Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Using economic regulation to prevent resource congestion in large-scale shared infrastructures

Xavier León [a,*], Tuan Anh Trinh [b], Leandro Navarro [a]

[a] Distributed Systems Group, Departament d'Arquitectura de Computadors, Universitat Politècnica de Catalunya, Barcelona, Spain

[b] Network Economics Group, Department of Telecommunications and Mediainformatics, Budapest University of Technologies and Economics, Hungary

ARTICLE INFO

Article history:
Received 3 August 2009
Received in revised form
6 November 2009
Accepted 16 November 2009
Available online xxxx

Keywords:
Resource congestion
Economic resource allocation
Economic regulation
Incentive mechanism
Large-scale shared computational infrastructures

ABSTRACT

In this paper we study the problem of large-scale resource congestion from the control and regulation point of view. Applications and services running in large-scale shared infrastructures like Grids or PlanetLab have different resource usage profiles and different resource consumption strategies according to their specific requirements. However, users of these types of infrastructure tend to prefer a subset of available nodes to execute their tasks. As a result, this pattern of user behaviour usually leads to an unfair distribution of work between nodes — i.e. some nodes are highly loaded while the others remain almost idle. We find that most current research focuses on short-term and per-resource scheduling, and the issue of efficient resource allocation in the long term, and system wide, is not yet appropriately studied. Thus, there is a need for controlling, distributing and limiting the capacity of each participant to consume resources considering the state of the system as a whole. Our main contribution is the introduction of a novel macro-scheduling (long-term and system-wide) mechanism for resource capacity self-regulation in which virtual currency or money is used as a tool to govern resource and service usage in massively distributed settings, which are otherwise hard to control. We show by simulation that our approach successfully redistributes the load in a fair and economically efficient manner.

1. Introduction

The current and future Internet, as an open shared network and service infrastructure, is used by more people and more diverse applications every day. However, the growth in usage, capacity and diversity makes it increasingly difficult to provide users with sustainable, high-quality services.

In addition, shared computing infrastructures rely on the individual contributions of participants to create an infrastructure with enough power to run large-scale applications and services. A key characteristic of this type of shared infrastructure is its peer-to-peer nature, in which participants are both consumers and resource providers acting in their own interests. Examples include scientific collaboration grid networks [1,2] or network testbeds such as PlanetLab [3] or EmuLab [4].

To solve the resource allocation problem, traditional schedulers optimize usage, throughput or response time at a cost of centralizing components and compromising scalability. An opposite approach is to implement an economic solution in which computational markets give users control over the service levels they require in large-scale resource sharing. These economics-inspired systems have been shown to be a decentralized, scalable and efficient way of allocating resources according to user preferences.

However, the combination of resource-intensive applications and user preferences can lead to levels of demand that saturate the infrastructure, negatively affect other users or compromise the overall stability of the system. In addition, saturation usually affects only a subset of the infrastructure resources, known as hotspots, whereas the load remains low over the rest of the system.

In this paper, we present a set of mechanisms for self-regulation of resource and service exchange to ensure that work is distributed in a fair and stable way. Besides, current economics-inspired models focus only on short-term (i.e. micro-economic[1]) interactions between participants, and the effects of long-term, system-wide (i.e. macro-economic[2]) interactions have not yet been analysed in depth. Although economic scheduling is efficient regarding the

* Corresponding address: Jordi Girona, 1-3. Building D6-001, Campus Nord, Barcelona 08034, Spain. Tel.: +34 93 40 16783.

E-mail addresses: xleon@ac.upc.edu (X. León), trinh@tmit.bme.hu (T.A. Trinh), leandro@ac.upc.edu (L. Navarro).

URLs: http://dsg.ac.upc.edu (X. León), http://netecon_group.tmit.bme.hu (T.A. Trinh), http://dsg.ac.upc.edu (L. Navarro).

[1] A branch of economics that focuses on the ways in which individuals, households and firms determine how to allocate limited resources, typically in markets where goods or services are being bought and sold.

[2] A branch of economics that focuses on the behaviour of an economy at the aggregate level and the effects of government actions (such as laws or taxation levels).

0167-739X/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.future.2009.11.004

social welfare of a system, it can also lead to system-wide performance penalties.

We stress the importance of introducing regulatory mechanisms for controlling and limiting user demand for shared resources. Consequently, our main *contribution* is the *introduction of a novel mechanism for self-regulation of resource capacity based on virtual currency management*. Besides, we show that our macroeconomic mechanism is a powerful tool that (i) provides users with incentives to distribute their tasks to prevent the emergence of hotspots in large-scale infrastructures and (ii) enables the redistribution of wealth to improve the social fairness of the system. Additionally, the money-based infrastructure introduces an economic incentive for enforcing regulatory standards to improve the overall governability and sustainability of the network.

The rest of this paper is organized as follows: in Section 2 we present the motivation and problem statement; in Section 3 we present related work on resource allocation mechanisms; in Sections 4 and 5 we describe the system model and present our regulatory mechanism; in Section 6 we analyse the simulation results; in Sections 7 and 8 we conclude the study by discussing the applicability of our solution and proposing areas for future work.

2. Motivation and problem statement

Free access and unrestricted demand for finite resources ultimately leads to over-exploitation and degrades quality of service (QoS). This problem arises because the benefits of exploitation are received by individuals – each of whom is determined to maximize their use of the resource – whereas the costs of the exploitation are distributed between all of those who have access to the resource. This, in turn, increases demand for the resource to such an extent that the resource is exhausted and becomes useless. This is a clear example of the well-known problem *tragedy of the commons* [5] or *free-riding*.

As a motivation, in this section we analyse a reference system like PlanetLab using data from CoMon, a monitoring infrastructure for PlanetLab [6]. We measured PlanetLab usage over a one-month period,[3] restricting our study to available nodes (i.e. nodes to which users actually had access) and discarding those nodes that were inoperative due to maintenance work, network connectivity problems, or for other reasons.

Fig. 1(a) shows the load quartiles (first, second and third quartiles) during the measurement period for all working nodes in PlanetLab. Despite daily variations, the nodes appear to be very highly loaded: the mean for each day is over 3, which is considered to be overloaded [7]. This is not necessarily a problem, since it is a sign of the usefulness of the infrastructure for the users. However, Fig. 1(b) shows that almost 50% of nodes are persistently overloaded whereas the others are underloaded. The overloaded nodes provide lower QoS to applications, whereas the other nodes are mainly idle. In Fig. 1(c), we can see that the distribution has a positive skew, which indicates that the mass of the distribution is unbalanced.

If we consider a reasonable scenario in which tasks are distributed uniformly among resources, the load distribution (i.e. the number of tasks running on a given resource) should follow a normal distribution, with low variance and a skew value close to zero, which indicates that most of the nodes support the same workload. We assume that our target load distribution should be similar to a normal distribution because if tasks are distributed randomly among resources following an unbiased and uniform distribution, all resources are treated equally. Therefore, according to the law of large numbers, if there are a large number of resources, the sum of independent identical variables (the sum of tasks on a resource equals the load) constitutes a normal distribution.

