# Selfish Grid Computing: Game-Theoretic Modeling and NAS Performance Results

**Yu-Kwong Kwok, ShanShan Song, and Kai Hwang**
University of Southern California, Los Angeles, CA 90089, USA

*Abstract*— **Selfish behaviors of individual machines in a Grid can potentially damage the performance of the system as a whole. However, scrutinizing the Grid by taking into account the non-cooperativeness of machines is a largely unexplored research problem. In this paper, we first present a new hierarchical game-theoretic model of the Grid that matches well with the physical administrative structure in real-life situations. We then focus on the impact of selfishness in intra-site job execution mechanisms. Based on our novel utility functions, we analytically derive the Nash equilibrium and optimal strategies for the general case.**

**To study the effects of different strategies, we have also performed extensive simulations by using a well-known practical scheduling algorithm over the NAS (Numerical Aerodynamic Simulation) workload. We have studied overall job execution performance of the Grid system under a wide range of parameters. Specifically, we find that the Optimal selfish strategy significantly outperforms the Nash selfish strategy. Our performance evaluation results can serve as valuable reference for designing appropriate strategies in a practical Grid.**

*Keywords*—**Grid computing, non-cooperative games, virtual organizations, selfish behaviors, online scheduling, Nash equilibrium, optimal strategies, performance evaluation.**

## I. Introduction

The lofty goal of Grid computing is to leverage on the interconnection of a large number of geographically distributed machines to solve computational problems faster at a gigantic scale [3], [24]. However, this goal is based upon the premise that the interconnected machines are *cooperative* in the sense that they are willing to execute remote jobs. We believe that as the Grid scales up, this premise may no longer hold. Notice that the Grid is a large scale peer-to-peer (P2P) system at the server-level (rather than at the desktop level as in file-sharing P2P applications). Thus, the "peers", i.e., the Grid sites, owned and managed by different organizations, may not always want to cooperate with each other. Indeed, the various computers *within* a Grid site may not even cooperate with each other. This scenario resembles the situation in the non-cooperation among states of a large country, or the non-cooperation among departments in a large organization.

Thus, it is an important research problem to model the Grid and its constituents by taking into account the potential non-cooperativeness at various levels. With such modeling, we can then study the impact of *selfishness* and subsequently design proper strategies to avoid its adverse impacts. This can in turn lead to a much more efficient utilization of the Grid processing resources. However, despite that there have been several recent attempts in scrutinizing the Grid from a so-called "market" oriented perspective [6], [8], [27] (as detailed in Section II), the modeling problem of the Grid with realistic selfishness concepts is relatively unexplored.

In this paper, we propose a new *game theoretic* modeling of the Grid and present our analytical as well as simulation results. Specifically, we make three contributions:

1) **A Hierarchical Game Theoretic Grid Model:** We consider that to manage the scalability of a Grid, a hierarchical structure must be used. Essentially, the hierarchy consists of three levels: the global scheduling level, the inter-site level, and the intra-site level. We believe that this hierarchical structure matches well with the physical administrative structure of Grid sites.
   Based on this hierarchy, we introduce three different game theoretic scenarios: the intra-site job execution game, the intra-site bidding game, and the inter-site bidding game. Due to space limitations, we focus on the intra-site job execution game in this paper and the other games, and most importantly, the interplay among the three games, will be presented in subsequent papers.
2) **Mathematical Analysis of the Intra-Site Job Execution Game:** We first propose a novel but realistic utility function for each participating machine within a Grid site. We then formally derive the equilibrium strategies and the optimal strategies. Based on these analytical results, we design algorithms for the machines to achieve a high utility as well as high performance, despite that the machines are selfish.
3) **Extensive Performance Evaluation of the Model:** We conducted extensive simulations to study the behaviors of the Grid under different strategies: heterogeneous strategies, Nash strategies, and optimal strategies. Specifically, based on a well-known practical scheduling algorithm, namely the MinMin algorithm [4], we studied the utility and job execution performance (in terms of makespan and slowdown ratio) of the Grid system under a wide range of parameters. Our performance evaluation results can serve as valuable reference for designing appropriate strategies in a practical Grid.

The rest of the paper is organized as follows. Section II presents a brief review of related work. In Section III, we de-

This research was supported by an NSF ITR Research Grant under contract number ACI-0325409. Corresponding author: Kai Hwang, Email: kaihwang@usc.edu, Tel: 213-740-4470, Fax: 213-740-4418. Y.-K. Kwok participated in this project while he was visiting USC, on sabbatical leave from the University of Hong Kong (HKU). Kwok's research at USC was also supported by a research grant from the Research Grants Council of the HKSAR and the 2003-2004 Outstanding Young Researcher Award given by HKU.

scribe our proposed hierarchical Grid model and the associated game theoretic research problems. We then describe in detail our modeling and analytical formulations of strategies for the intra-site job execution game in Section IV. Section V contains a detailed discussion about our simulation setup and the parameters used. We present the extensive simulation results and our interpretations in Section VI. The last section concludes the paper.

## II. Related Work

Recently we have witnessed an intensive interest in using game theoretic and market-oriented approaches in the analysis and design of distributed computing and networking algorithms [1], [16], [20], [21], [23], [28]. There have also been some recent results on game theoretic job allocation and scheduling reported in the literature [17]. Regev and Nisan [22] suggested the so called POPCORN market for trading online CPU time among distributed computers. In their system a virtual currency called "popcoin" was used between buyers and sellers of CPU times. The social efficiency and price stability were studied using the Vickrey auction theory [19]. Similar approaches were also proposed by other researchers [5], [7], [8], [26].

Wolski *et al.* [27] proposed a model called G-Commerce in which computational resources among different Grid sites are traded in a barter manner. The efficiency of two different market conditions—commodities markets and auctions—were studied by simulations. They concluded that a commodity market is a better choice for controlling Grid resources compared with auctions. Ghosh *et al.* [9] study the load balancing issues in a mobile computational Grid. In their model, there was a wireless access point (WAP) which mediates the requests from different mobile devices constituting the Grid. Using Nash Bargaining Solution (NBS), they devised a framework for unifying network efficiency, fairness, utility maximization and pricing.

