Market-Oriented Meta-Scheduling for Utility Grids

Saurabh Kumar Garg[1], Srikumar Venugopal and Rajkumar Buyya
Grid Computing and Distributed Systems (GRID) Laboratory
Department of Computer Science and Software Engineering
The University of Melbourne, Australia
{sgarg,srikumar,raj}@csse.unimelb.edu.au

Abstract

As users increasingly require better quality of service from Grids, resource management scheduling mechanisms have to evolve in order to satisfy competing demands on limited resources**.** Traditional algorithms are based on system-centric approaches which do not consider user requirements and interests. These system-centric approaches for scheduling user applications aims to maximize system performance and thus do not consider user utility. Most importantly, these traditional techniques do not perform effectively when the demand for resources surpasses the supply. Therefore, we propose a complementary approach which uses basic economic and market principles to help allocate limited resources more efficiently and fairly by balancing supply and demand. This requires good valuation schemes for both grid resources and user applications. In this work, we design valuation and allocation mechanisms for mapping jobs with Quality of Service (QoS) requirements to heterogeneous grid resources. Finally, we propose a Meta-Broker framework for Grid Market to realize these scheduling strategies with different objectives.

1. Introduction

With the advances in Information and Communications Technology such as the emergence of multi-core processors and networked computing environments, the computing is transformed to a model consisting of services that are commoditised and delivered in a manner similar to utilities such as water, electricity, gas, and telephony. In such a model, users access services based on their requirements without regard to where the services are hosted. Grid Computing is one of the most promising paradigm which supports such a utility model for IT services. Grid computing has led to the creation of large computing infrastructures, or Grids, that aggregate clusters and supercomputers in different physical organizations. After almost a decade of research and development by the Grid community, many Grids, such as TeraGrid [1], NorduGrid [2], OurGrid [3] and EGEE [4] are being deployed around the world in both academic and commercial settings. A number of vendors such as IBM, Hewlett-Packard and Sun Microsystems are offering grid-related products and services. Thus we now believe that federated networked systems have grown to the scale where economic policies are the logical next step. We foresee that, for providing computing as utility, in few years Grids infrastructure will include a Grid Market/Exchange analogues to Stock exchange.

This Grid Market/Exchange will act as a market place where any grid user and grid resource can buy and sell their resources. This market place can simplify the work of a grid user and a resource provider who can discover their required services and their additional information at same place. In this exchange, client demand can be satisfied not only within one organization but also across multiple administrative domains. But to make the vision of computing services as utility a reality and to transform grids into a real computation market, still there are a lot of challenges to be faced from the areas of security, uniform access to all users, computational economy, resource allocation and management, resource discovery and data locality [11]. Among these challenging areas, our main focus of research is on resource management and computational economy which is managed by Meta-schedulers (Meta-broker) in the Grid Market.

One of the important challenge before the meta-schedulers will be to match the aggregated

[1]Student Author

resources with the various user applications with different QoS requirement. In such a scenario, each participant will compete with others and even willing to pay more for their required service. Thus it becomes a challenge to manage and allocate different services and resources effectively, cooperatively, fairly and efficiently. Service providers need to decide how to allocate services and clients need to determine how to acquire the services they want. Thus, there are three main goals that a Meta-scheduler (Meta-Broker) system seeks to achieve:

1. Balance supply and demand of the system according to Valuation of users.
2. To increase efficiency of resources by balancing load.
3. To allocate user application fairly and to satisfy as much users as possible with maximizing the aggregate utility of all users.

The traditional meta-schedulers such as Maob, GridWay[6], Glite[5] and CSF[7] for Grids are system centric and favour system performance over increasing user's utility. Thus, these schedulers are not suitable for environments with multiple users having different QoS requirements and competing for the same resources as they do not differentiate between users with different requirements. They particularly fail when the demand for resources is in excess of supply since it is not possible to completely satisfy all user requests. Consequently, researchers have been examining the appropriateness of 'market-inspired' resource management techniques to ensure that users are treated fairly.

In recent years, a number of researchers have proposed economy-based models for more efficient management of Grid resources [8][10][9]. Such models apply well-known and proven economic mechanisms such as markets and auctions to solve the challenges of resource allocations in shared distributed computing environments. In these models, users have to pay for accessing resources which are assigned prices that reflect the value placed on them according to their capability and availability. Users are limited by their payment capacity (or budget) in selecting appropriate resources. Auctions have been particularly preferred by many such projects — for example, Tycoon [9] and Bellagio [10] — as they provide a decentralized structure, are easy to implement, provide immense flexibility to participants to specify their valuations and are considered as the most efficient among current market management systems. But these economic-based systems have many limitations. First, while these approaches distribute services fairly, they limit the ability of customers to express fine-grained preferences for services. In addition to that, users may not be able to express their true valuations accurately as they may lack the sophistication to make decisions based on changing resource load and prices. Finally, users with low budgets and urgent requirements may not be able to gain resource allocation as the system may be monopolized by those with large budgets.

