Reputation-based Estimation of Individual Performance in Collaborative and Competitive Grids

Thanasis G. Papaioannou*, George D. Stamoulis

Department of Computer Science, Athens University of Economics and Business (AUEB), 76 Patision Str., 10434 Athens, Greece.

Abstract

Hidden information is a critical issue for the successful delivery of services in grid systems. It arises when the agents (hardware and software resources) employed to serve a task belong to multiple administrative domains, thus rendering monitoring of remote resource provision absent or unreliable. Therefore, the grid service broker can often observe only the outcome of the collective effort of groups of agents rather than their individual efforts, which makes it hard to identify cases of free-riding or low-performing agents. In this paper, we first identify cases of *hidden information* in grid systems and explain why they cannot be handled satisfactorily by the existing accounting systems. Second, we develop and evaluate a reputation-based mechanism enabling the grid service broker to deal effectively with hidden information. Our mechanism maintains a reputation metric for each agent; we propose and evaluate several approaches on how to update this metric based only on the observations of collective outcomes. We also provide recommendation on which such approach is preferable for a grid service broker in collaborative or competitive environments.

Key words:
hidden information, collective outcome, incentive, rational agent, competitive virtual organizations

*Tel: +30 210 8203549, Fax: +30 210 8203686

Email addresses: pathan@aueb.gr (Thanasis G. Papaioannou), gstamoul@aueb.gr (George D. Stamoulis)

1. Introduction

A grid virtual organization (VO) allows the seamless aggregation of computational resources in multiple administrative domains into a single pool. Thus, users have the opportunity to lease for a certain time-frame bundles of software, CPU cycles, bandwidth and storage space in order to execute computationally intensive tasks, according to a service level agreement (SLA) between the user and the grid service broker. The inclusion of strict or statistical QoS guarantees in the SLAs and the conformance of grid services to them are considered very important for the commercial viability of grid service provision [1], especially in competitive environments. The user however is unaware of the exact resources that execute his task. Thus, the grid service broker must select an appropriate set out of the pool of available resources and offer them to the user as a bundle, henceforth referred to as group (or cluster). This selection is subject to the adverse effects of hidden information on the effort exerted by each agent representing a resource owner in the group. Even if each agent does have the *incentive* to perform as agreed (e.g. due to the revenue to be earned) it may also prefer to free-ride and save on the associated cost of the effor. The broker often has reliable information only on the outcome of the *collective* effort rather than the individual effort of the agents constituting such a group (see Section 3). However, assessing an individual agent's performance is important for selecting the resources to be employed in future service instances, which in general are to be executed by different group of agents each, overlapping only in part with previous ones. Therefore, a complementary mechanism is needed to aid the broker estimate individual performance, despite the lack of relevant accurate information.

The contribution of this paper is twofold: a) We identify and classify cases of *information asymmetry* in grids. b) We propose and evaluate several approaches for estimating the individual performance of agents based on collective outcomes and in the presence of ratings or not. The importance of our subject is also emphasized in [2], where it is stated that "one customer's poor behavior can affect the reputation of the cloud as a whole". To the best of our knowledge both our contribution on information asymmetry in grids and our reputation-based approaches for alleviating this are innovative.

The remainder of this paper is organized as follows: We first discuss (in Section 2) general issues about hidden information and reputation as well as related work. Then, in Section 3, we identify cases of hidden information in grid systems and explain why and when they cannot be handled satisfactorily by the available grid accounting systems. In Section 4, we formulate the objective of the broker and that of each individual agent. Then, in Section 5, we develop our reputation-based mechanism. We employ a reputation metric for each of the agents and propose and evaluate certain *estimation approaches* on how to update this metric on the basis of the observations of collective outcomes. In Section 6, the effectiveness of our estimation approaches for individual performance is evaluated by means of simulation experiments in three scenarions: a) a single VO employing each of the approaches, b) multiple VOs employing collaboratively each of the approaches, and c) multiple competitive VOs each employing a different estimation approach. In systems where individual performance can be observed by other agents of the group, reputation may be based on their *ratings* for this performance. We show that the availability of even a few such ratings can considerably improve the accuracy of performance estimation, if the agents in the group can achieve *Byzantine agreement* against liars. Otherwise, a simple deterministic approach is preferable. Clearly, if a broker can estimate accurately enough the *expected* individual performance of the agents in its VO, then he can increase its own reputation in the grid market and thus be more competitive. Based on our results, we propose which estimation approach should be used by the grid service brokers to estimate the individual performance of agents in competitive or collaborative environments. Finally, in Section 7, we conclude our work.

2. Background and Related Work

Many times in markets buyers and sellers share different portions of information for a service that is exchanged. This situation is referred to as information asymmetry. For example, sellers may be more informed on the quality of the product they offer than clients, or the former does not know the paying behavior of the latter. Depending on the specific information that is hidden, two different problems having different effects may arise in a market: *moral hazard* (i.e. post-contractual opportunism) and *adverse selection* (i.e. misreported inferior quality). In moral hazard, each party in a contract may have the opportunity to gain by violating the principle terms of the agreement [3]. It may lead to *market decomposition*, as agents may be better off to leave this market for a more robust one. In adverse selection, true service quality is misreported to clients. Such a situation will eventually drive all sellers out of the market, except for the lowest quality ones, thus leading to a "market of lemons", as explained by Akerlof in [4]. Reputation is a proper means for revealing hidden information [5].

Previous works applying reputation in grids are based either on user's rating of the agent's individual performance [6], [7], or on event monitoring the deviation of the actual agent's individual performance from the promised one [8]. Also, the approach in [9] calculates indirect reputation based on Resource Usage Records (RUR) [10] and decreases reputation in their absense. In [11], trust for grid nodes is calculated based on the exchange of vectors of direct experience among nodes for various service contexts. The grid nodes are organized per institution and multiple institutions belong to a VO. The global trust of an entity is derived from the reliability of its institution and the trust of the entity as perceived within the institution. In case of collective service outcomes in grid environments, group reputation becomes relevant. According to Tirole [12], collective reputations are history-dependent. Group reputation results to a certain characterization (or stereotype) of a group, which is long-lasting; i.e., members inherit the collective reputation of their elders. Also, according to Levy [13], the higher the information transparency for group members' performance, the higher the member's incentive to perform better. Otherwise, there is potential for group members to free-ride on the high performance of other members and put the blame on others for a collective service of low quality.

Also, [14] deals with the problem of trust and sabotage-tolerance in volunteer grid computing environments. Sabotage-tolerance is very related to our work as we also care about successful collective outcomes. Three categories of sabotage-tolerance techniques are identified: a) replication and voting, b) sampling, c) checkpoint-based techniques. Under replication, the results of replicas are compared and a majority voting is applied. The results not agreeing with the majority are marked as erroneous. Under sampling, the supervisor sends some test samples along with the application tasks and checks the results sent back by the workers. There are several sampling techniques, but it is hardly feasible to generate indistinguishable tests for generic computations. Under checkpoint-based techniques, the supervisor asks at a checkpoint-time the state of computation from the worker, executes the task up to the next checkpoint-time, hashes the results and compares the hash with the corresponding hashcode sent by the worker. However, centralized checkpoint verification bear too much communication and processing overhead to the supervisor, while in the decentralized version redundancy of workers used as verifiers and direct communication capability among workers are necessary. In this paper, we investigate the effectiveness of "cheaper" approaches for revealing individual performance that are also applicable to the case of complex workflow tasks.