Larson and Sandholm [13] pioneered the consideration of the computation cost involved in determining the valuations which are essential inputs to the auction system. They defined the notion of "miscomputing ratio" which characterizes the impact of selfishness on the efficiency of the auction. Nisan and Ronen [17], [18] formally defined the job allocation in a distributed system using a truthful mechanism framework.

Grosu and Chronopoulos [10] recently designed a load balancing system based on the Vickrey-Clarke-Groves (VCG) mechanism [19] in which each computer optimizes its "profits" by considering the payment and cost involved in handling a job. Volper *et al.* [25] proposed a game-theoretic middleware called GameMosix. Selfish behaviors are modeled by "friendship relationships" in that computers will help each other only when they have established friendship relationships before.

With reference to the above mentioned related work, our proposed models and analytical formulations are novel in that we consider the hierarchical relationships among individual computers in a gigantic computational Grid. Our work is also the first of its kind in investigating the selfishness issues *within* a Grid site.

## III. A Hierarchical Semi-Selfish Grid Model

As mentioned earlier in Section I, the ultimate scale of a computational Grid is gigantic, and thus, the Grid, pretty much like the Internet itself, will cross organizational and national boundaries. An open question is that how such a gigantic distributed computing platform, which is likely to be composed of hundreds of thousands of processors, is to be structured and maintained. We believe that a hierarchical structure, as depicted in Figure 1, is the only feasible solution.

In our study, we envision that each "Grid site" is not going to be a single computer but rather a network of computers[1], each of which is a cluster of machines or a tightly-coupled massively parallel machine. Thus, eventually we may have hundreds of Grid sites, each of which consists of tens of multiprocessors (i.e., clusters and parallel machines). Indeed, such a structure, again resembling the Internet itself, closely matches the "administrative" structure of computing resources in organizations.

For instance, the computer science department of a university might own a large cluster of PCs, the electrical engineering department might possess another, and the physics department might manage a massively parallel supercomputer. Yet all these computing resources participate in the global Grid community according to the university's mandate. Thus, at the intra-site level, the participating computers, each of which is autonomous, form a *federation*. At the inter-site level, the participating Grid sites form another level of federation.

With the hierarchical structure shown in Figure 1, there are also two levels of job scheduling and dispatching, depicted in Figure 2. Specifically, the job submission system, which is implemented as a global middleware, channels user submitted jobs to the global scheduling system. We envision that such a job submission middleware can be easily constructed using Web services tools (e.g., WSDL and SOAP messages [2]). Equipped with a global Grid processing resources registry (again could be based on the UDDI protocol), the global scheduler performs job allocation, according to a certain scheduling algorithm.

Most importantly, at the inter-site level, the scheduler has only the knowledge of the processing capability of each Grid site as a whole, without regard to the details within the site. In this manner, the scalability of scheduling at the global Grid level can be efficiently handled. Furthermore, again this scheduling model conforms well to the administrative structure of the Grid community in the sense that the global scheduler probably should not "micro-manage" the execution of jobs down to the machine level. The global scheduler makes use of the "capability parameters" supplied by the Grid sites as the inputs to the scheduling algorithm. These capability parameters are, in turn, mediated by the local job dispatcher at each Grid site based on its information about the local participating machines.

As described above, our hierarchical model, while capturing the realistic administrative features of a real-life large-scale distributed computing environment, is also generic in nature. Indeed, this federation-based Grid model opens up a large variety of interesting research issues. Firstly, any efficient online

[1]Throughout this paper, we use the term "computer" and "machine" to refer to a monolithic autonomous computing platform that possibly consists of multiple CPUs.

job scheduling algorithm can be used. Furthermore, it is an important study about how the various parameters are generated and communicated. Indeed, from the hierarchical model, we can formulate three different game theoretic job allocation and execution problems:

1) **Intra-Site Job Execution Strategies:** This problem concerns about the strategies of the participating computers inside a Grid site. Specifically, although the various computers participate in the making up of the Grid site, each individual computer is selfish in that it only wants to execute jobs from local users but does not want to contribute to the execution of remote jobs.
   For example, even though a cluster of PCs in the computer science department is designated as one of the member computer of a university-based Grid site, the cluster's administrators and/or users may still prefer to dedicate the computing time to process local requests as much as possible. However, if every participating computer does not contribute, the Grid site as a whole will fail to deliver its promise as a serving member of the Grid community, thereby defying the original motive of forming the Grid. Thus, one of the participating computer eventually has to take up a job assigned to the Grid site by the global scheduler.
   This problem is interesting in that we need to determine how a participating computer should formulate its job execution game theoretic strategy so as to maximize its own utility (i.e., execute more local jobs) without rendering the whole site non-operational. We focus on this problem in this paper.

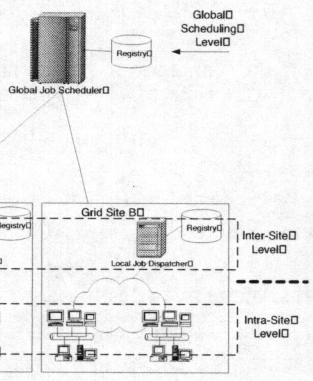


Fig. 1. A hierarchical structure of a large-scale open Grid computing platform.

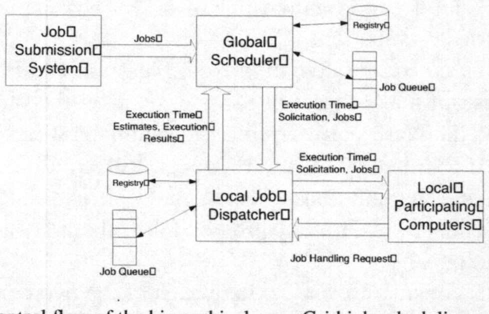

Job Submission System
Jobs
Global Scheduler
Registry
Job Queue
Execution Time Estimates, Execution Results
Execution Time Solicitation, Jobs
Execution Time Solicitation, Jobs
Registry
Local Job Dispatcher
Local Participating Computers
Job Queue
Job Handling Request

Fig. 2. Control flow of the hierarchical open Grid job scheduling.

