# Reputation-based Estimation of Individual Performance in Grids*

Thanasis G. Papaioannou and George D. Stamoulis
*Department of Computer Science, Athens University of Economics and Business (AUEB),*
*76 Patision Str., 10434 Athens, Greece.*
*{pathan, gstamoul}@aueb.gr*

## Abstract

*Hidden information is a critical issue for the successful delivery of SLAs 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.*

## 1. Introduction

A grid virtual organization (VO) allows the seamless aggregation of computational resources in multiple administrative domains into a single pool. 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. A SLA is agreed between the user and the grid service broker, containing the terms for this leasing of resources and the broker's obligations regarding performance, security etc. 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]. The user is unaware of the exact resources, i.e. software modules, hardware, etc., that execute his task. On the other hand, the grid service broker must select an appropriate set of resources and present them to the user as a bundle, henceforth referred to as group (or cluster). This selection is made out of the pool of available resources constituting the VO represented and coordinated by a grid service broker.

Unfortunately, this selection process is subject, to the adverse effects of *hidden information*. Indeed, an agent representing a resource owner is required to contribute a certain effort, in order for the group to meet the SLA. We assume that it is beneficial for an agent to be part of a group that delivers successfully a SLA; e.g., the members of such a group share part of the VO's revenue, to be compensated for their effort, while this does not apply in case of failure. Thus, each agent does have the *incentive* to perform as agreed; however, it may still have the incentive to free-ride; i.e. exert less effort and save on the associated cost. The broker often observes only the outcome of the *collective* effort rather than the individual effort of the agents constituting the group serving the user (see Section 3). Thus, due to this case of hidden information, whenever a SLA is not granted, the broker might not be able to objectively and safely identify which agents did not perform adequately. However, predicting an individual agent's performance is important for future service instances. For example, inclusion of a free-riding agent in a group in the future affects adversely the *probability* of this group to *successfully meet a SLA*. Note also that, in general, each task is executed by a different group of agents, which in general overlap only in part. Hence, the broker should somehow be able to place lower priority to a low-performing or a free-riding agent when selecting the ones to form a group. Thus, a

* This work has been carried out as part of the project GridEcon (FP6–2005–IST5–033634) funded by the EU.

complementary mechanism is needed to aid the broker estimate individual performance, due to the lack of relevant accurate information. Note that, as explained in Section 3, although accounting mechanisms that accurately meter individual performance in multi-organizational grid VOs have been proposed, they are in fact costly, while rational agents may have the incentive and the capability to mislead them.

In this paper, we study such a complementary mechanism to aid the broker, which is based on *reputation*. In particular, we first discuss (in Section 2) general issues about hidden information and reputation. Then, in Section 3, we identify cases of hidden information in grid systems and explain why they cannot be handled satisfactorily by the available grid accounting systems, particularly, when the VO spans multiple domains in which the broker does not have the authority to collect accounting information directly. In Section 4, we formulate the objective of the broker and that of each individual agent. Then, in Section 5, we develop the reputation-based mechanism. We employ a reputation metric for each of the agents and propose and evaluate certain approaches on how to update this metric on the basis of the observations of collective outcomes. Each approach is evaluated experimentally in terms of the achieved probability of successfully delivering SLAs (in Section 6). Moreover, in systems where individual performance can be observed by other agents of the group, reputation may be based on their *ratings* for this performance. Thus, we also investigate how such ratings can be exploited, whenever available in a grid system, for improving the accuracy of the reputation metric for individual agents. We show that the availability of even a few such ratings can considerably improve the accuracy of performance estimations and thus the VO's overall success probability, if the agents participating can achieve *Byzantine agreement*; otherwise, a simple deterministic approach is preferable.

For a market with multiple competing brokers, we can assume that a reputation mechanism for grid VOs (i.e. brokers) would also be used, aiding the users select the best broker for a new task. Clearly, if a broker can estimate accurately enough the expected individual performance of his associated agents, then he can increase his own reputation in the grid market and thus be more competitive. However, in this paper, we do not deal directly with this issue. We only consider one broker trying to estimate agents' individual performance as accurately as possible and serve new requests without affecting the QoS of the jobs already scheduled. To the best of our knowledge both our contribution on information asymmetry in grids and our reputation-based policies for alleviating this are innovative. Previous works applying reputation in grids are based either on user's rating of the agent's individual performance [2], [3], or on event monitoring the deviation of the actual agent's individual performance from the promised one [4].