The presence of overloaded nodes is a result of correlated user preferences (i.e. users tend to prefer similar nodes in terms of reputation or technical characteristics) and the lack of incentives to behave considerately in these types of collaborative environment.

Even if micro-economic schedulers are used, the correlated preferences could lead to overloading certain nodes with higher preference weights. However, the higher revenue generated by these overloaded nodes leads to an unfair distribution of wealth among participants.

This scenario is very similar in real-world economies, where wealth is distributed following a Pareto distribution (i.e. a "long-tail" distribution). Although a Pareto wealth distribution is not inherently unhealthy, a fair distribution of wealth (in open testbed infrastructures like PlanetLab) would be one that gives researchers the same chance to test their proposals in similar conditions, regardless of their actual incomes.

To address the unbalanced load between nodes and the unfair wealth distribution between users, we propose a macroeconomic approach (i.e. a long-term, system-wide strategy) based on regulation through virtual currency management, which can be viewed as a capacity management mechanism, that gives resource providers and consumers incentives to redistribute the load in large-scale shared infrastructures to prevent congestion on hotspot nodes and to distribute wealth equitably — i.e. accommodate user demand in a more reasonable scenario (see Fig. 1(b)). Our solution is based on automatically detecting and forcing the redistribution of tasks by taxing resource prices until we reach a reasonable distribution of work.

3. Related work

In this section we present related resource allocation studies in which both economic and non-economic mechanisms are used.

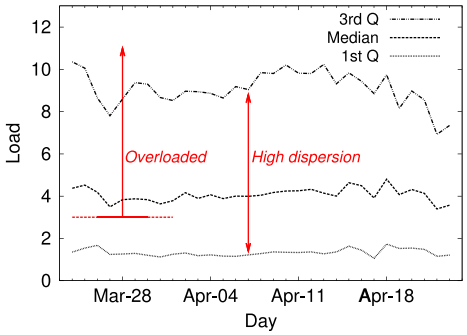
Studies which do not consider economic concepts analyse the resource allocation as a scheduling problem (see the survey by Pinedo [8]). These proposals range from simple, centralized scheduling algorithms like First Come First Serve (FCFS), to Shortest Job First (SJF), which provides efficient allocations but does not take into account the different values of the users' tasks.

On the other hand, extensive research has also been carried out into the application of economic models to resource allocation in large-scale shared infrastructures. Computational markets have been shown to allocate resources efficiently in a decentralized way in the presence of selfish utility-optimizing resource consumers and selfish profit-optimizing resource providers. Shirako [9] is a toolkit for building utility services for dynamic on-demand sharing of networked resources through programmatic interfaces. Shirako is based on a common, extensible resource leasing abstraction [10] similar to those used in the allocation of airline seats. It combines elements of lifetime management and mutual exclusion. Although Shirako mainly uses flexible mechanisms for trading resources between clients through a series of leases, there is no regulation mechanism between resource brokers, which can lead to over-provisioning of resources. Consequently, negotiated service level agreements (SLAs) might be broken and the infrastructure, or a subset of it, may suffer congestion.

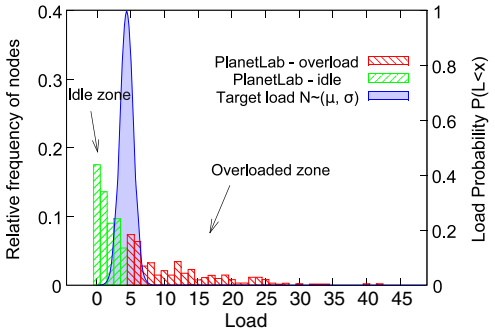
Bellagio [11] is a market-based resource allocation system for federated distributed computing infrastructures like PlanetLab. The Bellagio architecture is based on a centralized auctioneer which allocates resources periodically and determines the corresponding user payments. Users specify resources of interest by bidding in a combinatorial auction [12]. The amount of virtual currency owned by a site is determined directly by the central authority, which establishes the share of virtual currency assigned to each site. Although the final allocation is proportional to the bids advertized by users, Bellagio does not provide software agents that can

[3] Observed from 24 March to 23 April 2009, with samples of each node taken every 5 min. Traces are publicly available following the instructions at http://comon.cs.princeton.edu/#DataAccess.

Please cite this article in press as: X. León, et al., Using economic regulation to prevent resource congestion in large-scale shared infrastructures, Future Generation Computer Systems (2009), doi:10.1016/j.future.2009.11.004



(a) Load evolution.



(b) Load distribution.

	Real Load	Target Load
Mean	4.393	$\simeq$ 4.393
StDev	6.565	min
Median	5.935	$\simeq$ 4.393
Skewness	2.118	$\simeq$ 0.000

(c) Load (uptime).

Fig. 1. Measures of system overload for all working nodes in PlanetLab from 24 March to 23 April 2009. (a) The evolution of load quartiles. (b) The frequency histogram of the load average for each node and the load CDF for the same period. The green zone represents idle resources and the red zone represents overloaded resources. (c) The load distribution statistics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

act on behalf of their users to maximize their utility; consequently, the system can be sub-optimal, because human users may not behave in an economically rational way under certain circumstances.

Buyya et al. developed Nimrod-G [13], a resource broker that supports deadline and budget constrained scheduling algorithms [14] for task-intensive applications in clusters. However, their algorithms are only designed to make efficient local resource allocations between competing users, and do not establish a coordination between schedulers or fully analyse the long-term effect of their algorithms on the overall infrastructure. Also, studies of double auctions (see the survey by Friedman [15]) or combinatorial auctions (see the survey by De Vries [16]) allocate users to resources on the basis of short-term economic efficiency and do not take into account infrastructure-wide metrics such as the distribution of work among resources.

Finally, Tycoon [17] is a distributed market-based resource allocation system in which every node in the system runs an independent auction for its local resources. Auctioneers conduct a proportional share-based auction in which users receive a proportional amount of a single resource (virtualized CPU and memory) determined by the size of the bids made by all users for the same resource. Users are assigned a fixed amount of currency to spend over time to allocate their tasks. Although this model is similar to the one presented in Section 4, it does not consider system-wide metrics associated with correlated resource preferences—e.g. the uneven distribution of work among nodes. In addition, the Tycoon model considers symmetric systems in which all users have the same budget, whereas our model takes into account the behaviour of the system in the presence of variable budget constraints between participants.

We believe that most current research focuses on the short-term allocation and maximization of user utility and does not address the behaviour and health of the system as a whole or system-wide metrics like the distribution of work load, the proportionality between consumption and contribution, or the impact of different budget constraints on user utility.

