

Are Federated Cloud Sharing Systems Sustainable?: On Dynamic Sharing Markets and Their Stability

Ranjan Pal¹, Member, IEEE, Sung-Han Lin², Member, IEEE, Aditya Ahuja, Student Member, IEEE, Nachikethas Jagadeesan³, Member, IEEE, Abhishek Kumar⁴, Student Member, IEEE, and Leana Golubchik, Senior Member, IEEE

Abstract—The recent emergence of the small cloud (SC), both in concept and in practice, has been driven mainly by issues related to service cost and complexity of commercial cloud providers (e.g., Amazon) employing massive data centers. However, the resource inelasticity problem faced by the SCs due to their relatively scarce resources might lead to a potential degradation of customer QoS and loss of revenue. A proposed solution to this problem recommends the federated sharing of resources between competing SCs to alleviate the resource inelasticity issues that might arise. Based on this idea, a recent effort proposed *SC-Share*, a performance-driven static market model for competitive small cloud environments that results in an efficient market equilibrium jointly optimizing customer QoS satisfaction and SC revenue generation. However, an important question with a non-obvious answer still remains to be answered, without which SC sharing markets may not be guaranteed to sustain in the long-run - *is it still possible to achieve a stable market efficient state when the supply of SC resources is dynamic in nature?* In this article, we take a first step to addressing the problem of efficient market design for single SC resource sharing in dynamic environments. We answer our previous question in the affirmative through the use of Arrow and Hurwicz's disequilibrium process in economics, and the gradient play technique in game theory that allows us to iteratively converge upon efficient and stable market equilibria.

Index Terms—Small cloud, dynamic market, stability

1 INTRODUCTION

CLOUD computing is becoming increasingly popular and pervasive in the information technology (IT) marketplace due to its on-demand resource provisioning, high availability, and elasticity. These features allow cloud end users (e.g., individuals, small-scale companies, world-wide enterprises) to access resources in a pay-as-you-go manner and to meet varying demands sans upfront resource commitments [5]. Cloud service providers (Amazon AWS [6], Google Compute Engine [7], and Microsoft Azure [8]) allow customers to quickly deploy their services without a large initial infrastructure investment.

1.1 The Rise of Small-Scale Data Centers

There are some non-trivial concerns in obtaining service from large-scale public clouds, including *cost* and *complexity*. Massive cloud environments can be costly and inefficient for some customers (e.g., Blipex [9]), thus resulting in more and more customers building their own smaller data centers [10] for better control of resource usage; for example, it is hard to guarantee network performance in large-scale public clouds due to their multi-tenant environments [11]. Moreover, smaller data center providers exhibit greater flexibility in customizing services for their users, while large-scale public providers minimize their management overhead by simplifying their services; e.g., Linode [12] distinguishes itself by providing clients with easier and more flexible service customization. *The use of small-scale clouds (SCs) is one approach to solving cost and complexity issues.*

Despite the potential emergence of small-scale clouds, due to their moderate sizes, they are likely to suffer from resource under-provisioning, thus failing to meet peak demand at times. This leads to a resource provisioning dilemma where the SCs have to make the tradeoff between request loss and the cost of over-provisioning. One way out of this dilemma is for such small clouds to cooperate with each other via a federation to help meet each others' user demand via resource sharing at low costs¹, thereby increasing their individual

- R. Pal is with the University of Cambridge, CB2 1TN Cambridge, United Kingdom. E-mails: ranjanpal19@gmail.com; rp631@cam.ac.uk.
- S-H. Lin is with NetApp, Sunnyvale, CA 94089 USA. E-mail: SungHan.Lin@netapp.com.
- A. Ahuja is with the Indian Institute of Technology, New Delhi, Delhi 110016, India. E-mail: aditya.ahuja@iitd.ac.in.
- N. Jagadeesan is with Google, Mountain View, CA 94043 USA. E-mail: nachikethas@google.com.
- A. Kumar is with the University of Helsinki, 00100 Helsinki, Finland. E-mail: abhishek.kumar@cs.helsinki.fi.
- L. Golubchik is with the University of Southern California, Los Angeles, CA 90007 USA. E-mail: leana@usc.edu.

Manuscript received 27 Nov. 2018; revised 11 Oct. 2019; accepted 14 Nov. 2019. Date of publication 22 Nov. 2019; date of current version 8 Dec. 2020.

(Corresponding author: Ranjan Pal.)

Recommended for acceptance by A. Zomaya.

Digital Object Identifier no. 10.1109/TSUSC.2019.2955093

1. This approach can be used to share other virtualization based resources like containers, unikernels as well.

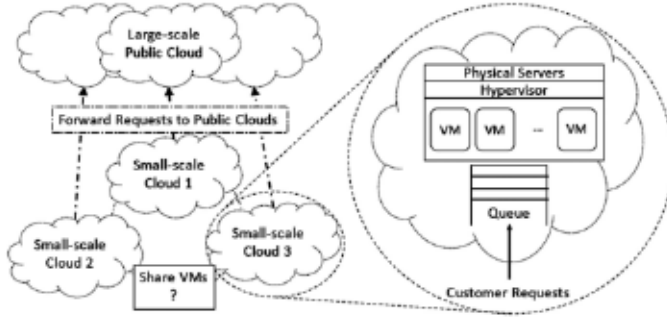


Fig. 1. SC system overview.

resources when in need without having to significantly invest in more. Such cooperation is analogous to *Business Clusters* described in mainstream economics which emerge due to, among other factors, shared interests and geographical proximity [13].

1.2 Research Motivation

Ideally, an SC would want to service all its customers solely using its own resources. However, the primary barrier to this goal is its individual resource capacity which might not be enough to service peak customer demand. In such a case, the SC can either resort to peer SCs for additional resources, thereby incurring borrowing costs, and/or buy the services of a big public cloud (e.g., Amazon). An example of this procedure with three SCs is illustrated in Fig. 1. The latter option is potentially more expensive than the former because a big public cloud might simply treat an SC as a customer and it is harder (or more expensive) to regulate where a big public cloud places their data, resulting in more privacy threatening. Thus, from an SC's viewpoint, its challenge is to satisfy two conflicting objectives: (i) to serve as many customer demands to generate more revenue, and (ii) to incur as low as possible, borrowing and/or buying resources from other clouds in order to reduce costs. Without losing generality, we assume that buying resources from big clouds (e.g., Google, Amazon) is the last resort for an SC in events of low resource availability, and in such events it would try its best to get resources from peer SCs, in order to maximize the profit. Another challenge is to ensure that at market equilibrium (see below), the SCs and their customers ideally operate on parameters (see Section 2) that allow the market to be *efficient*, a condition commonly characterized in microeconomics by certain popular functions (see Section 2.3) of market stakeholder utilities, and one that entails optimal social welfare allocation amongst the SCs and their customers. This is a non-trivial and challenging task as the existence of a market equilibrium does not necessarily imply market efficiency [14]. In addition to the above mentioned challenges, the SC market is dynamic in nature due to the non-static nature of the supply of SC resources, as well as due to the variations in customer demand over time, and failures. This dynamic nature of the SC market is likely to lead to frequent market equilibrium perturbations and potentially a state of market disequilibrium. Conditioned on the achievability of a market static efficient equilibrium, a state of eventual disequilibrium will threaten the long-term sustainability of SC markets, i.e., a deviation from efficient market equilibria. Here, the term 'market equilibrium' refers to a situation in which all market stakeholders mutually

satisfy their interests, in which case an important challenge is to design a stable market that is robust to perturbations and always returns to its equilibrium point(s) when perturbations do occur.