Therefore, both the scheduling systems have their pros and cons. In this work we propose a novel meta-scheduler that unifies the advantages of both the systems for benefiting both users and resources. In order to do that, we design a valuation metric for user's applications and computational resources based on multi-criteria requirements of users and resource load. The meta-scheduler maps user applications to suitable distributed resources using a Continuous Double Auction (CDA). Through simulation, we compare our scheduling mechanism against other common mechanisms used by current meta-schedulers. The results show that our meta-scheduler mechanism can satisfy more users than the others while still meeting traditional system-centric performance criteria such as average load and deadline of applications. We plan to extend this work by integrating various pricing functions for computing services and also different workload models.

2. Grid Market Architecture

In this section we discuss Grid Market Architecture which acts as a market place where Grid users and resource/service providers can find their desired clients. In general, a market can be understood as the location where demand

and supply meet. The Grid Market thus consists of the service consumers and providers, representing demand and supply, and of the software system that implements the following market functionalities:

1. **Grid Information Service (GIS) and Resource Catalogue** - support resource discovery, allocation and scheduling.
2. **Reservation Service**: Grid Resource Reservation System can do the advanced reservation to assure availability of resources at the right price & time.
3. **Pricing System** – keeps update of current service prices for usage of compute resources.
4. **Meta-Brokers (Meta-scheduler)** – provide interface between resources and end-user. They have responsibility to match user application with appropriate resources. It can also integrate various economic models within the concept of current community grids (resource sharing).
5. **Accounting System** - presents and reports on usage of compute resources. It also calculates money that needs to be charged from user based on these usage reports.

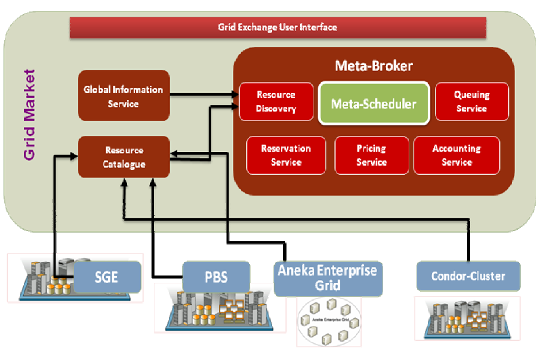

Figure 1. Grid Market

Meta-Scheduler is the most important component of the Grid Market which aims to match user needs to available resources, fairly and efficiently. The efficient allocation of Grid resources requires an adequate scheduling mechanism from matching demand and supply and incentive compatibility. Thus, we analysed and designed a valuation scheme with allocation mechanism for mapping jobs with Quality of Service (QoS) requirements to heterogeneous grid resources.

3. DAM: Double Auction-based Meta-Scheduling Mechanism

Figure 2 shows the elements of the meta-scheduler, which can be divided into three parts: (1) collection: meta-scheduler collects queue information from the resources, (2) valuation: assign values to the user applications and resource queues, and finally, (3) matching using CDA. In a CDA, both sellers and buyers submit bids to an auctioneer who continually ranks them from highest to lowest in order to generate demand and supply profiles. From the profiles, the maximum quantity exchanged can be determined by matching selling offers or asks, starting with lowest price and moving up, with the demand bids, starting with highest price and moving down. This format allows buyers to make offers and sellers to accept those offers at any particular moment. In the Figure 1, U_n represents user application, a_k and b_n represent ask and bid, and ${}^{m}Q_k$ represents resource queue. At regular intervals (henceforth referred to as scheduling intervals), the meta-scheduler matches the jobs to the resource queues if the deadline constraint of the application is satisfied. If a job cannot be matched, then it is considered in the next scheduling interval.

4. Pricing Mechanism

The Grid services may be priced based on the cost of infrastructure, and economic factors like supply and demand. However, user needs and urgency, and simultaneously, efficient utilization of Grid services must be reflected through pricing (valuation) of user applications and resources. Therefore, the meta-scheduler must generate a pricing metric for both users and resources that takes into account all these constraints. This pricing is dynamic, that is, in each scheduling cycle; it gets updated based on various parameters, and the dynamic demand and supply of system.