2) **Intra-Site Bidding:** This problem concerns about the determination of the advertised "execution capabilities" for jobs submitted to the global scheduler. Recall that for the scheduler to allocate jobs using a certain scheduling algorithm, it needs to know all the sites' execution capabilities—in our study, these are modeled as the execution times needed for the pending jobs. To determine the execution time needed for a certain job, within a Grid site each participating computer can make a "declaration"—a notification to the local job dispatcher specifying the time needed to execution the job.
   The local job dispatcher can then "moderate" all these declarations to come up with a single value to be sent to the global scheduler. For example, if the local job dispatcher is aggressive in job execution, it could use the "minimization" approach—taking the minimum value of the declarations from all the member computers. On the hand, a conservative approach is to perform "maximization"—taking the maximum value instead.
   This problem is also interesting in that we need to analyze, possibly using auction theory, to determine the best strategies for each member computer in "bidding" (i.e., making execution time declarations). Specifically, we need to determine whether *truthful revelation* is the best approach in the bidding process.
3) **Inter-Site Bidding:** Similar to the intra-Site situation, at the inter-site level, the various local job dispatchers also need to formulate game theoretic strategies for computing the single representative value of the job execution time to be sent to the global scheduler.

Another exciting avenue of research is to study the inter-play of these three games, i.e., how the selfishness of each individual computer affects the intra-site bidding, which, in turn, will impact the inter-site bidding in a complicated manner.

Indeed, different combinations of the above games will result in different Grid structures. For a *semi-selfish* Grid, the intra-site games are non-cooperative while the inter-site game is cooperative. This model fits most nowadays' Grid situation because a Grid is usually formed after some cooperative negotiations at the organization level. However, the individual machines operated by bottom-level departments may not cooperate among each other. For a *fully-selfish* Grid, the games are assumed to be non-cooperative at all levels. This model is the most general model. Finally, the *ideal* Grids are modeled by cooperative games at all levels.

Due to space limitations, in this paper we only present our formulation, analysis, and results on the first problem introduced above. Specifically, to simplify the model, we assume that the inter- and intra-site bidding processes, truthful mechanisms [14] are used. In subsequent papers, we will present our results on the untruthful revelation of participating machines within each Grid site and the inter-site auction problem.

## IV. Semi-Selfish Cooperation Mechanisms and Mixed Strategies

In this section, we present our analytical formulation of the game theoretic framework for the intra-site job execution mechanism. We first describe the job model and execution policies.

We then formulate the 2-player case, followed by the general $n$-player case. Game theoretic algorithms induced by our analysis are formalized at the end of this section.

In our game theoretic study of Grid job scheduling, we consider a class of malleable jobs [11], each of which has following execution time model: $T(J_k) = a_k + \frac{b_k}{P}$, where $a_k$ is the serial portion of the job $J_k$ and $b_k$ is the parallel portion that can be shortened (hence, malleable) if more processors are available. That is, the execution time decreases in a linear manner as the number of processors allocated to the job increases. Thus, we assume that each job is a parallel application that can be executed using multiple processors. Consequently, the "cost" for each participating computer (e.g., possibly a cluster of PCs) in executing a job is the number of processors, $P$, devoted to the job, during the job's execution time.

To model the "selfish" behavior of each participating computer (i.e., each player) in a Grid site $j$, we propose the following *utility function*:

$$U_i = \frac{P_i^t}{P_i^r} \tag{1}$$

where $U_i$ is the utility of player $i$, $P_i^t$ is the total number of processors in player $i$, and $P_i^r$ is the total number of processors it used for a remote job. Here, we assume that $P_i^r > 0$ because there is always some overhead for a computer to participate in the Grid (e.g., need to expend some processing resources to monitor the Grid status, or to advertise its capabilities, and so on). Essentially, each machine is selfish in the sense that it does not want to contribute to the Grid community if possible by minimizing the utilization of the machine by remote jobs. However, the Grid site as a whole would like to maximize its *Reputation Index* (RI) which quantifies the contributions of the site.

Specifically, the RI value $R_j$ will be incremented if an assigned job is successfully executed at site $j$ and decremented if the job fails (the failure of a job will be elaborated below). In the following, we propose our novel formulation of this assigned job execution mechanism as a non-cooperative game [19] to study the dynamics of the conflicting goals of the selfish machines and the Grid site as a whole.

In our model, we assume that after a job is assigned to a Grid site, the job is associated with an *execution deadline* in that the job can be held in the job queue at the local job dispatcher for a certain period of time. Let us denote this time by $2\tau$. We elaborate the rationale behind this policy in Section IV-B. Thus, in the execution game, there are two rounds of "moves". Within each round, each computer acts according its selfish strategy and it can choose to either ignore the job or take it up.

We consider *mixed strategies* [19] in our study. Essentially, each computer uses a probabilistic "wait-and-see" approach—try to avoid the work by waiting, with a certain probability, for some other computer to take it up. Now, consider that if a job is taken up immediately after it is assigned, the amount of resources occupied is given by: $P^r = P_o + Q$, where $P_o$ is the fixed overhead component of resources but $Q$ is a variable component which depends on how much time is left for the job (here, the player index indicated by the subscript is dropped for clarity). Specifically, if the job is taken up immediately after assignment, then $Q = P$ where $P$ is the number of processors needed in order to finish the job using the amount of time advertised by the Grid site to the global scheduler. On the other hand, if the job is executed after one round (i.e., $\tau$ units of time) because no computer takes it up in the first round, then the number of processors involved becomes: $P^r = P_o + Q = P_o + P + P_w$. That is, the waiting time $\tau$ has to be compensated by "throwing in" $P_w$ more processors to the job. Let us consider a simple scenario first—only two computers are involved.

