Reservation Schemes for Grid Markets

Marcel Risch

International University in Germany, Campus 3,
76646 Bruchsal, Germany
Marcel.Risch@i-u.de

Abstract. Grid markets today have very limited capabilities with regards to reservations. However, it is vital that customers are given the ability to purchase resources far in advance of the actual usage which necessitates the existence of a sold reservation model. While this is not an issue of great importance in today's Grid markets, reservation schemes must be developed to enable a sustainable operation of Grid markets. This paper explores a number of different reservation models that are in use today and evaluates their usefulness both from the point of view of resource sellers as well as from the point of view of resource buyers. It then goes on to propose a new reservation scheme for grid markets.

Keywords: Grid Markets, Reservations.

1 Introduction

One of the major issues that must be addressed if open Grid markets are to become a reality is the issue of reservations. Since grid customers will not only wish to buy resources shortly before they are needed but also far in advance of actual usage, reservations are vital to keep such an economy operating. While many markets have been developed and some have reached the production-level stage, reservations still remain largely ignored.

Such an option is a requirement if customers are to be allowed to order resources far in advance of the actual usage. This requirement, in turn, leads to the important question of which type of reservation should be used. There are a number of reservation types which are used in various settings. These will be analyzed with regards to a grid market so that the appropriate reservation scheme can be chosen.

The paper is structured as follows: in the next section, two currently existing grid markets will be explained with a special emphasis on their reservation schemes. In the following section, a number of reservation schemes which are currently in use in other fields are analyzed with regards to their suitability for grid markets. We then develop an improved reservation scheme which is based on a prepaid model. This section will be followed by an analysis of the future work.

2 Existing Reservation Schemes

2.1 Existing Grid Markets

The Distributed European Infrastructure for Supercomputing Applications (DEISA) [1] is an organization in which eleven high performance computing (HPC) centers are

L. Telesca et al. (Eds.): DigiBiz 2009, LNICST 21, pp. 1–8, 2009.

interconnected to provide computing power mainly to researchers from participating countries. In addition to researchers, private enterprises can also use some of the computing power for which they have to pay. The procedure for gaining access to these resources is very complex, as can be seen from [2].

While the DEISA consortium reserves the resources, it does so at its own convenience. Customers do not have any influence on the time at which the resources are available but must wait until their assigned time. Therefore, this model cannot be considered to be customer friendly and is not suitable for a grid market in which customers can obtain resources for usage at their convenience.

The Amazon Elastic Compute Cloud (EC2) [3] is the first grid market in which anybody can buy resources. The resources can be bought in bundles for a fixed price and are charged for according to the actual usage. This system allows users to create images which are then uploaded to virtual computers and run on the Amazon.com servers. While this approach has many benefits, it is nevertheless not perfect from an economic point of view.

Users who want to reserve resources find themselves unable to do so. While they are able to add multiple resources to an existing resource pool, they cannot reserve the resources in advance. Thus, it can happen that when resources are scarce the customer cannot add urgently needed resources to his pool. Furthermore, once the resource is obtained, it can be used without any time constraints, which implies that Amazon.com is unable to determine when resources will become available for further use.

Tsunamic Technologies [4] provides clustering services to paying customers. This company provides a number of services:

- On-Demand Cluster: A simple clustering scheme without any requirements. Users can use as much or as little compute power as they require.
- Subscription Clusters: In this scheme, users agree to use a minimum number of CPU-hours per month, in exchange for a lower CPU-hour rate.

The first scheme is identical to the Amazon EC2 scheme. The second scheme can be seen as a middle ground between a reservation and a reservation-less system. By requiring the user to use a certain amount of CPU-hours, the company can perform some basic planning tasks. On the other hand, it is not clear whether the resources are guaranteed to be available for the required amount of hours per month or whether they can be reserved for certain time periods.

2.2 Reservations in Other Markets

Since current markets do not handle reservations well, we have to look at other areas in which reservations have been widely used. There are a number of reservation models that currently exist which could all be applied to grid markets.