Valuation (Pricing) of Resources: In order to balance load across independent grid services, the meta-scheduler tries to submit more jobs to the least loaded resources. Also, the most urgent job must be matched to the fastest queue.

Therefore, the valuation of resources should be such that the resource with minimum load should get minimum value (as in CDA, the maximum bid is matched to minimum ask). Therefore, $P_R(t)$, the price of a resource at time t, is determined by the following:

$$P_R(t) = K \times w_{R(t-1)} \times c_R \times l_{(t-1)}, \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where $w_{R(t-1)}$ is average queue waiting time, *Demand* is total number of tasks to allocate, *Supply* is total number of CPUs in all resources, c_R is initial price given by the resource, and $l_{(t-1)}$ load of resource.

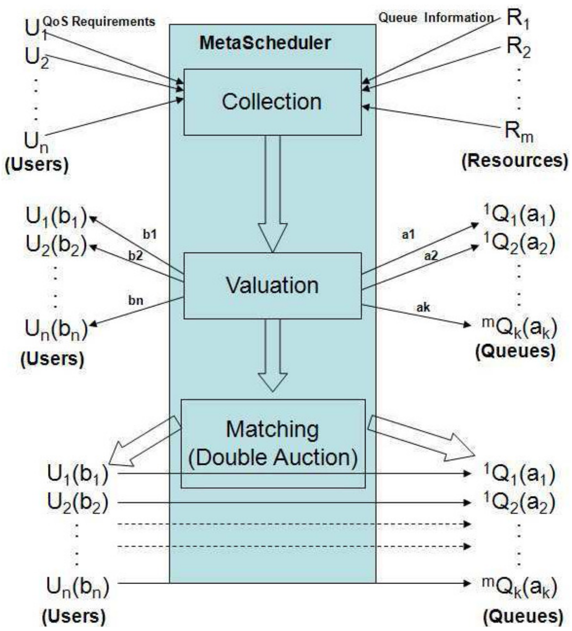

Figure 2. Main Steps in DAM Mechanism

Valuation (Pricing) of User Application: As discussed previously, each user submits to the meta-scheduler his/her budget (b_u), deadline (d_u), application length (l_u), and the number of nodes required (n_u). Let $P_u(t)$ be the valuation of the user application. Following is the metric for pricing of user applications,

$$P_u(t) = k_u \times \frac{b_u}{d_u - T_t} \times \frac{Demand}{Supply} \times (T_t - S_t) \ldots\ldots\ldots (2)$$

5. Experimental Results

We have evaluated DAM mechanism through extensive simulations using real workload traces gathered from existing supercomputers [12]. For the experiments, we have selected a subset of 500 applications (associated with each user) from the trace of the Linux cluster (Thunder) at LLNL for the duration between February and June 2007 [12], competing for 8 Grid Resources (simulated European Data Grid 1 test bed). The user application is modelled as a Bag-of-Task application, i.e., about 30,000 jobs were submitted to the Meta-scheduler. The initial valuation of user applications and resource are randomly generated using uniform distribution. We compared our scheduling mechanism (DAM) with two traditional scheduling mechanisms (SJF and FCFS) and two market-based scheduling mechanisms (HBFQ and FairShare) using performance metric such as number of Deadline missed, Deadline urgency and Budget per job.

