Sharing Resources and Costs in Computational Grids

Costas Coucoubetis* and Richard Weber†

1. INTRODUCTION

The growing popularity of Grid and cloud computing infrastructures arises from the fact that they enable computing resources to be shared more efficiently than do traditional stand-alone infrastructures. While the communications and engineering technology with which to implement these infrastructures is now well-developed, there remain many unanswered economic questions about operations. In this paper we investigate questions about how efficiently to allocate resources amongst a population of participants, each of whom has some private information about the value it places upon the resources it obtains. Each participant would like to benefit as best as possible from use of the shared resources, net of any contribution that it is required to make to the costs of the infrastructure's formation and maintenance. This means that the participants' individual aims are not aligned with that of maximizing system efficiency as a whole. We address the issues of how to operate a shared facility, and to charge participants for their use of it, so that its overall value to the participants is maximized, subject to being able to pay for its initial and running costs.

In a resource sharing facility there is unavoidable conflict between participants; each cares only for the resource that he receives whenever he requests some from the facility. This means that there must be a well-defined *sharing policy* for resolving conflicts. This could be an egalitarian policy that simply allocates each participant an equal share of the resource; or it could be more sophisticated policy that makes use of information that the participants provide about the values they place on obtaining resource. It is also necessary to cover the cost of the building and operating the facility. One way to cover the cost of building a facility is to require participants to pay fees. Another way is to add together actual resources that participants contribute; in this case we might operate a policy in which we ask each participant to choose for himself a quantity of resource that he will contribute to a shared pool of resource, and then say that it will be shared, at each instant, amongst any participants who wish to draw on it, in proportion to the sizes of those participants' contributions. The participant who contributes will receive more. This begs questions: might it be better to share the resource in some other way, say in proportion to the squares, or cubes, of the participants' contributions? How does the declared sharing policy affect the sizes of the contributions that participates will decide to make? It is questions like this that we address.

If a facility is already in place then its running cost must be recovered by collecting fees from the participants. We might begin by asking each participant how much value he places on obtaining resource from the facility. If one participant (which we shall henceforth call an 'agent') states that he places a greater value on obtaining resource from the facility than does another agent, then he should probably receive a greater share of the resource whenever those two agents are simultaneously wishing to draw on it, but he should also pay a greater fee. The challenge is to come up with a way of charging fees, and of sharing the resource, which incentivizes agents to be truthful and which generates fees sufficient to cover the cost of the facility. Other important factors include what accounting information is available to the policy maker and can be used as input to the policy decisions.

The problem of policy design is certainly not trivial, as observed in [3], [5], [7]. As we see in what follows, the choice of sharing policy affects the strategies that agents will use for reporting their actual resource needs, or in choosing the sizes of their contributions to a shared resource pool. Simple policies may perform very badly and induce a great amount of free riding (i.e., they can encourage agents to adopt strategies of seeking 'something for nothing'). There is recent work in [2] and [11] regarding the definition of accounting requirements for Grids, which in turn would influence the types of poli-

*Department of Computer Science, Athens University of Economics and Business, Athens 11362, GREECE, courcou@aueb.gr

†Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB2 0WB, UK, rrw1@cam.ac.uk

cies that can be implemented. In this paper we look at a number of models, making different assumptions about the parameters that can be measured, and obtain optimal policies for each model. For instance, the frequency with which a participant requests resources may be an important parameter. If this parameter can be measured, then we can incentivize a participant to declare it truthfully by threatenting to fine him if measurements of his his parameter do not match up with what he had declared. However, if the facility is not able to make measurements by which to police participants' declarations this way, then we may take another approach. The form of the policy itself can be designed to incentivize truthful revelations. This is done indirectly, by offering each participant a choice of options and then observing which of them he chooses.

What is new in this paper is the formulation of the models for designing optimal policies, the connection with optimal auction design, results for a large number of participants, and comparison of the various approaches. We show that simple policies are not optimal and that efficient policy design should not be trivialized; policy parameters play important role in the final outcome. Our results are for simple models with few parameters, and these are not intended to be fully realistic. However, they demonstrate general features that good policies should have. More work is needed to translate these to something practical and directly implementable. This work should refine the results in this paper and investigate their impact on job scheduling policies, see [6] and [15].

1.1 Grid technology and virtual resource sharing

Grid technology is about 'resource virtualization' [13], i.e., about providing a layer of abstraction between the physical computing resources and the applications that use them. This commoditizes computation since it does not matter either on what particular hardware an application runs, or where that hardware is physically located. An application requires only that it is run on a specific number and type of virtual machines, irrespectively on how these are implemented and where they are physically located. As a result, organizations who participate computational Grids can share virtual data centers, in which resources that are unused by one participating organization can be used by other participants. This aggregation of computing and software resources offers linear scalability: adding a computer in some organization results in increasing the total resource pool. It also allows for large economies of scale and scope since few large data centers can serve many individual organisations and reduce the cost of IT per organization. This is the driving force behind cloud computing and enables the new 'Internet of Services' vision in which software services are offered by competitive providers that charge per use and are run somewhere on the shared computing infrastructure.

Since organizations have computational demands that fluctuate over time and scope, there can be a great overall cost reduction when organizations obtain resources from a central shared facility (be it actual or virtual) as compared to each organization building its own smaller facility. A shared facility will have a size near the sum of the average resource requirements of the participants, whereas if they individually install capacity the sum of the sizes of their installations will be near the sum of their peak requirements. In addition to the savings in hardware there can be savings in software since the same programs can be reused by the participants in this virtual organization. Note, however, that the savings occurs because demand for computation fluctuates; if demand were constant, then no statistical multiplexing would take place and the advantage of using a shared facility would be greatly reduced.

1.2 Economic issues

Suppose that n organizations are considering their participation in a facility (possibly virtual) which provides a pool of shared resource. Our model allows each organization to value resources usage differently. Specifically, we suppose that if organization i obtains a quantity of resource x it obtains benefit $\theta_i u(x)$. The function $u(x)$ is the same for all participating organizations; we suppose it is increasing and concave. However, the value of the parameter θ_i is known only by organization i; we say that it is this organization's 'private information'. The parameter θ_i captures the importance that organization i places on obtaining resources. The economic problem for the facility designer is to efficiently share the virtual resources in a context that each organization is behaving strategically in order maximize its own net benefit. Key decisions must be taken about (a) how to incentivize organizations to participate, (b) what fees, or actual amounts of resource, the participants should be required to contribute, (c) how resources should be allocated when more than a single organization wants to draw on the shared pool simultaneously, and (d) how the cost of running the facility should be shared. In this work we address the above issues in the context of *collaborative grids*. These are shared facilities that are managed with the objective of maximizing the sum of the total benefit obtained by the participants who share the facility; this is in contrast to maximizing profit of individuals or of the facility manager.

One possible approach to sharing computational resources is to form an open market for computation, see [16], [14]. In this market providers (sellers) and consumers (buyers) of computing resources go to trade. In this case incentives for resource contribution are rather clear and driven by direct profit. The market may operate similarly to the stock market, except that commodities are perishable. For instance, an organization might go to the market and say that it needs 10 virtual machines of a certain type for 8 hours and state that the maximum price it is willing to pay is 100 euros. This corresponds to a 'bid' in this market. Similarly, an organization that has excess computing capacity can make a similar posting in the market but by specifying the minimum price it is willing to sell. This corresponds to an 'ask' in this market. The market matches the asks and bids, just as in the stock market, and allocates resources accordingly. If this market is relatively competitive, then it will also result in efficient allocation of resources, see [17], [18]. Organizations will base their decision on how much infrastructure to self-procure and how much to get from the market based on the equilibrium market price and on the statistics of their demand for computation. Alternatively, large sellers/buyers of capacity might participate in specialized auctions (like in e-Bay) to sell/buy resource contracts for immediate or future use.

Our approach differs from the above, but is complementary. It is not based on a competitive market; rather it *regulates* the system by setting rules to which participants must abide and a policy for sharing the resource pool. It is appropriate when *a given set of organizations* decide to collaborate over a long period of time, to do one of the following.

(i) share the cost of running a given facility;

(ii) create a new shared virtual facility by each contributing actual computing resources from the start (or by providing finance for purchasing and maintaining those resources).

Case (ii) is common in large e-science projects, e.g., [1], [4], [9], [12], and in other virtual facility building projects like OneLab [8] and PlanetLab [10]. This approach may be preferred to the free market approach when organizations prefer long-run predictable contracts and to make contributions in kind (infrastructure), rather than participating in dynamic markets in which prices fluctuate and yearly expenses are not predictable.

Our approach for problems of types (i) and (ii) is based on the theory of optimal auctions [19]. We design an auction in which agents' bids determine resource-sharing contracts. These contracts specify what quantities of resource each agent will obtain in each possible circumstance that some subset of agents wish to draw on the resource pool simultaneously. The parameters of the contracts become finalized only after all agents have made their bids. The auction is engineered so that each participant is incentivized to bid truthfully, i.e., to reveal the true values of his personal parameter θ_i. The resulting contracts provide optimal resource sharing, subject to a constraint that the fees paid by the agents will cover the cost of the system. In this model the rules of running the system are defined as functions of the bids of of the participants. Thus we are engaged in what is known as 'mechanism design' [20]. We are effectively seeking to design rules for a game (in which the agents are invited to play strategically) such that at the resulting Nash equilibrium of this game th economic efficiency is maximized, subject to covering the cost of running the facility. The optimal schemes can be nonintuitive. In Section 6 we give an example to show that it can be optimal to allocate the resource amongst agents in a different way than would occur if the agents were permitted to trade the resource amongst themselves within an internal market.

Let us add a word of caution about the application of auction theory. In optimal auction design one asks that certain constraints be satisfied ex-ante rather than ex-post. This means that they are to be satisfied in expectation *before the auction takes place*, but that they may not hold at the end. For instance, an agent who decides to participate because (ex-ante) he has a positive expected benefit may find (ex-post) that after the auction takes place he winds up with a negative benefit. This happens, for example, in horse-race betting. A gambler backs a horse because he forecasts positive expected benefit in doing so, but he winds up out of pocket if his horse does not win. This may be of concern if the auction takes place once at the start and defines the benefit of a participant for the long run. If this benefit turns out to be negative ex-post, then the participant will have good reason to break his contract and leave the system.

Fortunately, the above issues are not a problem when our optimal policy requires the same set of participants to run the auction repeatedly for a large number of times. By the law of large numbers each participant will experience his expected net benefit and this is positive by construction. This is the case in model (i) mentioned above. Unfortunately, this does not hold in model (ii). Although agents access the shared facility in a repeated fashion, the auction that defines the sharing rules is played just once at the beginning of time when this facility is formed. Hence a participant who has ex-ante positive expected gain may find that ex-post this is no longer true. He is motivated to leave the system and take back his resource contribution. Fortunately, these issues become of lesser concern when number of participants becomes large. This is precisely the case we analyze fully in Section 5.