Finally, [15] proposes a superscheduler in a federation of trusted or monitored clusters utilizing a commodity market-based approach. Jobs are assigned to fastest or cheapest clusters that have the necessary resources (i.e. number of processors, processing speed, communication bandwidth) to complete the job within the time and budget constraints. However, in this federated environment, clusters can significantly gain by pretending to have the fastest resources, so as to be assigned the jobs and then just complete them slower but within the time constraints for successful service outcome. In our approach, we make no assumptions on the trustworthiness of processors or clusters and we do not rely on the advertised quality of processors. Instead, we estimate their true quality and assigned the jobs to the required number of the most efficient processors.

3. Information Asymmetry in Grids

In this section, we analyze information asymmetry issues that arise in grids and how they can be alleviated by means of reputation. First, we discuss incentive issues on the collection of accounting information in grids.

Grid Accounting and Incentives: There are available solutions for secure aggregation of accounting information on remote resources' consumption in grids. When mobile agents [16] or web services [17] are employed for metering and secure communication channels are used for transferring the relevant data, then much credibility can be attained for accounting information. However, in an open environment (i.e. not an enterprise grid), credible metering can be tricky, since a resource owner (i.e. agent) may have the incentive to *manipulate* accounting information on his resources' consumption. For example, the owner may claim higher resource consumption than the actual one, in order to hide his low effort or to earn more money. Also, if there exist resource owners that *aggregate* or relay accounting information of others, then they may attempt to manipulate this too, in order to serve their individual objectives (e.g. demote competitors) or for malicious purposes. Due to these *conflicting incentives*, information asymmetry on resource consumption in open grids may still arise despite the employment of sophisticated and costly accounting mechanisms. For example, even if the CPU time is accurately metered, the processor speed of the resource owner may be different than agreed with the user, who cannot verify the CPU type. Another example arises if host-level rootkits (kernel, library or application ones) that manipulate the output of accounting/monitoring tools for resource consumption in remote hosts are employed by a resource owner. Also, authentication procedures that are involved in the various grid accounting solutions [16, 17, 18, 10] do not reveal hidden type (i.e. quality) or opportunistic behavior of agents. Consequently, different degrees of information asymmetry may arise in grid accounting. In DGAS [18] and in GridBank [10] accounting architectures the use of direct (i.e. money) and indirect (i.e. reciprocation) rewards provides agents with incentives for misbehavior. In case of pre-payment, an agent has no incentive for job completion, while in case of post-payment he has the incentive to manipulate accounting information. This also applies to the case of the "pay-as-you-go" charging scheme used in Gridbank. Collusions among participating entities provide further incentives for such behaviors and for negative/ positive discrimination of price and QoS to others. Therefore, both adverse selection and moral hazard problems may arise in grids. We describe the various cases below:

Observable Individual Effort: First, note that in case of accurate and credible metering, hidden information on the quality of resources offered by agents and their behavior (i.e. effort spent, honesty etc.) in a grid environment could be revealed directly to the *user* by means of reputation, if requested for decision making on selection of a suitable grid virtual organization (VO) for his needs, as in [7]. The type or behavior of agents may remain fixed or vary dynamically. Thus, a proper reputation metric should be employed for each case, i.e. Bayes or Beta [19] respectively. Reputation can also be used by a provider coordinating a grid VO in order to select the group of available agents to assign a complex task.

Observable Individual Outcome not Effort: A different case of hidden information arises when only the service provision outcome of a task assignment is observable, due to the fact that the consumption of the associated individual resources (or, even more, the associated effort or quality) cannot be *accurately* metered. This paradigm fits better to the case of *bag-of-task* jobs, where the objective for high throughput of task execution places demanding performance requirements to accounting mechanisms. Another such case of hidden information arises when resource providers reside in remote administrative domains and therefore the output of accounting mechanisms for resource consumption can be unreliable, as explained earlier in this section. Then, reputation (based on user ratings) can be a proper means for approximating the *expected* performance outcome of an agent performing individual tasks. However, the actual quality of resources or the performance strategy of the agent will remain hidden information.

Observable but Unverifiable Individual Outcome: A more interesting case of hidden information can arise when there is *no reliable metering* at the individual agent's level; this applies when inter-VO clusters can be formed or in an open grid resource market. Handling this case effectively is very important, as tamper-proof metering of consumption of the resources of a remote agent can be very costly in terms of communication and computational overhead. Tamper-proof accounting can be even impossible due to either the local security policies of a third party that owns the individual agent or insecure system/kernel modules installed in the remote agent. Moreover, hidden information can arise when there is *no detailed specification* of the expected performance and intermediate by each individual agent, which is a common case for complex *workflow* tasks. Subtasks may depend on other subtasks, which in turn depend on others, etc. Also, an agent may be assigned multiple subtasks, thus the relevant dependency graph at the level of agents (i.e. which agent has to wait for which ones to complete their tasks etc.) would very likely contain circles that complicate accountability. Since there is no detailed specification of the expected performance per agent, it is very hard to employ verification approaches, such as those in [14] except for replication. Sonnek *et al.* [20] also employs replication of each subtask assignment to multiple agents selected on the basis of reputation; the outcome for this subtask is determined by means of the *majority* rule. In this paper, we also assume that there is some redundancy employed, leading to tolerance of few under-performing agents, yet more economically than replication of each individual agent as in [20]. The focus of our work in Sections 4-6 is to show how reputation can be exploited in this context. Also, the prediction of the future individual performance of agents based on distributed monitoring of resource consumption is performed in [8], [21]. However, this approach cannot be employed in environments where the monitoring information may be manipulated by malicious agents.

Observable Collective Outcome: In systems where individual performance can be observed (e.g. peer-to-peer and e-markets), reputation is based on the *rating* of this performance by the user. Clearly this is not possible in the grids of our focus, where the user can only rate collective outcomes. On the other hand, certain agents of the group may be able to rate the performance of *others*, if they exchange intermediate results of a workflow task and they can verify their correctness. However, the *credibility* of such ratings is questionable, especially, if the agents are strategically rating each other when competing in being assigned tasks. This can be tolerated to a certain extent; see Sections 5-6.

In the rest of the paper, we employ reputation in the most interesting case of asymmetric information among those discussed above; namely that of verifiable group outcome and hidden/unverifiable individual effort. There are very interesting examples of grid services to which this case of information asymmetry applies. E.g.: a) A distributed search engine implemented by a group of grid agents that search in different sets of data: a query may fail due to an agent that is under-performing either inherently or intentionally; even if a query is served satisfactorily it may be hard to determine whether all group members performed as required. b) A group of agents serving as a sequence of video streaming servers forming paths from the root to the leaves of a multicast tree: the failure of an individual agent may lead to the collapse of a sub-tree; identifying the source of the problem may not be easy.