## 2. Hidden information and reputation

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. Another example is when clients do not pay for the services that they obtain but providers do not know it prior to service provision. Depending on the specific information that is hidden, two different problems having different effects may arise in a market: *moral hazard* and *adverse selection*. These problems are discussed below.

Moral hazard can be present any time two parties come into agreement with one another. Each party in a contract may have the opportunity to gain by violating the principle terms of the agreement. For example, in an e-marketplace, the buyer typically sends money to the seller before receiving the goods. Hence, the seller is tempted to keep the money and not ship the goods, or to ship goods that are inferior to those advertised. This lack of trust among participants may lead to *market decomposition*, as they may be better off to leave this market for a more robust one.

Adverse selection is present in situations where sellers possess information (on some aspect of their innate ability, product quality, etc.) that buyers do not (or vice versa). Such situations often arise in markets for experience goods. Consider, for example, an online hotel booking site where hotels of different qualities advertise rooms. Consumers cannot be certain about the true quality offered by each hotel until having actually stayed there. Also, hotels do not have an incentive to advertise any of their weak points. Knowing this, consumers will assume that all hotels are of average quality and will not be willing to pay more than the average price. 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 [5]. The most important distinction between (pure) moral hazard and (pure) adverse selection is that, in the former case, all sellers are capable of the same type of behavior (e.g. cooperate, cheat), whereas in the latter case seller behavior is completely constrained by their innate "type".

Many distributed systems rely on cooperation among self-interested nodes. For example, in peer-to-peer file-sharing systems, overall download latency and failure rate increase when peers do not share their resources. Similar examples also apply to distributed systems (including the Web), mobile ad-hoc networks, etc. *Reputation*-based mechanisms are proper means for discovering hidden information about participants in such systems (e.g. see [6], [7]). In particular, such mechanisms can deter moral hazard by serving as *sanctioning* mechanisms. Such mechanisms can also alleviate adverse selection, by *signaling* to users the quality of service that can be delivered by each agent.

Reputation in an e-marketplace or a peer-to-peer system with one-to-one transactions is calculated as follows: After observing the *outcome* of his transaction, a client *rates* the providing agent for his performance. Binary rating (i.e., success vs. failure) is known to be adequate for calculating a reputation value that is representative of the expected outcome of a transaction with a specific provider [6], [7]. All ratings that a specific agent got for all services he provided are aggregated in a single reputation value. The two main aggregation approaches are as follows: a) Aggregation based on *Bayes'* rule [7]; according to this approach, reputation of each provider expresses the *a posteriori* belief that he belongs to a certain performance type given the history of his service provisions. b) *Beta* aggregation [8], according to which, 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.

As explained in Section 1, it is necessary for the grid service provider to infer hidden individual performance of agents from the collective outcomes they produce. According to Tirole [9], 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 [10], 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.

Finally, an important issue of reputation is that its accuracy and, thus, its effectiveness in offering the right incentives are vulnerable to the inclusion of untruthful ratings in its calculation. Matters are worse in competitive environments, such as an inter-domain grid, where each member has the incentive to demote its competitors obtaining an advantage over them in future selections. As we showed in [11], this issue can be effectively dealt with by a specific mechanism providing the incentives for reporting *truthful ratings*. For simplicity, in this paper, we assume that no such mechanism is employed in the system, although some agents may rate others untruthfully.

## 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. There are available solutions for secure aggregation of accounting information on remote resources' consumption in grids. When mobile agents [12] or web services [13] 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 [12-15] 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 [14] and in GridBank [15] accounting architectures the user 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.

We have already discussed the main issues of hidden information in Section 2. In grids, both adverse selection and moral hazard problems may arise. 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 VO for his needs, as in [3]. The type or behavior of agents may remain fixed or vary dynamically. Thus, a proper reputation metric should be employed for each case. Reputation can also be used by a provider coordinating a grid VO in order to select the group of agents that will execute a task successfully. Since this group varies per task, an individual reputation metric for each agent should be employed.