2. THE FULL INFORMATION CASE

We begin by considering a problem of efficiently sharing a fixed quantity of computing resource Q amongst a set of agents, $N = \{1, \dots, n\}$. Suppose that the daily cost of operating the system is c. Time proceeds in discrete epochs, $1, 2, \dots$, such as hours or days. Suppose that if at epoch t an agent i is allocated resource x_i then he obtains utility (or can generate revenue) of $\theta_{i,t}u(x_i)$. At epoch t, the value of $\theta_{i,t}$ is realized as a i.i.d. sample from a probability distribution with a continuous density f_i and a distribution function F_i. The distributions $F_1, \dots, F_n$ are public knowledge, i.e., they are known to all agents and the system designer.

In general, $\theta_{i,t}$ is private knowledge, known only by agent i, and not until time t. However, in the 'full information' we suppose that the vector $\theta = (\theta_{1,t}, \dots, \theta_{n,t})$ is fully known to the system operator. Knowing it, he can then maximize the social welfare, defined as

$$\sum_{i=1}^{n} \theta_{i,t}u(x_i)\,, \tag{1}$$

by simply computing the optimal allocation vector

$$\begin{aligned} x(\theta) &= \big(x_1(\theta), \dots, x_n(\theta)\big) \\ &= \arg \max_{\substack{x_1,\dots,x_n \\ \text{s.t.} \sum_{i=1}^{n} x_i \le Q}} \left\{ \sum_{i=1}^{n} \theta_{i,t}u(x_i) \right\}. \end{aligned} \tag{2}$$

This allocation vector maximizes at time t the social welfare and generates benefit $\theta_{i,t}u(x_i)$ for participant i. If the facility is shared for a single step t, then assuming that $\sum_i \theta_{i,t}u(x_i) > c$, he can ask for payments p_i such that $p_i \le \theta_{i,t}u(x_i)$ and $\sum_i p_i = c$. If the same set of agents share the facility continuously, we compute the optimal allocation at each time t from (2) but ask agents to make constant payments at each t. We simply require that these satisfy, for each i,

$$E\left[\theta_{i,t}u(x_i(\theta))\right] - p_i \ge 0\,, \tag{3}$$

and cover the cost, i.e., $\sum_i p_i = c$. In (3) the expected net benefit of participant i is his long run average benefit when for each t we use the allocation computed in (2). If (3) cannot be satisfied, then there is simply no solution to our problem that can cover cost c.

The above full information solution to the problem is called the 'first-best' solution since it achieves the highest possible economic efficiency. Unfortunately, in practice the $\theta_{i,t}$ are private information of the agents and they will act strategically if asked to reveal them. An agent might choose to declare a value of $\theta_{i,t}$ that is greater than its true value in order to obtain a larger resource share. This means that in practice a game takes place amongst the agents. Agent i declares $\theta_{i,t}$ and his payoff is his expected net benefit. As the designer of the rules of the game we wish to arrange that the Nash Equilibrium of this game is a point that is as economically efficient as possible. This amounts to finding appropriate functions $x(\theta) = (x_1(\theta), \dots, x_n(\theta))$ and $p(\theta) = (p_1(\theta), \dots, p_n(\theta))$. These define the rules of the game in respect to how resource allocations and the payments depend upon participants' declarations. At the equilibrium, these should satisfy the following properties.

1. Agents should find it in their interest to be truthful in declaring their $\theta_{i,t}$.
2. Agents should see positive expected net benefit from participation.
3. Expected total payments should cover the cost c.
4. Expected social welfare (total net benefit) should be maximized among over all possible choices of $x(\theta)$ and $p(\theta)$ that satisfy 1–3 above.

Thus far we have supposed that the facility is pre-existing, and of size Q. If there is the possibility of choosing Q then things are a bit different. Now the game is played once; the size of the facility is determined by the payments of the agents and then agents continuously share the facility.

3. SHARING A FACILITY AND COVERING A FIXED RUNNING COST

In this section we investigate the problem of how to efficiently share a facility's resources while simultaneously recovering its daily running cost, say c, from the agents who benefit All agents have private information in the manner described in Section 2. We look at two approaches, each of which is appropriate in certain circumstances.

In Section 3.1 we suppose that the same set of agents share the facility over a long period of time. As regards agent i, it is public information that $\theta_{i,1}, \theta_{i,2}, \dots$ are i.i.d. samples from a distribution F_i. The facility operator can therefore check that the values of $\theta_{i,1}, \theta_{i,2}, \dots$ that agent i declares have an empirical distribution equivalent to F_i. We refer to this capability of the system as *policing* the parameter θ_i. We show that under the combination of policing and an optimal sharing policy the agents are incentivized to truthfully report their parameters at all times. This means that if policing is possible then the system can be run just as efficiently as in the full information case of Section 2.

In Section 3.2 we suppose that policing is not possible. This might be because a virtual facility is formed for relatively short time and so there is not be enough time to detect whether or not an agent declares $\theta_{i,1}$, $\theta_{i,2}$, ... consistent with F_i. Or perhaps the set of participants changes over time and one cannot enforce such long term contracts. In these cases the solution requires the use of the theory of optimal auctions. We must solve the problem independently for each time t. However, we begin with the simplest of the two approaches, looking first at the what can be done with policing.

3.1 Policing declarations

The key idea in this section is that we can incentivize agents to be truthful by *policing* their declarations. We do this do by checking that the empirical distribution of agent i's declared $\theta_{i,1}, \theta_{i,2}, \ldots$ converge to F_i, and threatening to impose a very large fine if this is not the case. Once we know that the agents are truthful, the problem trivializes since we can then use (2) to make an optimal resource allocation for each vector $\theta = (\theta_{1,t}, \ldots, \theta_{n,t})$. Agent i obtains in the long run the average utility of $E\left[\theta_{i,t} u(x_i(\theta))\right]$. Any fixed per-period payment vector p can meet the demands of cost recovery, provided that $\sum_i p_i = c$ and $E\left[\theta_{i,t} u(x_i(\theta))\right] - p_i \geq 0$.

It remains to check that the combination of the allocation mechanism (2) and the policing mechanism described above does actually incentivize agents to be truthful. To check this, we start by noticing that the payment of p_i that is to be taken from agent i is fully determined by public knowledge of $F_1, \ldots, F_n$, and so it does not depend on the agent's declarations of $\theta_{i,1}$, $\theta_{i,2}$, $\ldots$. Let us now consider whether it could be advantageous for agent i to decide that whenever his parameter is $\theta_{i,t} = \theta_i$ he will declare it to be $\theta'_{i,t} = \theta'_i$ (possibly even randomizing this declaration). For convenience let us drop the subscript t and think about a typical period. The policing mechanism constrains the extent to which agent i can be untruthful about θ_i because θ'_i must have the same distribution as θ_i. Subject to this constraint, the agent wishes to maximize $E[\theta_i V(\theta'_i)]$, where $V(\theta'_i)$ denotes the expected value of $u(x_i(\theta))$, conditional on agent i declaring his parameter to be θ'_i.

To see that agent i does best by always being truthful, i.e., with $\theta'_i = \theta_i$, note that because $u(x)$ is concave increasing in x and $x_i(\theta)$ is determined by (2) the function $V_i(\theta'_i)$ must be nondecreasing in θ'_i. We can now apply the Hardy-Littlewood rearrangement inequality[1] to obtain

$$E[\theta_i V(\theta'_i)] \leq E[\theta'_i V(\theta'_i)] = E[\theta_i V(\theta_i)]\,.$$

This shows that there is no incentive for the agent to be other than truthful in his declarations.

3.2 Optimal auctions

Now we turn to the more difficult circumstance in which it is not possible to policy the parameters $\theta_{i,t}$. We suppose that at each time t there is a possibly new set of agents, say N_t, and if agent $i \in N_t$ is allocated resource x_i then he can generate revenue $\theta_i u(x_i)$. Again we assume that the value of θ_i is realized as a sample from a probability distribution with a continuous density f_i and a distribution function F_i, which is public knowledge. Once again, the facility operator wishes to design a mechanism that allocates the resource as efficiently as possible while simultaneously covering the daily running cost of c.

Let us focus upon what happens at some time t. The resource Q is to be determined by an auction-like mechanism amongst the set of agents N_t. Suppose $N_t = \{1, \ldots, n\}$. The agents declare their values of $\theta = (\theta_1, \ldots, \theta_n)$. The declared value of θ_i, say θ'_i, can be viewed as a bid by agent i, and he may declare θ'_i as different to θ_i if he thinks this will make him better off. The rules of the auction are described by functions of the bids: the agents pay amounts $p_1(\theta), \ldots, p_n(\theta)$, and receive amounts of resource $x_1(\theta), \ldots, x_n(\theta)$, where these must satisfy $\sum_i x_i(\theta) \leq Q$.

An alternative, and more practical, interpretation of the above mechanism can be described in terms of usage contracts that relate agents' payments to the amounts of resource they are allocated. These contracts are parameterized by θ and the agents' make bids (i.e., agent i bids by specifying θ_i) and these determine the choice of contract. A feature of this system of contracts is that the actual details of the implementation of a contract (e.g., amount of payment, resource share) become specific only after all participants have made their bids. Agents make bids inputs based upon their expectations of the bids that other agents will made. Before proceeding to some general theory we now describe an example.

3.2.1 Example: sharing a machine and covering its running cost

Suppose the resource that is to be shared is a single machine. We focus on a single day in which there are just two agents contending for this resource. Agent i has a utility for the machine of θ_i, $i = 1, 2$. The values of θ_1 and θ_2 are private information, but it is public knowledge that they are independent and uniformly distributed as $\theta_1 \sim U[0,1]$ and $\theta_2 \sim U[0,2]$. We wish to allocate the machine in a most efficient way while generating payments whose expected sum is c, which is the cost of running the machine for one day.

The theory of optimal auction provides the following prescription as to what is best to do. (Details of the theory are postponed until in the next section, when we put this solution in a more general context.) What we should do is to consider a class of schemes, parameterized by a number β, where $0 \leq \beta \leq 1/2$. In scheme(β) we ask the agents to reveal their θ_i, we calculate

$$\phi_1 = \max\{\beta, \theta_2 - \beta\} \text{ and } \phi_2 = \max\{2\beta, \theta_1 + \beta\}\,,$$

and then allocate the machine as follows (this being well-defined since it is impossible to have both $\theta_1 > p_1$ and $\theta_2 > p_2$):

- If $\theta_1 > \phi_1$ then the machine is allocated to agent

[1] The Hardy-Littlewood rearrangement inequality is a generalization to integrals of the simple fact that if $a_1 \leq \cdots \leq a_n$ and $b_1 \leq \cdots \leq b_n$ then $\sum_i a_{\pi(i)}\, b_i \leq \sum_i a_i b_i$ for any permutation π of the indices $1, \ldots, n$.