## A. The 2-Player Game

Let us consider two participating computers, denoted by $M_1$ and $M_2$, having mixed strategies $s_i$, where $0 \leq s_i \leq 1$ for $i = 1, 2$. Here, $s_i$, called the *degree of cooperation* (DoC) in our study, is the probability (i.e., the mixed strategy) that the assigned job is taken by computer $M_i$. Now, in the first round, if $M_1$ chooses not to take up the job, there are two possible outcomes: (1) $M_2$ takes it up; or (2) $M_2$ also does not take it up. Suppose that after the first round, if the job is not taken up, $M_1$ will take it up with probability 1. As such, we have:

$$Q = s_1 P + (1 - s_1)(1 - s_2)(P + P_w) \tag{2}$$

By symmetry, a similar expression can also be derived for $M_2$. Suppose $P_w = \alpha P$, where $0 < \alpha < 1$ (i.e., $\tau$ is not a long period of time with respect to the job's execution time; elaborated in Section IV-B). Here, $\alpha$ is called the *selfishness penalty factor* because it quantifies the amount of extra resources incurred should the machine refuses to take up the job earlier. Differentiating $U_1$ with respect to $s_1$ gives:

$$\frac{\partial U_1}{\partial s_1} = -\frac{P}{(P^r)^2}(s_2(1 + \alpha) - \alpha) \tag{3}$$

Depending on the value of $s_2$, $\frac{\partial U_1}{\partial s_1}$ takes on different values:

1) $s_2 < \frac{\alpha}{1+\alpha} \Longrightarrow \frac{\partial U_1}{\partial s_1} > 0$: $M_1$'s best "execution strategy" is "always do it", i.e., $s_1 = 1$.
2) $s_2 > \frac{\alpha}{1+\alpha} \Longrightarrow \frac{\partial U_1}{\partial s_1} < 0$: $M_1$'s best "execution strategy" is "always ignore", i.e., $s_1 = 0$.
3) $s_2 = \frac{\alpha}{1+\alpha} \Longrightarrow \frac{\partial U_1}{\partial s_1} = 0$: $M_1$'s best "execution strategy" is any one of the two possible actions, i.e., it is indifferent.

*Theorem 1:* The strategy combination $(s_1, s_2) = (\frac{\alpha}{1+\alpha}, \frac{\alpha}{1+\alpha})$ achieves a *Nash Equilibrium* [19] in the 2-player game, i.e., no player can benefit by unilaterally deviating from this strategy combination.

**Proof:** Theorem is true because $Q = P$ by Equation (2) and thus, $U_i$ does not depend on the value of $s_i$ (for $i = 1, 2$) under this symmetric combination. (Q.E.D.)

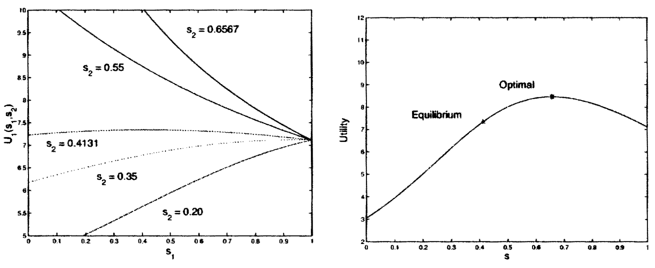
It should be noted that deviating from the Nash equilibrium strategy does not make the utility worse. In fact, the only requirement of the Nash equilibrium is that unilateral deviation does not lead to a better utility. However, this equilibrium is a weak one and the solution is degenerated [19] in that each player $i$ can choose any strategy provided the other player fixes its strategy to be $\frac{\alpha}{1+\alpha}$.

Now, let us consider the case where each of the two computers is patient enough to wait for one more time interval $\tau$ (i.e., the absolute deadline) before committing itself to take up the job. Thus, the variable component of the number of processors involved becomes:

$$
\begin{aligned}
Q &= s_1P + (1 - s_1)(1 - s_2)(s_1P + P_w + \\
&\quad (1 - s_1)(1 - s_2)(P + 2P_w))
\end{aligned} \tag{4}
$$

With some sample numerical values (i.e., $P^t = 256$, $P_o = 4$, $P = 32$, and $\alpha = 0.5$), Figure 3(a) shows the relationships among $U_1$, $s_1$, and $s_2$. We can draw a number of conclusions:

- If $M_1$ always takes up the job, i.e., $s_1 = 1$, its utility is independent of $M_2$'s strategy $s_2$;
- The maximum value of the utility increases with $s_2$;
- For small values of $s_2$, the optimal strategy for $M_1$ is to always take up the job;
- For large values of $s_2$, the optimal strategy for $M_1$ is to always wait;
- For some values of $s_2$, the optimal strategy for $M_1$ is the interior of the strategy space, i.e., $s_1 \in (0, 1)$.



(a) $U_1(s_1, s_2)$ versus $s_1$ with various values of $s_2$

(b) Utility versus $s$

Fig. 3. Relationships among the utility function $U_1$ and DoP $s_1$ of machine $M_1$, and the DoP $s_2$ of machine $M_2$.

Furthermore, it can be shown that the best execution strategy with variable $s_2$ for $M_1$ is:

$$s_1 = \frac{2(1 + 2\alpha)(1 - s_2)^2 - (1 - \alpha)(1 - s_2) - 1}{2(1 + 2\alpha)(1 - s_2)^2 - 2(1 - s_2)} \tag{5}$$

If we take $s_1 = s_2$ and $\alpha = 0.5$ so as to solve this cubic equation, we can get only one real root: $s_1 = s_2 = 0.4131$.