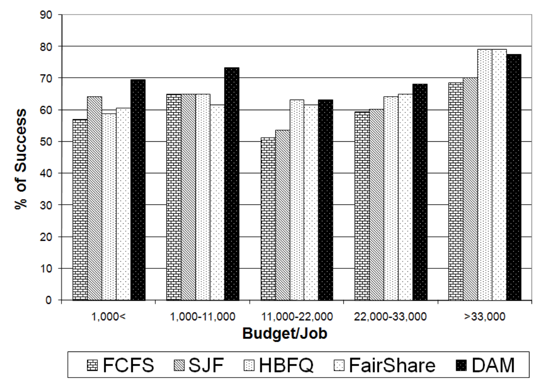

Figure 3. Effect of user budget

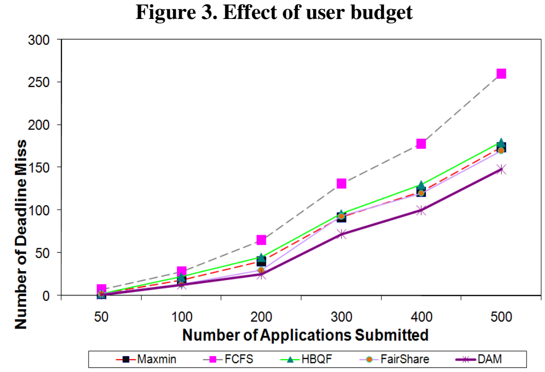

Figure 4. Number of Deadline Miss

Figure 3 clearly shows that our mechanism is not only able to satisfy more users, but also benefited all users in different budget ranges. DAM schedules almost equal number of users in low and high budget groups. From Figure 4, we can clearly see as the demand for resources (number of user applications) increases; the number of applications that missed their deadline also correspondingly increases due to the scarcity of resources. In this scenario, DAM is able to satisfy more number of users than other mechanisms as DAM is assigning valuation to user applications according to deadline. In FCFS, deadline misses is increasing rapidly due to starvation of many urgent applications. Figure 5 shows the increase in load of resource over time. It shows how resource demand is increasing with time with more jobs assigned to faster resources having more number CPUs.

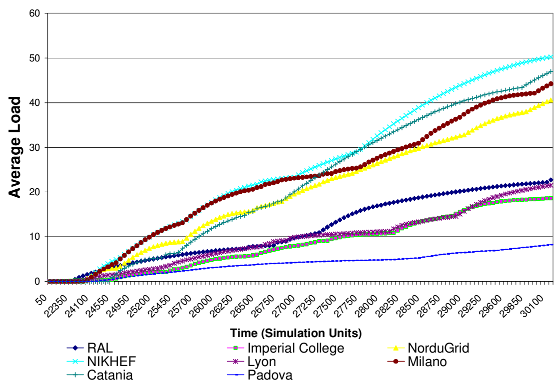

Figure 5. DAM: Load Variation

6. Conclusion

We presented the Grid Market Architecture and designed a Meta-Scheduling mechanism which integrates the capabilities of both market-based and traditional scheduling algorithms. The meta-scheduler uses valuation metrics to map user applications to resources consisting of independent resources in a fair and efficient manner to benefit both user and resource side. The simulation study clearly shows the effectiveness of the proposed mechanism and shows how the meta-scheduler manages different user requirements in a scenario where the demand for the resources exceeds the supply. In future, we want to analyse our mechanism using different pricing schemes for user applications and resources. In addition to that, we will also integrate them in real Meta-schedulers to further analyse the efficiency of DAM.

References

[1] Charlie Catlett, "The Philosophy of TeraGrid: Building an Open, Extensible, Distributed TeraScale Facility", *Proc. of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid*, May 21-24, 2002.

[2] Smirnova et al, "The NorduGrid Architecture and Middleware for Scientific Applications". *Lecture Notes in Computer Science*, v. 2657. pp. 264--273.

[3] Andrade et al, "OurGrid: An Approach to Easily Assemble Grids with Equitable Resource Sharing,", In *Proc. of the 9th Workshop on Job Scheduling Strategies for Parallel Processing*, June 2003

[4] Gagliardi et al., "Building an infrastructure for scientific Grid computing: Status and goals of the EGEE project," *Philos. Trans.: Math. Phys.Eng. Sci.*, vol. 363, 2005.

[5] Glite Workload Manager. http://glite.web.cern.ch/glite/wms/

[6] E. Huedo, R. Montero, and I. Llorente, "A framework for adaptive execution on grids", *Software - Practice and Experience*, vol. 34, no. 7, pp. 631–651, June 2004

[7] C. Smith, "Open source metascheduling for virtual organizations with the community scheduler framework (CSF)", *Technical whitepaper, Platform Computing*, 2003.

[8] B. N. Chun and D. E. Culler, "User-centric performance analysis of market-based cluster batch schedulers", *2nd IEEE International Symposium on Cluster Computing and the Grid*, May 2002.

[9] Lai et al, "Tycoon: An implementation of a distributed,market-based resource allocation system", *Multiagent Grid System*, vol. 1, no. 3, pp. 169-182, Aug. 2005.

[10] Alvin AuYoung, Brent N. Chun, Alex C. Snoeren, and Amin Vahdat, "Resource allocation in federated distributed computing infrastructures", *Proceedings of the 1st Workshop on Operating System and Architectural Support for the On-Demand IT Infrastructure (OASIS04)*, October 2004.

[11] R. Buyya, D. Abramson, and S. Venugopal, "The Grid Economy", Proc. of the IEEE, Vol. 93, No. 3, pp. 698-714, IEEE Press, New York, USA, March 2005.

[12] Parallel Workloads Archive, Available: http://www.cs.huji.ac.il/labs/parallel/workload, May 2005.