1, who pays ϕ_1,

i.e., $x(\theta) = (1, 0)$ and $p(\theta) = (\phi_1, 0)$.

- If $\theta_2 > \phi_2$ then the machine is allocated to agent 2, who pays ϕ_2,

  i.e., $x(\theta) = (0, 1)$ and $p(\theta) = (0, \phi_2)$.

- Otherwise the machine is not allocated, i.e., $x(\theta) = p(\theta) = (0, 0)$.

These rules produce maximal expected payment when $\beta = 1/2$ and this is $P = E[p_1(\theta) + p_2(\theta)] = \frac{31}{48} = 0.6458\dot{3}$. Let $U = E[\theta_1 x_1(\theta) + \theta_2 x_2(\theta)]$. The expected net benefit of the agents (total expected utility minus total expected payment) is $U - P = \frac{13}{48} = 0.2708\dot{3}$. If it is desired to cover a cost c, which is greater than $P = 0.6458\dot{3}$, then this cost cannot be covered. Notice that the machine is not always allocated to the agent who values it most. E.g., if $\theta_1 = 5/6$ and $\theta_2 = 1$ then the machine is allocated to agent 1.

If $c < P$, then we are collecting too great an expected payment from the agents if we take $\beta = 1/2$. It is possible to change the rules a bit, by moving β towards 0, so that the expected payment becomes $P' = c$ and the agents obtain a greater expected net benefit of $U' - P' > U - P$. As c decreases towards 0 it will eventually be small enough that we can cover the cost by allocating the machine to the agent with greatest θ_i, and requiring him to pay $\min\{\theta_1, \theta_2\}$. This is the rule with $\beta = 0$, and it is optimal for $c \leq \frac{5}{12} = 0.41\dot{6}$. The net benefit is then $\frac{13}{12} - c = 1.08\dot{3} - c$. It corresponds to the Vickrey mechanism in which the highest bidder wins and pays the second highest bid. That is, if $\theta_1 > \theta_2$ then agent 1 is allocated the machine and pays $p_1 = \theta_2$.

It is interesting to compare the above to a different scheme which might be used by a non-sophisticated facility operator. In this scheme (which we call A1) we post a price p and ask both agents whether they are willing or not to pay this price in return for use of the machine. If just one is willing then he is allocated the machine and pays p. If both are willing then the machine is allocated to agent 2 (since on average he has a greater θ_i) and he pays p. Otherwise the machine is not allocated and no payment is taken. The value of p is chosen so that the net benefit is maximized, subject to the expected payment being equal to c.

The graph below shows a plot of the expected net benefit to the participating agents against the cost that must be covered. We observe that our optimal auction-based policy can achieve a greater economic efficiency that A1, while covering the same cost. It also allows for covering a wider range of costs. This elementary example shows the complexity of the allocation rules that may be needed to share resources in a Grid environment.

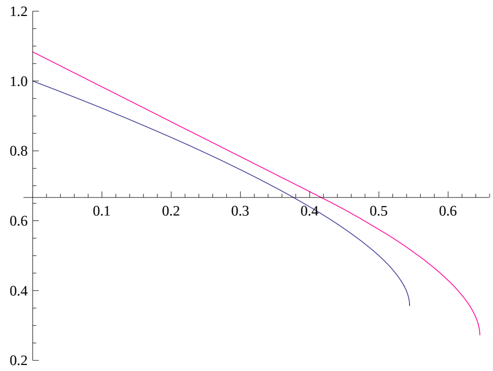


Figure 1: Plots of the expected net benefit against the cost that must be covered, c. The top graph shows the net benefit of an optimal policy. Above the horozontal axes the first best optimum can be obtained ($c \leq 0.41\dot{6}$). For a given c the optimal policy achieves greater net benefit than the policy A1, whose net benefit is shown in the bottom graph. Note that even the optimal policy cannot cover a cost $c \geq 0.6458\dot{3}$.

Remarks. Notice that the above scheme has the desirable property of being ex-post rational, in that we do not take any payment from the agent who is not allocated the machine, and the agent who does get the machine pays less than it is worth to him. This may be a reasonable requirement in Grids where the identity of participating agents changes at all times.

In collaborative Grids in which the same agents are sharing the facility over time and we need not impose such strong ex-post requirements. In this case we need only require weaker property of ex-ante rationality, i.e., that on each agent has positive net benefit when he uses the facility, by having a positive benefit in the long run average.

So far as recovery of the cost c is concerned, note that above we have $E[p_1(\theta) + p_2(\theta)] = c$, but not $p_1(\theta) + p_2(\theta) = c$ for all θ. So the daily running cost may not be covered exactly on a given day; however, since the facility runs in a repeated fashion, we know by the Law of Large Numbers, it will also be covered in the long run, and this is all that we require.

Taken together with Section 3.1, the previous discussion shows that operating policies may be very sensitive to modelling details. Policing is a practical option only when the same set of agents with known profiles are sharing the facility in a repeated fashion; this gives a special structure to the problem that allows us to achieves the same efficiency as in the full-information case. However, if agents change from day to day, or policing is not possible then we must to do something more interesting and nontrivial. We have given a simple example in this subsection, and in the next subsection give details of some general theory.

3.2.2 How to efficiently share a facility and cover its running cost

In traditional auction design the usual aim is to maximize the expected revenue obtained by the seller, i.e., to maximize the expected sum of payments $E\left[\sum_i p_i(\theta)\right]$. However, we wish to do something slightly different. Rather than maximizing $E\left[\sum_i p_i(\theta)\right]$, we only require that the payments are sufficient to meet the daily running cost c. Moreover, since the facility is to be used repeatedly over many days we need not satisfy $\sum_i p_i(\theta) \geq c$ on each day, but only a weaker long-run average constraint, which can be expressed

$$E\left[\sum_i p_i(\theta)\right] \geq c\,. \tag{4}$$

Subject to (4), we seek to maximize the expected daily net benefit of the users. Focusing on a particular day, this means we wish to choose the functions $x_1, \dots, x_n$ and $p_1, \dots, p_n$, to maximize the expected sum of net benefits

$$E\left[\sum_i \theta_i u(x_i(\theta)) - p_i(\theta)\right]\,, \tag{5}$$

subject to (4). There is also an additional constraint that our choice of functions $x_1, \dots, x_n$ and $p_1, \dots, p_n$, must induce agent i to truthfully reveal his parameter θ_i when he acts in his own best interest. That is, for agent 1, we must have that for all θ_1,

$$\begin{aligned}\theta_1 = \arg\max_{\theta_1'} E_{\theta_2,\dots,\theta_n}\big[&\theta_1 u(x_1(\theta_1', \theta_2, \dots, \theta_n)) \\ &- p_1(\theta_1', \theta_2, \dots, \theta_n)\big]\,,\end{aligned}$$

and similarly for the other agents. (We may impose this constraint because the 'revelation principle', which tells us that there exists an optimal policy within the class of policies that induce agents to be truthful about their private information.) A last constraint is that agents are individually rational, i.e., agent i is not forced to participate in this resource sharing game if he does not think it is profitable for him to do so; at some initial time he can decide whether or not to participate, and do so only if his net benefit is guaranteed to be positive in the long run. This constraint can be expressed as requiring that for all θ_i,

$$E_{\theta_{-i}}\left[\theta_i u(x_i(\theta)) - p_i(\theta)\right] \geq 0\,, \tag{6}$$

where $E_{\theta_{-i}}$ denotes expectation over all θ_j, $j \neq i$, while keeping θ_i constant.

It turns out that the solution to this problem can be described as follows. First, define

$$g_i(\theta_i) = \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}\,. \tag{7}$$

We suppose that this is a nondecreasing function of θ_i. For example, if θ_i is distributed uniformly on $[0, \xi_i]$ then $g_i(\theta_i) = 2\theta_i - \xi_i$.

Then there exists a $\lambda \geq 0$ such that the optimal allocation is

$$\begin{aligned} x(\theta) &= (x_1(\theta), \dots, x_n(\theta)) \\ &= \arg \max_{\substack{x_1,\dots,x_n \\ \text{s.t.} \sum_i x_i \leq Q}} \left\{\sum_i (\theta_i + \lambda g_i(\theta_i)) u(x_i)\right\}. \end{aligned} \tag{8}$$

This comes about because the problem can be reduced to maximizing a Lagrangian of

$$L = E\left[\sum_i (\theta_i + \lambda g_i(\theta_i)) u(x_i)\right] - (1+\lambda)c\,.$$

This is maximization is carried out for every θ, within the expectation.

The optimal payments can be taken to be

$$p_i(\theta) = P_i(\theta_i) = P_i(0) + \theta_i V(\theta_i) - \int_0^{\theta_i} V(w)\,dw \tag{9}$$

where

$$V(\theta_i) = E_{\theta_{-i}}\left[u(x_i(\theta))\right]\,,$$

and $P_i(0)$ is some subsidy such that $P_i(0) \leq 0$. Note that (9) implies the constraint (6), since it implies the even stronger constraint $\theta_i V(\theta_i) - P_i(\theta_i) \geq 0$ for all θ_i.

It is interesting to compare solutions with $\lambda > 0$ and $\lambda = 0$.

(i) If $\lambda > 0$ then $E\big[\sum_i P_i(\theta_i)\big] = c$ and $P_i(0) = 0$ for all i. Note that the resource is not necessarily allocated in the same way that an internal market would allocate it. For example, suppose $n = 2$ and $\theta_1 \sim U[0, 1]$, $\theta_2 \sim U[0, 2]$. Suppose that c is such that we cover the cost when taking $\lambda = 1$. Then if $\theta_1 = 5/6$ and $\theta_2 = 1$ we will have that x_1, x_2 should be chosen to maximize $\frac{3}{2}u(x_1) + u(x_2)$. Assuming u is concave this will mean we should take $x_1 > x_2$, even though $\theta_1 < \theta_2$.

(ii) If $\lambda = 0$ then we see from (8) that the grid resource is always allocated in the most efficient way, i.e., to maximize $\sum_i \theta_i u(x_i)$. This is now the same way an internal market would allocate it. The expected sum of payments can create a surplus, say $s = E\big[\sum_i P_i(\theta_i)\big] - c > 0$. In this case we may take $P_1(0), \dots, P_n(0)$ as any quantities summing to $-s$; for instance we could share the surplus equally amongst the agents by setting $P_i(0) = -s/n$.

Remarks. Although the above gives a methodology, it is not one that is easy to apply, even in simple cases. It is not even easy to say whether or not $\lambda = 0$, although we know this depends on the size of c.