However, again this Nash equilibrium is sub-optimal, as evident by the following analysis. Let us take $s_1 = s_2 = s$ in Equation (4) and substitute it into the utility function $U$ (here, the subscript is dropped because of symmetry). We then consider the case of setting $\frac{\partial U}{\partial s} = 0$ under this "enforced" symmetrical strategy combination. We have:

$$4(1+2\alpha)s^3 - 3(3+8\alpha)s^2 + 2(4+13\alpha)s - 2(1+5\alpha) = 0 \tag{6}$$

Solving this cubic equation with $\alpha = 0.5$, we also get only one real root: $s = 0.6567$. Figure 3(b) shows the variation of the utility function with symmetrical strategies (i.e., $s_1 = s_2$). We can see that the optimal strategy is $s_1 = s_2 = \hat{s} = 0.6567$, while the Nash equilibrium strategy is $s_1 = s_2 = 0.4131$. Thus, the Nash equilibrium utility is *Pareto inefficient* [19], which is a common characteristic in non-cooperative game models. Fortunately, under our hierarchical scheduling model, we can make use of the local job dispatcher to guide the players (i.e., the participating computers) to use the optimal strategy. This is elaborated in Section IV-D below.

Here, we assume that there are only exactly two rounds of moves. We can easily extend the analysis to the general $\infty$-round case where there are infinite number of rounds [12]. However, as explicated below, in practice a 2-round policy is more appropriate.

## B. The Two-Round Policy

In the above analysis, the selfishness penalty factor $\alpha$ is defined as: $\alpha = \frac{P_w}{P}$. It can be shown that:

$$\frac{\alpha}{1+\alpha} = \frac{\tau}{\Gamma} \tag{7}$$

where $\Gamma$ is the execution time for the parallel fraction of the job. Here, first of all, we can see that with a fixed value of $\alpha$, $\tau$ varies from one job to another. Secondly, as $\alpha$ gets larger, $\tau$ becomes a larger fraction of $\Gamma$.

Indeed, with $\alpha = 0.1$ (i.e., 10% more processors are needed to finish the job after each round), $\tau$ is equal to 9.1% of $\Gamma$. On the other hand, with $\alpha = 0.5$ (i.e., 50% more processors are needed to finish the job after each round), $\tau$ is equal to 33.3% of $\Gamma$. Thus, with $\alpha = 0.5$, after two rounds of waiting, 66.7% of originally useful execution time is wasted and 200% of the originally needed resources are needed to finish the job. In view of this, it is deemed to be reasonable to consider that the job is rejected if it is not taken up by any player after two rounds. As detailed in Section IV-D, a rejected job is re-scheduled by the global scheduler to a possibly new site in the next batch.

## C. The $n$-Player Game

We can easily extend the 2-player 2-round game to the $n$-player scenario, i.e., there are $n$ participating computers in a Grid site. Specifically, we have the following theorem.

*Theorem 2:* The variable component of the number of processors involved in the $n$-player game is:

$$Q = P(s_i + (s_i + \alpha)\prod_{j}^{n}(1 - s_j) + (1 + 2\alpha)\prod_{j}^{n}(1 - s_j)^2) \tag{8}$$

Thus, the symmetric Nash equilibrium strategy is given by:

$$\hat{s} = \frac{2(1 + 2\alpha)\xi^2 - (1 - \alpha)\xi - 1}{2(1 + 2\alpha)\xi^2 - 2\xi} \tag{9}$$

where $\xi = (1 - \hat{s})^{n-1}$.

**Proof:** Can be easily shown by mathematical induction on $n$. (Q.E.D.)

On the other hand, the optimal strategy is given by the following theorem.

*Theorem 3:* The optimal symmetric strategy $\hat{s}$ is given by the unique legitimate real root[2] of the following equation:

$$1 - n(s+\alpha)(1-s)^{n-1} + (1-s)^n - 2n(1+2\alpha)(1-s)^{2n-1} = 0 \tag{10}$$

[2]When $n$ is odd, there is only one real root; when $n$ is even, there are three real roots but only one of them is within the (0, 1) interval.

**Proof:** Can be easily shown by mathematical induction on $n$. (Q.E.D.)

Again, we can also easily extend the 2-round case to the $\infty$-round case for this $n$-player scenario [12]. Indeed, we use the generalized $\infty$-round equations in the algorithms formalized below as well as in our simulation study.

## D. Algorithms for Intra-Site Game Players

Using Algorithm 1, the local job dispatcher continuously communicates with the global scheduler (e.g., via some SOAP messages) to check if the global scheduler is soliciting execution time estimates for pending jobs. If so, the local job dispatcher will in turn solicit such estimates within its jurisdiction. Here, we assume that the local job dispatcher uses a conservative approach in that it uses the maximum value of the local estimates as the representative value for the global scheduler.

**Algorithm 1** Local Job Dispatcher

```
1: if global scheduler solicits "execution time estimates" for a job, $J_k$ then
2:     Solicit execution time estimates from all participating computers, $T_i(J_k)$;
3:     Return $\max\{T_i(J_k)\}$ to the global scheduler;
4: end if
5: if a job is assigned to the site then
6:     Check the currently active number of participating computers, $n$;
7:     Broadcast the value of optimal strategy $\hat{s}$ according to the optimal equation of the $\infty$-round case [12] to all the participating computers;
8:     for round = 1, 2 (i.e., a total of $2\tau$ units of time) do
9:         if a computer $M_i$ takes up the job (ties are broken randomly) then
10:            Send the job to $M_i$;
11:            Declare that the job is unavailable;
12:        end if
13:    end for
14:    if no computer takes up the job then
15:        Declare that the job fails;
16:    end if
17: end if
```

If there is a job assigned to the Grid site, the local job dispatcher will then coordinate the intra-site job execution game. Specifically, to enforce the participating sites to use the optimal strategy, it first determines the value of $\hat{s}$ based on the current number of active participants (i.e., $n$) in the site. Then, it waits to see if within two rounds (i.e., $2\tau$ units of time) the job is taken up by some participant. If so, the job is handed over to the volunteer; otherwise, the job is declared as failed and the global scheduler is notified. Consequently, the global scheduler needs to re-schedule the job in the next batch, inevitably leading to a longer overall makespan for the whole set of jobs. Furthermore, the global scheduler will deduct the RI value of the concerned Grid site, which in turn will hurt the reputation of Grid site. Corresponding to Algorithm 1, each participating machine uses the Algorithm 2 to play the intra-site game.