The restaurant-style is modeled after the reservation service offered in restaurants. In the case of grid markets, the customer orders a resource far in advance of the actual usage, telling the provider a starting time and date for the usage, as well as the duration for which the resources are to be used. The provider then reserves the resources ordered. Should the customer use the resources, they will remain reserved for the duration specified by the customer when the reservation was made. However, if the customer is not ready by the starting time, the reservation is canceled and the provider can attempt to sell the idle resources on the grid market.

For the client, this model offers a number of distinct advantages over buying resources for in-house installation. Firstly, this model allows the customer to reserve resources before he knows if he actually needs them. Secondly, if the client wishes to cancel the reservation, there will be no consequences for the customer.

For the provider, this model has few advantages. The main advantage lies in the fact that the provider can see how high the demand for a certain resource is. On the other hand, the provider is not able to accurately forecast the income generated, due to the risk of cancellations which requires the provider to apply some strategies to offset the consequences of these cancellations, such as overbooking.

As opposed to the restaurant reservation scenario, where the customer had nothing to fear in the case of late cancellations or expiration of reservations, the airline-style reservation makes it the customer's responsibility to have all data ready on time. The reservation scheme is modeled after the airline reservation system in which the customer reserves a seat, pays in advance and then gets a ticket and reservation. In the case of cancellations, there are three different models to consider: in the first case, the customer is refunded the entire payment; in the second case, the customer is returned the payment minus a fixed fee; and in the last case, the customer cannot cancel the ticket at all. Thus, the risk incurred by a cancellation is not entirely shouldered by the provider in the last two cases. The first is identical to restaurant-style reservations.

The two remaining cases are much more advantageous for the provider as compared to the previous reservation scheme. Firstly, since the customer pays in advance, the provider can be certain of the income generated by the resources. If the customer has the job delivered on time, he can have access to all resources. Should the client not be ready by the starting time of reservation, the provider has the option of selling the resources on the grid market. In addition, the provider can be certain of a small amount of income being generated. Furthermore, if the provider is able to sell the idle resources, the income generated is increased by the cancellation fee, thus giving the provider an added bonus for allowing reservations.

For the resource user, this model is more risky, since a cancelled reservation can be costly. In the case that the resources are still needed and the reservation has been cancelled, the cost will be even higher, since the customer is forced to purchase the required resources on the grid market. As opposed to the restaurant-style reservation scheme, this reservation model puts a fixed amount of the costs for missing a reservation on the shoulders of the customer.

Since the first two reservation schemes have been focused on either party carrying the financial burdens of cancelled reservations, the third alternative is a mixture of these two. Its first point of change is the fact that reservations do not expire if the customer is late. Instead, the length of time that the resources will be kept available depends on the non-refundable down payment the client is willing to pay. An improved solution is a proportional model which can be described mathematically as follows:

$$\frac{P_{Down}}{C_{Total}} = \frac{T_{Available}}{T_{Reserved}} \quad (1)$$

In the equation, P_{Down} denotes the down payment made by the customer, C_{Total} denotes the total cost for the resources for the entire reservation period, $T_{Available}$ denotes the time for which the resources will be kept available without customer action, $T_{Reserved}$ denotes the total reservation period. Thus, the duration for which the resources will be kept available even if the customer does not take any action is proportional to the size of the down payment. In other words, if the client pays 50% of the total resource cost, the resources will be reserved for half the time that the client requested them for.

The improvement of the model lies in its fairness and its firm guarantees. The client can be certain that the resources will be available for a certain amount of time which is influenced by the size of the down payment. The provider, on the other hand, can determine the likelihood that the resources will be actually used by the customer. A large down payment would indicate a high probability that the resources will be needed; while a low down payment would indicate that the customer is not certain if the reserved resources will be needed. This is proportional model is somewhat similar to the truth telling reservation scheme proposed in [5].

Additional advantages for the provider are that the customers are encouraged to take the reservation seriously, since the resources have been partially paid for. Furthermore, this model gives the provider the ability to plan the cash flow because he knows a lower bound on the income thanks to the prepaid reservations.