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 matching case for this hidden information case 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, as in [2]. However, the actual quality of resources or the performance strategy of the agent will remain hidden information. Yet, it is not obvious at all how to apportion performance liability to the agents that provide a service.

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 by each individual agent, which is a common case for complex *workflow* tasks. In particular, consider a group of agents of a VO, whose resources possibly belong to different domains, interact with each other in order to accomplish a demanding task. 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 verify the correctness of an individual outcome. To deal with this problem, Sonnek *et al.* [16] employ 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 our paper, we also assume that there is some redundancy employed, leading to tolerance in under-performing agents, yet more economically than replication of each individual agent as in [16]. 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 [4], [17]. However, this approach cannot be employed in environments where the monitoring information is manipulated by malicious agents.

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 and the provider 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 competing with each other in being assigned tasks. Then, there is potential for strategic ratings of agents by others. 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 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- $a_i$. An agent $i$ has an inherit performance capability, i.e. quality $q_i$, which is *private* information. The individual performance outcome of a node $i$ 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 $\vec{x} =$ $(x_k, \ldots)$ s.t. $k \in S$, the collective outcome of the group is a function $g(\vec{x})$. The function $g(.)$ depends on the effectiveness of the integration of the grid resources, their level of heterogeneity etc. In general, we can reasonably assume that $g(\vec{x})$ is increasing in every $x_i$. For example, if all agents provide CPU cycles and the user is interested in the total processing capacity offered, then we can simply use $g(\vec{x}) = \sum_{k=1,\ldots,|S|} x_k$. 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 $q$ in order for the task to be considered as successful. Besides being conformant 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. This objective corresponds to the optimization problem

$$\min_S \mathrm{E}[g(\vec{x})], \text{ s.t. } \mathrm{E}[g(\vec{x})] \geq q \tag{1}$$

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 broker's optimization problem:

$$\max_S \mathrm{E}[g(\vec{x})] \tag{2}$$

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 2, 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 $r_i / \sum r_j$ , i.e. is proportional to its reputation. This is a more fair approach than the selection of the agents with top reputation. Nevertheless, 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 [6] (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 highest possible individual expected 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 aggregation rules discussed in Section 2, according to which $r_i$ expresses the probability of successful individual performance of agent $i$. We have also studied how to update other arbitrary scoring metrics that also lead to the ordering of agents according to their performance. Such approaches can be equally effective as those studied in this paper. Due to space limitation, we omit these studies and results.

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 $q_i$ that each member $i$ of the group $S$ should meet in order for the collective performance to meet threshold $q$. 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, that each agent $i$ submits a rating vector $v_i$ for all other members of its group to the broker. The rating vector is given by the following formula:

$$v_i[j] = \left\{ \begin{array}{l} 1, \text{if } x_j \geq q_j \\ -1, \text{if } x_j < q_j \end{array} \right\}, \forall\, j \in S \tag{3}$$

Then, the broker sums the rating of each of the members and updates its reputation, according to Beta rule (see Section 2), as follows:

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

$t_i$ 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.

If the agents are competitive to each other (e.g.

belong to two competing CPU owners) then the *credibility* of the rating of one by the other cannot be taken for granted. In particular, in such a case, an agent may have the incentive to submit false or malicious ratings for some of its co-members in a group to which a task was allocated. As will be seen in Section 6, this phenomenon reduces the effectiveness of reputation. Another alternative for the broker is to employ the approaches introduced below.

The exchange of intermediate outcomes among agents cannot be assumed in certain cases, such as that of inter-organizational VOs (due to low trust and long interconnection lines) and that of low replication of standalone tasks. Then, the broker has to infer individual performance on the basis of collective ones 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 $r_i/\Sigma_j r_j$, 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 $(1-r_i)/\sum_j(1-r_i)$.