4. The Participants' Optimization Problems

In this section, we formulate the optimization problems expressing the decision-making process of the grid broker (who selects and schedules the agents to provide their resources per task) and of the participating agents. We consider an inter-domain virtual organization (VO), where a large number of agents can offer their resources to users. Workflow tasks are assigned to groups of agents. At any given time of a task request, an agent i is either available (i.e. has available resources), with probability a_i, or unavailable with probability $1-a_i$. This availability is also related to the quality and the stability of its hardware resources. An agent i has an inherit performance capability, i.e. quality q_i, which is *private* information, e.g. q_i can be the pair (CPU speed, link bandwidth) of resource i. The individual performance outcome of an agent i is given by $x_i(q_i)$. If the agent belongs to a fixed performance type j, then x_i follows the corresponding distribution; e.g. individual performance $x_i(q_i)$ can be at some acceptable level (implying that the agent exerted the effort required thereby) with a fixed probability p_j, independently of other events in the system. On the other hand, if agent i is *rational*, then it follows a dynamic strategy according to which $x_i(q_i)$ is selected each time. The performance outcome of a group of grid agents is determined by the collective effort of its members. In particular, for a group S with size $|S|$ whose members' performance capabilities are given by the vector $\vec{x} = (x_i, \ldots)$ s.t. $i \in S$, the collective outcome of the group is a function $g(\vec{x})$. The function $g(.)$ expresses the effectiveness of the integration of the grid resources, their level of homogeneity etc. In general, we can reasonably assume that $g(\vec{x})$ is monotonically increasing in every x_i s.t. $i \in S$, i.e. more resources lead to increased effectiveness. For example, if $g(.)$ expresses total processing capacity and $x_i(q_i)$ is equal to the processing speed, then we can simply use $g(\vec{x}) = \sum_{i=1,\ldots,|S|} x_i$. If the resources offered by the agents are heterogeneous, then defining $g(\vec{x})$ is more complicated; e.g., if we deal with both CPU cycles and storage space, then $g(\vec{x})$ is a function of the two corresponding partial sums. Furthermore, according to the SLA, the broker should offer to the user at least a minimum quality level ϕ in order for the task to be considered as successful. Besides being compliant to the SLA, the broker may be interested in allocating resources efficiently in order to serve successfully as many tasks as possible, i.e. attain high throughput. In such a case, the broker should select the group S of agents so that the expected performance both exceeds the minimum requirement and matches it as closely as possible for cost-effectiveness. This objective corresponds to the optimization problem:

$$S = \arg\min_{\vec{x}} E[g(\vec{x})], \text{ s.t. } E[g(\vec{x})] \geq \phi \tag{1}$$

In general, this is a combinatorial problem with exponential (in the number of agents) computational complexity. A simpler approach is for the broker to select the group S of nodes so that the expected performance for the present task is maximized yet in an economic way, i.e. reducing the opportunity cost allowing the concurrent provision of services to other users as well. To this end, we assume that up to fixed number N of agents are employed for each task. We henceforth adopt this unconstrained broker's optimization problem:

$$S = \arg\max_{\vec{x}} E[g(\vec{x})] \tag{2}$$

In order to solve either of the above optimization problems, the broker should be aware of the performance capabilities or of the relevant strategy of each agent. Note that, under this approach, working groups have a fixed size N, as only then the objective function of equation (2) is maximized. As explained in Section 3, this constitutes hidden information. Nevertheless, reputation leads to the revelation of the expected individual performance (as opposed to the exact computational capabilities and strategies of the agents). Thus, we have to adopt a heuristic approach to solve the above optimization problem, in which reputation is employed as a proxy of x_i. In particular, we assume that each time a new task arises, the broker sorts the available agents in descending order of reputation and selects an adequate number starting from the *top*; that is, the broker employs the best amongst the awaiting agents, this approach is employed in the experiments of Section 6. Under this greedy approach, agents with low or even medium reputation cannot be selected. An alternative approach would be to give all agents a chance to be selected, while favoring the ones with higher reputation. This is attained by means of the following randomized selection rule: when a new agent is to be added to a group each agent's selection probability equals $r_i/\sum_j r_j$, i.e. is proportional to its reputation. This is a more fair approach than the selection of the agents with top reputation. Although, when viewed in a long series of selections, the two approaches do not differ considerably w.r.t. the expected number of times each agent is selected. This applies when, for each task request, each agent i has a considerable probability $1 - a_i$ to not be available to participate in the corresponding group.

Next, we deal with the optimization problem faced by each *individual agent*. We assume that the broker offers the agents the *incentives to participate* to working groups, by sharing the revenues of collective service outcomes to them. However, individual performance demands costly effort. We take that individual performance is approximated by means of reputation and the broker selects the agents that form working groups for each task on the basis of a reputation-based policy. Then, as argued in [22] (yet for a different setting), when appropriately selected, such a policy provides each agent with the incentive to exert the *maximum possible effort* in collective service provision. Thus, it is *beneficial* for the broker to employ a reputation metric that reveals each agent's expected individual performance.

5. Our Approach

Next, we study how individual expected performance of agents can be revealed by associating to each agent i a reputation metric r_i that is properly updated on the basis of collective outcomes. In particular, we study how this update can be done under both the Bayes and the Beta [19] aggregation rules, according to which r_i expresses the probability of successful individual performance of agent i. According to Beta aggregation, each agent's reputation equals the fraction of the "weighted number" of her successful service provisions over the "weighted total number" of her service provisions, with the weight of each service provision being a negative exponential function of the elapsed time. We have also studied other update approaches based on arbitrary scoring metrics that also lead to the ordering of agents according to their performance. Such approaches constitute hybrid methods among those studied in this paper and thus we omit these studies and results for clarity. We study two different cases of available information in a group of agents:

- intermediate outcomes of the computations may be exchanged among the nodes during the collective computation;
- no intermediate outcomes are exchanged among individual nodes; e.g. when individual outcomes are sent to a coordinating agent that composes the final outcome delivered to the user.