4. System model

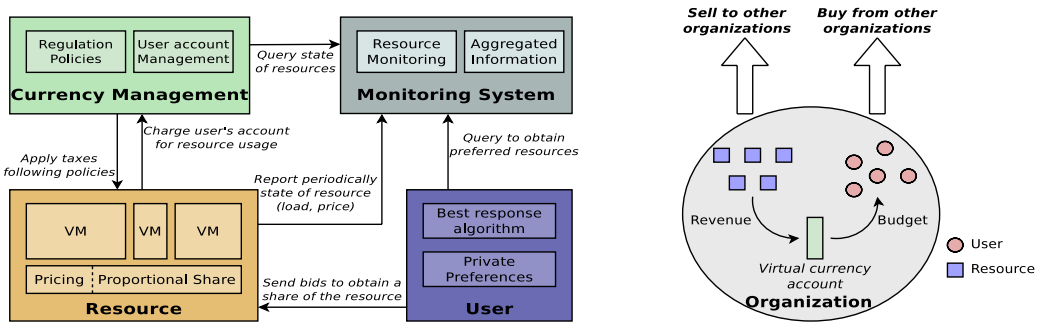
Our solution is designed for a system consisting of an arbitrary large set of nodes (physical or virtual machines) at diverse locations which communicate via message passing over a network such as the Internet. The system is dynamic in the sense that nodes and networks can be added or removed and can degrade (overload) or fail at any time. The nodes are resources owned by different organizations and, although there are common protocols and rules, there is no need for a central executive authority to carry out the day-to-day management of the system. Each organization can freely determine the number of resources it contributes to (or shares with) the system beyond a specified minimum. Participants in the system are human users (or software agents participating on their behalf) who usually belong to a single organization and execute their tasks across a subset of available nodes (see Fig. 2(b)).

In view of the above scenario and the benefits of market-based resource allocation in decentralizing resource scheduling in a end-to-end way, the system requires a short-term micro-economic resource allocation foundation to enable users to express their preferences as prices. Extensive research has been carried out in this field. The most popular approach is to use some form of auction to extract the market price directly from the users' bids, for example an English auction, a Dutch auction, a double auction or combinatorial auction [18,15,16,19]. The main drawbacks of auction-based systems are that the response time is slow (bidders have to wait for auction clearing) and they are unsuitable for divisible resources because of the complexity involved in determining the most efficient way to divide a resource.

4.1. Proportional share allocation

The simplest and most appealing mechanism for shared divisible resources is to use proportional share auctions, which have already been proposed for OS process scheduling [20], I/O disk scheduling [21] or task scheduling in grid environments [22]. In this case, allocations are proportional to the consumer's weight

Please cite this article in press as: X. León, et al., Using economic regulation to prevent resource congestion in large-scale shared infrastructures, Future Generation Computer Systems (2009), doi:10.1016/j.future.2009.11.004



(a) System architecture. (b) Virtual organization model.

Fig. 2. System architecture and system model overview. (a) The interactions between components. (b) A schematic representation of a virtual organization based on contributed resources that earn virtual currency and users who consume resource from other organizations. Resources and users from the same organization share the same virtual account.

(or preference for a resource) and inversely proportional to the sum of all other users' weights for the same resource. Therefore, we base the mode for our system on the price-anticipating mechanism proposed in [23], in which each user submits a bid for resources and the price of the machine is determined by the total bids submitted. More formally, the price of resource j is set to $Y_j = \sum_{i=1}^{k} x_{ij}$, where k is the number of bids on resource j and x_{ij} is a non-negative user's i bid for resource j. Following the proportional share allocation mechanism, user i receives a fraction $r_{ij} = \frac{x_{ij}}{Y_j}$ of resource j.

A game theory analysis of the aforementioned price-anticipating mechanism can also be found in [23]. Feldman et al. propose an algorithm for finding the *best response*[4] of an agent to the system. Given a fixed budget X and a pool R of divisible resources, the algorithm finds the distribution of bids across resources that yields the highest utility for an individual player i by solving the following optimization problem (1):

$$\text{maximize } U_i\left(\forall_{j\in R}\frac{x_{ij}}{Y_j}\right) \text{ subject to} \tag{1}$$

$$\sum_{j=1}^{m} x_{ij} = X_i \quad \text{and} \quad x_{ij} \geq 0.$$

The computational cost of the optimization algorithm is $\theta(n \log n)$, which is acceptable considering the computational power of current hardware and the input size of the problem. Finally, Feldman et al. show that there is always a Nash equilibrium when the players' utility functions are strongly competitive, i.e. when there are at least two users competing for each resource, which is a reasonable assumption in systems like PlanetLab. They also show that the Nash equilibrium resulting from the best response dynamics is efficient and fair.

4.2. Metrics

The *utility* of each user i is represented by a function U_i of the shares obtained by the user from each machine. An important issue in representing utility is the notion of preference for resources, since each user could have a different preference for the same machine. Consequently, we consider a linear utility function $U_i(r_{i1}, \ldots, r_{in}) = w_{i1}r_{i1} + \cdots + w_{in}r_{in}$, where w_{ij} is the private preference of user i for resource j. This utility function is suitable for heterogeneous environments in which resources are valued differently by each participant.

To study the behaviour of our proposed mechanism, we consider the following metrics.

- Load uniformity and dispersion: The load distribution is an interesting metric for measuring the overall health of the system in terms of congestion. In this model we assume that if users are willing to spend virtual money on a resource, they will eventually execute a process. Therefore, we consider the load L on a resource to be the number of positive bids on that resource. To measure the distribution of load among nodes, we define two different metrics. Firstly, the *load uniformity* is represented by $\frac{\min L_i}{\max L_i}$, which is the ratio between the minimum and maximum loads. The higher the ratio, the greater the distribution of the load among resources, since each resource supports a similar workload. Secondly, we measure the dispersion of the load as the standard deviation $\sigma(L)$ of the node loads.
- Efficiency (price of anarchy): For an allocation scheme ω at equilibrium, the efficiency is computed as $\pi(\omega) = \frac{U(\omega)}{U^*}$, where $U(\omega) = \sum_{i=0}^{n} U_i(\omega_i)$ and $U^* = \max(U(\omega))\forall\omega$. In our case, it is easy to compute the social optimum U^* because it is achieved when we allocate a whole node to the user with the highest preference weight on that node. Thus, it represents the loss in efficiency as a result of user's selfishness and decentralization.
- Fairness (uniformity, envy-freeness): To represent the fairness of our system, we consider two different metrics: utility uniformity and envy-freeness. *Utility uniformity* is represented by $\upsilon(\omega) = \frac{\min U_i(\omega_i)}{\max U_i(\omega_i)}$, which is the ratio between the minimum and maximum utilities. The higher the ratio, the fairer the mechanisms, since users obtain similar utility from the system. Another way to measure the fairness of an allocation in Economics is to determine the *envy-freeness* [24], which is represented by $\rho(\omega) = \min(\min_{ij} \frac{U_i(\omega_i)}{U_i(\omega_j)}, 1)$, where $U_i(\omega_i)$ is the utility of user i and $U_i(\omega_j)$ is the utility that user i would have if it was allocated the resource shares of user j. In other words, envy is related to user i's perception of its own allocation with respect to those received by the other users.

