Decision Support Systems 52 (2012) 562–576

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Modeling and analysis of the effects of QoS and reliability on pricing, profitability, and risk management in multiperiod grid-computing networks

Jose M. Cruz [a,*], Zugang Liu [b]

[a] Department of Operations and Information Management, School of Business, University of Connecticut, Storrs, CT 06269-2041, USA
[b] Department of Business and Economics, Pennsylvania State University, Hazleton, PA, 08071, USA

ARTICLE INFO

Article history:
Received 28 July 2010
Received in revised form 31 August 2011
Accepted 11 October 2011
Available online 19 October 2011

Keywords:
Grid computing
Market network equilibrium
Pricing mechanism
Quality of service
System reliability
Risk management
Decision support systems

ABSTRACT

In this paper we develop a network equilibrium model for optimal pricing and resource allocation in Computational Grid Network. We consider a general network economy model with Grid Resource Providers, Grid Resource Brokers and Grid Users. The proposed framework allows for the modeling and theoretical analysis of Computational Grid Markets that considers a non-cooperative behavior of decision-makers in the same tier of the grid computing network (such as, for example, Grid Resource Providers) as well as cooperative behavior between tiers (between Resource Providers and Grid Brokers). We introduce risk management into the decision making process by analyzing the decision-marker's reliability and quality of service (QoS) requirement. We analyze resource allocation patterns as well as equilibrium price based on demand, supply, and cost structure of the grid computing market network. We specifically answer the following questions with several numerical examples: How do system reliability levels affect the QoS levels of the service providers and brokers under competition? How do system reliability levels affect the profits of resource providers and brokers in a competitive market? How do system reliability levels influence the pricing of the services in a competitive environment? How do users' service request types, QoS requirements, and timing concerns affect users' behaviors, costs and risks in equilibrium? How does the market mechanism allocate resources to satisfy the demands of users? We find that for users who request same services certain timing flexibility can not only reduce the costs but also lower the risks. The results indicated that the value of QoS can be efficiently priced based on the heterogeneous service demands.

1. Introduction

Grid computing is a form of distributed system wherein computing resources are shared across networks. These resources include CPU hours, storage, and applications and they are shared based upon their availability, capability, and cost, as well as the user's quality of service requirements. Consequently, grid technology can help organizations accelerate application performance, improve productivity and collaboration, and reduce total cost of IT infrastructure ownership. Grid computing has evolved from a niche technology associated with scientific and technical computing into a business-innovating technology that is driving increased commercial adoption. The growth of the grid computing market coupled with the increase in the number of market participants has generated many new challenges. In particular, since the grid-computing market is essentially a decentralized decision-making network, system reliability and quality of service (QoS) are among the major concerns of the market participants and users. In this paper, we study how the issues of quality of service, reliability, risk and standards can be resolved in this emerging market, and how these issues will affect the decision process and support systems as well as a company's bottom line. In particular, we investigate the following questions:

- How do system reliability levels affect the QoS levels of resource providers and brokers in a competitive market?
- How do system reliability levels affect the profits of resource providers and brokers in a competitive market?
- How do system reliability levels influence the pricing of services in a competitive environment?
- How do users' service request types, QoS requirements, and timing concerns affect users' behaviors, costs and risks in a competitive market?
- How does the market mechanism allocate resources to satisfy the demands of users with different QoS requirements?

Recently grid computing market design and resource allocation have been a theme of many studies. A significant contribution to establish a market vocabulary and structure that can be associated with grid resource markets is from the work of Buyya [6], Wolski et al. [52], and Bapna et al. [3]. Wolski et al. [52] considers a multi-

* Corresponding author.
E-mail address: jcruz@business.uconn.edu (J.M. Cruz).

0167-9236/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2011.10.012

commodity (CPU and storage) resource market and suggests that a tâtonnement-based commodities exchange market structure is a better choice for controlling grid resources than auction strategies previously defined in Regev and Nisan [41] and Waldspurger et al. [48], particularly with respect to price stability and resource utilization efficiency. Buyya [6], on the other hand, focuses on the implementation issues of scheduling and resource brokering and suggests a variety of market mechanisms, including models based on commodity markets, posted-prices, auctions and bargaining. Lin and Lin [34] examine the optimal service priority selection problem for a grid computing service user, who submits a multi-subtask job for the priced services in a grid computing network. Bapna et al. [3], however, were the first to model grid resource allocation based on economic principles. They develop a decentralized market model using a combinatorial auction approach and consider only two classes of decision makers, buyers and sellers. Zhang et al. [57] use real option valuation technique and simulation to analyze economic decision criteria for a grid computing provider wishing to provide such a service to businesses. These studies, however, did not consider QoS, reliability, and risk management in a multiple criteria, multiple resource, multiple decision-makers and multiperiod grid computing network optimization framework.

We propose what we believe is a novel approach to the modeling and analysis of grid computing markets. We model the multicriteria decision-making behavior of various decision-makers, which includes: the maximization of net profit and the minimization of risk. The proposed framework which is based on network equilibrium theory, allows for explicit modeling of decentralized decision-making behavior of the market participants. This approach differs from the recent models [3,6,48,51,52] in that, first and foremost, we consider several different types of decision makers and model their behavior and interactions explicitly. Second, we introduce risk management into the decision making process by analyzing the decision-marker's reliability and QoS requirement. Finally, the proposed framework provides the tools for computing efficient resource allocation patterns as well as equilibrium price patterns based on demand, supply, and cost structure of the grid computing network.