Our Goal. In this paper, our goal is to formulate the joint 'stakeholder satisfaction problem' in dynamic SC environments as an efficient, stable, and sustainable dynamic market/ecosystem design task, and propose an effective solution for it. Here, we consider three different stakeholders: (i) the SC customers, (ii) profit maximizing autonomous SCs, and (iii) a regulatory agency (e.g., local government or governing bodies, [15],[16]) overseeing certain functioning aspects of the autonomous SCs (e.g., ensuring customer data privacy).

1.3 Research Contributions

We make the following research contributions in this paper.

- We propose a utility theory based small cloud competitive market model comprising of federated SC customers, profit maximizing autonomous SCs, and a regulatory agency overseeing some functionality aspects of the federated SCs, as the market stakeholders. The model mathematically expresses the stakeholder interests in terms of utility functions and paves the path for analyzing SC markets for market equilibrium properties (see Section 2).
- We analyze our proposed market model via a convex optimization framework for the existence and uniqueness of a static market equilibrium at which (i) the *utilitarian social welfare function* (see Section 3 for a definition) is maximized, i.e., the market equilibrium is socially efficient, (ii) the market equilibrium is *Pareto efficient* (see Section 3 for a definition), (iii) the market is cleared, i.e., the SC supply balances customer demand, and (iv) no stakeholder has any incentive to deviate from the equilibrium. We show that there exists a unique static competitive market equilibrium jointly satisfying (i), (ii), (iii), and (iv), however there are several static market equilibria jointly satisfying (ii), (iii), and (iv). (see Section 3).
- Using the notion of a *disequilibrium process* proposed by Arrow and Hurwicz [3], [4], we apply the *gradient play* technique in game theory [17] that is based on the theory of differential equations, to investigate the dynamic market setting where a static market equilibrium (conditioned on their existence) is potentially subject to perturbations that might lead to market disequilibrium. In this regard, we show (in theory) that static market equilibria achieved in small cloud markets (see Section 3 for details on static markets) is *asymptotically stable* in dynamic market settings (see Section 4). Our use of the gradient play technique is motivated by the fact that in many practical market environments stakeholders (i) find it behaviorally difficult or computationally expensive to play their *best responses* [18], (ii) have zero or incomplete knowledge of the utilities of other stakeholders in the market, and (iii) cannot even observe the actions of other stakeholders in the worst case. In such environments, gradient play is a suitable technique to achieve static market equilibrium stability iteratively [19], from a

state of disequilibrium. More specifically, for our market setting the occurrence of (i)-(iii) is quite likely. Gradient play also works when issues (i)-(iii) do not arise (see Section 2.3).

2 COMPETITIVE MARKET MODEL

In this section, we propose a utility theory based small cloud *Walrasian* competitive market model comprising of profit maximizing autonomous SCs in a federation, their customers, and a regulatory agency overseeing some functionality aspects of the SCs. A Walrasian competitive market [14] represents a *pure exchange economy* without production, where there are a finite number of agents, i.e., SCs in our work, endowed with a finite number of commodities, i.e., computing resources in our work, that are traded with SC customers and peer SCs. *The aim behind proposing the model is to pave the path for mathematically analyzing SC markets for market equilibrium properties, and derive their practical implications.*

In this paper, we consider each SC customer to deal with three job types, where each job comprises multiple tasks: (i) Type I jobs that need to be serviced *wholly/entirely* when they arrive (e.g., a user could invoke a regular MapReduce batch job that defines a set of Mappers and Reducers to be executed for the job to complete in its entirety.), (ii) Type II jobs that can be *curtailed* to fewer tasks (e.g., an approximate computation job as in [20]), where the curtailment decision primarily arising from (a) the nature of VM instance prices, (b) the unnecessary of the job to continue executing beyond a certain accuracy already achieved, and (c) the unnecessary of the job to continue executing beyond a certain deadline, and (iii) Type III jobs where certain tasks can be *shifted* over time for future processing, the remaining job tasks requiring service as they arrive (e.g., analyzing a DNA sequence, re-running partially/entirely a current job later when it gets killed in a spot cloud environment due to momentary unavailability of resources.). Next, we model the stakeholders in the SC market.

2.1 Modeling the SCs

Let there be n autonomous profit maximizing SCs. Each SC can be geographically distributed. Customer demand for SC i is a set of processing tasks from its customers (both end-users and peer SCs) that require the use of virtual machines. We assume that each SC i reserves (allocates) a total of vm_i^r virtual machines (VMs) in its data center to service demand from its customers. We term such VMs as *reserved* VMs. The value of vm_i^r is pre-determined by SC i based on the statistics of customer demand patterns observed over a period of time. *For simplicity, we will focus on VMs representing a single resource type in this paper. The case for multi-resources will be dealt in future work.* In the event that vm_i^r machines are insufficient to satisfy consumer demand, SC i borrows vm_i^b VMs from peer SCs. Here, vm_i^b is the number of *borrowed* VMs available to SC i from its peers. In the event that both reserved and borrowed VMs are insufficient to meet customer demand, SC i resorts to a *public cloud* for vm_i^{pc} VM instances. We assume here that a public cloud is large enough to provide any required number of VM instances to SCs. We do not consider communication network bandwidth issues between private clouds to be a bottleneck to customer service satisfaction in this paper.