In the case where intermediate outcomes are available, the agents participating in a group that performs a certain task may have information on the individual outcome of some others and on the lowest performance threshold ϕ_i that each member i of the group S should meet in order for the collective performance to meet threshold ϕ. Therefore, nodes are able to rate the performance of others. We consider two different rating approaches the applicability of which depends on the amount of information available to the members of a group about other members: *RATE ALL* and *RATE ONE*. According to the *RATE ALL* approach, each agent j submits a rating vector v_j for all other members i of its group to the broker. The rating vector is given by the following formula:

$$v_j[i] = \left\{ \begin{array}{ll} 1, & \text{if } x_i \geq \phi_i \\ -1, & \text{if } x_i < \phi_i \end{array} \right\}, \forall i \in S \tag{3}$$

Then, the broker sums the rating of each of the members and updates its reputation, according to Beta rule, as follows:

$$r_i' = \left\{ \begin{array}{ll} \frac{\beta r_i t_i + 1}{\beta t_i + 1}, & \text{if } \sum_j v_j[i] > d \\ \frac{\beta r_i t_i}{\beta t_i + 1}, & \text{if } \sum_j v_j[i] < -d \\ r_i, & \text{if } \left|\sum_j v_j[i]\right| \leq d \end{array} , \forall i \in S \right\} \tag{4}$$

t_i is the number of task computations of agent i before the present task, $\beta \in (0,1)$ is the factor that discounts the weight of past transactions, and

d is a confidence threshold; the higher the level of confidence required for the outcome of the rating procedure, the higher d. Clearly, formula (4) corresponds to a majority rule.

On the other hand, according to the *RATE ONE* approach, each agent rates the performance of another randomly selected member of its group according to (3) and submits this rating to the broker. Again, the broker sums the ratings for each member of the group and updates reputations according to a majority rule such as that of (4). The two approaches defined above constitute "extreme" cases w.r.t availability of rating information. Intermediate cases are also conceivable, but are not further studied.

The exchange of intermediate outcomes among agents cannot be assumed in certain cases, such as that of inter-organizational VOs (due to potentially low trust, long interconnection lines that introduce high communication cost/latency and NAT/Firewalls among organization) and when the assigned tasks are replicas of standalone ones. Indeed, trust can be low in the case that agents belonging to competing organizations constitute the VO and in open grid environments where agents are competitive to each other. If the agents are competitive, then the credibility of their ratings to each other cannot be taken for granted. For example, an agent may have the incentive to submit false or malicious ratings and promote co-members in the working group that belong to the same organization and demote co-members that belong to competitive organizations. As will be seen in Section 6, in this case, the effectiveness of rating-based reputation update approaches is reduced. As we showed in [23], this issue can be effectively dealt with by a specific mechanism providing the incentives for reporting truthful ratings exploiting disagreement of ratings as an evidence of lying. For simplicity, in this paper, we assume that no such mechanism is employed in the system, although some agents may rate others untruthfully. If no intermediate outcomes are available, the broker has to infer individual performance on the basis of collective outcomes only. For simplicity, in each case of a *successful* collective outcome, we take that all group members have exerted high effort and we increase the reputation values of all group members. We propose the following approaches for updating reputation of group members after a *failed* service provision:

- Punish all equally (*PUNISH ALL*): decrease the reputation of each agent of the group as in the middle case of (4).
- Punish probabilistically fairly to reputation (*PROP*): The reputation

of a member i is updated according to the middle case of (4) but with probability $r_i/\sum_j r_j$, where the j's are all group members. Thus, the group members share the blame of the collective failure according to their expected performance.

- Punish probabilistically "inversely" to reputation (*INVERSE*): The reputation of a member i now decreases according to the middle case of (4) with probability $(1-r_i)/\sum_j(1-r_j)$.

Furthermore, if we can assume that the grid broker has some idea of *how many* agents did not perform adequately, thus leading the whole group to a failure, then there are more possibilities. To illustrate them, while keeping our study simple, we adopt the following assumption: A group S fails in a service provision (i.e. has lower collective performance than acceptable) if at least M of its members do not exert the necessary effort. Then, the broker can punish only M members, by lowering their reputation, while leaving intact that of all other members of the group. We propose the following two punishing approaches:

- Punish the worst M (*WORST M*): sort the reputation values in descending order and decrease the reputation of the M members with the lowest reputation values according to the middle case of (4).

- Punish random M (*RANDOM M*): Decrease the reputation values of M random group members as above.

So far we have only dealt with reputation in accordance to Beta rule. Next, we employ Bayes' rule for updating agents' individual reputation values based on their collective outcome. Bayes approach is a standard one for approximating hidden variables. Our setting can be considered as a special case of a Bayesian network of star topology with the broker being the hub node and the agents being one hop away. This approach (*BAYES*) is applicable when there are specific fixed performance types of nodes with known success probabilities. For simplicity, we assume that there are two performance types High (H) and Low (L) of agents that have success probabilities p_H, p_L respectively. Then after a failure collective service outcome (F), the reputation r_i of the individual agent i is computed according to the formulas below (6).

$$
\begin{aligned}
r_i' &= \Pr[A_i \in H | F] \\
&= \frac{\Pr[F|A_i \in H] \cdot \Pr[A_i \in H]}{\Pr[F|A_i \in H] \cdot \Pr[A_i \in H] + \Pr[F|A_i \in L] \cdot \Pr[A_i \in L]} ,
\end{aligned} \tag{5}
$$

where $\Pr[F|A_i \in H]$, $\Pr[F|A_i \in H]$ are given by the formulas below:

$$\Pr[F|A_i \in H] = p_H \cdot \Pr[\geq M+1 \text{ agents fail}] + (1-p_H) \cdot \Pr[\geq M \text{ agents fail}] ,$$

$$\Pr[F|A_i \in L] = p_L \cdot \Pr[\geq M+1 \text{ agents fail}] + (1-p_L) \cdot \Pr[\geq M \text{ agents fail}] ,$$

$\Pr[\geq M \text{ agents fail}] = \sum_{m=M}^{N-1} \sum_{\sum_{j \neq i} v_j = m} \Pi_{j \neq i} p_j^{v_j} (1-p_j)^{1-v_j}$ and

$p_j = r_j \cdot p_H + (1-r_j) \cdot p_L$.

A formula similar to (5) applies to the case of a successful collective outcome. In the next section, we compare by means of simulation experiments this standard approach with the previous ones for updating reputation that incur less processing overhead.

6. Experimental Results

6.1. The Simulation Model

In order to examine the effectiveness of the various heuristic approaches for estimating the hidden performance of an individual agent, we perform a series of simulation experiments. Initially, we evaluate the estimation approaches employing a single virtual organization (VO) in a monopolistic scenario. Then, we evaluate the performance of each estimation approach employed by a consortium of collaborative grid service brokers that share the user requests and seek to achieve the highest possible social welfare, i.e. overall success ratio. This scenario can be realistic in settings where a single VO is not capable of handling all user requests due to limited capacity. Finally, in a third scenario, we assume competition among grid service brokers that follow different estimation approaches and evaluate the latter in terms of achievable throughput of successfully served requests in this case.