In the above, we assumed that the size of Q is given and that there is a cost c that must be covered. It is also possible to take a similar approach under the assumption that the size Q of the system is not given, but that we can choose it initially at a cost of $c(Q)$. What is the optimal size of Q? In this case, following the same ideas as above, the value of the optimal expected sum of net benefits is

$$\min_{\lambda} \max_{Q} \Bigg\{ E\Big[\max_{\substack{x_1,\dots,x_n \\ \text{s.t} \sum_i x_i \le Q}} \Big\{ \textstyle\sum_i (\theta_i + \lambda g_i(\theta_i)) u(x_i) \Big\} \Big] - (1+\lambda) c(Q) \Bigg\} \qquad (10)$$

Where we make use of the fact that the appropriate value of λ can be determined my minimizing (i.e., by solving the Lagrangian dual problem). Notice that we do not need the agents to reveal any information in order that we can determine the best Q, since it is assumed that we know the $F_1, \dots, F_n$. However, we do rely on the fact that on a daily basis the total total resource Q will be shared by the method of maximizing $\sum_i (\theta_i + \lambda g_i(\theta_i)) u(x_i)$.

4. FORMING A SYSTEM OF SHARED INFRASTRUCTURE

In this section we consider a rather different way of forming and operating a system of shared computing resource. Rather than ask agents to declare their daily parameters $\theta_1, \dots, \theta_n$, we ask each agent i, right at the start, to choose a size of contribution, say q_i, that he would like to make to a shared pool of computing resource. The size of this contribution acts as a sort of bid by agent i, and will help to determine what quantity of resources he is allocated on future days. It also serves as the payment he makes to participate in this system. This model can be used in Grids in which institutions form virtual organizations: each institution allows part of its infrastructure to be shared by other partners. This corresponds to its contribution to the virtual organization which inherits all the contributions of the participants. How much should an agent contribute in such system? Which parameters affect this decision?

We will see that the sharing policies that will be used while the system runs have a great effect on the amounts agents are willing to contribute. Essentially, unless one cleverly designs these policies, the equilibrium will be inefficient in terms of the total system size. In this section we demonstrate how this can occur, even when we have full information regarding the θ_i of the agents. We look at how the sharing policy affects agents' decisions about their initial contributions. We show it is rather subtle to design rules that incentivize agents to make appropriate contributions. Regarding the design of optimal policies, we have been able to do so only for systems with a very large number of participants, where the law of large numbers makes the analysis tractable. These results are in Section 5.2 where we analyze a case in which the θ_i are private information.

To keep things simple we consider a model in which the sequence $\theta_{i,1}, \theta_{i,2}, \dots$ are simply i.i.d. Bernoulli random variables, distributed $B(1, \alpha_i)$. The idea is that on any given day, agent i may be either 'active' or 'inactive', i.e., have either full or no interest in using the system. These two events occur with probabilities α_i and $1 - \alpha_i$ respectively, independently each day and independently of the active/inactive states of the other agents. All active agents have the same utility function for resource amount x, namely $u(x)$. This is simply a special case of our model above, in which $\theta_i = 0$ or $\theta_i = 1$ with probabilities $1 - \alpha_i$ and $\alpha_i = 0$ respectively. Assume that all the above parameters are public information. How should the common resource be shared in a period when k agents are active? This question seems trivial unless one makes the connection with the incentives it invokes when the facility is formed.

Let the set of active agents be S, where $|S| = k$. Suppose $Q = \sum_j q_j$. Since all contending agents have the same concave utility function $u(x)$, it would seem sensible to take $x_i(S) = Q/k$. But is this correct? Or should the sharing policy depend on the α_i and on the initial contributions of the agents? If participants know that this policy will be followed, what quantity of resources will they contribute to the common pool?

We need to evaluate the combined effect of a sharing policy, (i.e., the efficiency with which it shares resource while the system runs) and its influence upon the initial resource contributions. Then we can compare different policies and possibly choose the optimal one. One might expect, for example, that sharing resource amongst agents in proportion to their initial contributions provides better incentives and hence greater efficiency than does simply apportioning the resource equally amongst agents.

In Section 4.1 we analyze a simple numerical example that has just 2 agents, but in which all the above issues arise. In Section 4.2 we investigate what happens with an equal sharing policy when there are n agents with, $\theta_i = 0, 1$, but with different $\alpha_1, \dots, \alpha_n$. Finally, in Section 4.3 we look at subscription pricing, in which all participants are charged the same fixed fee.

4.1 An example with two identical agents

Suppose there are just two agents. When agent i is active and is allocated resource x_i then he obtains revenue of $u(x_i)$. Suppose the cost of buying resource is linear, $c(Q) = Q$ and that agents contribute q_1, q_2. Let x_i^S be the share of resource given to agent i when the set of active agents is S. The average net benefit of agent 1 per period is

$$\alpha_1(1-\alpha_2)u(x_1^{\{1\}}) + \alpha_1\alpha_2 u(x_1^{\{1,2\}}).$$

If we take $x_i^{\{i\}} = x_i^{\{1,2\}} = q_1$ then we model agents acting alone. The 'equal division' discipline is $x_i^{\{i\}} = q_1 + q_2$ and $x_i^{\{1,2\}} = \frac{1}{2}(q_1+q_2)$.

When there are just two identical agents, and $\alpha_1 = \alpha_2$, then we would expect that under any reasonable scheme they should contribute equally and share equally. However, it matters what the scheme is. Consider two agents with full information, each with $\alpha_1 = \alpha_2 = \alpha$. Suppose $u(x) = r - 1/x$. In this case the social optimum is achieved by a social planner that chooses the contributions $q_1 = q_2 = q$ to maximize

$$\alpha\left(r - \frac{1-\alpha}{2q} - \frac{\alpha}{q}\right) - q\,. \tag{11}$$

For $r = 10$ and $\alpha_1 = \alpha_2 = 0.8$ this has an optimum net benefit per agent of 6.3029, for $q = 0.8485$. Acting alone each agent would maximize

$$\alpha_i\left(r - \frac{1}{q}\right) - q\,, \tag{12}$$

an agent obtains expected net benefit of $\alpha_i 10 - 2\sqrt{\alpha_i} = 6.21115$, for $q = 0.8944$. However, sharing a facility with the 'equal division' mechanism, agents 1 and 2 maximize respectively

$$0.8\left(10 - \frac{1.8}{q_1+q_2}\right) - q_1\,, \quad 0.8\left(10 - \frac{1.8}{q_1+q_2}\right) - q_2\,,$$

with respect to q_1 and q_2. There is equilibrium for any (q_1, q_2) such that $q_1 + q_2 = 1.2$. If we require $q_1 = q_2$ then this is $q_1 = q_2 = 0.6$, and each agent has net benefit 6.2. This is less than the 6.21115 they obtain acting alone. In fact, when $n = 2$, two identical agents will prefer to act alone for all $\alpha_1 = \alpha_2 > 7/9$.

The problem becomes worse as the number of agents increases. With $n = 10$ identical agents each contributes $q_i = 0.256125$ and the net benefit per agent is 5.18263. Once n is as large as 98 then the equilibrium is driven to a point where agents no longer have positive net benefit. They will start deserting the system.

We have made a surprising observation: two identical agents can obtain greater net benefit by acting on their own than by participating in a shared system in which their contributions are determined as the Nash equilibrium of a nonzero-sum game. As we have seen above, the social welfare obtained by 'equal shares division' can be less than stand alone for $\alpha > 7/9$. With $\alpha = 0.8$ the stand alone welfare is 6.21115 and the grid welfare is only 6.2. This is because the incentives are wrong and each agent tries to be a partial free-rider. How might we improve things? One way is with proportional sharing.

Proportional sharing

Let us try a different mechanism: $x_i^{\{i\}} = q_1 + q_2$ and $x_i^{\{1,2\}} = q_i$. This corresponds to a scheme in which the resource is divided amongst agents in proportion to their contributions, rather than in equal shares. Now the equilibrium is at $q_1 = q_2 = 0.824621$ and the social welfare is $8 - 7/\sqrt{17} = 6.30225$, which is better than the stand alone welfare. The only change is that we have replaced $x_i^{\{1,2\}} = (1/2)(q_1+q_2)$ with $x_i^{\{1,2\}} = q_i$.

By solving (11) we find the central planner optimum achieves social welfare is $8 - 6\sqrt{2}/5 = 6.30294$ at $q_1 = q_2 = 0.848528$. This is 0.0006931 better than in the scheme we described immediately above.

Consider now a scheme that shares resource proportionally to sth powers of the contributions. That is,

$$x_i^{\{i\}} = q_1 + q_2\,, \quad x_i^{\{1,2\}} = \frac{q_1^s}{q_1^s + q_2^s}(q_1+q_2)\,.$$

Equal division is $s = 0$. Proportional division is $s = 1$. What about other values of s? It turns out that the equilibrium point is increasing in s. At $s = 9/8 = 1.125$ the equilibrium point is exactly the same as that of the social optimum. (Note if $q_1 = q_2 = q_0$ then $x_i^{\{1,2\}} = q_0$, for any s.) In fact, this works for any α when we take $s = \frac{1}{2}(1 + 1/\alpha)$. Note that this means taking $s \geq 1$. The reason is that for $u(x) = r - 1/x$, the optimal q for the social planner is where

$$\frac{d}{dq}\left[\alpha\left(r - \frac{1-\alpha}{2q} - \frac{\alpha}{q}\right) - q\right] = \frac{\alpha + \alpha^2 - 2\,q^2}{2\,q^2} = 0\,,$$

and under the s-powers proportionate scheme q_1 is found by solving

$$\frac{d}{dq_1}\left[\alpha\left(r - \frac{1-\alpha}{q_1+q_2} - \frac{\alpha}{\frac{q_1^s}{q_1^s+q_2^s}(q_1+q_2)}\right)\right]_{q_1=q_2=q}$$
$$= \frac{\alpha + \alpha^2 + 2\,s\,\alpha^2 - 4\,q^2}{4\,q^2} = 0\,.$$

These are the same condition when $s = \frac{1}{2}(1 + 1/\alpha)$.

Let us summarise in the following table. The numbers are for $r = 10$, $\alpha = 0.8$. Note that for all α the optimal values of q_1, q_2 are decreasing as we go through the schemes in the order stand alone, central planner, proportional division, equal division.