Let $c(vm_i^r)$ be the associated operating cost to SC i for reserving vm_i^r virtual machines to serve its customers. We define $c(vm_i^r)$ via a separable equation of the following form.

$$c(vm_i^r) = f_1(vm_i^r) + f_2(vm_i^r), \quad (1)$$

where $f_1(\cdot)$ (a linear function) and $f_2(\cdot)$ (a non-linear function) are functions such that the *marginal operating cost* for SC i is a *general decreasing linear function* of the number of VM instances, i.e., the additional operating cost, $\frac{dc}{dvm_i^r}$, due to a unit increase in the number of VMs required to service customer demand varies in a negative linear fashion with the number of VMs. Such marginal cost functions are also popular in economics to model diminishing costs/returns [14]. We approximate the number of VMs as a non-discrete quantity. Specifically, for the purpose of analysis, we assume the cost function $c(\cdot)$ to be *concave*, *quadratic*, and *twice continuously differentiable*, i.e., the marginal costs become decreasing linear functions of the number of VM instances. We can define one such $c(vm_i^r)$ function as follows.

$$c(vm_i^r) = \alpha_r^i vm_i^r + \frac{\beta_r^i}{2} (vm_i^r)^2, \quad (2)$$

where α_r^i (a positive value) and β_r^i (a negative value) are SC i 's cost coefficients for its reserved resources, i.e., virtual machines, such that the marginal operating cost for SC i is a negative linear function. The above quadratic form of the cost function, apart from satisfying the property of negative linear marginals, not only allows for tractable analysis, but also serves as a good second-order approximation for the broader class of concave payoffs [21]. We define π_i^r to be the profit that SC i makes through its reserved VMs for servicing customers, and define the maximum profit that SC i can make, via the following optimization problem.

$$\max_{vm_i^r} \pi_i^r = \max_{vm_i^r} [\rho_i vm_i^r - c(vm_i^r)]$$

subject to

$$vm_{\min_i}^r \leq vm_i^r \leq vm_{\max_i}^r,$$

where ρ_i is the per-unit VM instance price charged by SC i to its customers, and $vm_{\min_i}^r$ and $vm_{\max_i}^r$ are the lower and upper bounds for the number of VM instances reserved by SC i for its customers. We assume that each SC i is small enough not to be able to exert market power over its peer SCs and strategically influence the prices they charge their customers. i.e., each SC is a *price taker* [14]. The prices that individual SCs charge their customers are determined by individual SCs in price competition with one another in the process of maximizing their own net utilities.

Let $c(vm_i^b)$ be the associated operating cost to SC i for borrowing vm_i^b virtual machines from peer SCs to serve customers, when the reserved VMs are not enough to satisfy customer service demands. Like in the case of formulating $c(vm_i^r)$, we formulate $c(vm_i^b)$ in a manner such that the associated marginal operating costs for borrowing an additional VM instance decreases in a negative linear fashion with the number of VMs. Mathematically, we represent $c(vm_i^b)$ by the following equation:

$$c(vm_i^b) = \alpha_i^b vm_i^b + \frac{\beta_i^b}{2} (vm_i^b)^2, \quad (3)$$

where α_i^b (a positive quantity) and β_i^b (a negative quantity) are SC i 's coefficients for its borrowed virtual machines. We denote by π_i^b the profit that SC i makes when borrowing VMs from peer SCs for servicing customers, and define the maximum profit that SC i can make, via the following optimization problem:

$$\max_{vm_i^b} \pi_i^b = \max_{vm_i^b} [\rho_i vm_i^b - c(vm_i^b) - c(vm_i^{pc})]$$

subject to

$$vm_{\min_i}^b \leq vm_i^b \leq vm_{\max_i}^b.$$

Here, (i) $vm_{\min_i}^b$ and $vm_{\max_i}^b$ are the lower and upper bounds for the number of VM instances borrowed by SC i for its customers, from peer SCs, (ii) $c(vm_i^{pc})$ is the cost to SC i to offload vm_i^{pc} VM instances worth of customer demand to a public cloud in the event that vm_i^c and vm_i^b VM instances together are not enough to service i 's total customer demand. We represent $c(vm_i^{pc})$ in the same manner as $c(vm_i^c)$ and $c(vm_i^b)$, and express it via the following equation:

$$c(vm_i^{pc}) = \alpha_{pc}^i vm_i^{pc} + \frac{\beta_{pc}^i}{2} (vm_i^{pc})^2, \quad (4)$$

where α_{pc}^i (a positive quantity) and β_{pc}^i (a negative quantity) are SC i 's coefficients for the resources the public cloud uses to service i 's offloaded customer demand portions. We do not assume any constraints on the resources available to the public cloud for servicing offloading requests by SCs.

2.2 Modeling SC Customers

For a customer j who has a Type I job, we express this customer's utility for that job as a concave, quadratic, and twice continuously differentiable separable function, $U_j(\cdot)$, defined as follows.

$$U_j(vm_j^c) = \alpha_j^c vm_j^c + \frac{\beta_j^c}{2} (vm_j^c)^2, \quad (5)$$

where vm_j^c is the amount of VM instances required to process j 's entire job. Similar to the motivation and rationale behind the concave quadratic cost functions for SCs, the utility function of an SC customer is designed such that the marginal utility for the customer is a *decreasing linear function* of the number of VM instances, i.e., the additional utility increase due to a unit increase in the number of VMs varies in a negative linear fashion with the number of VMs. α_j^c (a positive quantity) and β_j^c (a negative quantity) in the above equation are j 's utility coefficients.

As in the case of a customer with a Type I job, for a customer j who has a Type II job, we express his utility for that job as a quadratic twice continuously differentiable function, $U_j(\cdot)$, defined as follows:

$$U_j(vm_j^c) = \alpha_j^c vm_j^c + \frac{\beta_j^c}{2} (vm_j^c)^2, \quad (6)$$

where vm_j^c is the amount of VM instances required to process j 's curtailed job, and is expressed as

$$vm_j^c = \kappa_j^1 vm_j^c + \kappa_j^2 vm_j^c, \quad \kappa_j^1, \kappa_j^2 \in (0, 1).$$

Here, α_j^c (a positive value) and β_j^c (a negative value) are j 's utility coefficients for Type I jobs. The interpretation of vm_j^c is as follows: $\kappa_j^1 vm_j^c$ is the number of VMs required to accomplish j 's curtailed task, whereas $\kappa_j^2 vm_j^c$ is the additional number of unused VMs that contribute to j 's extra utility when its job is curtailed, and provides it with an overall *perceived* satisfaction greater than that obtained from the utility derived solely using $\kappa_j^1 vm_j^c$ used VMs for the curtailed job.

For a customer j who has a Type III job, similar to the case of Type I and Type II jobs, we express his utility for those tasks as a quadratic twice continuously differentiable function, $U_j(\cdot)$, defined as follows:

$$U_j(vm_j^s) = \alpha_j^s vm_j^s + \frac{\beta_j^s}{2} (vm_j^s)^2, \quad (7)$$

where vm_j^s is the amount of VM instances required to process j 's time-shiftable tasks, and α_j^s (a positive value) and β_j^s (a negative value) are j 's utility coefficients for time-shiftable jobs.

