Economic Issues in Shared Infrastructures

Costas Coucoubetis* and Richard Weber†

ABSTRACT

We define some interesting incentive issues that arise in the management of virtual infrastructures. We demonstrate that participants' decisions about the quantities of infrastructure that they will choose to contribute to a virtual organization can be greatly affected by the resource sharing policy that they know will be deployed when the system operates. Unless this policy is well-designed, agents will attempt to free-ride by contributing less resource than is desirable. What is new is our formulation of models for designing optimal management policies, an analysis that demonstrates the inadequacy of simple sharing policies, and proposals for some better ones. We find an optimal policy in a limit as the number of participants becomes large. We learn that simple policies may be far from optimal and that efficient policy design is not trivial; policy parameters play important role in optimizing the efficiency of virtual facility formation.

1. INTRODUCTION

Infrastructure virtualization is a powerful tool towards the creation of a global computing and communication infrastructure. It allows organizations to cooperate and contribute physical resources to the creation of a virtual slice of a network or of a computing and storage facility like a computational Grid. This virtual infrastructure is shared by the participating organizations and supports specific services and applications or scientific experiments. Although virtualization technology has made significant progress in this direction, there remain many interesting and unanswered economic questions about the business models that can make such virtual infrastructures viable. In particular, what are the incentives for an organization to contribute (virtualized) resources to the common slice? How should the

costs be shared? How should the new virtual facility be shared among its contributors?

In this paper we make the fundamental assumption that the virtual infrastructure is shared and generates economic value to the participants that contribute to its formation. We assume that each participant will make use of the shared virtual facility by running experiments or applications, possibly for profit. We address the question of how to efficiently form and share such a virtual infrastructure amongst a number of participants, each of whom has some private information about the value it places upon being allocated a quantity of resources. Each participant would like to obtain for himself as much as possible of the shared infrastructure (or service), while contributing minimally to the costs of its formation and maintenance. The result is that the participants individual aims are not aligned with overall system efficiency. This is an important observation and suggests that unless the appropriate incentives are in place, the economic performance of the resulting system may be greatly reduced. This raises the issue of infrastructure management.

How should a virtualized infrastructure be managed? Since such an infrastructure is shared, there is unavoidable conflict between participants in respect of the sizes of shares of it they obtain each time they need to use it. What policy should one use to manage such conflicts? How does such a policy affect the incentives that the participants have for releasing resources to the common slice? Viewing this virtualized infrastructure as a common resource pool, what should be the rules for accessing it? It could be an egalitarian policy that simply provides each participant with an equal share of the resources of the common pool; or it could be a more sophisticated policy that makes use of information that the participants provide about the values they place on obtaining resources. It should also provide a means of covering the cost of 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 instead of monetary payments; this is a more practical

model for virtualized infrastructures where participants decide on how many physical resources to release on a virtual infrastructure slice, and the sum of these contributions defines the size of the actual slice. 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 resources (the virtualized infrastructure slice), and then say that at all future instants the resource pool will be shared amongst any participants who wish to draw on it in proportion to the sizes of their contributions. The participant who contributes more will receive more. But might the system work better if the resource is shared in proportion to the squares of their contributions? It is questions like this that we address.

The problem of policy design is certainly not trivial, as is observed in [3], [7], [5]. As we see in what follows, the choice of sharing policy will affect the way in which agents will choose the sizes of their contributions to the shared resource pool. Simple policies may perform very badly and induce a great amount of free riding (i.e., they can encourage agents to seek 'something for nothing'). There is recent work in [2], [11] regarding the definition of accounting requirements for Grids, which in turn would influence the types of policies 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 becomes active and requests resources may be an important parameter. If this parameter can be measured, then we can incentivize a participant to declare it truthfully by threatening to fine him if measurements of his his parameter do not match up with what he had declared. Other parameters may be known only to a participant and cannot be measured. Then the form of the policy itself can be designed to incentivize truthful revelation. This is done indirectly, by offering each participant a choice of options and then observing which of them he chooses.