Many authors have previously addressed the issues of QoS, reliability, and risk in grids. In terms of QoS, Dogan and Ozguner [13] consider the problem of scheduling a set of independent tasks with multiple QoS requirements. Golconda and Ozguner [28] compare five QoS-based scheduling heuristics. Ernemann [15] addresses the idea of applying economic models to the scheduling task. Chunlin and Layuan [10] consider scheduling finite resources to satisfy the QoS needs of various grid users with multiple dimensional QoS requirements. In this paper, however, we take a network market approach to QoS and assume that the user's transaction cost is a function of QoS requirement. Moreover, we assume that QoS is a function of the individual decision-maker's system reliability.

In terms of reliability, Hwang and Kesselman [22] present a failure handling system based on work flows. Jin et al. [27] propose a framework for the adaptive deployment of failure detectors and, based on it, a policy-based failure handling mechanism to choose the appropriate failure recovery method. Kola et al. [32] provide a classification of faults in large distributed systems (with the main focus on grids). Huedo et al. [21] evaluate the reliability of computational grids from the end user's point of view. The book of Xie et al. [55] present many models for measuring reliability of grid computing system. In this paper, in addition to modeling the reliability of individual decision nodes and links, we incorporate reliability into the decision making processes.

Most of the research concerned with grid computing risk have addressed some form of risk in computing jobs. Irwin et al. [24] and Popovici and Wilkes [40] minimize the risk of paying penalties to compensate users so as not to reduce the profit of service providers. Kleban and Clearwater [30,31] determine the risk of completing jobs later than expected based on either the makespan (response time) or the expansion factor (slowdown). Yeo and Buyya [56], in contrast, provide a way to evaluate reliability and risk with respect to the required objectives, such as the deadline to complete the job, the budget and the penalty for any deadline violation. Nevertheless, these research consider the case of a single decision-maker and did not consider the grid network market competition and risk.

The grid computing network market equilibrium approach, developed in this paper, permits one to represent the interactions between decision-makers in the market in terms of network connections, reliability, QoS, risk, flows, and prices. We consider the relationships among QoS, reliability and risk. In addition, we consider non-cooperative behavior of decision-makers in the same tier of the grid computing network (such as, for example, grid resource providers) as well as cooperative behavior between tiers.

The paper is organized as follows. In Section 2, we present the model setup. In Section 3, we model the various decision-makers and their behavior and analyze the equilibrium among the heterogenous decision makers. In particular, we develop a variational inequality formulation governing the entire grid computing network. In Section 4, we conduct computational studies to investigate the five questions raised in the beginning of the paper. In Section 5, we discuss the managerial insights. In the concluding Section 6, we summarize our results and suggest directions for future research.

2. Grid computing network market model setup

In this Section, we develop an equilibrium model for a grid computing network market in which decision-makers operate in a decentralized manner. We develop an economic model of the computational grid market that provides an optimal pricing and resource allocation mechanism. Grid computing is exemplified in the Globus project [19] and GrADS project [4]. We note that grid computing is closely related to the fasting emerging business model of “cloud computing”. In particular, Aymerich et al. [2] pointed out that “in computational terms cloud computing is described as a subset of grid computing concerned with the use of special shared computing resources”. In addition, Foster et al. [17] argued that “cloud computing relies on grid computing as its backbone and infrastructure support.” In this case, various services (e.g. financial and accounting, database storage and analysis, scientific computing, risk management, customer relationship management, etc.) are offered by the “software as a service”, to customers (users). The revenue of the SaaS industry in 2010 has grown by 15.7% worldwide and was forecasted to reach $9.2 billion [18]. These SaaS providers may or may not own the resources necessary for the computations, and can purchase the resources from resource providers such as Amazon (Compute Cloud Ec2 and Data Could S3) and Sun Microsystems [3].

The structure of the grid presented in this paper generalizes those of Nimrod/G, World Wide Grid [9], Legion [33], NetSolve [43], and DISCWorld [11]. In these grids there are three sets of entities involved in the grid computing — job owners (users), resource owners (resource providers), and the scheduler (broker). Job owners specify their job requirements, the deadline, the budget available for their job, and the time at which the job will be ready to be processed [52]. The resource owners set the price for using their resources, and specify their capacities and availabilities. Examples of resource pricing can be found at Sun Grid [20,47] and the cluster of Tsunamic Technologies Inc. [23]. Note that our network is general enough and therefore a two-tiered network with providers and buyers is a special case of our network where the providers also manage resources.

The economic model presented in this paper reflects those of the Nimrod/G, World Wide Grid [9] and in Buyya et al. [8]. The World Wide Grid consists of computers in five continents and has been used for drug design [5]. The economic model in World Wide Grid considers deadline, budget, and dollar cost of resources to decide the assignment of jobs to resources. Buyya et al. [8] proposed an economic approach for scheduling jobs in a grid. They considered budget,