A customer j can have jobs of all three types. Thus, his aggregate tasks are worth $vm_j^{ag} = vm_j^c + vm_j^c + vm_j^s$ VM instances. Therefore, customer j 's aggregate utility takes a similar form to his utility for a specific job type, and is given by

$$U_j(vm_j^{ag}) = \alpha_j^{ag} vm_j^{ag} + \frac{\beta_j^{ag}}{2} (vm_j^{ag})^2, \quad (8)$$

where α_j^{ag} (a positive quantity) and β_j^{ag} (a negative quantity) are j 's utility coefficients for his job aggregate.

We denote π_j^{type} to be the net utility that customer j generates through getting service for a given job type = $\{c, e, s\}$ from its contracted SC, and define the maximum net utility that customer j can generate, via the following optimization problem:

$$\max_{vm_j^{type}} \pi_j^{type} = \max_{vm_j^{type}} [U_j(vm_j^{type}) - \rho_j vm_j^{type}]$$

subject to

$$vm_{\min_j}^{type} \leq vm_j^{type} \leq vm_{\max_j}^{type}.$$

Here, $vm_{\min_j}^{type}$ and $vm_{\max_j}^{type}$ are the lower and upper bounds for the number of VM instances used up by customer j 's job type (be it whole, curtailed, shifted, or aggregate). ρ_j is the price paid by customer j to his chosen SC per VM instance used.

2.3 Modeling the Regulator

The role of the regulator (e.g., the government, a federated agency) as applicable to our work is to ensure (i) good privacy practices between SCs, (ii) the design of policies/mechanisms that enable autonomous SCs to price customers appropriately without making excessive profits through market exploitation, and (iii) an optimum level of social welfare allocation amongst the autonomous SCs at market equilibrium. (i) is specific to our problem setting and is one of the most important motivations for the presence of a regulator (see Section 1) in the first place.² However, the presence of a regulator brings in other important benefits through (ii) and

2. In practice, using mechanism design theory, the regulator can devise efficient economic mechanisms that enable SCs to find it incentive compatible in protecting the privacy of their customers. However, we do not focus on the design of such mechanisms in this paper.

(iii). (ii) is necessary to prevent any SC from exploiting its customers on service costs. In this work we do not focus on the design of such mechanisms, and assume the existence of one,³ whereas (iii) is important from an economic perspective as *maximizing* social welfare is a key objective in welfare economics because it leads to (a) a certain level of equitability of allocations (in resources or in net utility) amongst the stakeholders, (b) might guarantee *Pareto efficiency* at market equilibrium [14], and (c) an optimal social welfare state denotes the best possible operating point of an economic system. A Pareto efficient allocation of utilities amongst a set of stakeholders ensures that at market equilibrium none of the stakeholders can increase their net utility without decreasing any other stakeholder's net utility. The notion of equitability is important in the context of autonomous SC markets because they often operate in a decentralized fashion, and ideally, we would want a social welfare allocation at market equilibrium that does not result in considerable disparity amongst the players' allocations (despite being Pareto efficient).

In this paper, we define the social welfare function of the regulator to be the sum of the net utilities of the SCs and their customers at market equilibrium. We denote this function by SW , and express it as

$$SW = \sum_{j \in C} U_j(v m_j^{ag}) - \sum_{i \in SC} (c(v m_i^c) + c(v m_i^b) + c(v m_i^{pc})), \quad (9)$$

where C is the set of consumers, SC is the set of small clouds, the first term is the sum of the utilities of the consumers, and the second term is the sum of the costs faced by the SCs in SC for servicing customer demands. The aforementioned social welfare expression is the standard Bergson-Samuelson *utilitarian social welfare function* in economics [23], [14] whose optimality does not focus on equality of resource or utility allocations amongst stakeholders, i.e., the SCs and the customers, but only on Pareto efficiency of resource allocations amongst the stakeholders, and equality of marginal utility allocations amongst the stakeholders. Note that due to our autonomous SC setting, the regulator in practice might not have enough say in welfare maximizing resource allocation, and can only expect to have the social welfare function maximized in the best case because it cannot directly enforce optimal strategy choices on the SCs like in a centralized control setting. *The important question here is whether the utilitarian social welfare function is indeed the most appropriate choice for this work.*

We choose to work with the utilitarian function over two other popular Bergson-Samuelson social welfare functions used in economic applications: the *egalitarian function*, and the *Rawl's function*, for the following reasons:

- The parameters corresponding to the unique optimal solution of the maximum utilitarian social welfare problem *coincide* with those obtained at the unique equilibrium of a purely distributed market comprising autonomous SC's *without* the presence of a

regulator, and are Pareto optimal. This result is due to Arrow-Debreu's first and second fundamental theorems of welfare economics [14]. In addition, at market equilibrium, there is equitability in the marginal utilities of all the autonomous SCs (in case of SCs, the utility is represented by cost and is thus a negative utility) and their customers. The parameter coincidence property does not necessarily hold for non-utilitarian social welfare functions.

- The Rawl's social welfare function focusses on maximizing the minimum resource/utility allocation to any stakeholder (e.g., SC) within the class of market stakeholders. A major drawback of adopting this social welfare function is that it will in general discourage SCs from sharing their resources (even at Pareto optimal system settings) with other SCs (consequently affecting customer QoS satisfaction), thereby challenging the core philosophy behind an SC market, and will not likely be popular with either the SCs or the regulator. A maximin utility allocation among SCs would favor, for example, a regime that reduces every SC to complete "misery" if it promotes the well-being of the most "miserable" SC by even a very small amount.
- The egalitarian social welfare function focuses on equalizing the utilities of all market stakeholders in the absolute sense. Similar to the case of Rawl's function, it suffers from the major drawback that it will in general discourage SCs from sharing their resources (even at Pareto optimal system settings) with other SCs. Likewise, it is unlikely to be popular amongst either the regulator or autonomous SCs. For example, if we had to choose between two allocation policies, one under which all SCs would have a cardinal utility of 100, but one SC would have a utility of 99; the second policy under which every SC is "miserable" and will have a cardinal utility of 1 unit. The egalitarian regulator would prefer the latter because under this option, every SC has exactly the same utility level.