For this purpose, we first define in Mathematica a simulation model with a single grid VO consisting of 100 agents. We assume that the population of agents belongs to equal shares of two fixed performance types, namely "High" and "Low" that succeed in service provision with probabilities 0.9 and 0.1 respectively. The performance type of an agent is approximated by a reputation metric that is updated according to the various approaches introduced in the previous section. Initially, each agent is assigned an intermediate reputation value r_0=0.5. (This is a reasonable choice as agents are not supposed to change names easily in grids. Otherwise, a low r_0 should be employed to render reputation building more difficult.) Time is assumed to be slotted. At each time slot, users submit workflow tasks to the grid VO, which are allocated to groups dynamically formed on the spot by the grid service broker. Specifically, the grid service broker sorts the agents based on their reputation values and selects the N agents with the top reputation values that are eligible to participate to form the working cluster. Ideally, this selection should be done based on the true success probability of the nodes. This approach provides an upper bound on the achievable efficiency of any reputation-based selection and is referred to as *IDEAL*. Note that, without any reputation metric, random selection could only be employed, achieving the expected success ratio of a randomly selected group, i.e. 0.05. For simplicity, we assume that service instances are completed in one time slot. However, we take that the probability of an agent being available to perform tasks at a time slot is 0.3. This availability value is selected so as the group of eligible agents per time slot to be significantly different. The group size is considered fixed, i.e. with N=10. Also, when rating is applicable, we consider that each agent has a certain rating type, which is orthogonal to its performance type: "honest" and "collusive". Honest agents always rate truthfully the performance of other agents. On the other hand, collusive agents always demote honest ones for their performance and promote their colleagues. All results presented constitute mean values and confidence intervals of 10 runs.

6.2. A Single Virtual Organization

Initially, we take that the working cluster fails in collaborative task accomplishment if more than M=3 of its members fail. If agents are truthful, the relative effectiveness of the various approaches for revealing hidden individual performance based on the collective outcome is shown in Figure 1. The two rating-based approaches perform close to *IDEAL*, while *BAYES* approach follows. Note that when the effectiveness of the rating-based approaches is similar, then *RATE ONE* approach should be employed as being more economical in terms of communication overhead. Another result depicted in Figure 1 is that the simple *PUNISH ALL* approach is more effective than the other more sophisticated ones that do not employ ratings and achieves success ratio very close to rating-based approaches.

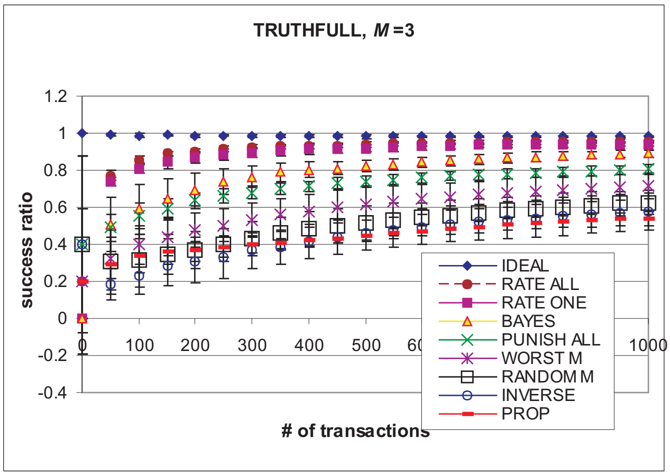


Figure 1: Agents are truthful and M=3. The approaches are ordered as in the legend.

Next, we examine how the tolerance threshold (i.e. the number of redundant agents) M affects the effectiveness of the various reputation update approaches. As shown in Figure 2, with truthful agents and M=2, the effectiveness of the various reputation update approaches that do not employ ratings diminishes considerably for lower M values. This is due to the increased difficulty in identifying low-performing agents in the group, e.g. more agents are punished than those that deserve so with *PUNISH ALL* and fewer than those that deserve so with *WORST M* and *RANDOM M*. On the other hand, rating-based approaches are still very effective, as the aggregation of truthful ratings of other group agents identify low-performing ones.

Henceforth, for clarity reasons, we only depict the effectiveness of the best reputation update approaches, namely *RATE ALL*, *RATE ONE*, *BAYES* and *PUNISH ALL*. If the tolerance threshold M is relatively high (M=3), then the effectiveness of the rating-based approaches remains high for collusive agents consisting up to $N/3$ of the population of the grid VO, which is the theoretical threshold for Byzantine agreement, as depicted in Figure 3. Note that non-rating-based approaches are not affected by the presence of liars. On the other hand, if the tolerance threshold M is lower, then rating-based

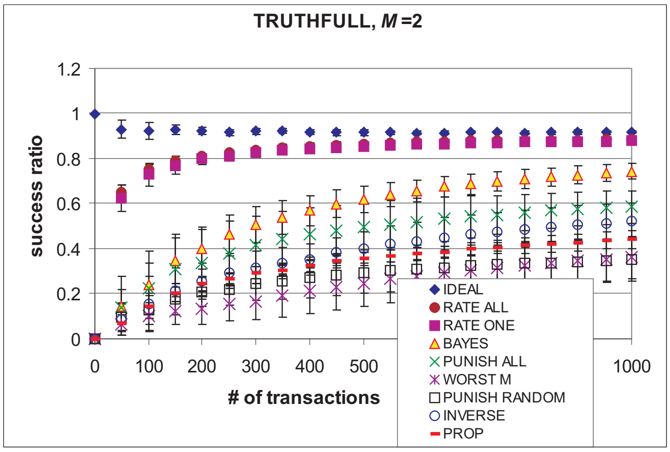


Figure 2: Agents are truthful, but M=2.

approaches become ineffective in the presence of collusive agents, as depicted in Figure 4 for 30% collusive agents and M=2. This is because fewer guilty agents have to be identified among the N agents of the group and, as now majorities of false ratings can be formed, the performance of rating-based approaches is negatively affected. The effect of lying to the rating-based approaches is even more severe for M=1. Note that, as experimentally found, the effectiveness of *BAYES* and *PUNISH ALL* approaches is irrelevant to the percentage of collusive liar agents and increases with the tolerance threshold. This is because the more the guilty agents for a collective service provision failure the more agents correctly take the blame for the collective failure. However, *BAYES* is only applicable in presence of fixed pre-known agent performance types in the VO. We omit these results for brevity reasons.

We have also considered the case that the distribution of the fixed performance types of agents is Uniform in $[0, 1]$. As depicted in Figure 5, the *RATE ALL* approach still performs very close to *IDEAL*. On the other hand, *RATE ONE* is no longer effective, as the limited ratings of this approach severely affect the accuracy of the estimation of the individual performance of agents. The other approaches fall behind from *RATE ALL* in terms of effectiveness and only *PUNISH ALL* achieves a success ratio above 0.5%. Note that *BAYES* approach is not applicable for this distribution of agent performance.