There are other decentralised schemes that are better than sharing in proportion to contributions. For example, let $q_0 = \sqrt{\alpha(1+\alpha)/2}$ be the minimum contribution that a central planner would choose for agents 1 and 2. We could take

$$x_i^{\{1,2\}} = q_1 + q_2 1_{\{q_1 \geq q_0\}}$$
$$x_i^{\{i\}} = q_1$$

That is, when agent 1 is the only who is active he is allowed to use agent 2's contribution, but only if he

scheme	social welfare	values of q_1, q_2
stand alone	$r\alpha - 2\sqrt{\alpha}$	$\sqrt{\alpha}$
	6.21115	0.894427
central planner	$r\alpha - \sqrt{2\alpha(1+\alpha)}$	$\sqrt{\alpha(1+\alpha)/2}$
$s = \frac{1}{2}(1 + 1/\alpha)$	6.30294	0.848528
proportional	$r\alpha - \frac{\sqrt{\alpha}\,(3+5\,\alpha)}{2\,\sqrt{1+3\,\alpha}}$	$\frac{1}{2}\sqrt{\alpha(1+3\,\alpha)}$
division $s = 1$	6.30225	0.824621
equal division	$r\alpha - \frac{3}{2}\sqrt{\alpha(1+\alpha)}$	$\frac{1}{2}\sqrt{\alpha(1+\alpha)}$
$s = 0$	6.2	0.6

Table 1: Comparing the social welfare obtained through self-provisioning, s-proportional sharing where s is optimally chosen by a central planner with full information, proportional sharing, and equal sharing when $u(x) = r - 1/x$ and the activity frequency is α. The numbers are for $r = 10$, $\alpha = 0.8$. Note that for all α the optimal values of resource contributions q_1, q_2 are decreasing as we go through the schemes in the order stand alone, central planner, proportional division, equal division. A surprising property is that by optimally choosing s the central planner achieves the same solution as choosing the optimal q in (11).

contributes at least q_0. In the example, $q_0 = 0.848528$. This scheme achieves the same social welfare as does a central planner. The difficulty is that the scheme needs to know the parameters α_1, α_2 in order to compute q_0 (as was also the case in choosing $s = 1.25$ above). Is there a scheme we could invent that does not need this information?

4.2 The wrong incentives of equal sharing

Suppose that there are n participants and linear resource cost $c(Q) = Q$. Suppose agents have types $\theta_1 = \cdots = \theta_n$ and $\alpha_1 > \cdots > \alpha_n$. Consider again the policy of sharing resource equally. It turns out that this policy does not work well, because most agents are free-riders. Only agent 1 will have any incentive to contribute resources to the grid.

To see this, note that the problems that are to be solved by agents 1 and 2, respectively, are to maximize

$$nb_1(q) = \alpha_1\Big[\alpha_2 Eu\left(\frac{q_1 + q_2 + \cdots + q_n}{M+2}\right) + (1-\alpha_2)Eu\left(\frac{q_1 + q_2 + \cdots + q_n}{M+1}\right)\Big] - q_1$$

with respect to q_1, and to maximize

$$nb_2(q) = \alpha_2\Big[\alpha_1 Eu\left(\frac{q_1 + q_2 + \cdots + q_n}{M+2}\right) + (1-\alpha_1)Eu\left(\frac{q_1 + q_2 + \cdots + q_n}{M+1}\right)\Big] - q_2$$

with respect to q_2, where M is a random variable denoting the number of agents $3, \ldots, n$ that are present. From the fact that $\alpha_1(1-\alpha_2) > \alpha_2(1-\alpha_1)$ it follows that

$$\frac{\partial}{\partial q_1} nb_1(q) = 0 \implies \frac{\partial}{\partial q_2} nb_2(q) < 0\,.$$

This means that the only possible Nash equilibrium is one in which the contribution that agents $2, \ldots, n$ will choose to make are $\bar{q}_2 = \cdots = \bar{q}_n = 0$.

Now let M' be the number of the agents $2, \ldots, n$ who are present. For an equilibrium to exist with $\bar{q}_1 > 0$ and $q_2 = \cdots = \bar{q}_n = 0$ it would have to be that

$$\alpha_1 \frac{\partial}{\partial q_1} \sum_{m=0}^{n-1} \phi_m u\left(\frac{q_1}{m+1}\right)\Bigg|_{q_1=0} - 1 > 0$$

where $\phi_m = P(M' = m)$. This is equivalent to

$$\alpha_1 u'(0) E\left(\frac{1}{M'+1}\right) - 1 > 0\,.$$

Clearly, $E\left(\frac{1}{M'+1}\right) \to 0$ as $n \to \infty$. This means that if $u'(0) < \infty$ and n is sufficiently large then no agent will wish to make any contribution.

4.3 Subscription pricing with equal sharing

One possible scheme is to charge a flat subscription fee to any agent who wishes to participate in the system. We purchase the greatest amount of resource that the collected fees allow, and in each epoch share it equally amongst any agents who are active. This is the same as requiring an an equal size of resource contribution from all participants. Such schemes are very commonly used in practice due to their simplicity. Let us investigate how well one can do with such a scheme.

Suppose that all agents have the same activity frequency, α. We ask them each to contribute q and build a facility of size $Q = nq$. Taking the same model as in Section 4.1, namely $u(x) = r - 1/x$, we find that an agent's expected net benefit per period is

$$nb = \alpha Eu\left(\frac{Q}{M+1}\right) - q = \alpha\left(r - \frac{1+(n-1)\alpha}{nq}\right) - q\,,$$

where M is the number of other $n-1$ agents that wish to use the resource in the same period that a given agent wishes to use it. So $M \sim B(n-1, \alpha)$. This can be maximized by a social planner to

$$\alpha r - 2\alpha\sqrt{1 + \tfrac{1-\alpha}{\alpha n}}\,, \quad \text{by} \quad q = \alpha\sqrt{1 + \tfrac{1-\alpha}{\alpha n}}\,.$$

For $\alpha = 0.8$, $r = 10$, $n = 1$ and $n = 2$ this gives the same result as in the first two lines of Table 1.

Now suppose that agents differ in their activity frequency α_i, and that a priori they are uniformly distributed on $[0, 1]$. We ask every participating agent to make a fixed subscription q. There will be a minimum α, say α_q, for which it is advantageous for an agent to participate. By considering the marginal agent we have that

$$\alpha_q E_N\left[r - \frac{1 + \left(\frac{1+\alpha_q}{2}\right) N}{(N+1)q}\right] - q$$
$$= \alpha_q\left(r - \frac{1 - \alpha_q^n + (1+\alpha_q)n}{2nq}\right) - q = 0\,,$$

where N is the number of the other $n-1$ agents who have their α_i greater than α_q. So $N \sim B(n-1, 1-\alpha_q)$. The expected net benefit of all the agents is

$$\tfrac{1}{2}(1-\alpha_q^2)n\left(r - \frac{1-\alpha_q^n + (1+\alpha_q)n}{2nq}\right) - (1-\alpha_q)nq$$

For $r = 10$ we find

n	q	α_q	net benefit per agent
1	0.7366	0.0852	3.616
10	0.5418	0.0697	3.939
100	0.5184	0.0578	3.982
∞	0.5158	0.0575	3.987

Of course it would be even better to ask for a subscription fee that depends on α, which could then be policed. For example, we might charge a subscription of $\alpha\, q$. When n is large it is optimal to take $q = 1$, $\alpha_q = 0$ and the expected net benefit is about $4n$. Notice this is almost the same as is achieved by the above scheme which charges the same subscription of $q = 0.5158$ to all agents.

5. OPTIMAL INCENTIVES IN LARGE SYSTEMS

In this section we analyze different incentive issues when forming Grids with large number of participants. Using the law of large numbers allows to obtain interesting solutions of a simple and intuitive form for how to run such systems. We consider two different cases, each of which is of independent interest. In the first case we assume that the activity frequencies of the individual participants are private information, but have known distribution along the population. In the second case we assume that these frequencies are known, but the parameter θ of each participant is unknown. Again it is assumed that the distribution of the θs are known, and for simplicity we assume these fixed over time.

5.1 Optimal incentives for declaring activity frequencies

We consider the exact solution of the optimum design for systems with a large number of agents that are of the 'active/inactive' type, i.e., $\theta_{i,t} = 0$ or $\theta_{i,t} = 1$, with probabilities $1 - \alpha_i$ and α_i respectively. Let us now suppose that the types, i.e., the values of $\alpha_1, \ldots, \alpha_n$, are unknown to the system designer. He would like to elicit these as part of an incentive compatible scheme that optimally sizes a system whose cost is covered by the payments of the agents. In practice this corresponds to a case in which the central planner does not want to use accounting mechanisms for estimating, and thereby policing the α_is by penalizing agents for mis-reporting. Here it is structure of the tariffs that is to incentivize participating agent to truthfully reveal their α_is.

To obtain some interesting qualitative results regarding the optimum incentive payments we treat analytically the case of agents having $u(x) = \sqrt{x}$, and assume that a priori that the α_i are distributed uniformly on $[0, 1]$. That is, there are approximately equal numbers of agents with each value of α in the range $[0, 1]$. We also assume that cost is linear $c(q) = q$. The number of agents is very large, so we may suppose (by the law of large numbers) that we can meet demands from the common resource pool provided the total amount contributed trough payments equals the cost of the system covering the average demand. The numbers obtained in this section can be viewed as upper bounds on performance for a system with a small number of agents.

The go-it-alone solution

If agents must each go-it-alone, then an agent with parameter α will choose to build his own facility of size x to maximize

$$\alpha u(x) - x$$

and therefore take $x = \frac{1}{4}\alpha^2$. The average social welfare per agent is then

$$\int_0^1 \tfrac{1}{4}\alpha^2\, d\alpha = \tfrac{1}{12} = 0.083\dot{3}\,.$$

The full information solution

Suppose a system designer having full information decides to provide an agent with parameter $\alpha = t$ with resource $x(t)$. The expected social welfare (per agent) is

$$\int_0^1 t\, u(x(t))\, dt - \int_0^1 t\, x(t)\, dt\,.$$

So the optimum is $x(t) = 1/4$ for all t, and the resulting social welfare is $1/8 = 0.125$. It is perhaps somewhat surprising to see that a system designer will wish to allocate the same resource of $1/4$ to any agent on those occasions he is present. This is because every time any agent is present he presents an opportunity to earn benefit $\sqrt{x}$. It makes sense (by concavity of u) that the resource allocation should be the same every time.

5.1.1 *The partial information solution using optimal tariffs*

Now the designer of the system wishes to optimize the system by designing appropriate incentive compatible tariffs. Each agent chooses the tariff that is most beneficial to him. A tariff specifies the amount of resource an agent will receive each time he is active and the corresponding payment he must make initially in order to participate in such a system.

We consider the set of tariffs $q(t)$, $x(t)$ parametrized by t, the type of the customer. According to these tariffs an agent who contributes $q(t)$ gets $x(t)$ whenever he is active. The set $\{q(t), x(t) : t \in [0,1]\}$ is the set of possible choices. Thus the agent maximizes his net benefit to $f(\alpha)$, where

$$f(\alpha) = \max\left\{\max_s \left[\alpha u(x(s)) - q(s)\right], 0\right\}$$

The maximum of linear functions of α is convex in α; this is how we know $f(\alpha)$ is convex. Similar to the arguments in Section 3, we must have

$$\alpha u'(x(\alpha))x'(\alpha) - q'(\alpha) = 0\,.$$

So if an agent with parameter α^* has net benefit 0, then

$$q(\alpha) = \alpha u(x(\alpha)) - \int_{\alpha^*}^{1} u(x(s))\,ds$$

and

$$\int_{\alpha^*}^{1} q(\alpha)\,d\alpha = \int_{\alpha^*}^{1} (2\alpha - 1)u(x(\alpha))\,d\alpha\,.$$

The resource constraint is

$$\int_{0}^{1} [\alpha x(\alpha) - q(\alpha)]\,d\alpha \le 0\,.$$

and so we seek to maximize a Lagrangian of

$$L = \int_{\alpha^*}^{1} \Big[(s + \lambda(2s-1))u(x(s)) - (1+\lambda)sx(s)\Big]\,ds\,,$$

where $\alpha^* = \lambda/(1+2\lambda)$ (the value of s such that $s + \lambda(2s-1) = 0$).