3 STATIC MARKET ANALYSIS

As a precursor to analyzing dynamic markets, in this section we first derive and analyze perfectly competitive static SC market equilibria. We assume perfect competition amongst SCs due to their lack of economic power in influencing other SCs based on their quantity of VM availability. Since prices in perfect competition are *strategic complements* (in the terminology of Bulow, Geanakoplos and Klemperer [24]), i.e., the decrease in an SC's customer price results in the decrease of customer prices charged by other SCs in competition, we are going to eventually converge to a stage where a single uniform customer price will prevail in the SC market [24]. We are interested to know whether such a price results in social welfare optimality. Equivalently, if a federated agency were to centrally impose a customer charging price on all SCs (thereby breaking their autonomy) that would maximize social welfare, what would be the relationship between such a price (quantity) and the market equilibrium price (quantity) outcome of the price-quantity competition game? In this regard, we (a) formulate and solve an optimization problem for a regulator who wishes to achieve socially optimal

3. Economists Laffont and Tirole have proposed *principal-agent* models in this regard [22] which will enable autonomous SCs to charge appropriate prices to customers purely out of self-interest.

market equilibria that maximizes utilitarian social welfare amongst the market stakeholders, (b) characterize market equilibria in the absence of a regulator and draw comparative relationships between the equilibria obtained, with socially optimal market equilibria. In practice, the competition between SC firms is likely to be imperfect in nature, and Laffont and Tirole have addressed models [25] under such settings which result in market efficiency.

Optimization Problem Formulation. Here, we formulate a regulator's optimization problem so as to achieve socially optimal market equilibria. The primary goal of the formulation is to maximize the net utilities for the SC customers, and minimize the net cost of operation of SCs to reach a net maximum social welfare situation amongst the SCs and their customers. We define this problem mathematically as follows:

$$\text{OPT : } \max SW$$

subject to

$$\sum_{j \in C_i} vm_j^{ag} - (vm_i^r + vm_i^b + vm_i^{pc}) = 0, \forall i \in SC,$$

where the objective function is to maximize social welfare SW (see Equation 9 in main paper) and the constraint is the supply-demand balance equation, with $\sum_{j \in C_i} vm_j^{ag}$ representing total customer demand, and $(vm_i^r + vm_i^b + vm_i^{pc})$ representing total SC i supply. C_i is the set of customers served by SC i . A potential solution to the above optimization problem indicates the parameters at which the SC market can ideally operate and (i) make all stakeholders satisfied to a point that no one has an incentive to deviate, and (ii) maximize the total satisfaction of all the stakeholders together. We denote such an ideal state of market operation as a *static socially efficient market equilibrium*.

Dual Problem Formulation. We will solve OPT using the *primal-dual* approach [26]. The advantage of using the primal-dual approach is that the dual optimization problem of the primal is always convex [26], and its solution results in global optima which can be related back to the optimal solution of the primal problem. Before deriving the dual optimization problem, we first define the Lagrangian function of OPT as follows:

$$L = \sum_{i \in SC} (c(vm_i^r) + c(vm_i^b) + c(vm_i^{pc})) - \sum_{j \in C} U_j(vm_j^{ag}) + \sum_{i \in SC} \rho_i \left(\sum_{j \in C_i} vm_j^{ag} - \rho_i (vm_i^r + vm_i^b + vm_i^{pc}) \right),$$

where $\rho = (\rho_1, \dots, \rho_n)$ is the vector of Lagrange multipliers for the constraint in OPT. The dual optimization problem, DOPT, is then defined as follows.

$$\text{DOPT : } \max_{t := \{vm^e, vm^c, vm^s, vm^r, vm^b, vm^{pc}, \rho\}} \inf L,$$

where vm^e, vm^c , and vm^s are vectors of customer VM types and vm^r, vm^b , and vm^{pc} are vectors of SC VM types. Note that vm_i^{ag} for any customer i equals $vm_i^e + vm_i^c + vm_i^s$. Thus, the goal here is to find an optimal tuple t , that is an optimal solution to both OPT and its dual.

Solving the Dual. The dual optimization problem is convex and its optimal solution is found by applying the *Karush-Kuhn-Tucker* (KKT) conditions [26] that are stated through Equations ((10a), (10b), (10c), (10d), (10e), (10f), and (10g)). Solving these equations, we obtain the optimal solution to DOPT. The optimal solution to DOPT is the static market equilibrium. We denote this solution by the tuple $\{vm^{e*}, vm^{c*}, vm^{s*}, vm^{r*}, vm^{b*}, vm^{pc*}, \rho^*\}$. We now state the KKT conditions in the form of Equations ((10a), (10b), (10c), (10d), (10e), and (10f)) as follows.

$$\frac{d(c(vm_i^r))}{dvm_i^r} | vm_i^{r*} - \rho_i^* = 0, \forall i \in SC. \quad (10a)$$

$$\frac{d(c(vm_i^b))}{dvm_i^b} | vm_i^{b*} - \rho_i^* = 0, \forall i \in SC. \quad (10b)$$

$$\frac{d(c(vm_i^{pc}))}{dvm_i^{pc}} | vm_i^{pc*} - \rho_i^* = 0, \forall i \in SC. \quad (10c)$$

$$\rho_i^* - \frac{\partial(U_i(vm_i^e))}{\partial vm_i^e} | vm_i^{e*} = 0, \forall i \in C. \quad (10d)$$

$$\rho_i^* - \frac{\partial(U_i(vm_i^c))}{\partial vm_i^c} | vm_i^{c*} = 0, \forall i \in C. \quad (10e)$$

$$\rho_i^* - \frac{\partial(U_i(vm_i^s))}{\partial vm_i^s} | vm_i^{s*} = 0, \forall i \in C. \quad (10f)$$

$$\sum_{j \in C_i} vm_j^{ag} (1 - \kappa_j^1 - \kappa_j^2) = (vm_i^r + vm_i^b + vm_i^{pc}), \forall i \in SC. \quad (10g)$$

Equilibrium in Autonomous Settings. The key question is whether the solution to DOPT can be realized as a market equilibria in a distributed autonomous setting. Based on the general equilibrium theory in microeconomics [14], market equilibria in a perfectly competitive autonomous setting of firms is known as *Walrasian equilibria*. It turns out from general equilibrium results in [14] that the unique optimal solution to DOPT (i) is a competitive Walrasian equilibrium that is Pareto efficient, (ii) satisfies *Arrow-Debreu's* first and second fundamental theorems of welfare economics that establishes the *if and only if* relation between the existence of a Walrasian equilibrium and its Pareto efficiency [14], (iii) maximizes utilitarian social welfare (again derived from *Arrow-Debreu's* first and second fundamental theorems), and (iv) clears the market by balancing total SC resource supply with consumer and SC resource demand. Thus, in view of points (i) - (iv), a regulator's social welfare maximization objective coincides with the welfare state obtained at market equilibrium in a distributed autonomous firm setting. We consider this unique equilibrium state to be the benchmark at which the SC market would be willing to always operate. However, in practice, for a perfectly competitive market with non-utilitarian social welfare functions, there may be multiple Pareto efficient Walrasian market equilibria that are not socially efficient.