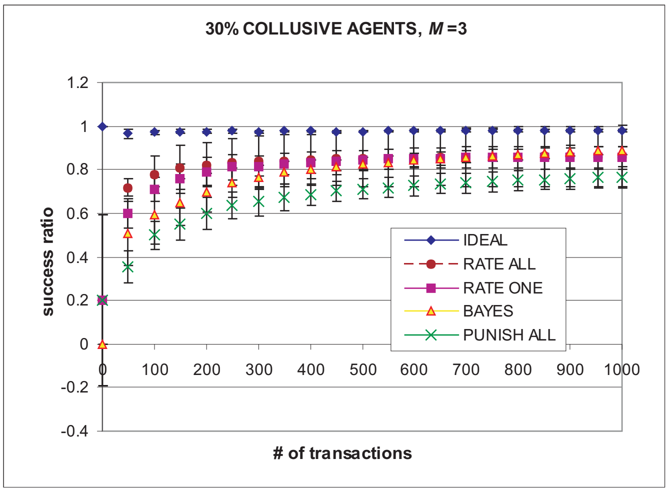


Figure 3: 30% of agents are collusive liars and M=3.

6.3. Multiple Collaborative VOs

We now evaluate the estimation approaches when they are employed by multiple collaborating grid service brokers that share the requests among them, aiming to collectively achieve the highest possible overall success ratio, rather than each of them to optimize some individual performance object. According to this scenario, we now simulate a consortium of 8 VOs that collaboratively serve user requests. Each VO consists of 100 agents that belong to equal shares of two fixed performance types “High” and “Low” as before. The performance of a grid service broker is estimated by means of reputation (i.e. group reputation) and especially using Beta aggregation [19]. We assume that the user employs grid service broker selection probabilistically fair with respect to their reputation values. This choice avoids the concentration of all service requests to the most reputable provider. Note that this reputation-based grid service broker selection policy would have been almost equivalent with the Highest Reputation one [22], according to which the broker with the available resources that had the highest reputation would be selected, if we had assumed that grid service brokers were not always capable of serving new requests. In this scenario, we are interested in the overall success ratio achieved by the collaborative brokers when they all employ the same approach for estimating the performance of their individual agents. In Figure 6, depicted are these success ratios collaboratively achieved by the grid service brokers for each of the estimation approaches. In the following

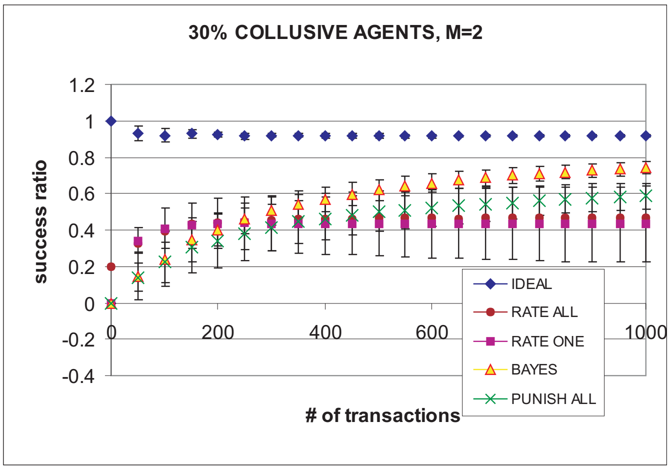


Figure 4: 30% of agents are collusive liars and M=2.

graphs, we refrain from depicting the confidence intervals for clarity reasons. Comparing Figure 6 with Figure 3, we observe that the performance of the estimation approaches is not very different when observing fewer transaction outcomes per policy. Specifically, rating-based approaches are the most effective and followed by *PUNISH ALL*. However, the effectiveness of all approaches decreases, as expected due to the lower number of collective outcomes per VO and the subsequent slower convergence of the reputation values of individual agents to their performance. On the other hand, *BAYES* is the most negatively affected by the sharing of service assignments and respectively of service outcome observations among VOs.

6.4. Multiple Competitive VOs

In this subsection, we simulate 8 grid service brokers each representing one VO that compete each other for service requests and each employing a different estimation approach for the individual performance of their agents. The performance of each grid service broker is again estimated by means of reputation, i.e. Beta aggregation. Also, the users select brokers probabilistically fair to their reputation, as in the previous scenario. In order to avoid any random bias toward an approach, we allow for a period of 160 transactions in the beginning of the experiments the transactions to be equally shared among brokers. We first assume that the agents of each VO belong to two performance types “High” and “Low”, as above. As depicted in Figure 7, the

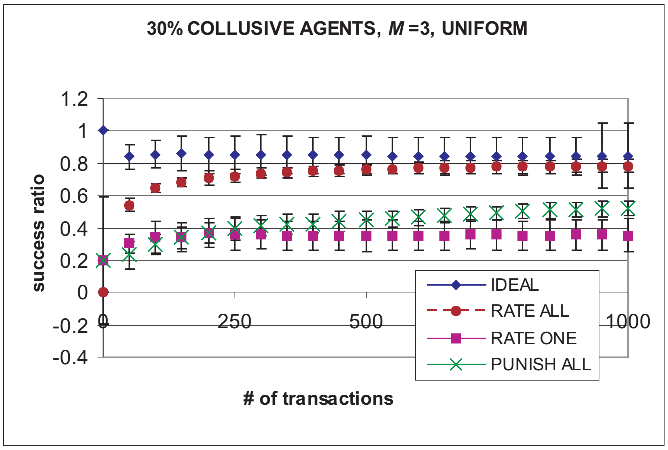


Figure 5: Uniform distribution of agent performance, while 30% of agents are collusive liars and M=3.

brokers that employ *BAYES* and *PUNISH ALL* approaches achieve the highest success ratios. Also, as depicted in Figure 8, the brokers with the same two estimation approaches receive the highest numbers of service requests. Although rating-based approaches are still effective, they fall behind *BAYES* and *PUNISH ALL* due to the presence of collusive liars. Therefore, although *BAYES* does not perform effectively when employed collaboratively, it is very effective when employed against other approaches. This is because *BAYES* and *PUNISH ALL* perform better in distinguishing performance types of agents earlier than other approaches and, as a result, they then concentrate more service requests. However, if all competitive brokers employ *BAYES*, then they would achieve lower efficiency and thus at least one of them would have the incentive to deviate toward a rating-based or the *PUNISH ALL* approach. Therefore, *BAYES* is not a *stable* choice for a grid service broker in a competitive environment. The most effective stable solutions are the rating-based ones, if ratings are available and the presence of collusive liars is not very high. Otherwise, *PUNISH ALL* is the most effective stable approach in a competitive environment.

Finally, we consider that, in the same scenario, the performance types of the agents of each competitive VO are uniformly distributed in $[0, 1]$. In this case, the *RATE ALL* and *PUNISH ALL* estimation approaches are equivalently effective and achieved the highest throughputs of successfully delivered services, as depicted in Figures 9 and 10. *RATE ONE* is not very

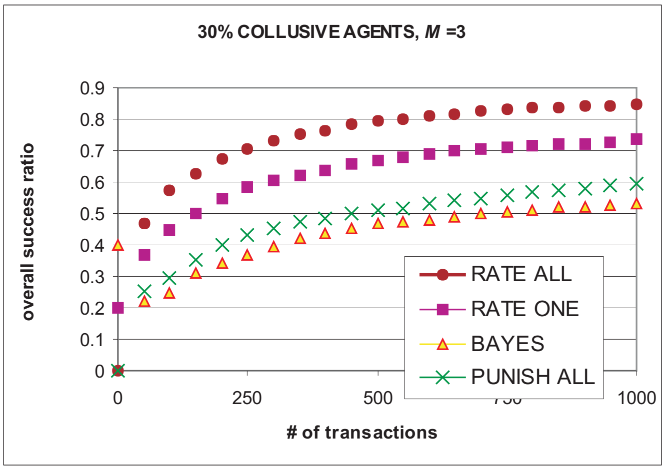


Figure 6: Same estimation approach collaboratively employed by 8 brokers, while 30% of the agents are collusive liars and M=3.

effective in this case, as was also expected by Figure 5. However, their achieved success ratios are lower than those in Figure 7, because discovering the performance of individual agents is more difficult in this case. Note that *BAYES* is not applicable for this performance distribution of agents.