Recall that we take as an example, $u(x) = \sqrt{x}$. Then the maximizing $x(s)$ is

$$x(s) = \left(\frac{2\lambda+1}{2(\lambda+1)} - \frac{\lambda}{2(\lambda+1)s}\right)^2$$

This means that L is maximised to

$$\frac{1 - 2\,\lambda^2 \log\left(\frac{\lambda}{1+2\lambda}\right) - \lambda^2}{8(\lambda+1)}\,.$$

We find the correct λ by minimizing this with respect to λ, giving $\lambda = 0.232206$. This gives for a solution in which for $\alpha \ge 0.158566$,

$$q(t) = 0.173521 + 0.0942239 \log t$$
$$x(t) = \left(0.594224 - \frac{0.0942239}{t}\right)^2$$

and $q(t) = x(t) = 0$ for $t < 0.158566$ $(= \lambda/(1+2\lambda))$.

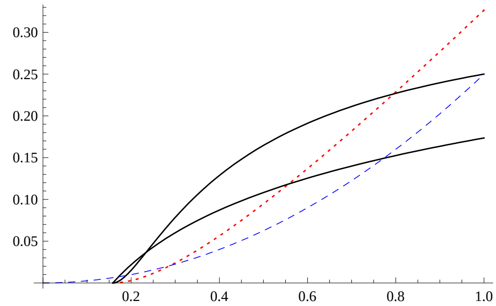


Figure 2: The black lines show $q(t)$ and $x(t)$, with $q(t) < x(t)$ when $t > 0.2339$. The dotted red line is net benefit $f(t) = tx(t) - q(t)$ and the blue dashed line is $t^2/4$, the net benefit obtained by an agent acting alone.

Remarks.

1. The social welfare obtained is 0.116121 and this is just a bit less than the social welfare of 0.125 that could be obtained by a system designer having full information.

2. The optimal scheme is one in which agents with small α (less than $\alpha^* = 0.1586$) are prevented from participating. Intuitively, the reason we need to do this is so we can incentivize the agents with greater α to make more substantial contributions. Another way to think about this is that agents with small α are the ones that are likely to be free-riders. We do better by preventing them from participating.

3. The black lines in Figure 2 show $q(t)$ and $x(t)$ (the amounts that agents will contribute and receive when declaring $\alpha = t$). Most agents receive more than they contribute. But agents with values of $\alpha \le 0.23389$ receive less than they contribute. However, if go-it-alone is not possible (because they cannot purchase and install resource for themselves) then they will still take up this scheme, since their net benefit is positive.

4. The dashed blue line is $t^2/4$, which is the net benefit an agent could obtain if he were to go-it-alone, by taking $x(t) = q(t) = t^2/4$. The dotted red line is $tu(x(t)) - q(t)$, the net benefit that an agent obtains in the shared system. This is convex, so there would be no

benefit to an agent with parameter α masquerading as being two agents with parameters $\alpha/2$. Notice that the dotted red and dashed blue lines cross; an agent does better by going alone if $\alpha \leq 0.2884$.

In the next subsection we rework this analysis under the assumption that agents can go-it-alone.

5.1.2 When agents can go-it-alone

In this section we obtain a solution in which for small α the agent obtains the same net benefit as if he were stand-alone. Although such as agent does not directly benefit from participation, his presence benefits the system because his contribution is available to others when he does not need it. So for $\alpha = t \leq t^*$ the expected benefit of an agent is the same as he obtains stane-alone, namely $f(t) = t^2/4$. Incentive compatibility implies that in this range we must have $x(t) = q(t) = t^2/4$. We wish to maximize the net benefit of

$$\left(\int_0^{t^*} t\sqrt{\tfrac{1}{4}t^2}\,dt + \int_{t^*}^1 tu(x(t))\,dt\right) - \left(\int_0^{t^*} t\tfrac{1}{4}t^2\,dt + \int_{t^*}^1 tx(t)\,dt\right)$$

subject to

$$\int_0^{t^*} \tfrac{1}{4}t^2\,dt + \int_{t^*}^1 q(t)\,dt = \int_0^{t^*} t\tfrac{1}{4}t^2\,dt + \int_{t^*}^1 tx(t)\,dt$$

where for $t > t^*$,

$$q(t) = q(t^*) + tu(x(t)) - t^*u(x(t^*)) - \int_{t^*}^t u(x(s))\,ds\,.$$

Integrating this we have

$$\begin{aligned}\int_{t^*}^1 q(t)\,dt &= (1-t^*)[q(t^*) - t^*u(x(t^*))] \\ &\quad + \int_{t^*}^1 (2s-1)u(x(s))\,ds \\ &= -\tfrac{1}{4}{t^*}^2(1-t^*) + \int_{t^*}^1 (2s-1)u(x(s))\,ds\end{aligned}$$

The Lagrangian that we wish to maximize is therefore

$$\begin{aligned}L &= \tfrac{1}{6}{t^*}^3 - \tfrac{1}{16}{t^*}^4 \\ &\quad + \lambda\Big(\tfrac{1}{12}{t^*}^3 - \tfrac{1}{16}{t^*}^4 - \tfrac{1}{4}(1-t^*){t^*}^2\Big) \\ &\quad + \int_{t^*}^1 \big[(s + \lambda(2s-1)u(x(s)) - (1+\lambda)sx(s)\big]\,ds\,.\end{aligned}$$

Within the integral, the maximizing $x(t)$ is, as above,

$$x(s) = \left(\frac{2\lambda+1}{2(\lambda+1)} - \frac{\lambda}{2(\lambda+1)s}\right)^2$$

and this must equal ${t^*}^2/4$ at $s = t^*$. Now L must be stationary with respect to t^*. These two conditions give and $t^* = \lambda/(1+\lambda)$. Substituting this value of t^* and minimizing over λ gives $\lambda = 0.279688$, $t^* = 0.21856$. The social welfare value is 0.115444, which is a 38.5% improvement on go-it-alone welfare and 92.4% of the central planning optimum. The full scheme is

$$\left.\begin{aligned} q(t) &= \tfrac{1}{4}t^2 \\ x(t) &= \tfrac{1}{4}t^2 \end{aligned}\right\} \quad 0 \leq t \leq 0.21856$$

$$\left.\begin{aligned} q(t) &= 0.178124 + 0.109281\log(t) \\ x(t) &= \left(0.609281 - \tfrac{0.109281}{t}\right)^2 \end{aligned}\right\} \quad 0.21856 \leq t \leq 1$$

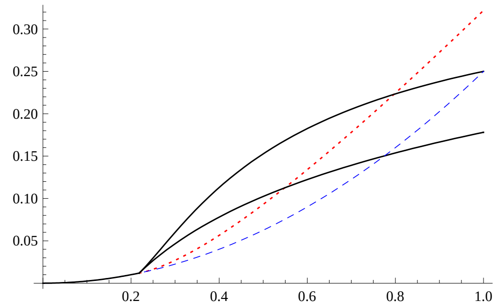


Figure 3: The black lines show $q(t)$ and $x(t)$, with $q(t) < x(t)$. The dotted red line is net benefit $f(t) = tx(t) - q(t)$ and the dashed blue line is $t^2/4$, the net benefit obtained by an agent acting alone. Unlike Figure 2, the net benefit of all agents is at least as great as they would obtain by go-it-alone.

Remarks.

1. Notice that in both Figures 2 and 3 we see $x(1) = 1/4$, which is the same as an agent with $\alpha = 1$ would choose for himself to obtain on a go-it-alone basis. Such an agent is always present and therefore the resources he contributes can never be used by other agents. At first, this might make one think that there is nothing to be gained by including him in the grid. However, we see that it is still advantageous to include him, since we can increase social welfare by letting him use some of the resources that others do not always need, and so he needs only supply resources $q(1) = 0.178124$ (Figure 2), or 0.173521 (Figure 3) rather that 0.25.

2. It is interesting that both $x(t)/q(t)$ and $f(t)/(\frac{1}{4}t^2)$ are maximized at $t = 0.548867$. This makes sense, since agents with α near 0.5 are worthy to receive the greatest proportional increase in benefit through participating in the grid. They contribute quite a lot of resources, but are still sufficiently absent that the resources they contribute can be used by others. Agents with α near 0 or 1 do not contribute much in the way of resources that others can use.

3. These calculations are for a particular u and under the assumption that the α are uniformly distributed on $[0, 1]$. We have seen that under these assumptions the agents with small α make little contribution. However, in realistic applications the distribution of α is probably not this way. It is more likely that most agents have small α. In that case, we may not find that agents with small α make no contribution to the performance of the grid. It would be interesting to repeat this analysis under the assumption that α is uniform on say $[0, 0.2]$.

5.1.3 A one contract approximation

Suppose there is only one contract on offer. An agent can decide to join the grid, in which case he must contribute q, and then gets use of $x > q$ whenever he needs resources. Essentially, we are trying to approximate the optimal $f(x)$ by

$$f(t) = \max\{\tfrac{1}{2}t^2, t\sqrt{x} - q\}$$

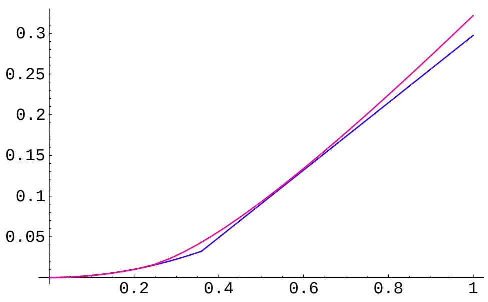
An agent with parameter α will wish to take out this contract if by doing so he has positive net benefit, and this is more than he would obtain buying resource on his own. Thus we have the requirement

$$\alpha\sqrt{x} - q \geq \max\{0, \tfrac{1}{2}\alpha^2\}$$

This means that an agent will participate if and only if

$$2(\sqrt{x} - \sqrt{x - q}) \leq \alpha \leq \min\{2(\sqrt{x} + \sqrt{x - q}), 1\}$$

Under the constraint $\int tx(t)\,dt = \int q(t)\,dt$, we find that the social welfare is maximized to 0.109503 at $q = 0.1169$ and $x = 0.1713$. Note that this is not much less than the social welfare of 0.116121 that we found for an optimal design above. The figure below compares the optimal f (red) and the optimal 'one contract' approximation (blue). It is interesting that the optimal approximation lies strictly below the optimal curve. That is, every type of agent is worse off when we make this simple approximation (except those for small $\alpha \leq t^*$). (Note that in the region where the approximation is $t\sqrt{x} - q$, the blue curve is closest to the red around $\alpha = 0.56$, but there is still a gap. These curves do not touch.)