Computing Socially Optimal Equilibrium. The optimal solution to the dual optimization problem, DOPT, can be obtained in an iterative manner using a gradient approach, the principle behind which is the *Primal-Dual Interior Point Method* [26]. We adopt the *Primal-Dual Interior Point method* in our work because it has a polynomial-time complexity to

arrive at the optimal solution to convex programs [27]. The basis of the method is to *progressively* change the argument vector of DOPT so that minimal-Lagrange multiplier ρ satisfies the KKT conditions.

Denote by v , DOPT's argument vector sans the Lagrange multiplier ρ , $\{vm^e, vm^c, vm^s, vm^r, vm^b, vm^{pc}\}$. Applying the Interior Point method to DOPT gives us the the following equations:

$$v(t + \epsilon) = v(t) - k_v \nabla_x L \cdot \epsilon. \quad (11a)$$

$$\rho(t + \epsilon) = \rho(t) + k_\rho \nabla_x L \cdot \epsilon. \quad (11b)$$

Here, k_v and k_ρ are positive scaling parameters which control the amount of change in the direction of the gradient. Letting $\epsilon \rightarrow 0$, we get

$$\tau_v \dot{v}(t) = -\nabla_v L, \quad (12a)$$

$$\tau_\rho \dot{\rho}(t) = -\nabla_\rho L, \quad (12b)$$

where $\tau_y = \frac{1}{k_y}$ for $y = v, \rho$. The Interior Point Method converges in polynomial time when the duality gap approaches zero, due to the linear and super-linear convergence rate of the method [26].

4 DYNAMIC SC MARKETS

In this section we describe the notion of dynamic SC markets, provide a formalized model to describe such markets, and finally investigate their stability aspects.

The Notion of Dynamic SC Markets. In practice, an SC market can be dynamic in nature due to the non-static nature of the supply of SC resources and variability over time of customer demand. This dynamic nature of the SC market is likely to lead to frequent static market equilibrium (see Section 3 for the analysis of static market equilibria) perturbations, which in turn might (not always) lead to a state of market disequilibrium. Here, the term 'disequilibrium' refers to a state when market supply does not equal market demand due to perturbations in market parameters (e.g., customer prices), and as a result all stakeholders do not mutually satisfy their interests.

To provide a clear understanding of the stark distinction between static and dynamic markets, in the former, once a unique market equilibrium state (a vector characterized by optimal customer price charged, optimal VM resources allocated between the stakeholders, etc.) is reached (call it A) there is no deviation from that system state in inertia, i.e., a stable point, unless the system variables change - but they do not as the system is assumed to be static. Note that A is in the set of feasible points the system could operate in. Now considering the contrary situation of dynamic markets, system variables change over time due to the intermittent nature of resources driving them (e.g., variability in the equilibrium VM resources allocated due to policy constraints of meeting client Qos as utmost priority). As a result, the market system operating at stable operating point A will need to exit A once resource variations are inevitable. Assuming A to be a benchmarked point (as social welfare is maximized in this unique state - see Section 3) The question here is whether the market system can ideally return to A eventually. Note that if A is non-unique, the system will be equally willing to be in a welfare maximizing market equilibrium state B .

In such a case, an important challenge is to design a stable market that is robust to perturbations and always returns to its equilibrium point(s) when market disequilibrium results. Inspired by the notion of disequilibrium process [3], we propose a dynamic market mechanism for SCs. The concept of disequilibrium pertains to a situation where a static market equilibrium is perturbed, potentially to a disequilibrium state, and the underlying players (stakeholders) work together to re-attain the equilibrium. The main idea behind the disequilibrium process is an iterative sequence of action and state profiles (see below), i.e., information exchange between the dominant market stakeholders, of VM instance supply and demand levels, and per-unit VM instance prices, to arrive at a desired static equilibrium. Such an iterative process essentially implies an overall dynamic model with feedback. Our proposed dynamic market mechanism can also be used to re-attain a specific preferred equilibrium point from a given equilibrium point. We first present our dynamic market model and then follow it up with its stability analysis.

4.1 Dynamic Model

Our dynamic model of SC markets consist of a *state space*, $X \subset \mathbb{R}^n$, where each state, $\{\rho_i\} \in X$, is the profile of per-unit VM instance prices at each SC i . The *state dependent payoff*, i.e., profit function for each SC from its reserved resources is given by

$$\pi_i^r = \rho_i vm_i^r - c(vm_i^r).$$

The state dependent payoff for each SC from its borrowed resources is given by

$$\pi_i^b = \rho_i vm_i^b - c(vm_i^b).$$

Similarly, state dependent payoff for each SC from resources borrowed from a public cloud is given by

$$\pi_i^{pc} = \rho_i vm_i^{pc} - c(vm_i^{pc}).$$

The payoff function for the SC customers for a given job type $\in \{e, c, s\}$, is given by

$$U_j(vm_j^{type}) - \rho_j vm_j^{type}.$$

Each SC is assigned a *state dependent action* that permits the SCs and their customers to change their VM instance generation and consumption levels respectively. We assume a *perfect competition* [18] of VM instance prices amongst the SCs in competition, and following that the action for each SC i consists of committing a certain amount of VM instances that influences the market-clearing process. In this paper, we use the *gradient play* technique in game theory [17] to derive the state dependent actions of the SCs and their customers. Our use of the gradient play technique is motivated by the fact that in many practical market environments stakeholders (i) find it behaviorally difficult or computationally expensive to play their *best responses* [18], (ii) have zero or incomplete knowledge of the utilities of other stakeholders in the market, and (iii) cannot even observe the actions of other stakeholders in the worst case. In such environments, gradient play is a suitable technique to achieve static market equilibrium stability iteratively [19]. More specifically, for our market setting the occurrence of (i)-(iii) is quite likely. Gradient

play also works when issues (i)-(iii) do not arise. *The main idea behind the gradient play technique is the use of ordinary differential equations (ODEs) to describe the path of a perturbed system state to the static market equilibrium state.* Using gradient play, the action for the the i th SC is given by

$$\tau_i^r \dot{vm}_i^r = \rho_i - \beta_i^r vm_i^r - \alpha_i^r. \quad (13a)$$

$$\tau_i^b \dot{vm}_i^b = \rho_i - \beta_i^b vm_i^b - \alpha_i^b. \quad (13b)$$

$$\tau_i^{pc} \dot{vm}_i^{pc} = \rho_i - \beta_i^{pc} vm_i^{pc} - \alpha_i^{pc}. \quad (13c)$$