**Algorithm 2** Participating Machine (Player $i$)

```
1: if local job dispatcher solicits "execution time estimates" for a job, $J_k$ then
2:     Return the local estimate of $T_i(J_k)$ using the value of desired number of processors $P$ to the local job dispatcher;
3: end if
4: if a job is assigned to the site then
5:     Receive from the local job dispatcher the broadcast value of optimal strategy $\hat{s}$ derived from the generalized $\infty$-round case [12];
6:     for round = 1, 2 (i.e., a total of $2\tau$ units of time) do
7:         if Random < $\hat{s}$ then
8:             Declare that this machine takes up the job;
9:             Execute the job;
10:        end if
11:    end for
12: end if
```

## V. Simulation Setup

This section describes the experimental setup in our performance evaluation of the three strategies: *Optimal*, *Nash*, and *Random*. The Optimal strategy is based on Algorithms 1 and 2. The Nash strategy is also based on the same algorithms but with the $s_i$ values computed according to Nash equation of the $\infty$-round case instead of using the optimal $\infty$-round equation. The Random strategy models the situation where all the players are completely uncoordinated and use *heterogeneous* $s_i$ values randomly generated from a uniform distribution in the range $[0, 2s_{\text{Nash}}]$.

A hierarchical semi-selfish Grid infrastructure is simulated using a discrete event-driven simulator. At the inter-site level, $m$ cooperative Grid sites are simulated. At the intra-site level, a variable number of selfish players are simulated for each site, as governed by $n$ which is the mean number of players in our simulation platform.

Jobs are submitted to a centralized job scheduler. Each site reports their required processing times, in a "truthful" manner [19], to the centralized scheduler. The well known Min-Min scheduling heuristic [4] is used for the inter-site scheduling. Specifically, for each job, the Grid site that gives the earliest *Expected Time to Completion* (ETC) is identified first. Then the job that has the minimum earliest ETC is selected and then assigned to the identified Grid site. According to the 2-round policy, a job may be rejected repeatedly without being executed even after multiple scheduling batches. Thus, we incorporate a policy in our simulator that enforces a selected Grid site to execute a job which has been rejected three times.

We use three months of accounting records for the 128-node iPSC/860 located in the Numerical Aerodynamic Simulation (NAS) Systems Division at NASA Ames Research Center [15]. This trace contains 92 days data, gathered in year 1993. There are 16,000 jobs in the whole trace. For testing the job execution performance under a high-throughput Grid environment, the 92 days trace data is proportionally squeezed to 46 days. We map the 128 nodes to 12 Grid sites—four of the sites each contain 16 nodes, and the other eight sites each contain 8 nodes. Our simulations are based on the arrival time, job size, and runtime data provided by the trace. This trace was sanitized to remove the user specified information and pre-processed to correct for system downtime [15].

In our experiments, the initial values of RI at all sites are set to 0.5. The RI at each site is then updated for every batch process. Using the following equation, the new RI value of site $j$ is calculated from the old value plus some new input value gathered from that batch.

$$\mathrm{RI}_j^{\mathrm{new}} = \mathrm{RI}_j^{\mathrm{old}} + \left(\beta_1 \frac{W_j^1}{W_{\mathrm{total}}} + \beta_2 \frac{W_j^2}{W_{\mathrm{total}}} - \gamma \frac{W_j^r}{W_{\mathrm{total}}}\right) \quad (11)$$

where $W_{\mathrm{total}}$ is the total workload processed by all sites in a batch, $W_j^1$, $W_j^2$, and $W_j^r$ represent the workload accepted in the first round, accepted in the second round, and rejected eventually by site $j$, respectively. The corresponding weighting factors are $\beta_1$, $\beta_2$, and $\gamma$, respectively, which are all positive real numbers. In our study, they are set to 1, 0.5, and 1. Based on the RI updating rule above (i.e., Equation (11)), those sites that accept more jobs (in particular, accept jobs at the first round) will increase their reputation quickly, and those sites that reject more jobs will have their RIs declining rapidly.

## VI. Performance Evaluation Results

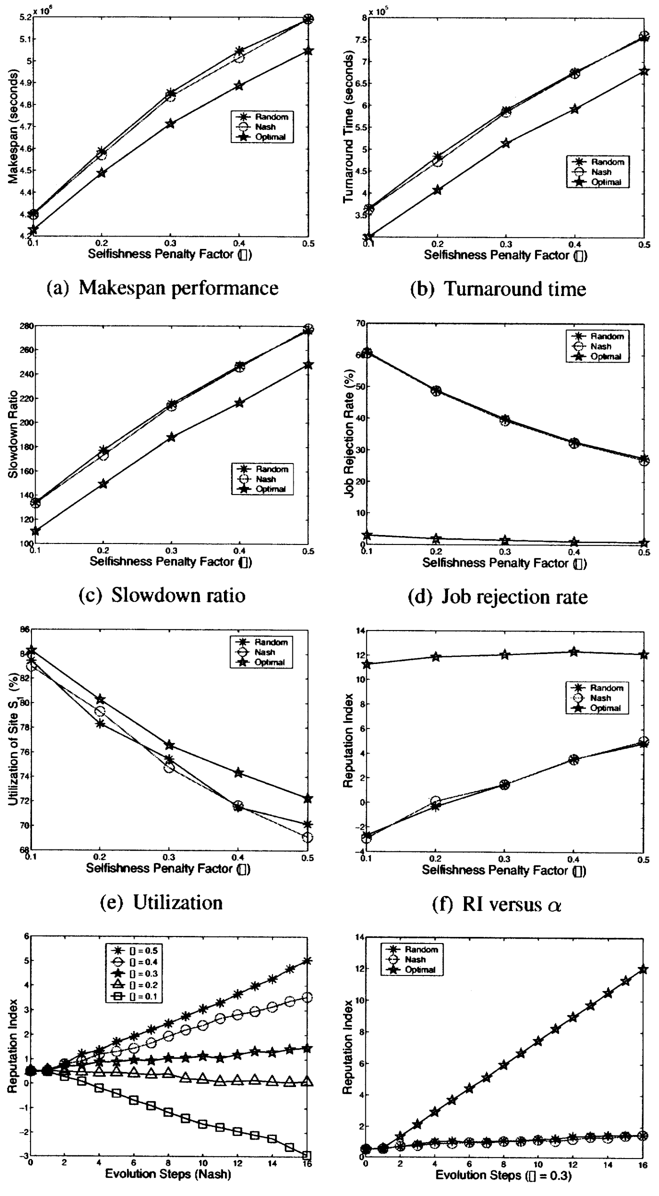
In this section, we present our simulation results over the NAS workload for the three strategies: *Random*, *Nash*, and *Optimal*. We evaluate the overall system performance using the following metrics:

- *Makespan*: The largest finish time among all the jobs;
- *Turnaround Time*: The average time spent by a job in the system;
- *Slowdown Ratio*: The ratio of the average turnaround time to the average waiting time of all jobs;
- *Reputation Index*: Defined in Section IV;
- *Utilization*: The fraction of resources used by remote jobs; and
- *Job Rejection Rate*: The percentage of jobs rejected by a site.

Due to space limitations, we can only show the results of the effects of varying values of the selfishness penalty factor $\alpha$. The full-set of results can be found in [12]. As indicated in Figure 4, a major observation is that the Optimal strategy consistently outperforms the Random and Nash strategies by a considerable margin. For example, the Optimal strategy's performance is 5 to 18% better than the other two strategies in terms of makespan, turnaround time, and slowdown ratio. Furthermore, the job rejection rate of Optimal is only around 2 to 3% but those of Random and Nash are around 40% on average.

Thus, we have an important conclusion: it is not necessarily bad for the machines to behave selfishly provided that they all use the same optimal mixed strategy values $s_i$ computed by our analysis. Another interesting conclusion is that the Nash equilibrium strategy is quite poor—almost the same as the Random strategy. Indeed, although there is no incentive for each player to deviate from the Nash equilibrium, the resulting equilibrium performance is not much different from that of a totally uncoordinated job execution scenario.

As to the effects of $\alpha$, we can see that as a larger $\alpha$ is used (i.e., heavier penalty), the makespan, turnaround time, slowdown, and utilization increase, while the job rejection rate decreases. This can be explicated by the fact that as the penalty is heavier, more time is needed to compensate for the refusal of job execution. For the RI, we can see from Figure 4(f) that the Optimal strategy is robust to the variation of $\alpha$, while a larger $\alpha$ leads to a higher RI in the Random and Nash strategies. On the other hand, as the RI evolution illustrated in Figures 4(g) and 4(h) shows, a larger selfishness penalty factor is needed for the Nash strategy but the Optimal strategy generates a linearly increasing RI.



(a) Makespan performance (b) Turnaround time

(c) Slowdown ratio (d) Job rejection rate

(e) Utilization (f) RI versus $\alpha$

(g) RI (Nash strategy) (h) RI evolution

Fig. 4. The effects of the selfishness penalty factor $\alpha$ on the performance of non-cooperative Grid machines under the NAS workload.

## VII. Conclusions

We have presented a novel and general hierarchical Grid computing model by taking machine selfishness into account. Our model matches well with real-life administrative structure of a practical Grid computing platform which is open and owned by a large number of autonomous management units organized at the inter-site and intra-site levels.

In this paper, we focus on the intra-site level to present a detailed mathematical analysis of the selfish behavior of individual machines within a Grid site. Nash equilibrium and optimal strategies are analytically derived. Using the analytical results obtained, we have performed extensive simulations to study the overall system performance under a wide range of parameters.

We have reached two important conclusions. Firstly, it is not necessarily bad for the machines to behave selfishly provided that they all use the same optimal mixed strategy values $s_i$ computed by our analysis. Secondly, the Nash equilibrium strategy is quite poor—almost the same as the Random strategy. Indeed, although there is no incentive for each player to deviate from the Nash equilibrium, the resulting equilibrium performance is not much different from that of a totally uncoordinated job execution scenario.

Finally, the question as to how to link the selfish factors in these Grids with appropriate trust negotiation models for Grid computing is still a wide open research problem. We are currently working on integrating the three different games into a unified framework which is then incorporated into our previously proposed trust binding methodology [24].

## References

[1] A. Akella, S. Seshan, R. Karp, and S. Shenker, "Selfish Behavior and Stability of the Internet: A Game-Theoretic Analysis of TCP," *Proc. ACM SIGCOMM 2002.*

[2] D. K. Barry, *Web Services and Service-Oriented Architectures*, Morgan Kaufmann, 2003.

[3] F. Berman, G. Fox, and T. Hey (Eds.), *Grid Computing: Making the Global Infrastructure a Reality*, John Wiley & Sons, 2003.

[4] T. D. Braun, H. J. Siegel *et al.*, "A Comparison of Eleven Static Heuristics for Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems," *J. Parallel and Distributed Computing*, vol. 61, 2001, pp. 810–837.

[5] J. Bredin, R. T. Maheswaran, C. Imer, T. Basar, D. Kotz, and D. Rus, "A Game-Theoretic Formulation of Multi-Agent Resource Allocation," *Proc. ACM Agents 2000.*

[6] R. Buyya, *Economic-based Distributed Resource Management and Scheduling for Grid Computing*, Ph.D. Thesis, Monash University, Melbourne, Australia, Apr. 2002.