An economically healthy resource allocation scheme should enforce a Nash equilibrium with high efficiency and high fairness. We also aim to guarantee high load uniformity and low dispersion, to distribute the load evenly while maintaining the high efficiency and fairness produced by the proportional share model.

5. Currency Management System: An economics-inspired self-regulation mechanism

As explained in Section 2, the main problem is the unfair distribution of tasks among nodes in heavily loaded systems. As in the real economy, free-market mechanisms sometimes fail to address such problems and central governments impose restrictions (regulations) on the system that are designed to act as incentives to

[4] In game theory, the best response is the strategy (or strategies) that produces the most favourable outcome for a player, taking other players' strategies as given.

Please cite this article in press as: X. León, et al., Using economic regulation to prevent resource congestion in large-scale shared infrastructures, Future Generation Computer Systems (2009), doi:10.1016/j.future.2009.11.004

behave in a certain way. The existence of a central authority with a degree of global and aggregated knowledge of the system does not necessarily restrict its scalability, as discussed in Section 7.

We propose a self-managed regulatory body (i.e. a virtual bank or *Currency Management System (CMS)*) which manages through simple policies the virtual currency used by participants to bid for resources. The CMS has a two-fold aim: (i) to limit the amount of currency each user can spend on resources, thereby restricting their long-term purchasing power; and (ii) to introduce long-term, system-wide macro-economic policies that act as a self-regulation mechanism by taxing resource prices according specified policies and redistributing wealth (virtual currency) among participants to improve the overall fairness of the system.

Resource taxation (following specified policies) increases prices and encourages users to bid for alternative resources. For example, one policy would be to *improve availability* by imposing taxes on those resources with low availability; consequently, resources with low availability would generate less revenue because users would tend to change their bids to resources with lower taxes, which would, in turn, provide an incentive to improve the availability of resources. Similarly, another policy would be to tax overloaded nodes to attenuate hotspots.

It is important to note the difference between: (i) the resource *price*, which represents the cost required to gain possession of a resource considering user's preferences and competition; and (ii) the *tax* price, determined dynamically by our mechanism as a means for solving the uneven distribution of work, similar to other taxes on consumption such as *value-added tax* (VAT). The former is a micro-economic mechanism for the regulation of access to resources and the latter is a macro-economic mechanism which introduces a correcting factor based on an observed effect seen at macroscopic level (uneven distribution of work among congested resources).

The architecture of our proposal is shown in Fig. 2(a). The CMS gathers system-wide and aggregated statistics from the *monitoring infrastructure*, such as average load and its dispersion, the effective contribution and consumption of users, etc. After a certain period of time, or *epoch*, the CMS uses predefined policies to determine the appropriate taxes on each resource (a factor to be applied on the price). During the epoch, users can freely evaluate their needs and spend their budgets according to their own strategies, without any other external restriction. Once the taxes have been determined, the price of resource j is computed according to Eq. (2), where k is the number of bids on resource j, b_i is the bid of consumer i and $\text{tax}_j \in [-1, \infty) \subset \mathbb{R}$ is the tax applied by the CMS to resource j.

$$Y_j = \left(\sum_{i=1}^{k} b_i \right) * \left(1 + \text{tax}_j\right). \tag{2}$$

Since the aim of our proposal is to redistribute the load as evenly as possible, we impose a higher tax on resources with higher loads, which increases the price and encourages participants to use spare resource instead (those with a lower tax and, consequently, a lower price).

The CMS does not know the users' preferences in advance, so it cannot anticipate user behaviour in response to a specific set of taxes. We therefore use a heuristic based on the ratio between the load and the target load to move towards the set of target taxes under which the load is distributed equally among all nodes (e.g. the load dispersion is minimal).

Algorithm 1 shows the procedure for determining the tax to apply during the next epoch. The tax does not jump straight to the target value but instead moves towards it at a rate determined by a learning rule, which prevents oscillations and produces a smooth approximation to the target value. The learning rule used is the Widrow–Hoff rule, which is a well-known learning mechanism used for back-propagation in neural networks [25] and used to move towards the target price in different economic agents [26]. It contains a parameter β (learning rate) that represents the speed with which the adjustment takes place.

Therefore, once the CMS has gathered information about the load and the current tax for each resource, it computes the uniformity metric to determine the current load dispersion in the system and adjusts the learning rate of the algorithm: a lower learning rate is used at higher uniformity to produce a smooth approximation to the target load. The CMS then applies the heuristic described above to the current tax to determine the new tax to apply to each resource.

Algorithm 1 Pricing tax regulation algorithm

Require: $\tau \leftarrow$ target load
Require: $S \leftarrow \{\forall j \in R(load_j, tax_j)\}$ // Set of resource info
  uniformity $\leftarrow \frac{\min load_j}{\max load_j}$
  $\beta \leftarrow 1.0 - \text{uniformity}$
  for all $(load_j, tax_j) \in S$ **do**
    $\Delta_{tax,j} \leftarrow \frac{load_j}{\tau} * \beta$
    $tax_j^{t+1} \leftarrow tax_j^t + \Delta_{tax,j}$
  end for

6. Performance analysis

We use simulations to evaluate the long-term impact of our system. To determine the effectiveness of our regulatory mechanism in improving the load distribution in comparison with the free-market (unregulated) scenario, we compare the best-response dynamics from the game theory analysis of the price-anticipating model (Section 4) with the best response dynamics under our regulatory mechanism (Section 5).

Method. The set-up of the simulations consists in fixing the number of resources m to 100 and varying the number of users n (from 10 to 200) to assess the scalability of the solution as more users (and, therefore, more load) are added to the system. We do not present the results for a variable number of resources because the simulations showed that different executions with the same ratio of resources to users produce similar results. The best-response algorithm is updated after each time step (1 simulated minute) and the epoch (at the end of which the tax regulation algorithm is executed) is defined as 60 time steps (1 simulated hour).

User preferences. Some nodes are persistently more loaded than others (see Fig. 1(b)) due to correlations of user preferences. To capture these correlations, we experiment with the following user preference model. For each user, we create a list of weights that are independently and identically distributed according to a uniform distribution $U \sim (0, 1)$. Next, we arrange the list in descending order to create a user's preference weight on resources so that $p_i = (p_{i1}, \ldots, p_{in})$, p_{ik} represents user i's weight on resource j and $p_{ik} > p_{i(k+1)}$. We then normalize the expression so that $\sum_{j=0}^{m} w_j = 1$. Consequently, we expect to have a high load on the first resources and a lower load as the weights decrease.

Note that we only consider positive weights on resources, so every user obtains a certain positive utility from each resource. If we had included resources with weights equal to zero, those users following the best-response algorithm would not have bid on these resources because the utility provided is also zero. In practical terms, this means that those resources with a weight equal to zero are not available. Therefore, and with no loss of generality, we only consider available nodes in our simulations.