Here, the parameters τ_i^r , τ_i^b , and τ_i^{pc} are time constants that describe the speed with which the action of VM instance commitment by SC i can be adjusted, and are free parameters to be determined. The goal of SC i 's action is to drive the solution vm_i^r , vm_i^b , and vm_i^{pc} to vm_i^{r*} , vm_i^{b*} , and vm_i^{pc*} , the solution to Equations (10a), (10b), and (10c) (see Section 3) at static market equilibrium. It can be seen that the RHSs of (10a), (10b), and (10c) (see Section 3) are proportional to the gradient $\nabla_{vm_i^r} L$, $\nabla_{vm_i^b} L$, and $\nabla_{vm_i^{pc}} L$ respectively, where L is the Lagrangian of OPT. The suite of Equations (10a), (10b), and (10c) (see Section 3) can be solved independently by SC i . In a similar fashion, using gradient play, the state dependent action for any SC customer $i \in C$ is given by

$$\tau_i^{ag} \dot{vm}_i^{ag} = \beta_i^{ag} vm_i^{ag} + \alpha_i^{ag} - \rho_i. \quad (14)$$

τ_i^{ag} is a free parameter to be determined that denotes the speed with which the consumption action of SC customer i can be adjusted. The goal of the SC customer action here is to drive the solution vm_i^{ag} to vm_i^{ag*} , the solution to Equation (10d) (see Section 3) at static market equilibrium. It can be seen that the RHS of 15 is proportional to the gradient $\nabla_{vm_i^{ag}} L$, $i \in C$, where L is the Lagrangian of OPT. Equation (15) can be solved independently by each SC customer i .

The dynamics of the pricing mechanism can be expressed via the following equation.

$$\tau_{\rho_i} \dot{\rho}_i = \sum_{j \in C_i} vm_j^{ag} (1 - \kappa_j^1 - \kappa_j^2) - (vm_i^r + vm_i^b + vm_i^{pc}), \quad (15)$$

where the goal is to drive the solution ρ_i , $\forall i \in SC$ to ρ_i^* , the solution of Equation 10e (see Section 3) at static market equilibrium. Here, τ_{ρ_i} is the free parameter denoting the speed with which ρ_i can be adjusted. Equations (13), (14), and (15) represent a dynamic model of the overall SC market. It resembles a repeated negotiation process where SC i responds with a commitment of vm_i^x , $x \in \{r, b, pc\}$ to suggested prices ρ_i received from the regulator; SC customer i responds with a consumption amount of vm_j^{ag} , $j \in C_i$, to the same prices. The regulator in turn adjusts its prices to these actions by the SCs and their customers, and returns new prices, $\{\rho_i\}$, and the process continues till convergence to the static market equilibrium. A compact representation of the above-mentioned dynamic SC market is presented in Section 2. This representation paves the way for analytically analyzing the stability of such markets.

4.1.1 A Compact Representation

We need to compactly represent the above dynamic SC market model to pave the way for analyzing the stability of such markets via the Arrow-Hurwicz criterion that is based on the

theory of Lyapunov stability (see Section 4.2). Using Equations (13), (14), and (15), our proposed dynamic market mechanism can be compactly represented in the matrix form via the following equation:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} A_1 + \Delta A_1 & A_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} \bar{a} \\ f_2(x_1, x_2) \end{bmatrix} \quad (16)$$

Definition of Equation Parameters. We now describe the parameters of Equation 16. We have

$$x_1(t) = [VM_{SC}^r \ VM_{SC}^b \ VM_{SC}^{pc} \ VM_C^e \ VM_C^c \ VM_C^s \ \Delta \rho]^T,$$

that is a vector of dimension $(|SC| + |C| + 2|SC| - 1) \times 1$. Here, $|SC| = n$. We also have

$$x_2(t) = [0]_{n-1 \times 1},$$

and

$$A_1 = \begin{bmatrix} -M_1 & 0 & 0 & M_2 \\ 0 & M_3 & 0 & -M_4 \\ 0 & 0 & 0 & -M_5 \\ -M_6 & M_7 & M_8 & 0 \end{bmatrix},$$

$$A_2 = [0 \ 0 \ -M_9 \ 0].$$

We define matrices M_1 to M_9 as follows: $M_1 = \text{Diag}(\frac{1}{\tau_i^{rpe}} \beta_i^{rpe})$, $\text{type} \in \{r, b, pc\}$. We assume that all for a given type, τ_i^{type} s are equal for all $i \in SC$. $M_2 = \text{Diag}(\frac{1}{\tau_i^{rpe}} A_{SC}^T)$, $\text{type} \in \{r, b, pc\}$, where $A_{SC} = \text{Diag}(1)$. $M_3 = \text{Diag}(\frac{1}{\tau_i^{cpe}} \beta_i^{cpe})$, $\text{type} \in \{e, c, s\}$. $M_4 = \text{Diag}(\frac{1}{\tau_i^{cpe}} A_C^T)$, $\text{type} \in \{e, c, s\}$, where $A_C = \text{Diag}(1)$. $M_5 = \text{Diag}(A'^T B A)$, where A' is an $(n) \times (n-1)$ matrix of 1's except for the 0 diagonal elements, B is an $n \times n$ matrix with all entries 1 except for entries of the form B_{ii} that take a value of zero, and A is an $n \times n-1$ matrix. $M_6 = \text{Diag}(\frac{1}{\tau_i^{rpe}} A_{SC})$, $\text{type} \in \{r, b, pc\}$. $M_7 = \text{Diag}(\frac{1}{\tau_i^{cpe}} A_C)$, $\text{type} \in \{e, c, s\}$. $M_8 = \text{Diag}(\frac{1}{\tau_i^{cpe}} A'^T B A')$, where A is an $(n-1) \times n$ matrix. $M_9 = [1]_{n \times n}$.