Overall, *RATE ALL* and *PUNISH ALL* are consistently very effective and stable choices for a broker in a competitive environment. However, in large groups with high collusion the last resort would be the *PUNISH ALL* approach, which achieves fair-enough effectiveness in all cases, provided that a large-enough tolerance threshold is employed. Recall from Section 4 that the more effective is an approach in estimating individual performance the stronger are the incentives for higher performance to individual agents inside a broker that employs a reputation-based selection policy, as proved in [22]. This is also true for dynamically rational performance strategies according to [22]. Although two simple behavioral models for agents have been considered in the experiments, the results would be similar for any static behavioral model. The estimation of dynamic behavior is in general more difficult and it depends on the percentage of agents that follow rational strategies and the speed of convergence of the various reputation update approaches. However, it has to be noted that if rational agents are aware of the specific reputation update approach for the estimation of their individual performance, then there may exist performance strategies to avoid detection of low performance

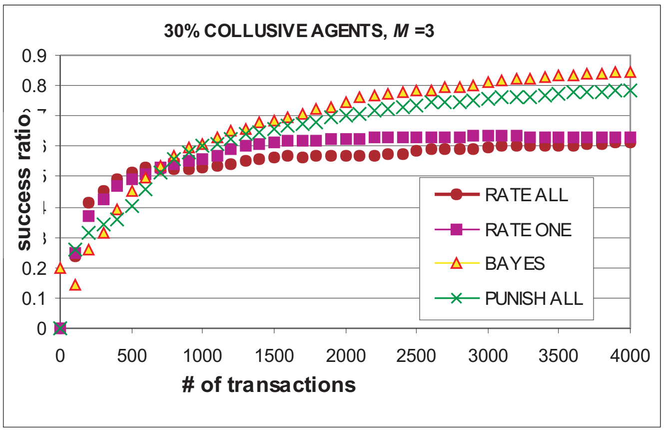


Figure 7: Different estimation approaches employed by 8 competitive brokers, while 30% of the agents are collusive liars and M=3.

from that particular approach. If several computationally "cheap" reputation update approaches are employed in parallel, then it will be trickier for agents of the rational performance strategy to avoid detection of their low performance. Specifically, any different trends in the reputation value of an agent from the various approaches would indicate strategic behavior. Further investigation of alleviating rational strategies that try to circumvent specific reputation update approaches is left for future work.

7. Conclusion

In this paper, we have identified and classified the different information asymmetry cases of grid systems. The most interesting one arises when only the "collective outcome" of a group of agents can be observed, as opposed to the individual performance of each agent. We argued how a proper reputation metric can facilitate the solution of the task assignment problem faced by the grid service broker in case that individually rational strategies are employed by grid agents. We have proposed several reputation-based approaches to deal with this issue. We experimentally found the ranking of these reputation-based approaches based on their accuracy for estimating individual agents' performance. If intermediate outcomes are exchanged among truthful agents, rating-based approaches are very efficient in identifying low-performing agents. Yet, if high collusion arises among agents that do not belong to known performance types, then the simple deterministic

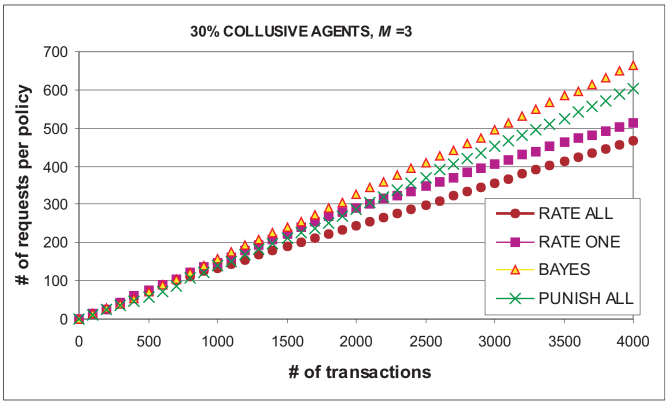


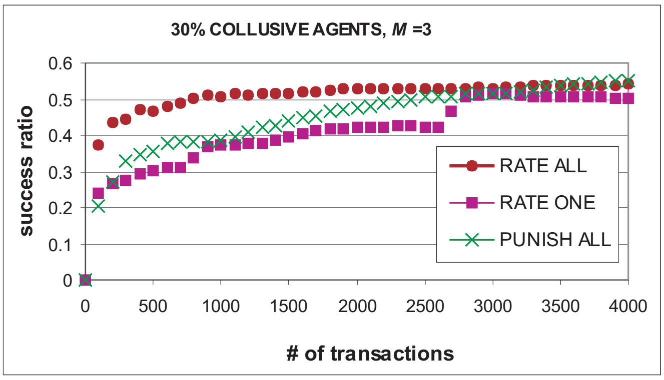
Figure 8: Different estimation approaches employed by 8 competitive brokers, while 30% of the agents are collusive liars and M=3.

PUNISH ALL approach can provide an effective solution to this information asymmetry problem, provided that a large enough tolerance threshold is employed.

Also, we experimentally found the effectiveness of the estimation approaches in collaborative and competitive scenarios among multiple virtual organizations. According to our results, rating-based and *PUNISH ALL* approaches are the most effective *stable* ones for estimating the performance of individual agents and provide the right incentives for exerting high individual performance. However, between the rating-based approaches, *RATE ONE* is involves less communication overhead, but only *RATE ALL* is efficient in case that the performance distribution of agents is unknown.

Therefore, in an actual service paradigm, a grid service broker should employ coupled the *PUNISH ALL* and *RATE ALL* approaches, but initially only employ reputation values calculated with the *PUNISH ALL* approach for agent selection. Then, the broker should periodically estimate the level of collusion in the VO by comparing against each other the performance rankings of agents resulting by these approaches. If they differ below a threshold, then the grid service provider should employ *RATE ALL* for calculating reputation values of individual agents for maximizing the success ratio of the VO in service provision; otherwise, the broker should continue employing *PUNISH ALL*.

As a future work, we intend to investigate the dynamics for VO formation in competitive environments where rational agents of hidden individual



[ht]

Figure 9: Uniform distribution of agent performance, different estimation approaches employed by 8 competitive brokers, while 30% of the agents are collusive liars and M=3.

performance are able to change VOs.