Convergence criteria and results. The convergence time is a measure of the speed with which the system reaches an equilibrium. As in [23], the price-anticipating system converges to a Nash equilibrium when the difference in the best-response utility between two time steps is less than ε (0.001 in our experiments). However,

Please cite this article in press as: X. León, et al., Using economic regulation to prevent resource congestion in large-scale shared infrastructures, Future Generation Computer Systems (2009), doi:10.1016/j.future.2009.11.004

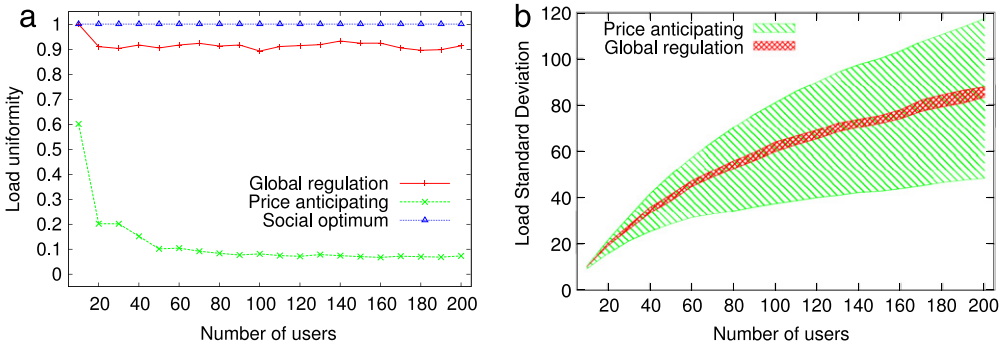


Fig. 3. (a) Load uniformity and (b) load dispersion.

when we apply our regulatory mechanism, we change the environment (i.e. the prices of resources) at the end of each epoch, and each epoch evolves iteratively to a different Nash equilibrium. Therefore, we consider that our tax regulation mechanism has converged when the value of $\Delta_{tax,j}$ (see Algorithm 1) is less than δ (0.1 in our simulations). The results presented in these sections are taken when the system has converged.

The simulation results show that the best-response dynamics converge after 5 iterations (5 simulated minutes), as in [23], whereas our regulatory mechanism converges in a range of 3–5 epochs (3–5 simulated hours). This shows that, although our mechanism is designed to be executed in a long-term time window, it is also able to converge in few iterations. Therefore, this mechanism can be executed frequently (i.e. using short epochs) when the load conditions are dynamic but the epoch can be longer when the load conditions are in a steady state.

Load uniformity and dispersion. The effectiveness of our proposal is based on distributing the load effectively among resources. We measure the load uniformity and its dispersion for the price-anticipating model and the regulatory mechanism, once they have converged. As shown in Fig. 3(a), the price-anticipating mechanism produces low load uniformity because users tend to bid for their preferred resources, which creates a large difference between the maximum and minimum loads. However, under our tax regulation mechanism, when the CMS detects that there is a subset of heavily loaded nodes and it begins to increase the corresponding tax (and, therefore, the price), users following the best-response algorithm tend to distribute their bids to cheaper (less loaded) resources to maintain as high a utility as possible. Our results show that the system encourages users to redistribute the load more evenly (similar load on nodes) regardless of the number of users.

Similarly, Fig. 3(b) shows the dispersion of the statistical variable load L ($\mu(L) \pm \sigma(L)$) in the system. As the number of users increases, the load on the system also increases. However, without tax regulation the dispersion is higher and increases with the number of users. Conversely, when taxes are applied the dispersion is maintained at similar values, which demonstrates that our proposal is scalable independently of the number of users.

Fig. 4(a) shows the empirical CDF of the load distribution for a simulation with 100 resources and 100 users. We can see that load is highly dispersed without regulation; approximately 50% of nodes are highly loaded, at levels above the mean load (>60 users bidding on them). However, when our regulatory mechanism is applied, the load distribution is centered at the target load (the mean load in our simulations) and the CDF shows that the variance is very low, because the load values of each resource fall within a small range (between 55 and 65).

Specifically, Fig. 4(b) shows that the standard deviation is very high when no regulations are enforced. However, under regulation, the average load remains similar but the standard deviation and skewness decrease. The Kolmogorov–Smirnov normality test for the global regulation case shows a significance value of 0.232 given the assumed significance level of 0.05, so normality cannot be ruled out. The normality assumption is also supported by the normal Q–Q plot in Fig. 4(c). The results show that our mechanism provides users with an incentive to redistribute their workloads evenly and ensures a reasonable distribution of work among resources.

Importantly, the behaviour of the price-anticipating algorithm (without regulation) is similar to the high load variance illustrated in Fig. 1(b) and (c) in Section 2, which shows the relationship between our simulations and real observed results from PlanetLab. These results clearly demonstrate that our regulatory mechanism redistributes the load among nodes and achieves a similar distribution to our target distribution, where the load is centered at the mean and shows low dispersion.

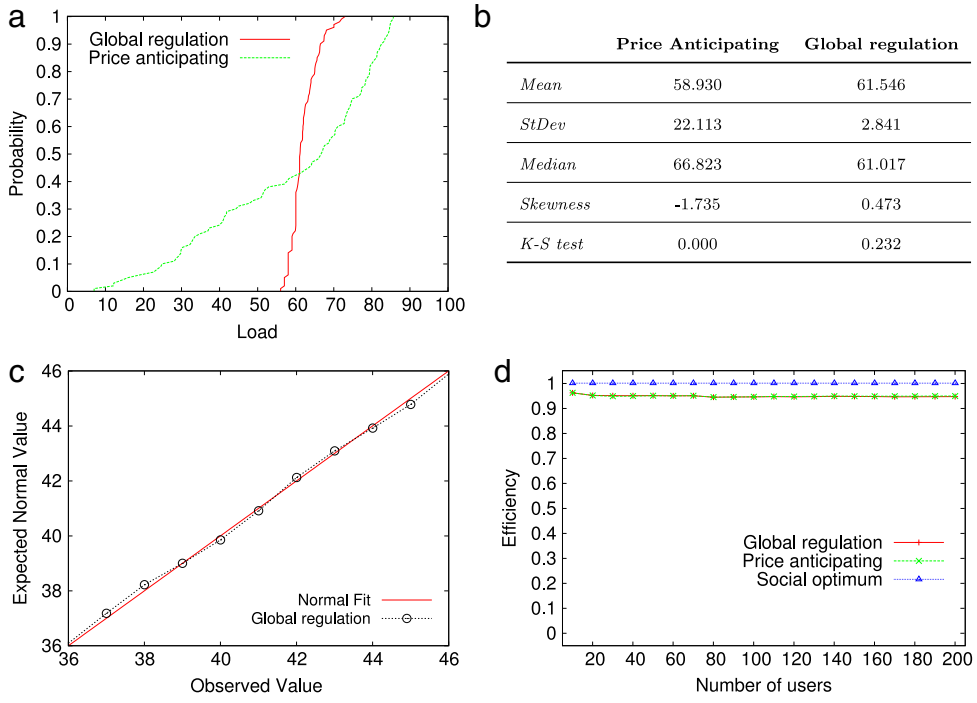
Efficiency (price of anarchy). Fig. 4(d) shows the efficiency as a function of the number of users. The efficiency achieved by the price-anticipating algorithm is very high (approximately 0.95) and the tax regulation mechanism does not lower the efficiency, irrespective of the number of users in the system; in other words, the system provides users with the same level of efficiency but the load is effectively redistributed to prevent hotspots.

