Jim Weichel

Introduction

Most Virtual Organizations (VOs) are projects or institutions that own computing resources that become part of the Open Science Grid (OSG). Each VO sets its own policy related to how these resources are used, including the portion, if any, that are open for use by other VOs.

A previous study[1] developed an economic value of opportunistic use of OSG that ranged from $2.1M to $5.5M. This economic value flowed to the users of opportunistic resources in 2008. The portion of opportunistic hours to total hours of use of OSG has remained fairly consistent at around 30% over the last several years. In fact over the last year, it has increased to 35%[2]

To get a broad gauge of the proportion of VOs with resources that share a portion of their resources broadly to other VOs, one can compare the number of VOs that allow the Engage VO access to the number of VOs with resources. Based on MyOSG results, there are 17 of 27 VOs that share their resources with Engage.

The purpose of this paper is to posit the reasons behind this sharing, to look at sharing from the economic point of view of the providers of the shared resource, and to propose possible mechanisms that could increase the economic value of sharing to those who share in order to encourage even more sharing.

Reasons for Opportunistic Sharing

Let us look at possible incentives and disincentives for opportunistic sharing.

Incentives:

1. A general collegial or noble feeling that could result from one's contribution to the overall OSG community. This can be institutionalized by a VO leader who makes it a policy to provide, say 10%, of the VO's resource available for opportunistic use.
2. A realization that one's project can make a broader contribution to science by allowing use of available resource at a relatively small cost to the project. This is similar to item 1 but is a benefit to a different community and begins to consider the benefit/cost ratio.
3. Belonging to OSG provides the opportunity to exceed the total resource owned by one's project as long as someone offers opportunistic use to your project. This offers some "weak" quid pro quo. It is weak since one doesn't

---

[1] OSG Executive Board, "OSG Value Proposition", January 12, 2009.
[2] See pie chart: Opportunistic use 11/2009 to 11/2010

have to offer opportunistic resources to be able to use opportunistic resources.

Disincentives:

1. Offering "unused" resources to others does have real costs to one's project. This includes actual expense for electricity since computers reduce their electricity use when they are idle by sleeping or even hibernating. This reduction gets multiplied due to the reduced need for cooling for these sleeping processors. In addition, the load on support resources are increased due to support provided to the opportunistic projects and increased troubleshooting and loss of resource to one's own project if one of the opportunistic users causes a problem that degrades the network or storage systems or never successfully completes execution. Igor Sfiligoi made a rough guess that this might be 0.3 FTE at UCSD.
2. If one's project has computing resources that are expected to meet the peak needs of the project, there is not even the "weak" quid pro quo incentive which implies one's project pays only the cost and will not receive any benefit.

Discussion

So by looking at these incentives/disincentives, it would appear that the system is working as well as it does now mostly due to the educational and scientific bias toward the greater good rather than any real economic drivers.

That raises the question, "Could an economic system be set up that would add an economic incentive to encourage VOs to provide opportunistic resources to others?"

Two Ideas[3]

Another way to state the question is, "How can we make in a VO's best interest to offer opportunistic resources?" Or, "How can I make an argument to a CFO/CIO on why we are offering opportunistic resources that cost real dollars for electricity and staff resources for support?"

1. Insurance Model

The concept is to maintain usage information that makes it apparent who is providing and who is using opportunistic resources on a pairwise basis in order to build a "credit" that can later be cashed in to get priority access to another VO's opportunistic capacity. For example, over the last 6 months, VO A has provided VO B with a net 200K hours of opportunistic computing. Then assume VO A needs to rerun an analysis or needs to speed up analysis to meet some important deadline.

---

[3] Based on discussions with Igor Sfiligoi.

VO A can then ask VO B to move them to the highest priority opportunistic user on VO B's resources so that VO A receives up to the limit of opportunistic cycles if VO A has the load to do that. This has the nice characteristic that VO B's capacity can still be efficiently utilized by the community whenever VO A fails to fully occupy the resource. This situation can be continued until VO B provides the 200K hours to VO A or could continue beyond that building a credit for VO B.

This may be improved by not keeping track of pairwise flow; and, instead, just the total debit or credit of usage by each VO. If you have a credit and need more capacity, you could request high priority on the opportunistic resources across OSG.

In an economic sense, this is somewhat analogous to an insurance policy. VO A pays a premium (in this case, 200K hours of computing and support) for the ability to cash in the policy if needed in the future. It differs from insurance because there is less leverage in that the payout is equal to the premium.

How important would such insurance be for a VO? For a VO with resources that easily exceed their maximum need, it seems like it would not be too important. But most projects needs grow over time, so excess capacity today may disappear in the future. In addition, as seen by projects today, there is sometimes a need to redo previous analyses when new information or capabilities become available. The excess capacity is unlikely to be twice the current load that would enable a redo in parallel with the ongoing load.