5.2 Forming a shared facility with many participants

This section is about a result for a large number of participants and incomplete information. It uses the theory developed in Section 3.2. The new model is similar to ones that we have discussed above, in that we suppose that on a given day agent i is active or inactive with probabilities α_i and $1 - \alpha_i$, where α_i is public knowledge. However, now we suppose that each agent also has a private parameter, θ_i, which he must be incentivized to reveal truthfully at the start through his choice of a contract (parametrized by θ) from the set of available contracts. Then the agent's benefit in any epoch in which he is active and allocated resources x_i is $\theta_i u(x_i)$. Unlike the model of Section 4 θ_i is not equal to 1 but is chosen randomly from a distribution F_i. Also unlike the model in Section 3, the parameter θ_i is not redrawn every day, but remains the same for all days. Interestingly enough, we do not require the distribution F_i to be public knowledge in order to construct our optimal policy for forming and running the facility.

A system is built of size $Q(\theta)$. Agents are charged payments $p_1(\theta), \ldots, p_n(\theta)$, and the sum of these covers the cost $c(Q(\theta))$. When agent i is contending for the resource amongst a group of active agents S he receives $x_i(\theta, S)$. Following the steps in the analysis presented in Section 3.2, the solution can be characterized as follows. For some $\lambda \geq 0$, the optimal allocation amongst agents in S is found in analogy to (19) by maximizing

$$\sum_{i \in S} (\theta_i + \lambda g(\theta_i)) u(x_i)\,.$$

Agent i has expected utility $\theta_i V_i(\theta_i)$, where

$$V_i(\theta_i) = E_{\theta_{-i}}\left[\sum_{S: i \in S} \alpha(S) u(x_i(\theta, S))\right],$$

and $\alpha(S)$ is the probability of the event that exactly the set of agents in S is active. His payment should be

$$P_i(\theta_i) = \theta_i V_i(\theta_i) - \int_0^{\theta_i} V_i(w)\,dw\,.$$

His expected usage is

$$E\left[\sum_{S: i \in S} \alpha(S) x_i(\theta, S)\right].$$

The above follows from the fact that we are maximizing pointwise with respect to (θ, S),

$$E\left[\sum_S \sum_{i \in S} \alpha(S)(\theta_i + \lambda g(\theta_i)) u(x_i(\theta, S)) - (1 + \lambda) c(Q(\theta))\right]$$

over $Q(\theta)$ and $x_i(\theta, S)$, and subject to $\sum_{i \in S} x_i(\theta, S) \leq Q(\theta)$ for all S. The Lagrange multiplier λ appears in order to impose the ex-ante constraint

$$E\left[\sum_i x_i(\theta_i)\right] = E\left[\sum_S \alpha(S)\sum_{i\in S} g(\theta_i)u(x_i(\theta,S))\right]$$
$$= E[c(Q)]\,.$$

If n is large, then when agent i is active the rest of the system will be in its typical average state, i.e., the proportion of agents of different types that are active will be the nearly the same (by the law of large numbers). Then it is reasonable to look for an approximate solution in which $x_i(\theta, S)$ is independent of S and we only need to satisfy the constraint

$$\sum_i \alpha_i x_i(\theta) \le Q(\theta)\,. \tag{13}$$

This constraint says that we can pick the x_is in a way that the average resource requirements of all the granted contracts does not exceed Q.

The Lagrangian is now

$$L = E\Big[\sum_i \alpha_i(\theta_i + \lambda g(\theta_i))u(x_i) - (1+\lambda)c(Q) \tag{14}$$
$$+ \mu\left(Q - \sum_i \alpha_i x_i\right)\Big]$$

in which must have $\mu = \lambda c'(Q)$. If $c(Q) = Q$ then $\lambda = \mu$ and this reduces to

$$L = E\left[\sum_i \alpha_i(\theta_i + \lambda g(\theta_i))u(x_i) - (1+\lambda)\sum_i \alpha_i x_i\right]\,.$$

This is to be maximized with respect to $x_i \ge 0$. Recall that λ is to be chosen so that the constraint is satisfied. In fact, in this special case the constraint (13) is satisfied by taking $\lambda = 0$. We can simply let

$$x_i(\theta) = x_i(\theta_i) = \arg\max_{x_i'}\{\theta_i u(x_i') - x_i'\}, \tag{15}$$

and achieve the first-best optimum. Agent i pays $\alpha_i x_i(\theta_i)$ and this exactly pays for his own average resource usage.

This optimal tariff has a very interesting interpretation. The optimal contract chosen by agent i secures the same amount of resources from the shared resource pool as the amount he would optimally choose to self-procure under the same resource cost if no shared infrastructure was available and he was always active. But he needs only pay for his average usage, namely pay only for $\alpha_i x_i$. And by construction, this tariff is incentive compatible, i.e., he will choose the tariff parameterized by his actual value of θ. Note that the above amount x_i is greater than the size of the facility he would form if he were to stand-alone since in this case he would solve the problem

$$x_i(\theta_i) = \arg\max_{x_i'}\{\theta_i \alpha_i u(x_i') - x_i'\}\,. \tag{16}$$

Thus in this large system an agent benefits from the existence of the other agents which are not always claiming resources; he uses the optimal amount when he is active and has to pay for it only when he uses it, since others pay for it when he is not.

Another important aspect of this tariff is that it is very simple to implement and does not require knowledge of the distributions of the θ_i of the agents in order to construct the optimal tariffs. Denoting by θ the contract parameter, one simply needs to construct the resource curve $x(\theta)$ by solving (16) for all values of θ, and the family of payment curves $ax(\theta)$, for all potential values of activity frequencies. An agent with activity frequency α_i (which is assumed known because of policing) will consider the pair $(x(\theta), \alpha_i x(\theta))$ and choose the contract t that is best suited for him. A property of the tariff is that he will choose $\theta = \theta_i$. Operating the system in the limit becomes very simple and achieves the performance of the full information first-best solution.

Example..

To illustrate the above, let us suppose that $c(Q) = Q$ and $u(x) = \sqrt{x}$. Then it is easy to calculate that for $\theta_i \sim U[0,1]$, the Lagrangian is maximized by

$$x_i = \begin{cases} 0 & \theta_i \le \lambda/(1+2\lambda) \\ \frac{(\theta_i + \lambda(2\theta_i - 1))^2}{4(1+\lambda)^2}\,, & \theta_i > \lambda/(1+2\lambda) \end{cases}$$

The maximized value of the Lagrangian is

$$E\left[\sum_i \alpha_i \frac{3(\theta_i + \lambda(2\theta_i - 1))^2}{16(1+\lambda)} 1_{\{\theta_i > \lambda/(1+2\lambda\}}\right]\,.$$

Recall that we can find the correct value of λ by minimizing the above with respect to λ. One can check that for all θ_i we minimize the term within the above expectation, namely,

$$\frac{3(\theta_i + \lambda(2\theta_i - 1))^2}{16(1+\lambda)} 1_{\{\theta_i > \lambda/(1+2\lambda\}}\,,$$

by taking $\lambda = 0$. Thus, we conclude that

$$x_i(\theta_i) = \tfrac{1}{4}\theta_i^2$$
$$V_i(\theta_i) = \alpha_i u(x_i) = \tfrac{1}{2}\alpha_i\theta_i$$

and so

$$P_i(\theta_i) = \alpha_i \tfrac{1}{2}\theta_i^2 - \int_0^{\theta_i} \alpha_i \tfrac{1}{2} w\, dw = \tfrac{1}{4}\alpha_i\theta_i^2\,.$$

6. THE NONOPTIMALITY OF A ‘MARKET SOLUTION’

We have seen in the example of Section 3.2.1 that the optimal allocation of resource may not be that which an internal market would achieve, which is (2), i.e.,

$$x(\theta) = \arg \max_{\substack{x_1,\dots,x_n \\ \text{s.t.} \sum_{i=1}^n x_i \le Q}} \left\{ \sum_{i=1}^n \theta_{i,t} u(x_i) \right\} . \qquad (17)$$

Instead we use (8), i.e.,

$$x(\theta) = \arg \max_{\substack{x_1,\dots,x_n \\ \text{s.t.} \sum_{i=1}^n x_i \le Q}} \left\{ \sum_{i=1}^n (\theta_{i,t} + \lambda g_i(\theta_{i,t})) u(x_i) \right\} . \qquad (18)$$

In the following example we explore this a bit further by looking at an example in which there is also the possibility of choosing the size of the installed resource Q.

We now illustrate what can happen by investigating an example in which the calculations are tractable. Suppose $u(x) = x$ and $c(Q) = \gamma Q$, with the constraint $Q \le 1$. We suppose that there are a number of scenarios, occurring with probabilities p_i and that in scenario k the utility that agent i has for a resource allocation x is $\theta_i \alpha_{ik} u(x)$. The α_{ik} are public information, but the $\theta_{i,t}$ are private to the agents. Focusing on a single time t, and dropping the suffix t in what follows, this leads to a problem in which for a given θ we seek to maximize, over x_{ik}, with $\sum_i x_{ik} \le Q$,

$$\sum_k \sum_i \big[\theta_i + \lambda g_i(\theta_i)\big] p_k \alpha_{ik} x_{ik} - (1+\lambda) Q \, .$$

Thus, $x_{ik} = x$ if both $\theta_i + \lambda g_i(\theta_i) > 0$ and $\alpha_{ik}(\theta_i + \lambda g_i(\theta_i)) > \alpha_{jk}(\theta_j + \lambda g_j(\theta_j))$ for all $j \ne i$. Otherwise $x_{ik} = 0$. Furthermore $Q = 1$ if

$$\sum_k p_k \max_i \big\{\alpha_{ik}\big(\theta_i + \lambda g_i(\theta_i)\big)\big\} - (1+\lambda) > 0 \, ,$$

and otherwise $Q = 0$.

Suppose there just two agents and these have θ_1 and θ_2 that are ex-ante known to be uniformly distributed on the intervals $[0, 1]$ and $[0, 2]$. For these distributions, $g_1(\theta_1) = 2\theta_1 - 1$ and $g_2(\theta_2) = 2\theta_2 - 2$. For convenience we shall let $h_i(\theta_i) = \theta_i + \lambda g_i(\theta_i)$.