The contribution of this paper is to define some interesting incentive issues that are related to the management of the virtual infrastructures and to business modeling. What is new is the formulation of the models for designing optimal management policies and their connection with optimal auction design, a rigorous analysis that demonstrates the inadequacy of simple sharing rules, and the proposition of specific policies like extensions of proportional sharing and minimum contribution schemes. We derive optimal policies in a limit in which the number of participants, n, becomes large, and compare the various approaches. The lessons to learnt are that simple policies may be far from optimal and that efficient policy design should not be trivialized; policy parameters play important role in the final outcome of virtual facility formation.

We must stress that the mathematics involved in constructing optimal policies in the context of incomplete information (the participants may not be truthfull about their actual needs in using the infrastructure) are very elaborate and rarely lead to simple analytic solutions even in the case of simple models with few parameters. For instance, we suppose in what follows that the size of the virtual facility is characterized by a one-dimensional parameter; this is a fairly good model for computational facilities, but is less good for virtualized networks. Hence our results 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 practical implementation, i.e, impact on job scheduling policies, see [6] and [15], and the incorporation of other important parameters that were left out in our modelling. In the rest of this paper we purposefully focus our discussions upon computational grids. These are simpler to treat than other infrastructures because computation and storage seem easier to commoditize. We continue with a discussion of the virtues of virtualization and the resulting economic issues we address in this work.

1.1 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 resource units, the virtual machines, irrespectively on how these are implemented and where they are physically located. This aggregation of computing and software resources offers linear scalability: adding a virtual resource in some organization results in increasing the total resource pool of the virtual infrastructure. 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. Other positive aspects include the flexible creation of shared infrastructures in short time scales that can serve specific purposes like supporting large experiments, solving specific computational problems, etc., and security (although the last aspect of these is not a concern here). 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 formed by other infrastructure providers.

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 virtual facility 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 utility 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 resource. 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 and contribute virtual resources, (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 the business model 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. 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 can post in the market its excess computing resources with an 'ask' of the minimum price at which it is willing to sell. 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 [18], [19]. 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 (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 to participate 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 theory for optimal auctions [20]. We design an schemes 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' [21]. 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 the economic efficiency is maximized, subject to covering the cost of running the facility. Mathematical details ommitted here can be found in the full paper, [17].

2. THE FULL INFORMATION CASE

We begin with a problem of efficiently sharing a fixed quantity of computing resource Q amongst a set of n agents. Here Q corresponds to the total size of the shared infrastructure.

Suppose that the daily cost of operating the system is $c(Q)$, a known function of its size, and for simplicity let $c(Q) = Q$. Time proceeds in discrete epochs, $1, 2, \ldots$, which for convenience we will call days. If on day t an agent i is allocated resource x_i then he obtains utility (or can generate revenue) of $\theta_{i,t} u(x_i)$. On day t, the value of $\theta_{i,t}$ is realized as follows. With known probability α_i agent i contends for resources (is active) and then $\theta_{i,t} = \theta_i$, else $\theta_{i,t} = 0$. The value of θ_i captures the actual value of computation for agent i and is chosen once at the start from a probability distribution F_i. The types of the agents, as specified by their $\alpha_1, \ldots, \alpha_n$ and distributions $F_1, \ldots, F_n$, are public knowledge, i.e., they are known to all agents and the system designer. For instance (α_1, F_1) might be parameters that are typically associated with a certain size of the participating organization. The actual value of θ_i that characterizes a given agent is known only to agent i, as his *private information*. This is in contrast to the value of α_i which is assumed known since it can be monitored objectively by the system. Note that the 'system designer' of our business model may be a fictitious entity. Its function is to represent the incentives of the consortium of the participants as a whole. In practice it could be some common authority that is responsible for managing the shared infrastructure (perhaps a piece of software) after agreeing an a set of policies by the participants.