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. 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). A similar set of formulas applies to the case of a successful collective outcome.

$$\left.\begin{aligned}
&r_i = \Pr[A_i \in H \mid \mathrm{F}] = \\
&= \frac{\Pr[\mathrm{F} \mid A_i \in H]\cdot \Pr[A_i \in H]}{\Pr[\mathrm{F} \mid \mathrm{A_i} \in H]\cdot \Pr[A_i \in H] + \Pr[\mathrm{F} \mid \mathrm{A_i} \in L]\cdot \Pr[A_i \in L]} \\
&\Pr[\mathrm{F} \mid \mathrm{A_i} \in H] = \Pr[\mathrm{A_i}\ \text{succeeds}]\cdot \Pr[\geq M+1\ \text{agents fail}] + \\
&\Pr[\mathrm{A_i}\ \text{fails}]\cdot \Pr[\geq M\ \text{agents fail}] \\
&\Pr(\geq M\ \text{others fail}) = \sum_{\mathrm{m=M}}^{N-1} \sum_{\sum_{j\neq i} v_j = m} \prod_{j\neq i} p_j^{\,v_j}\cdot (1-p_j)^{1-v_j} \\
&\text{where}\ \ p_j = r_j\cdot p_H + (1-r_j)\cdot p_L
\end{aligned}\right\} \quad (6)$$

## 6. Experimental results

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 experiments. In particular, we consider one grid VO consisting of 100 agents. We consider 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 node 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.) Users submit workflow tasks to the grid VO that are allocated to groups 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.

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.

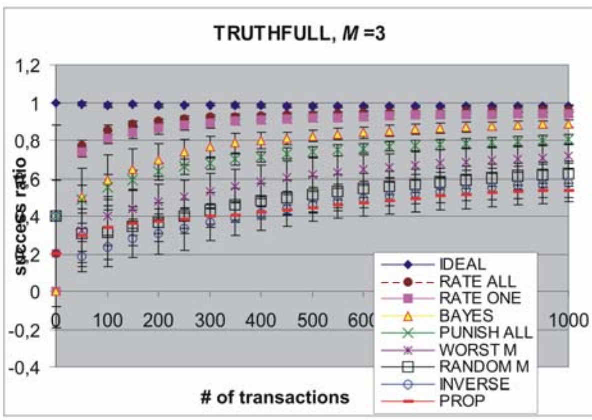


Figure 1. Agents are truthful and $M$=3. The approaches are ordered as in the legend.

Initially, we take that the working cluster fails in collaborative task accomplishment if more than $M$=3 of its members fail. Then, for truthful agents, 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.

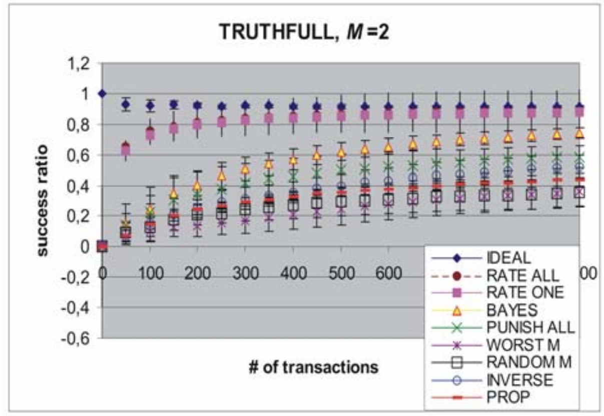


Figure 2. Agents are truthful, but $M$=2.

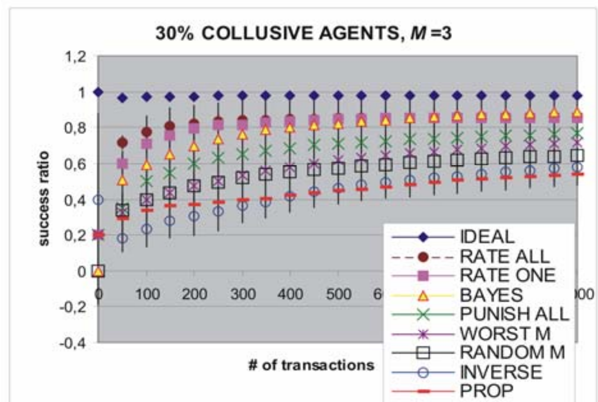


Figure 3. 30% of agents are collusive and $M$=3.