8. Acknowledgments

We thank Costas Courcoubetis and Eva Soukera for useful discussions on the subject of this paper.

References

[1] C. Kenyon, and G. Cheliotis, Grid resource commercialization: economic engineering and delivery scenarios, *Grid Resource Management: State of the Art and Future Trends*, Kluwer Academic Publishers, Norwell, MA, U.S.A., 2004, pp. 465-478.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica and M. Zaharia, Above the Clouds: A Berkeley View of Cloud Computing, EECS Department, Technical Report No. UCB/EECS-2009-28, University of California, Berkeley, Feb. 2009, http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html

[3] K. Arrow, Uncertainty and the Welfare Economics of Medical Care, *American Economic Review*, vol. 53, no. 5, 1963, pp. 941973.

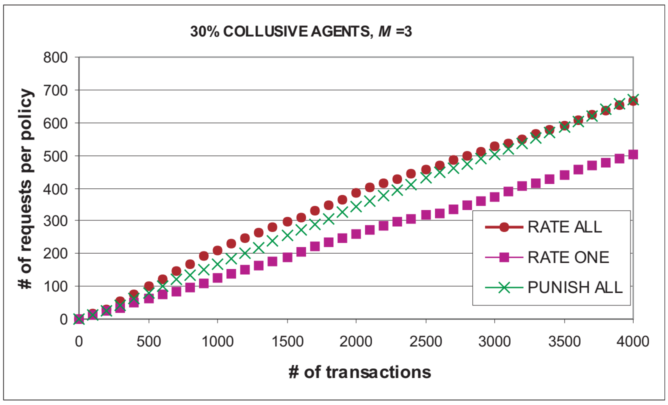


Figure 10: Uniform distribution of agent performance, different estimation employed followed by 8 competitive brokers, while 30% of the agents are collusive liars and M=3.

[4] G. A. Akerlof, The Market for 'Lemons': Quality Uncertainty and the Market Mechanism, *The Quarterly Journal of Economics*, vol. 84, no. 3, The MIT Press, Cambridge, MA, U.S.A., 1970, pp. 488-500.

[5] C. Dellarocas, The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms, Management Science, vol. 49, no. 10, Informs, Linthicum, Maryland, U.S.A., October 2003, pp. 1407-1424.

[6] J. D. Sonnek, and J. B. Weissman, A Quantitative Comparison of Reputation Systems in the Grid, in *Proc. of the IEEE/ACM International Workshop on Grid Computing*, Seattle, WA, U.S.A., 2005.

[7] F. Kerschbaum, J. Haller, Y. Karabulut, and P. Robinson, PathTrust: A Trust-Based Reputation Service for Virtual Organization Formation, in *Proc. of the International Conference on Trust Management*, Pisa, Italy, May 2006.

[8] G. C. Silaghi, A. E. Arenas, and L. M. Silva, A Utility-Based Reputation Model for Service-Oriented Computing, in *Proc. of the CoreGRID Symposium*, Rennes, France, August 2007.

[9] J. Echaiz, J. Ardenghi, and G. Simari, A Novel Algorithm for Indirect Reputation-Based Grid Resource Management, in *Proc. of the Inter-*

national Symposium on Computer Architecture and High Performance Computing, Gramado, Brazil, Oct. 2007.

[10] A. Barmouta, and R. Buyya, GridBank: A Grid Accounting Services Architecture (GASA) for Distributed Systems Sharing and Integration, in *Proc. of the International Symposium on Parallel and Distributed Processing*, Washington, DC, U.S.A., 2003.

[11] W. Xinhua; G. Xiaolin; C. Feifei, A Trust and Reputation Model of Grid Resources for Cooperating Application, in *Proc. of the International Conference on Semantics, Knowledge and Grid*, Beijing, China, Nov. 2005.

[12] J. Tirole, A theory of collective reputation (with applications to the persistence of corruption and to the firm quality), *Review of Economic Studies*, vol. 63, The Review of Economic Studies Ltd., Oxford, U.K., 1996, pp. 1-22.

[13] G. Levy, Decision Making in Committees: Transparency, Reputation, and Voting Rules, *American Economic Review*, vol. 97, no. 1, American Economic Association, Nashville, TN, U.S.A. March 2007, pp. 150-168.

[14] P. Domingues, B. Sousa, and L. Moura Silva, Sabotage-tolerance and trust management in desktop grid computing, *Future Generation Computer Systems*, vol. 23, no. 7, Aug. 2007, Elsevier, Amsterdam, The Netherlands, pp. 904-912.

[15] R. Ranjan, A. Harwood, and R. Buyya, A case for cooperative and incentive-based federation of distributed clusters, *Future Generation Computer Systems*, vol. 24, no. 4, Elsevier, Amsterdam, The Netherlands, Apr. 2008, pp. 280-295.

[16] O. Tomarchio, L. Vita, and A. Puliafito, Active monitoring in grid environments using mobile agent technology, in *Proc. of the Workshop on Active Middleware Services*, Pittsburgh, PENN, U.S.A., August 2000.

[17] E. Elmroth, P. Gardfjll, O. Mulmo, and T. Sandholm, An OGSA-Based Bank Service for Grid Accounting Systems, in Proc. of PARA 2004, LNCS 3732, Springer-Verlag, Berlin-Heidelberg, Germany, 2006, pp. 1051-1060.

[18] R. M. Piro, A. Guarise, and A. Werbrouck, An Economy-based Accounting Infrastructure for the DataGrid, in *Proc. of the International Workshop on Grid Computing*, Phoenix, Arizona, USA, 2003.

[19] A. Jøsang, S. Hird, and E. Faccer, Simulating the Effect of Reputation Systems on e-Markets, in *Proc. of the Conference on Trust Management*, Crete, Greece, 2003.

[20] J. Sonnek, M. Nathan, A. Chandra, and J. Weissman, Reputation-Based Scheduling on Unreliable Distributed Infrastructures, in *Proc. of the IEEE International Conference on Distributed Computing Systems*, Lisbon, Portugal, July 2006.

[21] R. Wolski, N. Spring, and J. Hayes, The Network Weather Service: A Distributed Resource Performance Forecasting Service for Metacomputing, *Future Generation Computing Systems*, vol. 15, no. 5-6, Elsevier, Amsterdam, The Netherlands, October 1999, pp. 757-768.

[22] T. G. Papaioannou, and G. D. Stamoulis, Reputation-based policies that provide the right incentives in peer-to-peer environments, *Computer Networks*, vol. 50, no. 4, Elsevier, Amsterdam, The Netherlands, March 2006, pp. 563-578.

[23] T. G. Papaioannou, and G. D. Stamoulis, An Incentives' Mechanism Promoting Truthful Feedback in Peer-to-Peer Systems, in *Proc. of the IEEE/ACM CCGRID Global P2P Computing Workshop*, Cardiff, U.K., May 2005.