In the 'full information' scenario we suppose that the vector $\theta = (\theta_1, \ldots, \theta_n)$ and so the vector $(\theta_{1,t}, \ldots, \theta_{n,t})$ is fully known to the system designer at each t. Knowing it, he can then maximize the social welfare (total system benefit), defined as

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

by simply computing the optimal allocation vector

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

That is, $x_1, \ldots, x_n$ maximizes (1) under the constraint $\sum_i x_i \leq Q$. It generates benefit $\theta_{i,t} u(x_i)$ for participant i. If the facility is shared for a single epoch t, then assuming that $\sum_i \theta_{i,t} u(x_i) > c(Q) = Q$, the system de-

signer can ask for payments q_i such that $q_i \leq \theta_{i,t} u(x_i)$ and $\sum_i q_i = Q$. If the same set of agents share the facility continuously, then 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,

$$E\left[\theta_{i,t} u(x_i(\theta))\right] - q_i \geq 0, \qquad (3)$$

and cover the cost, i.e., $\sum_i q_i = Q$. 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 the specific cost Q. Observe that since payments are constant over time, they can be made 'in kind', i.e., by agent i contributing a quantity q_i of the virtual resources that comprise the shared infrastructure Q.

The system designer chooses Q to maximize the social welfare, taking into account that the size of the system is not fixed and comes at a cost, i.e., he finds

$$Q^* = \arg \max_{Q} \left\{ E\left[\sum_i \theta_{i,t} u(x_i)\right] - c(Q) \right\}, \qquad (4)$$

where the allocations are computed using (2) and hence are also functions of Q.

The above is called the 'first-best' solution since it achieves the highest possible economic efficiency. Unfortunately, in practice the θ_i are private information of the agents and they will act strategically when asked to reveal them. An agent might choose to declare a value of θ_i that is greater than its true value in order to obtain a larger resource share. So in practice a game takes place amongst the agents. Agent i declares θ_i 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 by which to decide payments $q(\theta) = (q_1(\theta), \ldots, q_n(\theta))$, and resource allocations $x(\theta) = (x_1(\theta), \ldots, x_n(\theta))$. Note that $q(\theta)$ depends on the values of $\theta = (\theta_1, \ldots, \theta_n)$ declared at the start, and the resource allocations for day t, of $x((\theta_{1,t}, \ldots, \theta_{n,t}))$, depend on θ and the set of active agents at day t. These define the rules of the game as regards how payments and resource allocations 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 θ_i.

2. Agents should see positive expected net benefit from participation.

3. Expected total payments should cover the cost $c(Q)$.

4. Expected social welfare (total net benefit) should be maximized among over all possible choices of Q, $x(\theta)$ and $q(\theta)$ that satisfy 1–3 above.

3. SHARING POLICIES

Each agent's decision about the amount of resource that he is willing to contribute to a resource pool depends crucially on the policy that decides how resource will be shared amongst agents who are active on a given day. Unless one designs this policy cleverly, the total system size will not be optimal. This is because agents will attempt to free-ride by contributing less resource than is socially desirable. In this section we demonstrate how this can occur, even when we have full information regarding the agents' θ_i. It is rather subtle to design rules that incentivize agents to make appropriate contributions. However, things are relatively simple for systems with a large number of participants, where the law of large numbers makes the analysis tractable. We explain this in Section 4 for a case in which the θ_i are private information. These results are useful for designing approximately optimal policies for systems of practical sizes.

Consider the model of the previous section and for simplicity assume that $\theta_i = 1$ for all i. Hence agents differ only in respect of their activity frequencies α_i, assumed known. Let the set of active agents at day t be S. Suppose $Q = \sum_j q_j$. Since all contending agents have the same concave utility function $u(x)$, it would seem sensible to take $x_i(S) = Q/|S|$. But is this correct? Or should the sharing policy depend on the α_i and on the agents' contributions, q_i? *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, both on the efficiency with which resource is shared, and on the initial resource contributions that agents will make. 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 greater efficiency than sharing resource equally amongst agents.

In the next sections we analyze the effects of different policies for 2 agents, an equal sharing policy for n agents and subscription pricing, in which all participants are charged the same fixed fee.

3.1 Sharing a resource between two agents

Suppose there are just two agents. If 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 $c(Q) = Q$ 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

$$\alpha_1(1 - \alpha_2)u(x_1(\{1\})) + \alpha_1\alpha_2 u(x_1(\{1, 2\})) - q_1.$$

If we take $x_i(\{i\}) = x_i(\{1, 2\}) = q_1$ then we model agents acting alone, i.e., each building his own facility.