This model has very few disadvantages for both resource provider and client. The customer needs to ensure that the deadline can be met within the time frame specified by the down payment. The main disadvantage for the provider is that the hardware can remain idle because the customer doesn't need it but cannot be sold again on the grid market until the time reservation time has passed, depriving the provider of a second source of income.

3 An Improved Reservation System

Of the reservation schemes explained above, only the last shows some promise, since it strikes a balance between the risks shouldered by the customer and the risks shouldered by the provider. However, the main disadvantage is that the provider is very inactive when it comes to setting the size of the down payment. This new reservation model will rectify this shortcoming.

3.1 The Reservation Model

Since the airline-style reservation scheme is very advantageous for the provider, it should not be removed from the pool of possible reservation schemes in a grid resource economy. However, in order to level the playing field with regards to cancellations, the refund scheme must be changed to be more flexible than the currently existing schemes. In general, providers have three possible ways of calculating the size of the refund.

Ideally, the size of the refund should be proportional to the remaining time to the beginning of the intended resource usage. If customers cancel far ahead of the actual resource usage, they would get a large part of their money refunded. On the other hand, if the cancellation of the reservation is done shortly before the start of the reservation, only a small part of the down payment would be refunded.

This approach naturally leads to the question of how to determine the size of the refund. There are a number of possible mathematical solutions which will be detailed here. It is assumed that the customer makes the reservation at time T_0 and that the reservation will begin at time T_R while the cancellation is made at time T_C. The cost of the resources is assumed to be C_R while the refund is denoted by R.

The first and most basic of these refund models is the linear model in which the size of the refund is directly proportional to the time left to the start of the reservation.

$$R = C_R - \frac{C_R}{(T_R - T_0)} * T_C \ . \tag{2}$$

Thus the client is given an incentive to cancel the reservation as early as possible to maximize the refund. This equation is easy to understand, making the refund process very transparent for the customer. This can be seen in figure 1, where the refund percentage is linearly proportional to the time remaining till the reservation starts.

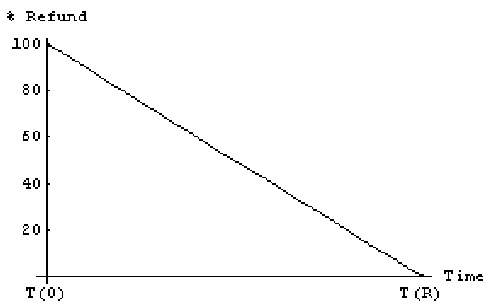

Fig. 1. Linear Model

An alternative strategy is to use a concave-up exponential function. In this case, the size of the refund would diminish very quickly but then fall slowly before the start of the reservation. This behavior can be modeled by the following equations:

$$\begin{aligned} R_1 &= \left(\frac{1}{1 + 100 * T_C}\right) * C_R \\ R_2 &= C_R \\ R &= Min[R_1, R_2] \end{aligned} \ . \tag{3}$$

Therefore, unless customers cancel reservations early, they will not get a large refund, since R_1 will decrease very quickly. R_2 avoids the problem of having an infinite return if the reservation is cancelled immediately after being placed. However, if the reservation is cancelled quickly after being placed, the customer will get a large refund. This is illustrated in figure 2.

Fig. 2. The Concave-up Model

It can be clearly seen that the refund is very large initially but then drops sharply, thus encouraging the client to either cancel a reservation quickly or not to cancel it at all.

The final strategy is very customer friendly, since it will not punish the customer for cancelling late. Instead, a large part of the down payment will be refunded if the reservation is cancelled early enough. However, he will be penalized for canceling reservations late. Mathematically, this concept can be described as follows:

$$R_1 = \frac{\left(\frac{T_R}{T_C - T_R} + 100\right)}{100} * C_R$$
$$R_2 = 0 \qquad .$$
$$R = Max[R_1, R_2] \tag{4}$$

The customer has an incentive to use the resources he has bought while at the same time is given some leeway to cancel the reservation. This is shown in figure 3.