Fairness (uniformity, envy-freeness). Fig. 5(a) shows the utility uniformity as a function of users for the correlated preferences presented above. Our regulatory mechanism achieves high utility uniformity (>0.8, all users obtain similar utility from the system), although with a small amount of uniformity lost in comparison with the price-anticipating simulation (approximately 15%, taking the highest difference). This is because, once our regulatory mechanism has been applied, those users bidding on the nodes with the highest preference weights obtain higher utility than those bidding on the less preferred nodes, as the load is similar for every node. However, user's perception is no longer envy-free, because agents who eventually bid on the nodes with the lowest preference weights (due to the increase in price on loaded nodes) would be *happier* with the allocation obtained by the users who bid on the more preferred nodes. Nevertheless, the envy-freeness index is still very high (>0.7) compared to the social optimum, which illustrates the trade-off between maintaining a highly efficient system, redistributing the load among nodes, and maintaining a high level of fairness.

Impact of users' preferences on utility. Finally, we compare the behaviour of our system without regulation – Fig. 6(a), (c) – and the system with regulation – Fig. 6(b), (d). Specifically, we compare the ratio of fitness[5] to revenue from the point of view of the resource

[5] The *fitness* of a resource is defined as the sum of weights of all users on that resource $\varpi(j) = \sum_{i=0}^{k} w_{ij}$, where k is the number of users and w_{ij} is the weight of user i on resource j.

Please cite this article in press as: X. León, et al., Using economic regulation to prevent resource congestion in large-scale shared infrastructures, Future Generation Computer Systems (2009), doi:10.1016/j.future.2009.11.004



	Price Anticipating	Global regulation
Mean	58.930	61.546
StDev	22.113	2.841
Median	66.823	61.017
Skewness	-1.735	0.473
K-S test	0.000	0.232

Fig. 4. (a) CDF of load distribution, (b) load distribution statistics, (c) normal Q–Q plot of the load for the global regulation case and (d) efficiency of the system.

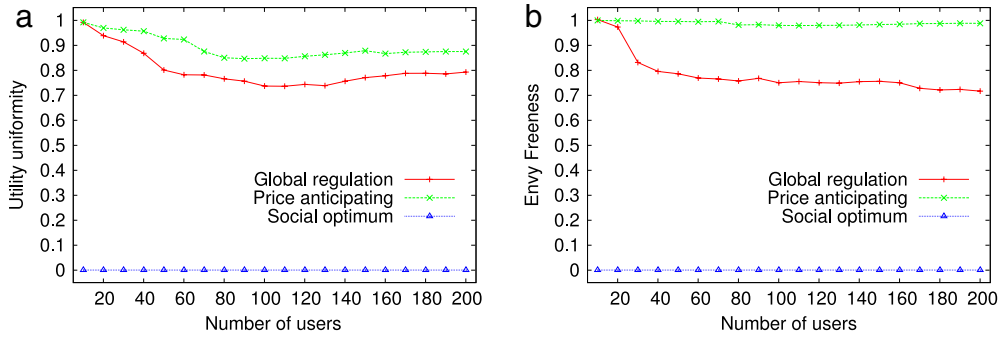


Fig. 5. (a) Utility uniformity and (b) envy-freeness.

provider –Fig. 6 (a), (b) – and the ratio of budget available to utility obtained from the point of view of the user — Fig. 6 (c), (d). [6]

Fitness and revenue show an almost linear relationship in the unregulated system. Consequently, the higher the preference weight of a resource, the more revenue it will generate, although it will also be affected by higher load (a resource obtains more virtual currency as more users bid on it). However, when regulations are enforced, the revenue generated by approximately 80% of resource providers is very similar (between 0.4 and 0.6 of normalized revenue), which means that wealth is distributed more evenly among organizations following application of regulatory taxes. This percentage represents those users "protected" by the control mechanism. However, those resources with higher fitness will still generate higher revenue because they attract higher user preference. This proves that our regulation system prevents the emergence of strong organizations (monopoly) that could dominate the resource market. As in real *welfare states*, those people with higher incomes are somehow "penalized" with higher taxes to increase the overall social welfare and the fairness of the state.

Because our model is decentralized (the revenue generated by a resource is spent by users belonging to the corresponding organization to obtain resources outside the organization), we can see that the revenue generated by resources (i.e. the budget from the user's point of view) is translated into higher utility as more budget becomes available. This is true for simulations with and without regulation; the only difference is that wealth is distributed more equitably when regulations are enforced and, therefore, there is a cluster of users with similar utilities (i.e. utilities are less dispersed when regulatory policies are applied).

7. Discussion

In this section, we discuss how our proposal could be integrated into a planetary-scale distributed system like PlanetLab to solve the problems explained in Section 2. Firstly, our regulatory mechanism (CMS) must be scalable to cope with large numbers of resources and users. Instead of a centralized solution like GridBank [27] or Tycoon Bank [17], which are susceptible to scalability problems, we have designed a prototype [28] on top of a structured peer-to-peer overlay similar to Chord [29], which enforces and manages a virtual currency account for each user in a scalable and efficient way.

[6] All values are normalized in the range [0, 1] considering maximum and minimum values obtained from the *price-anticipating* simulations.

Please cite this article in press as: X. León, et al., Using economic regulation to prevent resource congestion in large-scale shared infrastructures, Future Generation Computer Systems (2009), doi:10.1016/j.future.2009.11.004