Suppose $u(x) = r - 1/x$. Acting alone, agent i maximizes

$$\alpha_i(r - 1/q) - q, \tag{5}$$

For data $r = 10$ and $\alpha_1 = \alpha_2 = 0.8$ he obtains expected net benefit of 6.2112, for $q = 0.8944$.

Equal sharing. If $\alpha_1 = \alpha_2 = \alpha$, then we would expect that under any reasonable mechanism the agents should be incentivized to contribute equally and that resource should be shared equally when $S = \{1, 2\}$. However, it matters what the mechanism is. Consider an 'equal shares' policy of $x_i(\{i\}) = q_1 + q_2$ and $x_i(\{1, 2\}) = \frac{1}{2}(q_1 + q_2)$. Agent i has net benefit of

$$nb_i(q_1, q_2) = \alpha\left(r - \frac{1 - \alpha}{q_1 + q_2} - \frac{\alpha}{\frac{1}{2}(q_1 + q_2)}\right) - q_i.$$

The social optimum is achieved by a planner choosing $q_1 = q_2 = q$ to maximize $nb_1(q_1, q_2) + nb_2(q_1, q_2)$. This is achieved by $q = q_0 := \sqrt{\alpha(1 + \alpha)}$. The net benefit per agent is $w(q_0) = 6.3029$, for $q_0 = 0.8485$. Suppose agents have full information. Sharing resource with the equal shares policy, agent i maximizes $nb_i(q_1, q_2)$ with respect to q_i. If we require $q_1 = q_2$ then the equilibrium is $q_1 = q_2 = 0.6$, and each agent has net benefit 6.2. This is less than the 6.2112 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 $q_i = 0.2561$ and the net benefit per agent is 5.1826. 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 6.2112 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. The proportional sharing policy divides the resource between agents in proportion to their contributions. This gives $x_i(\{i\}) = q_1 + q_2$ and $x_i(\{1, 2\}) = q_i$. The equilibrium is at $q_1 = q_2 = 0.8246$ and the social welfare is 6.30225, which is better than the stand alone welfare. The only change is that we have replaced $x_i(\{1, 2\}) = \frac{1}{2}(q_1 + q_2)$ with $x_i(\{1, 2\}) = q_i$. This is just a bit less than the 6.30294 that a social planner could achieve.

Consider now a scheme that shares resource propor-

tionally 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 is exactly the same as that of the social optimum. 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 results are summarised in Table 1.

Other schemes can also be good. For example, recall that $q_1 = q_2 = q_0 = \sqrt{\alpha(1+\alpha)/2}$ achieves the first-best welfare. Let

$$x_1(\{1\}) = q_1 + q_2 1_{\{q_1 \geq q_0\}}$$
$$x_2(\{2\}) = q_2 + q_1 1_{\{q_2 \geq q_0\}} , \quad x_i(\{1,2\}) = q_i .$$

That is, when agent 1 is the one only who is active then he is allowed to use agent 2's contribution, but only if he contributes at least q_0. This scheme achieves the same social welfare as does a central planner. However, to compute q_0 we need to know the parameters α_1, α_2 (as when choosing $s = 1.25$ above). Is there a scheme we could invent that does not need this information?

scheme	social welfare	values of q_1, q_2
stand alone	$r\alpha - 2\sqrt{\alpha}$	$\sqrt{\alpha}$
	6.2112	0.8944
central planner	$r\alpha - \sqrt{2\alpha(1+\alpha)}$	$\sqrt{\alpha(1+\alpha)/2}$
$s = \frac{1}{2}(1 + 1/\alpha)$	6.3029	0.8485
proportional	$r\alpha - \frac{\sqrt{\alpha}(3+5\alpha)}{2\sqrt{1+3\alpha}}$	$\frac{1}{2}\sqrt{\alpha(1+3\alpha)}$
sharing $s = 1$	6.3023	0.8246
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: Social welfare per agent under different sharing policies, when $u(x) = 10 - 1/x$, $r = 10$, $\alpha = 0.8$. By optimally choosing s the system designer can achieve the first-best social welfare.