Also, it may not be well know that most projects are routinely using opportunistic capacity (or at least capacity owned by another VO). This must be providing faster results than if only owned resources were available. This is a form of economic return that can be very important to a project. To illustrate this, here is a table showing usage by six VOs for one week (11/3/10 – 11/9/10). [I think some of the UN graphs do not properly account for resources that have multiple owners and that makes it look like CMS and ATLAS are using more of Others' resources than they really are. For example, ATLAS use at the OU T2 and AGLT2 that is co-owned, corresponds to their Use on Others'.]

	Total Use	Use on Own	Total Owned Use	Use on Others'	Provided Others	Net Use by Other
Dzero	694,416	247,557	419,251	446,859	171,694	(275,165)
CMS	2,352,835	2,269,873	2,959,564	82,962	689,691	606,729
ATLAS	3,073,911	2,690,587	3,515,717	383,324	825,130	441,806
SBGrid	39,567	729	7,293	38,838	6,564	(32,274)
Ligo	506,032	68,461	87,849	437,571	19,388	(418,183)
CDF	1,008,938	985,856	1,512,561	23,082	526,705	503,623

Key:
Total Use – Hours of total use by the VO
Use on Own – Hours of use on resources owned by the VO

Total Owned Use – Total hours of use by the VO on its own resources
Use on Others' – Hours of VO use on resources it does not own (opportunistic)
Provided Others – Hours of use provided to other VOs
Net Use by Other – Net use of the VOs resource used by other VOs (negative implies the VO used more of other VO's resources than it provided to other VOs)

2. OSG' as a Broker Model

The concept is for OSG' to be a broker among the resource providers and users of OSG resources. A science project would come to OSG', with computing dollars, and request provisioning of resources for its use. OSG resource providers could offer to meet those needs using excess existing capacity or by installing new capacity to meet the needs. In either case, the resource provider would receive payment for those resources.

Here is an example of how this might work:

1. A science project proposal to a funding agency includes $ for the necessary computing resource. In the current world, they would require enough $ to house, install, maintain, expense, operate the computing - let's say for 500 CPUs to cover peak needs. Let us assume that they require 400 CPUs on average.
2. In the OSG' world, a funding agency would allow the science project to request a computing dollar amount that is, say half, the going Amazon EC2 price for the peak expected use - 500 CPUs. This is fewer $ than would be required for their own computing center (and a little more than 1/2 what it would cost to plan to use Amazon).
3. This is possible because OSG' offers to broker resources from their members. In exchange for an allocation of 400 average CPUs, the OSG resource provider gets the money allocated in the proposal. For these resources, the provider incurs the only the incremental cost to install and run the 400-500 CPUs, depending on the typical opportunistic capacity available to meet the peak loads.

How does this look to the parties involved?

A. The funding agency pays 3x-4x less to fund the computing for the proposal (guess - we would have to do the math) - big WIN.
B. The science project gets the needed computing resource without the headache of building/maintaining/operating a computing facility – WIN. (If associated with a university, the university might prefer to add to their computing, so this might be only a partial win. If the university already has a large resource associated with OSG, they could step up to provide the resource, so it still is a win for the university but only if they are an efficient provider.) Also, they are likely to get access to a higher peaking factor than if they had their own 500 CPUs. There is some loss of control of your computing resource but that is equivalent to outsourcing any function.

C. OSG gets a new science customer of their resources and provides more computing to science – WIN.

D. The resource provider gets funding to support providing resource for the science user. They can provide this very efficiently because 1) they may have enough available resource already, 2) or more likely, they will actually add 400 CPUs for the new user, 3) they are a large facility already so they operate near the efficient frontier of computing, 4) the science user will likely have a varied need day to day so sometimes they will use more than 400 CPUs and sometimes less, but generally there will be additional capacity others (including the VO that owns the facility) can use. In general, the incremental capacity is very low cost so they actually make money on the deal – WIN.

Discussion:

- Of course, it isn't quite this simple since the storage/networking needs also need to be met and would have to be included in the model.
- Note that Amazon has to be profitable, so it is reasonable to allocate only half the Amazon price in the non-profit world. The OSG Value Proposition did verify this to be a reasonable assumption.
- Not all resource providers may want to become paid computing suppliers to other projects. This requires additional staff and management, as well as the space, power, etc. for the computers and storage.
- The system would drive computing toward facilities that were most efficient in providing it because only they can provide resource for half the Amazon price.
- Large projects and universities with large computing facilities might choose to provide resource for related projects or projects run by people at that lab or university. As long as they are efficient, the resource can be co-located, under the direct control of the project.