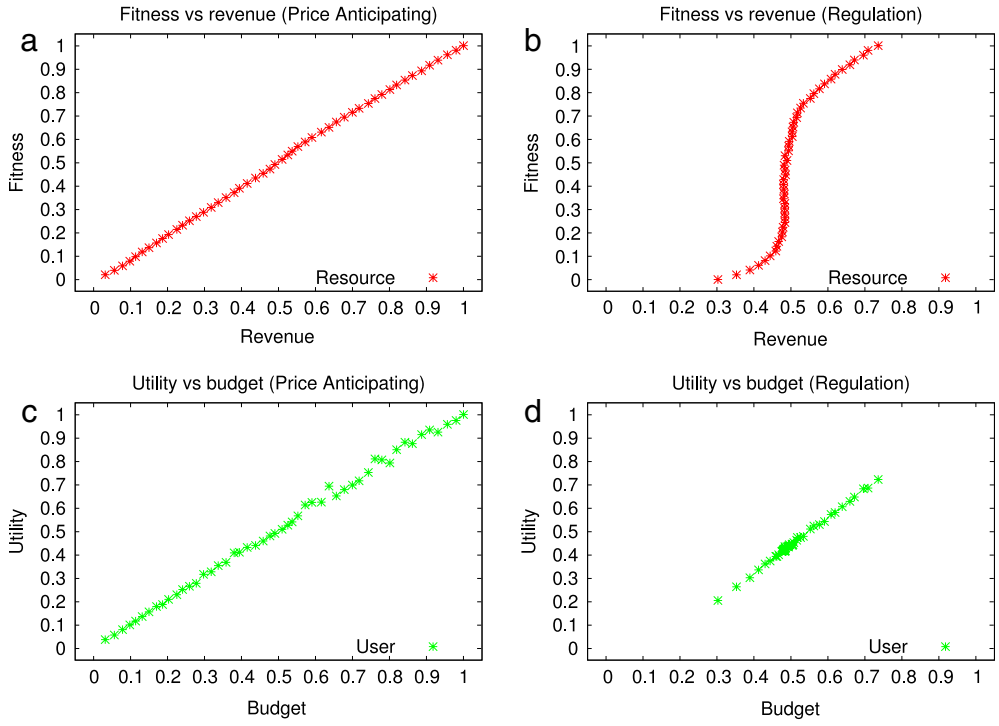


Fig. 6. System behaviour considering variable revenue and budget among participants and assuming that each resource has different fitness. Figures (a), (c) represent the behaviour of the system without regulation (price anticipating). Figures (b), (d) represent the behaviour when regulatory policies are enforced (Currency Management System, CMS).

In our proposal, the load derived from managing the users' virtual accounts is distributed among the set of nodes that make up the distributed hash table (DHT). In addition, our algorithm for calculating taxes can be executed independently by each node because it only needs information about the target load – which is not expected to change very frequently – and the load of the resources the node is responsible for managing (i.e. the resource's account is mapped to that node).

The system-wide information that the CMS uses to apply its policies – e.g. average load, effective consumption and contribution, availability, etc. – can easily be computed from the data already provided by PlanetLab's monitoring infrastructure CoMon [6], so no additional overhead is generated. Other information systems with higher scalability based on structured overlays, such as MIS [30], or discovery systems like SWORD [31], could also be used. In addition, we do not require instant information, which could be impossible to obtain in planetary-scale systems, but instead use aggregated information for relatively long time windows, in the order of hours, days or weeks depending on the system dynamics.

Finally, we discuss how our model could be integrated into the PlanetLab environment. PlanetLab's architecture and design allows the existence of resource allocation and brokering services through resource loans. Thus, a privileged slice (i.e. a service executing across several nodes with a specific share of reserved resources) might lend a subset of a node (e.g. a percentage of CPU or bandwidth) to other slices in exchange for a reward, in this case virtual money. The privileged slice would be responsible for executing and managing the proportional share mechanism explained in this paper, thereby allowing PlanetLab users to bid for a specific resource. It would also be responsible for applying the appropriate taxes to the resources price. Next, the privileged slice would transfer the actual payment from the user's account to the resource owner's account through our CMS. The specific details and exact implementation of the protocol are beyond the scope of this paper and will be covered in future work.

8. Conclusions

Large-scale shared and open infrastructures are based on resource contribution by organizations and are suitable for deploying planetary-scale public services and applications. Due to their diverse aims, these applications may have different, complex and often conflicting strategies. Without scalable and decentralized resource allocation, incentives and self-regulation mechanisms, these types of system can suffer from resource overloading and congestion, which reduce the QoS offered to users.

In this paper we have presented a mechanism in which virtual currency is used to give resource providers an incentive to work collaboratively and contribute to the efficient operation of the shared infrastructure. Currency management is also an incentive for consumers to use resources efficiently and a regulatory mechanism to ensure that they behave under certain rules designed to guarantee fairer access to the resources. Currency management limits the resource capacity to which each user is entitled and taxes resource prices according to a series of specific rules. Specifically, we present in this paper a regulatory policy for distributing work evenly between resources and preventing hotspots.

Simulations showed that our system provides a high level of efficiency in the presence of self-interested participants and decreases the load dispersion between nodes. In addition, the system offers a fair volume of resources (utility) according to the virtual currency owned by each organization, even under high-demand conditions.

Finally, we discussed how our proposal could be applied to a large-scale experimental facility such as PlanetLab at low cost and using existing tools and services. We considered the architecture and implementation of the PlanetLab testbed and explained how our solution could be successfully integrated.

In future research, we will conduct and in-depth analysis of the problems found in shared-resource infrastructures to determine how best to apply our approach and validate it as an extensible

Please cite this article in press as: X. León, et al., Using economic regulation to prevent resource congestion in large-scale shared infrastructures, Future Generation Computer Systems (2009), doi:10.1016/j.future.2009.11.004

solution to other system-wide problems. The federation of different public infrastructures remains a current popular area of research [32–34], and the ways in which a variety of resource markets from such federated infrastructures with different regulation policies might interact – e.g. through the exchange of virtual currency – is an important are for future work.

Acknowledgements

This work is being supported by the Universitat Politècnica de Catalunya (UPC) and partially supported by the Spanish Government through the P2PGRID project under contract TIN2007-68050-C03-01 and by the HSNLab, Department of Telecommunications and Media Informatics (BME).