3.2 Equal sharing provides wrong incentives

The inadequacy of equal sharing is true more generally. Suppose that there are n agents and $c(Q) = Q$. Again suppose $\theta_1 = \cdots = \theta_n = 1$ and $\alpha_1 > \cdots > \alpha_n$. Consider 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 agents 1 wishes 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 agent 2 maximizes a similar expression $nb_2(q)$ with respect to q_2, where M is a random variable denoting the number of agents $3, \ldots, n$ that are present. Since $\alpha_1(1 - \alpha_2) > \alpha_2(1 - \alpha_1)$ it follows that

$$\partial nb_1(q)/\partial q_1 = 0 \implies \partial nb_2(q)/\partial q_2 < 0 .$$

So the only possible Nash equilibrium is where agents $2, \ldots, n$ choose to contribute $q_2 = \cdots = q_n = 0$.

Now let M' be the number of the agents $2, \ldots, n$ who are present. For an equilibrium to exist with $q_1 > 0$ and $q_2 = \cdots = q_n = 0$ it would have to be that

$$\alpha_1 \partial E[u(q_1/(M' + 1))]/\partial q_1 - 1 = 0$$

for some $q_1 > 0$. This can happen if and only if

$$\alpha_1 u'(0) E[1/(M' + 1)] - 1 > 0 .$$

Clearly, $E[1/(M' + 1)] \to 0$ as $n \to \infty$. So if $u'(0) < \infty$ and n is sufficiently large then no agent will wish to make any contribution.

3.3 Equal sharing with subscription pricing

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 $\theta_1 = \cdots = \theta_n = 1$, but α_i differ, and that a priori these are uniformly distributed on $[0, 1]$. We ask every agent to make a fixed subscription q. There is a minimum α, say α_q, for which it is advantageous for an agent to participate. By considering the fact that the marginal agent's net benefit is 0, we have

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

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

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

For $r = 10$ we find optimal q and α_q as follows. For comparison, the final column shows the first-best that

could be obtained in the full information case. We can also calculate that under proportional sharing, as $n \to \infty$, agents of activity α are incentivized to contribute $\sqrt{0.6\alpha}$, and the net benefit per agent is 3.967. Stand-alone it would be 3.667.

n	q	α_q	net benefit/agent	
			subscription	first-best
2	0.6367	0.0726	3.770	3.827
10	0.5418	0.0697	3.939	3.966
∞	0.5158	0.0575	3.987	4.000

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

4. LARGE SYSTEMS

Now we look at Grids with a large number of participants and obtain a solution that is simple and intuitive. We take the model of Section 3, in which each agent i is characterized by known (α_i, F_i). He is active on day t with probability α_i and when active and allocated resources x_i his benefit is $\theta_i u(x_i)$. The parameter θ_i is private to agent i which he must be incentivized to reveal truthfully at the start through his choice from a set of available contracts.

For any n (not small) the optimal mechanism is like this: 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 $x_i(\theta, S)$. For lack of space, the proof of optimality is omitted here, but using optimal auction theory it turns out that the optimal mechanism can be characterized as follows. Define $g_i(\theta_i) = \theta_i - (1 - F_i(\theta_i))/f_i(\theta_i)$. There is a $\lambda \geq 0$, such that for all S the optimal way to share resource amongst a set of active agents S is to maximize

$$\sum_{i \in S}(\theta_i + \lambda g_i(\theta_i))u(x_i(\theta, S)), \quad (6)$$

over $\sum_i x_i(\theta, S) \leq Q(\theta)$. Here λ is a Lagrange multiplier for a constraint

$$E\left[\sum_i p_i(\theta)\right] \geq E[c(Q(\theta))].$$

This has an interesting limit when n is large, and it allows payments to made in kind. We note that when agent i is active the rest of the system will be in its typical average state. So 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

$$\sum_i \alpha_i x_i(\theta) \leq Q(\theta). \quad (7)$$

The problem reduces to one in which constraint (7) can

be satisfied by taking $\lambda = 0$ and has a solution in which

$$x_i(\theta) = x_i(\theta_i) := \arg\max_{x_i'}\{\theta_i u(x_i') - x_i'\}, \quad (8)$$

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

It is interesting that the optimal contract chosen by agent i secures the same amount of resources from the shared resource pool as he would optimally choose to self-procure if no shared infrastructure was available and he was always active. But he needs only pay for his average usage, namely for $\alpha_i x_i$. By construction, this scheme is incentive compatible, i.e., he will choose the tariff parameterized by his actual value of θ. Note that $x_i(\theta_i)$ exceeds the size of the facility he would form if he were to stand-alone, which would be

$$x_i^0(\theta_i) := \arg\max_{x_i'}\{\theta_i \alpha_i u(x_i') - x_i'\}. \quad (9)$$

Thus 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 but pays only when he uses it, since others pay for it when he is not.