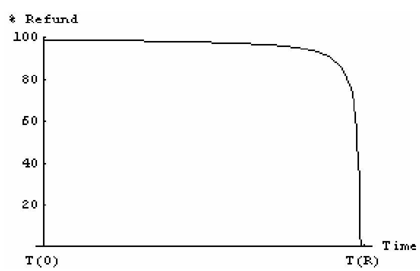

Fig. 3. The Concave-down Model

In this example, the user receives a large refund until shortly before the reservation begins. If such a scheme has been agreed upon with the provider, the customer can easily cancel the reservation without having to pay large fines.

3.2 Analysis of the Model

The prepaid model has a number of advantages for providers which will make this scheme very attractive to them. Since customers pay the moment the reservation is made, providers can be certain that they have a certain amount of income. The minimum income can be calculated by the time remaining to the start of the reservation. In addition, providers can choose from the three refund models the one which best fits the market situation and their own resource utilization. In essence, providers can choose the refund method as a way to attract customers or to reward returning customers.

For the customer, the mathematical model allows for a very transparent calculation of the size of the refund and thus allows the customer to confirm the provider's calculation. This ability to check the provider is another point in favor of such a scheme, since customers may no trust their providers.

As to which refund scheme is best under which circumstances is very difficult to say, due to the many parameters that must be considered: the demand, the willingness of customers to commit to resource purchases, the refund models used by other providers and the objectives of the provider. In the case of high demand and a high willingness to commit to resources, providers can choose the concave-up refund scheme, since customers will be concerned about being able to buy the resources they need. However, if the provider wishes to attract more customers, he may choose to use the linear model instead. This should attract customers who are not as certain of their resource requirements but still wish to participate in the grid market. It may also be the case that the provider chooses to offer the concave-down refund scheme to frequent customers as an incentive to return to this provider and not use other providers.

In the case of high demand and a low willingness to commit to resources, the provider can hope at best to use the linear model and may have to resort to the concave-down refund scheme. Using the concave-up model will only serve to discourage customers and thus be counterproductive when it comes to selling resources.

However, it is not only a matter of demand and the customer's willingness to commit to a purchase. If other providers offer mainly the concave-down or linear refund schemes, customers will prefer these, since they may not be entirely certain whether or not they need the resources. Furthermore, a provider needs to consider which refund policy other providers use. If most providers use the concave-up refund model, the provider may be able to attract more customers by using the less strict linear model. On the other hand, if most providers use the linear of concave-down models, using the concave-up model would be counter productive, since customers would choose to buy resources from the more flexible providers. Thus, determining the best refund scheme is up to the provider who must take into account the current market situation, the actions of other providers while taking his own goals into account.

4 Conclusions and Future Work

In this paper we have analyzed a number reservation schemes used in grid markets today, as well as analyzing some reservation schemes that are used outside of grids. The former showed that in today's grid economy, reservations are very much neglected.

Since the additional reservation schemes examined were not satisfactory, a new model was developed which is based on the prepaid reservations where the size of the refund is based on the time remaining to the start of the reservation.

The suggested refund model allows providers to choose the function with which to calculate the refund. The choice depends on a number of parameters and must be carefully made so that customers will not be turned away to the competition. This determination of which reservation scheme to use at which point in time is the topic of further research.

References

1. DEISA - Distributed European Infrastructure for Supercomputing Applications (2008), http://www.deisa.org
2. DEISA Extreme Computing Initiative, DECI (2008), http://www.deisa.eu/science/deci
3. Amazon Elastic Compute Cloud, Amazon EC2 (2008), http://aws.amazon.com/ec2
4. Tsunamic Technologies Inc. (2008), http://www.tsunamictechnologies.com
5. Wu, F., Zhang, L., Huberman, B.A.: Truth-telling Reservations (July 17, 2005), SSRN: http://ssrn.com/abstract=771804