-markets has been proposed by Jurca and Faltings in [20]. This approach, similarly to ours, employs disagreement in feedback messages for discovering potential lying. However, upon disagreement different fixed side-payments are fined to the transacting agents with the one fined to the seller being higher. This approach is not directly amenable to peer-to-peer systems since side payments require the existence of a bank for mediating the transactions, while sellers and buyers are not supposed to exchange roles. Also, in [20], strategic voting and collaborated lying agents are not considered.

## 7. Conclusions

In this paper, we have proposed a credibility mechanism providing strong incentives for truthful



reporting of ratings' information and, in general, of accounting information in peer-to-peer systems. Our credibility mechanism succeeds this by punishing two peers disagreeing in rating the outcome of their mutual transaction and by disregarding ratings' feedback upon such a disagreement. Thus, lying peers are punished. Moreover, malicious feedback for sincere peers cannot influence their reputation for performance, the calculation of which was shown to become very accurate. We demonstrated experimentally that the mechanism deals effectively with lying in dynamic environments even with large fractions of collaborated liar peers, which are much higher than those expected to emerge in real peer-to-peer systems with large populations. The credibility mechanism *always* results in higher benefit for sincere peers than for liar ones. Thus, truthful reporting of feedback is *incentive-compatible*. Also, participation in the peer-to-peer system is indeed beneficial for sincere peers; thus, the mechanism is *individually rational* too. We have also studied the impact of weighting the ratings employed in the calculation of reputation for performance with the credibility metric of the transacting peers. This approach results in somewhat improved effectiveness; we have omitted these results for brevity. We have also studied a methodology for fine-tuning the parameters involved in the mechanism. In future work, we shall analyze theoretically the efficient evolution of the credibility mechanism and its application in e-commerce.

## References

- [1] T. G. Papaioannou and G. D. Stamoulis. Effective Use of Reputation in Peer-to-Peer Environments. In *Proc. of the 4<sup>th</sup> IEEE/ACM International Symposium in Cluster Computing and the Grid*, Chicago, IL, USA, April 2004.
- [2] C. Dellarocas. Efficiency through feedback-contingent fees and rewards in auction marketplaces with adverse selection and moral hazard. In *Proc. of the 3rd ACM Conference on Electronic Commerce*, San Diego, CA, USA, June 2003.
- [3] B. Yu and M. P. Singh. Distributed Reputation Management for Electronic Commerce. *Computational Intelligence*, Vol. 18, Issue 4, pp. 535-549, 2002.
- [4] K. Aberer and Z. Despotovic. Managing Trust in a Peer-to-Peer Information System. In *Proc. of the 10th International Conference on Information and Knowledge Management*, New York, November 2001.
- [5] S. D. Kamvar, M. T. Schlosser and H. Garcia-Molina. EigenRep: Reputation Management in peer-to-peer Networks. In *Proc. of the Twelfth International World Wide Web Conference*, Budapest, Hungary, May 2003.
- [6] Jøsang, S. Hird and E. Faccer. Simulating the Effect of Reputation Systems on e-Markets. In *Proc. of the 1<sup>st</sup> International Conference on Trust Management*, Crete, Greece, May 2003.
- [7] P. Antoniadis, C. Courcoubetis, R. Mason, T. G. Papaioannou, G. D. Stamoulis and R. Weber. Results of Peer-to-Peer Market Models. project IST MMAPPs: Deliverable 8, 2004. Available at: <http://www.mmapps.org>
- [8] M. Feldman, C. Papadimitriou, J. Chuang and I. Stoica. Free-riding and whitewashing in peer-to-peer systems. In *Proc. of the ACM SIGCOMM Workshop on Practice and Theory of Incentives in Networked Systems*, Portland, Oregon, USA, September 2004.
- [9] J. H. Fowler. Altruistic Punishment and the Origin of Cooperation. Working paper no. 0410002, Economics Working Paper Archive at WUSTL, October 2004. <http://ideas.repec.org/p/wpa/wuwpga/0410002.html>
- [10] M. Schillo, P. Funk and M. Rovatsos. Using Trust for Detecting Deceitful Agents in Artificial Societies. *Applied Artificial Intelligence*, 14:825-848, 2000.
- [11] M. Feldman, K. Lai, I. Stoica and J. Chuang. Robust Incentive Techniques for Peer-to-Peer Networks. In *Proc. of the 4<sup>th</sup> ACM Conference on Electronic Commerce*, New York, NY, USA, May 2004.
- [12] C. Dellarocas. Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior. In *Proc. of the 2<sup>nd</sup> ACM Conference on Electronic Commerce*, Minneapolis, MN, USA, Oct. 2000.
- [13] T.-W. J. Ngan, D. S. Wallach, and P. Druschel. Enforcing Fair Sharing of Peer-to-Peer Resources. In *Proc. of the 2<sup>nd</sup> International Workshop on Peer-to-Peer Systems*, Berkeley, California, February 2003.
- [14] Pretty Good Privacy. <http://www.pgp.com>
- [15] M. Chen and J. P. Singh. Computing and Using Reputations for Internet Ratings. In *Proc. of the 3rd ACM Conference on Electronic Commerce*, New York, NY, USA, October 2001.
- [16] E. Damiani, S. De Capitani di Vimercati, S. Paraboschi and P. Samarati. Managing and Sharing Servents' Reputations in P2P Systems. *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, n.4, July/Aug.ust 2003, pp. 840-854.
- [17] L. Xiong and L. Liu. A Reputation-Based Trust Model for Peer-to-Peer eCommerce Communities. In *Proc. of the IEEE Conference on Electronic Commerce*, Newport Beach, CA, USA, June 2003.
- [18] N. Miller, P. Resnick and R. Zeckhauser. Eliciting Honest Feedback in Electronic Markets. Working paper presented at the *SITE* conference, June 2002. Available at: <http://www.si.umich.edu/~presnick/papers/elicit/>
- [19] R. Jurca and B. Faltings. An Incentive Compatible Reputation Mechanism. In *Proc. of the IEEE Conference on Electronic Commerce*, Newport Beach, CA, USA, June 2003.
- [20] R. Jurca and B. Faltings. Eliciting Truthful Feedback for Binary Reputation Mechanisms. In *Proc. of IEEE/WIC/ACM International Conference on Web Intelligence*, Beijing, China, September 2004.