Let $x_i = x_i(\theta_i)$, as defined above in (8). In practice, we need $\sum_i x_i(\theta, S) \leq Q(\theta)$ for all S. This is not possible if we try to take $x_i(\theta, S) = x_i$ for all S. However, we can modify things slightly. With agent i contributing q_i, we let $y_i = q_i/\alpha_i$ and $x_i(\theta, S) = y_i Q/\sum_{j \in S} y_j$, where $Q = \sum_j q_j$. Let us illustrate with $u(x) = r - 1/x$. Let $I_j \sim B(1, \alpha_j)$. Agent i has expected net benefit of

$$\alpha_i \theta_i E\left[r - \left(y_i + \sum_{j \neq i} I_j y_j\right)/(y_i Q)\right] - \alpha_i y_i$$
$$= \alpha_i(\theta_i(r - 1/y_i) - y_i) - \alpha_i(1 - \alpha_i)/Q.$$

The term $\alpha_i(1 - \alpha_i)/Q$ is small and varies little with y_i, and $\alpha_i(\theta_i(r - 1/y_i) - y_i)$ is maximized by $y_i = x_i(\theta_i)$. So agent i is incentivized to contribute $\approx \alpha_i x_i$ and the total welfare, which is $O(n)$, will differ from its first-best value by just $O(1)$. Note that the designer need know nothing about the θ_i. He constructs the resource curve $x(\theta)$ by solving (8) for all values of θ, and the family of payment curves $\alpha x(\theta)$, for all potential values of activity frequencies. Operations becomes very simple and efficiency is near the full information first-best solution.

5. LESSONS FOR PRACTICE

We have investigated policies for running shared computing resource infrastructures. We have assumed that participants will 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. The chief lessons from this study are as follows.

1. A participant's decision about the quantity of resources that he will choose to contribute to the 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 given quantity of resources may not be optimal. For example, if the resource will be shared equally amongst active participants then an agent may may choose to contribute nothing to the resource pool.

2. One way to incentivize potential participants to make significant contributions to the resource pool is to impose a rule that a participant will only be permitted to draw on the pool if he makes a minimum contribution to it a the point that it is formed, i.e., by contributing a minimum quantity of computing resources. We can further impose a sharing policy that ensures that an agent who contributes more resource will have greater priority for obtaining resource than an agent who has contributed less. Such rules 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.

3. 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 for resource shares according to their needs at each time, while generating enough payments to cover running cost. If the set of participants is constant, then simpler policies exist. We have proposed some, but at the added cost of implementing some accounting, such as policing the α_i.

We have obtained results for simple models under 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. Often the cost of the shared facility $c(Q)$ is not precisely known, and would take some non-trivial effort to define.

6. 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,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] J. Altmann, C. Courcoubetis, G. D. Stamoulis, M. Dramitinos, T. Rayna, M. Risch, and C. 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] C. Courcoubetis, M. Dramitinos, T. Rayna, S. Soursos, and G. D. Stamoulis. Market mechanisms for trading grid resources. See [14], pages 58–72, 2008.
[17] C. Courcoubetis and R. R. Weber. Economic issues in shared infrastructures. `nes.aueb.gr/publications/2009.virtualization_incentives.pdf`.
[18] M. W. Cripps and J. M. Swinkels. Efficiency of large double auctions. *Econometrica*, 74(1):47–92, 2006.
[19] G. K. Dhananjay and S. Sunder. What makes markets allocationally efficient? *Quarterly Journal of Economics*, 112(2):603–630, 1997.
[20] R. B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1), 1981.
[21] R. 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. `www.dictionaryofeconomics.com`.