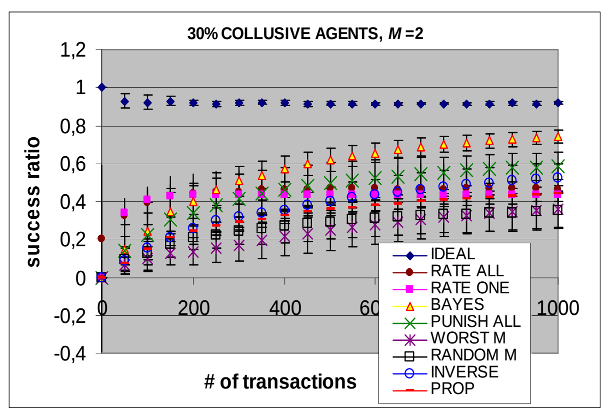


Fig. 4. 30% of agents are collusive and $M$=2.

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. 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.

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 approaches become ineffective in the presence of collusive agents, e.g. see Figure 4 for 30% collusive agents and $M$=2. The effect of lying is even more severe for $M$=1, again because fewer guilty agents have to be identified among the $N$ agents of the group. 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.

Finally, we have also considered the case that the distribution of the fixed performance types of agents is Uniform in [0, 1]. Experimental results produced the same ranking of proposed reputation-based approaches as in the previous performance scenario. Also, the rating-based approaches perform worse in the case of Uniform distribution of the performance types, as expected, while BAYES approach is not applicable. Thus, 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.

## 7. Conclusion

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

Therefore, in an actual service paradigm, a grid service broker should employ coupled the PUNISH ALL and RATE ONE 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, e.g. by comparing against each other the performance rankings of agents resulting by these approaches. If they deviate below a threshold then the grid service provider should employ RATE ONE for calculating reputation values of individual agents for maximizing the success ratio of the VO in service provision; otherwise, the broker should employ PUNISH ALL.

As a future work, we intend to investigate the combination of the reputation-based approaches introduced in this paper with a reputation mechanism evaluating the provided performance of grid brokers in competitive environments.

## 8. Acknowledgement

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

## 9. References

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

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

[3] 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.

[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] A. Jøsang, S. Hird, and E. Faccer, "Simulating the Effect of Reputation Systems on e-Markets," in *Proc. 1st International Conf. on Trust Management*, Crete, Greece, 2003.
[9] 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.
[10] 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.
[11] T. G. Papaioannou, and G. D. Stamoulis, "An Incentives' Mechanism Promoting Truthful Feedback in Peer-to-Peer Systems," in *Proc. of IEEE/ACM CCGRID Global P2P Computing Workshop*, Cardiff, U.K., May 2005.
[12] O. Tomarchio, L. Vita, and A. Puliafito, "Active monitoring in grid environments using mobile agent technology," in *Proc. Workshop on Active Middleware Services*, Pittsburgh, PENN, U.S.A., August 2000.
[13] E. Elmroth, P. Gardfjäll, 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.
[14] R. M. Piro, A. Guarise, and A. Werbrouck, "An Economy-based Accounting Infrastructure for the DataGrid," in *Proc. 4th International Workshop on Grid Computing*, Phoenix, Arizona, USA, 2003.
[15] A. Barmouta, and R. Buyya, "GridBank: A Grid Accounting Services Architecture (GASA) for Distributed Systems Sharing and Integration," in *Proc. 17th International Symposium on Parallel and Distributed Processing*, Washington, DC, U.S.A., 2003.
[16] J. Sonnek, M. Nathan, A. Chandra, and J. Weissman, "Reputation-Based Scheduling on Unreliable Distributed Infrastructures," in *Proc. IEEE International Conf. on Distributed Computing Systems*, Lisbon, Portugal, July 2006.
[17] R. Wolski, N. Spring, and J. Hayes, "The Network Weather Service: A Distributed Resource Performance Forecasting Service for Metacomputing," *Journal of Future Generation Computing Systems*, vol. 15, no. 5-6, Elsevier, Amsterdam, The Netherlands, October 1999, pp. 757-768.