Let there be four equally likely scenarios ($p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$), in which the values of $(\alpha_{1k}, \alpha_{2k})$ are $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$. This is a model in which each agent is equally likely to be active or inactive, independently of the other agent. Given a value for λ, the optimal value of the Lagrangian becomes

$$\int_{\theta_1=0}^1 \int_{\theta_2=0}^2 \Big[\max\Big\{ \tfrac{1}{4} \max[h_1(\theta_1), 0] + \tfrac{1}{4} \max[h_2(\theta_2), 0] + \tfrac{1}{4} \max[h_1(\theta_1), h_2(\theta_2), 0] - (1+\lambda)\gamma \, , 0 \Big\}\Big] \tfrac{1}{2} \, d\theta_1 d\theta_2 \, .$$

This can be computed for various $\lambda \ge 0$. By finding the value of λ for which the above is minimal (i.e., by solving the Lagrangian dual problem) we can find the value that produces exact satisfaction of the constraint that expected payments cover expected cost. For $\gamma = 1$ this turns out to be $\lambda = 0.5$. For $\gamma = 1/2$ it is $\lambda = 0.1737$. Thus, the solution for this example can be characterised as follows. We suppose that the agents have declared θ_1, θ_2. Let $a_1 = \lambda/(1 + 2\lambda)$, $a_2 = 2\lambda/(1 + 2\lambda)$. These are such that $h_i(\theta_i) > 0$ if and only if $\theta_i > a_i$.

1. Let $f_1 = (\theta_1 - a_1)^+$ and $f_2 = (\theta_2 - a_2)^+$.
2. Purchase resource $Q = 1$ (at cost 1) if

   $$f_1 + f_2 + \max\{f_1, f_2\} > \frac{(1+\lambda)\gamma}{(1/4)(1+2\lambda)} \, . \qquad (19)$$

   Otherwise $Q = 0$.
3. If (19) holds, so that the resource is purchased, then charge the agents appropriate amounts, $q_1(\theta)$ and $q_2(\theta)$.

   In each scenario, whenever it arises, allocate the total resource of 1 to the agent with greatest non-negative value of $\alpha_{ik} f_i$.

For $\gamma = 1/2$, we find the optimal solution is $\lambda = 0.1737$ and $(a_1, a_2) = (0.128915, 0.25783)$. We now make an interesting observation about this solution. Suppose $(\theta_1, \theta_2) = (0.87, 0.93)$. In this circumstance (19) is satisfied, so the resource will be installed ($Q = 1$). However in the scenario 4, where both agents can use the resource, it will be allocated to agent 1, despite the fact that $\theta_1 < \theta_2$. So the resource is not allocated in the same way that a market mechanism like (17) might do. In scenario 4 such a mechanism would allocate the resource to the agent with the greatest value of θ_i.

The optimized value of the objective function turns out to be 0.19583. If had not been necessary to incentivize truthful declaration of the θ_i (the full information case), then we could have obtained 0.201389, i.e.,

$$\int_{\theta_1=0}^1 \int_{\theta_2=0}^2 \Big(\tfrac{1}{4}\theta_1 + \tfrac{1}{4}\theta_2 + \tfrac{1}{4}\max[\theta_1, \theta_2] - \gamma\Big)^+ \tfrac{1}{2} \, d\theta_1 \, d\theta_2 \, ,$$

where $(x)^+ = \max\{x, 0\}$.

6.1 The ‘market solution’ to the problem

We now consider the maximum expected welfare that can be obtained if we impose the constraint that the resource (in the case $Q = 1$) is always allocated to the agent with with greatest value of $\alpha_{ik}\theta_i$. This is the prescription of (17) and what would happen if agents could trade the resource amongst themselves. The scheme would now be as follows.

1. The agents are asked to declare θ_1, θ_2.
2. As a function of $\theta = (\theta_1, \theta_2)$, a decision is made that $Q = 0$ or $Q = 1$.

   The agents are charged $q_1(\theta)$ and $q_2(\theta)$.
3. In scenario $(1, 0)$ the full resource is allocated to agent 1, and in scenario $(0, 1)$ the full resource is

allocated to agent 2. In the scenario $(1,1)$ the full resource is given to the agent who declared the greatest θ_i.

Let Θ be the set in which the resource will be installed. Then

$$V_i(\theta_i) = \int_{\theta_2 \,:\, \theta \in \Theta} \left[\tfrac{1}{4}u(x) + \tfrac{1}{4}u(x)1_{\{\theta_1 > \theta_2\}}\right] \, f_2(\theta_2)\, d\theta_2$$

In fact the solution will be nearly as above. We look at

$$E\left[\sum_k \sum_i \big[\theta_i + \lambda g_i(\theta_i)\big] p_k \alpha_{ik} u(x_{ik}) - (1+\lambda)c(Q)\right]$$

and impose the constraint

$$(x_{1k}, x_{2k}) = \arg\max_{x_{1k}, x_{2k}} \{\alpha_{1k}\theta_1 u(x_{1k}) + \alpha_{2k}\theta_2 u(x_{2k})\}$$

The only optimization is over Q.

For $\gamma = 1/2$ this we find cost covering where $\lambda = 0.2425$ and expected welfare of 0.19384, which is 0.002 less than the 0.19583 that was obtained with the non-market solution.

7. CONCLUDING REMARKS AND POLICY CONSIDERATIONS

In this paper we have investigated policies for running shared computing resource infrastructures. We have assumed that participants wil be strategic in disclosing private information about their actual resource needs and we have considered how best to share resources and take payments from the participants so as to maximize the overall efficiency of the system and while covering its costs. We can summarize some of the lessons that we learned as follows.

1. A participant's decision about the quantity of resources that he will choose to contribute to the common resource pool of a virtual organization can be greatly affected by the resource sharing policy that he expects will be deployed when the system operates. Thus, a sharing policy which simply optimizes the efficiency of the system for a certain given quantity of resources may not be optimal. For example, if a participant thinks that the resource will be shared equally amongst all participants, irrespective of the contribution that he makes, then be may may choose to contribute nothing to the resource pool.

2. One way we can incentivize potential participants to make significant contributions to a shared resource pool is to impose a rule that a participant will only be permitted to draw on the shared pool if he makes a minimum contribution to it a the point that it is being formed, i.e., by contributing a minimum quantity of computing resources. We can further impose a sharing policy which has the consequence that an agent who contributes more resource will have greater priority for obtaining resource than an agent who has contributed less. Rules like these will incentivize agents to make contributions that reflect their privately held beliefs about the benefits they expect to obtain. The result is a facility with an appropriately large quantity of resource, which is efficiently shared. Since contribution are made in kind there is no need for any internal money transfers.

It is necessary carefully to engineer the way in which the resource pool will be shared so that each participant knows that he will obtain a greater share of the resource pool if he makes a greater contribution.

3. In a large facility with many participants it may be best to make the participation requirement sufficiently great that some subset of agents, whose benefit from participating would be small, choose not to participate. Although we lose the contributions of these agents and whatever benefit they might have obtained, the remaining agents can be incentivized to make greater contributions, and the overall effect is greater benefit for these.

4. In a facility that is already built and so has a fixed size (such as NRNs, National Grid Infrastructures), the running cost must be shared by charging the participants. In general, if the identities of the participants change over time, then it is optimal to operate a specialized market in which participants bid and get resource shares according to their particular needs at each time, while generating enough payments to cover the cost of the facility. If the set of participants remains constant, then simpler policies exist for sharing the facility and recovering the cost. We proposed such a set of policies but with the added cost of imposing some specific accounting requirements in the running of the system.

We have obtained our results for simple models with certain economic assumptions that may not always hold. For instance, there are national infrastructures which cannot charge fees to participants and so services must be offered for free. It may be that participants cannot make payments in the form proposed in this paper simply because of internal accounting restrictions. Also in many cases the actual cost of the shared facility $c(Q)$ is not precisely known, and it would take some non-trivial effort to define it.

A challenging next step is to apply the theory developed in this paper in a specific real system. In this context one needs to translate the qualitative results regarding the shape of the policies and the incentive systems into specific quantities that will determine payments and job scheduling decisions.

8. REFERENCES

[1] Distributed European Infrastructure for Supercomputing Applications. www.deisa.org.

[2] The distributed grid accounting system (DGAS). www.to.infn.it/grid/accounting/main.html.

[3] e-Infrastructures Reflection Group. www.deisa.org, www.e-irg.eu/images/stories/publ/finnishpresidency-recommendationsanddecisions.pdf.

[4] Enabling grids for E-sciencE. www.eu-egee.org/.

[5] European grid initiative. web.eu-egi.eu/www.eu-egi.eu/blueprint.pdf.

[6] gLite middleware. glite.web.cern.ch/edms.cern.ch/file/722398/1.2/gLite-3-UserGuide.pdf.

[7] Joint policy security group. proj-lcg-security.web.cern.ch/proj-lcg-security/security_policy.html.

[8] The OneLab Project. www.onelab.eu.

[9] Partnership for Advanced Computing in Europe. www.prace-project.eu/.

[10] The PlanetLab project. www.planet-lab.org.

[11] User's guides for the DGAS services. https://edms.cern.ch/document/571271.

[12] WLCG MoU documents. lcg.web.cern.ch/LCG/planning/planning.html#res.

[13] The anatomy of the grid: Enabling scalable virtual organizations, IJSA, 2001. www.globus.org/research/papers/anatomy.pdf.

[14] Jörn Altmann, Costas Courcoubetis, George D. Stamoulis, Manos Dramitinos, Thierry Rayna, Marcel Risch, and Chris Bannick. GridEcon: A market place for computing resources. In D. Neumann J. Altmann and T. Fahringer, editors, *Grid Economics and Business Models: 5th International Workshop, GECON 2008, Las Palmas de Gran Canaria, Spain, August 26, 2008*, volume 5206, pages 158–196. Springer, 2008.

[15] G. Borges, M. David, J. Gomes, J. Lopez, P. Rey, A. Simon, C Fernandez, D. Kant, and K. M. Sephton. Sun grid engine, a new scheduler for EGCE middleware. pubs.doc.ic.ac.uk/egee-sge-integration/egee-sge-integration.pdf.

[16] Costas Courcoubetis, Manos Dramitinos, Thierry Rayna, Sergios Soursos, and George D. Stamoulis. Market mechanisms for trading grid resources. In D. Neumann J. Altmann and T. Fahringer, editors, *Grid Economics and Business Models: 5th International Workshop, GECON 2008, Las Palmas de Gran Canaria, Spain, August 26, 2008*, volume 5206, pages 58—72. Springer, 2008.

[17] M.W. Cripps and J.M. Swinkels. Efficiency of large double auctions. *Econometrica*, 74(1):47–92, 2006.

[18] Gode Dhananjay K. and Shyam Sunder. What makes markets allocationally efficient? *Quarterly Journal of Economics*, 112(2):603–630, 1997.

[19] Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1), February 1981.

[20] Roger B. Myerson. Mechanism Design. In Steven N. Durlauf and Lawrence E. Blume, editors, *The New Palgrave Dictionary of Economics Online*. Palgrave Macmillian, second edition, 2009. http://www.dictionaryofeconomics.com/article?id=pde2008_M000132>doi:10.1057/9780230226203.1078.