[7] K.-M. Chao, R. Anane, J.-H. Chen, and R. Gatward, "Negotiating Agents in a Market-Oriented Grid," *Proc. CCGrid 2002.*

[8] A. Czumaj and A. Ronen, "On the Expected Payment of Mechanisms for Task Allocation," *Proc. ACM PODC 2004.*

[9] P. Ghosh, N. Roy, S. K. Das, and K. Basu, "A Game Theory Based Pricing Strategy for Job Allocation in Mobile Grids," *Proc. IPDPS 2004.*

[10] D. Grosu and A. T. Chronopoulos, "Algorithm Mechanism Design for Load Balancing in Distributed Systems," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 34, no. 1, Feb. 2004, pp. 77–84.

[11] L. V. Kalé, S. Kumar, and J. DeSouza, "A Malleable-Job System for Time-Shared Parallel Machines," *Proc. CCGrid 2002.*

[12] Y.-K. Kwok, S. Song, and K. Hwang, "Non-Cooperative Grids: Game-Theoretic Modeling and Strategy Optimization," submitted to *IEEE Trans. Parallel and Distributed Systems*, Dec. 2004.

[13] K. Larson and T. Sandholm, "Miscomputing Ratio: Social Cost of Selfish Computing," *Proc. AAMAS 2003*, pp. 273–280.

[14] D. Lehmann, L. I. O'Callaghan, and Y. Shoham, "Truth Revelation in Approximately Efficient Combinatorial Auctions," *Journal of the ACM*, vol. 49, no. 5, Sept. 2002, pp. 577-602.

[15] V. Lo and J. Mache, "Job Scheduling for Prime Time vs. Non-Prime Time," *Proc. Cluster Computing 2002.*

[16] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, "Experiences Applying Game Theory to System Design," *Proc. ACM SIGCOMM 2004.*

[17] N. Nisan, "Algorithms for Selfish Agents: Mechansim Design for Distributed Computation," *Proc. 16th Symposium on Theoretical Aspects of Computer Science* (Lecture Notes in Computer Science, vol. 1563), pp. 1–17, 1999.

[18] N. Nisan and A. Ronen, "Algorithmic Mechanism Design," *Proc. ACM STOC 1999*, pp. 129–140.

[19] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, 1994.

[20] C. H. Papadimitriou, "Algorithms, Games, and the Internet," *Proc. ACM STOC 2001.*

[21] K. Ranganathan, M. Ripeanu, A. Sarin, and I. Foster, "Incentive Mechanisms for Large Collaborative Resource Sharing," *Proc. CCGrid 2004.*

[22] O. Regev and N. Nisan, "The POPCORN Market—An Online Market for Computational Resources," *Proc. ACM ICE 1998.*

[23] T. RoughGarden and E. Tardos, "How Bad is Selfish Routing?," *Journal of the ACM*, vol. 49, no. 2, Mar. 2002, pp. 236–259.

[24] S. Song, K. Hwang, and Y.-K. Kwok, "Security Binding for Trusted Job Outsourcing in Open Computational Grids," *IEEE Trans. Parallel and Distributed Systems*, revision submitted in Dec. 2004.

[25] D. E. Volper, J. C. Oh, and M. Jung, "GameMosix: Game-Theoretic Middleware for CPU Sharing in Untrusted P2P Environment," *Proc. PDCS 2004.*

[26] M. P. Wellman, J. K. Mackie-Mason, D. M. Reeves, and S. Swaminathan, "Exploring Bidding Strategies for Market-Based Scheduling," *Proc. ACM EC 2003.*

[27] R. Wolski, J. S. Plank, T. Bryan, and J. Brevik, "G-Commerce: Market Formulations Controlling Resource Allocation on the Computational Grid," *Proc. IPDPS 2001.*

[28] M. K. H. Yeung, *On Channel Adaptive Wireless Cache Invalidation and Game Theoretic Power Aware Wireless Data Access*, M.Phil. Thesis, The University of Hong Kong, Aug. 2004.

## Authors' Biographical Sketches

**Yu-Kwong Kwok** is an Associate Professor in the Department of Electrical and Electronic Engineering, University of Hong Kong (HKU). Dr. Kwok is currently on leave from HKU and is a Visiting Associate Professor at USC. His research interests include Grid computing, mobile computing, wireless communications, network protocols, and distributed computing algorithms. He received the 2003-2004 Outstanding Young Researcher Award from HKU. Dr. Kwok is a Senior Member of the IEEE. He can be reached at *ykwok@hku.hk*.

**ShanShan Song** received her B.S. degree in Computer Science from Special Class for Gifted Young in University of Science and Technology of China in 2001. She is currently pursuing the Ph.D. degree in the Department of Computer Science at University of Southern California. She specializes in P2P networks, network security, database systems, parallel and distributed computing, and knowledge management. Her current research activities cover the areas of trust management in Grid and P2P systems, security-driven scheduling algorithms, cooperative game strategies, and Grid computing systems. She can be reached via Email: shanshan.song@usc.edu or visit the web site: http://www-scf.usc.edu/~shanshas/.

**Kai Hwang** is a Professor and Director of Internet and Grid Computing Laboratory at the University of Southern California. He received the Ph.D. from the University of California, Berkeley. An IEEE Fellow, he specializes in computer architecture, parallel processing, Internet and wireless security, and distributed computing systems. He has authored or coauthored 7 scientific books and 180 journal/conference papers in these areas. Hwang is the founding Editor-in-Chief of the *Journal of Parallel and Distributed Computing*. Currently, he is also an Associate Editor of the IEEE Transactions on Parallel and Distributed Systems. He has performed advisory and consulting work for IBM Fishkill, Intel SSD, MIT Lincoln Lab., ETL in Japan, and GMD in Germany. Presently, he leads the NSF-supported ITR GridSec project at USC. The GridSec group develops security-binding techniques for trusted job scheduling in Grid computing, distributed IDS and pushback of DDoS attacks, fuzzy-logic trust models and selfish Grid Computing models, and self-defense software systems for protecting network-centric computing resources. Professor Hwang can be reached at *kaihwang@usc.edu* or through the URL: http://GridSec.usc.edu/Hwang.html.