References

[1] T. Scholl, B. Bauer, B. Gufler, R. Kuntschke, A. Reiser, A. Kemper, Scalable community-driven data sharing in e-science grids, Future Generation Computer Systems 25 (3) (2009) 290–300. doi:10.1016/j.future.2008.05.006.
[2] M. Sànchez-Artigas, P. García-López, escigrid: A p2p-based e-science grid for scalable and efficient data sharing, Future Generation Computer Systems, (2009), in press (doi:10.1016/j.future.2009.05.013). Corrected proof.
[3] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, M. Bowman, Planetlab: An overlay testbed for broad-coverage services, ACM SIGCOMM Computer Communication Review 33 (3) (2003) 3–12.
[4] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, A. Joglekar, An integrated experimental environment for distributed systems and networks, in: Proceedings of the Fifth Symposium on Operating Systems Design and Implementation, USENIX Association, Boston, MA, 2002, pp. 255–270.
[5] G. Hardin, The tragedy of the commons, Science 162 (3859) (1968) 1243–1248.
[6] K. Park, V. Pai, CoMon: A mostly-scalable monitoring system for Planetlab, ACM SIGOPS Operating Systems Review 40 (1) (2006) 65–74. URL: http://comon.cs.princeton.edu/.
[7] J. Peek, T. O'Reilly, M. Loukides, L. Mui, UNIX power tools, O'Reilly, 1997.
[8] M. Pinedo, Scheduling: Theory, algorithms and systems, Springer, 2008.
[9] L. Grit, D. Irwin, A. Yumerefendi, J. Chase, Virtual machine hosting for networked clusters: Building the foundations for autonomic orchestration, in: Proceedings of the 2nd International Workshop on Virtualization Technology in Distributed Computing, VTDC'06, IEEE Computer Society, Washington, DC, USA, 2006, p. 7. doi:10.1109/VTDC.2006.17.
[10] Y. Fu, J. Chase, B. Chun, S. Schwab, A. Vahdat, Sharp: An architecture for secure resource peering, SIGOPS Operating Systems Review 37 (5) (2003) 133–148. doi:10.1145/1165389.945459.
[11] A. AuYoung, B. Chun, A. Snoeren, A. Vahdat, Resource allocation in federated distributed computing infrastructures, in: Proceedings of the 1st Workshop on Operating System and Architectural Support for the On-demand IT Infrastructure, 2004.
[12] N. Nisan, Bidding and allocation in combinatorial auctions, in: Proceedings of the 2nd ACM conference on Electronic commerce, EC'00, ACM, New York, NY, USA, 2000, pp. 1–12. doi:10.1145/352871.352872.
[13] D. Abramson, R. Buyya, J. Giddy, A computational economy for grid computing and its implementation in the Nimrod-G resource broker, Future Generation Computer Systems 18 (8) (2002) 1061–1074. doi:10.1016/S0167-739X(02)00085-7.
[14] S.K. Garg, R. Buyya, H.J. Siegel, Time and cost trade-off management for scheduling parallel applications on utility grids, Future Generation Computer Systems, in press (doi:10.1016/j.future.2009.07.003). Corrected proof.
[15] D. Friedman, The double auction market institution: A survey, The Double Auction Market: Institutions, Theories, and Evidence, 1993, pp. 3–25.
[16] S. De Vries, R. Vohra, C. for Mathematical Studies in Economics, M. Science, Combinatorial auctions: A survey, INFORMS Journal on Computing 15 (3) (2003) 284–309.
[17] K. Lai, L. Rasmusson, E. Adar, L. Zhang, B.A. Huberman, Tycoon: An implementation of a distributed, market-based resource allocation system, Multiagent Grid Systems 1 (3) (2005) 169–182.
[18] J. Kagel, Auctions: A survey of experimental research, Internation Library of Critical Writings in Economics 113 (2000) 601–685.
[19] Hesam Izakian, Ajith Abraham, Behrouz Tork Ladani, An auction method for resource allocation in computational grids, Future Generation Computer Systems 26 (2) (2010) 228–235, doi:10.1016/j.future.2009.08.010.
[20] C. Waldspurger, Lottery and stride scheduling: Flexible proportional-share resource management, Ph.D. Thesis, Massachusetts Institute of Technology, 1995.
[21] Y.J. Nam, C. Park, Design and evaluation of an efficient proportional-share disk scheduling algorithm, Future Generation Computer Systems 22 (5) (2006) 601–610. doi:10.1016/j.future.2005.09.009.
[22] C. Li, L. Li, Competitive proportional resource allocation policy for computational grid, Future Generation Computer Systems 20 (6) (2004) 1041–1054. doi:10.1016/j.future.2003.11.029. (computational science of lattice Boltzmann modelling).
[23] M. Feldman, K. Lai, L. Zhang, A price-anticipating resource allocation mechanism for distributed shared clusters, in: Proceedings of the 6th ACM conference on Electronic commerce, ACM, NY, USA, 2005, pp. 127–136.
[24] T. Sandholm, Statistical methods for computational markets, Doctoral Thesis ISRN SU-KTH/DSV/R–08/6–SE, Royal Institute of Technology, Stockholm, 2008.
[25] C. Caux, Neural networks applied on identification of ship motions, in: Proceedings of the International Conference on Marine Simulation and Ship Manoeuvrability, MARSIM'96, Copenhagen, Denmark, Taylor & Francis, 9–13 September 1996, p. 577.
[26] P. Vytelingum, D. Cliff, N. Jennings, Strategic bidding in continuous double auctions, Artificial Intelligence 172 (14) (2008) 1700–1729.
[27] A. Barmouta, R. Buyya, Gridbank: A grid accounting services architecture (gasa) for distributed systems sharing and integration, in: Proceedings of the 17th Parallel and Distributed Processing Symposium, IPDPS'03, 2003, p. 8.
[28] X. León, L. Navarro, Currency management system: A distributed banking service for the grid, Tech. Rep. UPC-DAC-RR-XCSD-2007-6, Universitat Politècnica de Catalunya, Spain, July 2007. URL: http://gsi.ac.upc.edu/reports/2007/40/pfc-cms.pdf.
[29] I. Stoica, R. Morris, D. Karger, M. Kaashoek, H. Balakrishnan, Chord: A scalable peer-to-peer lookup service for internet applications, in: Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, ACM, NY, USA, 2001, pp. 149–160.
[30] R. Brunner, F. Freitag, L. Navarro, Towards the development of a decentralized market information system: Requirements and architecture, in: IEEE International Symposium on Parallel and Distributed Processing, IPDPS08, 2008, pp. 1–7. doi:10.1109/IPDPS.2008.4536461.
[31] D. Oppenheimer, J. Albrecht, D. Patterson, A. Vahdat, Design and implementation tradeoffs for wide-area resource discovery, in: Proceedings of the 14th IEEE International Symposium on High Performance Distributed Computing, HPDC'05, IEEE Computer Society, Los Alamitos, CA, USA, 2005, pp. 113–124. doi:10.1109/HPDC.2005.1520946.
[32] Global environment for network innovations, 2009. URL: http://www.geni.net.
[33] R. Ranjan, A. Harwood, R. Buyya, A case for cooperative and incentive-based federation of distributed clusters, Future Generation Computer Systems 24 (4) (2008) 280–295. doi:10.1016/j.future.2007.05.006.
[34] A.D. Stefano, C. Santoro, An economic model for resource management in a grid-based content distribution network, Future Generation Computer Systems 24 (3) (2008) 202–212. doi:10.1016/j.future.2007.07.014.

Xavier León is a Ph.D. student and a researcher at the Technical University of Catalonia (UPC). His research interests include computer networks, complex systems and self-organization of distributed and decentralized peer-to-peer systems through economics-based mechanisms. He received his M.Sc. in Computer Architecture, Networks and Systems from the Technical University of Catalonia. Contact him at xleon@ac.upc.edu.

Tuan Anh Trinh received his M.Sc. and Ph.D. degrees in Electrical Engineering from the University of Technology and Economics, Budapest, Hungary, in 2000 and 2005, respectively. Currently, he is the research leader of Network Economics Group (http://netecon_group.tmit.bme.hu/) at HSNLab, Budapest University of Technology and Economics, Hungary. His research interests include socio-economic issues and performance evaluation of communication networks.

Leandro Navarro is an assistant professor at the Technical University of Catalonia. His research interests include the organization and coordination of peer-to-peer grids and economics-inspired resource-allocation mechanisms. Navarro has a Ph.D. in telecom engineering from the Technical University of Catalonia. His is a member of the IEEE, the ACM, IFIP, the Association for Progressive Communications, and Computer Professionals for Social Responsibility (CPSR). Contact him at leandro@ac.upc.edu.

Please cite this article in press as: X. León, et al., Using economic regulation to prevent resource congestion in large-scale shared infrastructures, Future Generation Computer Systems (2009), doi:10.1016/j.future.2009.11.004