The expression $f_2(x_1, x_2)$ is a projection function onto the non-negative orthant, and is given by

$$f_2(x_1, x_2) = [cx_1 - VM^{\max}]_{x_2}^+, \quad (17)$$

where $c = BA'R$, R being a rotating matrix, of dimensionality $((|SC| - 1) \times |SC| + |C| + 2|SC| - 1) \times 1$, and $VM^{\max}$ denotes a vector of maximum VM instances committed by each individual SC. The n th row of the projection $[cx_1 - VM^{\max}]_{x_2}^+$ is denoted as

$$[cx_1 - VM^{\max}]_{x_2}^+ = \begin{cases} \max(0, [cx_1]_n - VM_n^{\max}), & \text{if } [x_2]_n = 0 \\ [cx_1]_n - VM_n^{\max}, & \text{if } [x_2]_n > 0 \end{cases} \quad (18)$$

ΔA_1 in Equation (16) represents the resource availability perturbations due to dynamics of the SC market. The value lies in a perturbation set E , where E is given by

$$E = \{\Delta A = \Delta_{SC} - \Delta_C | \Delta_{SC} \in E_{SC}; \Delta_C \in E_C\}. \quad (19)$$

Here,

$$\Delta_{SC} = \begin{bmatrix} M_{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ M_{11} & 0 & 0 & 0 \end{bmatrix},$$

where matrix M_{10} is given by $\text{Diag}(\frac{1}{\tau_i^{type}} \beta_i^{type} (\Delta_{SC})^2)$, $type \in \{r, b, pc\}$, and $\Delta_{SC} = \text{Diag}(\Delta_{SC}^{type})$. Matrix M_{11} is given by $\text{Diag}(\frac{1}{\tau_i^{type}} A_{SC}^T (I - \Delta_{SC}^{type}))$, and $A_{SC} = \text{Diag}(1)$. We also have E_{SC} expressed via the following:

$$E_{SC} = \left\{ \Delta_{SC} \mid \|\Delta_{SC}\| = \sqrt{\lambda_{\max}(\Delta_{SC}^T \Delta_{SC})} \leq \pi_{SC} \right\},$$

where π_{SC} is a finite constant. Similar to the expression for Δ_{SC} , we have

$$\Delta_C = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & M_{12} & 0 & 0 \end{bmatrix},$$

where the matrix M_{12} is given by $\text{Diag}(\frac{1}{\tau_i^{type}} A_C^T (I - \kappa_j^1 - \kappa_j^2))$. We also have

$$E_C = \left\{ \Delta_C \mid \|\Delta_C\| = \sqrt{\lambda_{\max}(\Delta_C^T \Delta_C)} \leq \pi_C \right\},$$

where π_C is a finite constant. Finally, we express $\bar{b}$ as

$$\bar{b} = \left[\text{Diag}\left(\frac{1}{\tau_i^x} \alpha_i^{type}\right) + \text{Diag}\left(\frac{1}{\tau_i^x} \alpha_i^x\right) \Delta_{SC}^{type} \text{Diag}\left(\frac{1}{\tau_i^y} \alpha_i^{type}\right) 0 \right]^T,$$

where $x \in \{r, b, pc\}$, and $y \in \{e, c, s\}$. We assume that for given x, y , the values of α_i^x and α_i^y are equal for all i .

4.2 Stability Analysis of Dynamic Markets

In this section, we derive results regarding the stability of static market equilibria in a dynamic SC market setting. Specifically, (i) we derive the dynamic market equilibria obtained through gradient play mechanics and compare it with the socially efficient static market equilibria, and (ii) study the region of attraction around dynamic market equilibria to derive stability connotations. The outcome of our study provides insights into operating zones for the federated cloud sharing markets sharing a single resource type, that will result in market stability in the presence of resource variations. We also emphasize here that our notion of stability is with respect to operating at the static equilibrium, as compared to the notion of forming stable coalitions in a federated cloud as very recently studied in [28].

Case - 1: We first consider stability aspects when κ_j^1, κ_j^2 equals zero, i.e., there are no curtailed jobs. In this case, the equilibria of the dynamic SC market described through Equations (10a), (10b), and (10c) (see Section 3) (via the use of the gradient play technique), lies in the set

$$E = \{(x_1, x_2) \mid A_1 x_1 + A_2 x_2 + \bar{\alpha} = 0 \cap f_2(x_1, x_2) = 0\}.$$

Let (x_1^*, x_2^*) be an equilibrium point in set E . We then have the following theorem stating the relationship between (x_1^*, x_2^*)

and the unique static SC market equilibrium obtained through Equations (10a), (10b), (10c), (10d), and (10e) (see Section 3).

Theorem 1. *The equilibrium (x_1^*, x_2^*) is identical to the unique static market equilibrium obtained from the solution of OPT.*

Proof. The equilibrium (x_1^*, x_2^*) when setting κ_j^1, κ_j^2 to zero, is a solution of the following.

$$\rho_i^* - \beta_i^r v m_i^{r*} - \alpha_i^r = 0, \forall i \in SDC. \quad (20a)$$

$$\rho_i^* - \beta_i^b v m_i^{b*} - \alpha_i^b = 0, \forall i \in SDC. \quad (20b)$$

$$\rho_i^* - \beta_i^{pc} v m_i^{pc*} - \alpha_i^{pc} = 0, \forall i \in SDC. \quad (20c)$$

$$\beta_i^{type} v m_i^{type} + \alpha_i^{type} - \rho_i^* = 0, \forall i \in C, type \in \{e, c, s, ag\}. \quad (20d)$$

$$\sum_{j \in C_i} v m_j^{ag} (1 - \kappa_j^1 - \kappa_j^2) = (v m_i^r + v m_i^b + v m_i^{pc}), \forall i \in SDC. \quad (20e)$$

Using Theorem 3 in [29], strong duality implies that equilibrium (x_1^*, x_2^*) exists is identical to the solution of the KKT conditions in Equations (10a), (10b), (10c), (10d), and (10e). It can be seen that Equation (20a) follows by replacing the cost function for SDCs in Equations (2), (3), and (4) (see main paper) in (10a). Similarly, Equation (20b) follows by replacing the utility function of SDC customers in Equations (5), (6), (7), and (8) (see main paper) in Equation (10d). Furthermore Equation (20c) is identical to Equation (10e). Thus, (x_1^*, x_2^*) is identical to the equilibrium in Equations (10a), (10b), (10c), (10d), and (10e). Hence, we proved Theorem 1.

Theorem Implications. The theorem suggests that in the absence of curtailed jobs, the equilibrium in a dynamic market setting is unique, and converges to the static market equilibrium in which the market existed initially before it was perturbed. Intuitively, when the SC market is perturbed from its equilibrium setting, a disequilibrium state might result, which will get resolved due to our proposed gradient-play based approach that rolls back the disequilibrium state to the original socially optimal static equilibrium state. In this paper, we are able to roll back to the original state in theory because of our assumptions regarding the nature of utility functions. *In practice, gradient play will guarantee a roll back of a disequilibrium market state to an equilibrium state not necessarily the original equilibrium state from which it was perturbed.*

We now investigate the stability of the dynamic market equilibrium to find the region of attraction around itself. We introduce a few definitions in this regard. Let $y_1 = x_1 - x_1^*$, $y_2 = x_2 - x_2^*$. Denote by $V(y_1, y_2)$ a scalar, positive definite Lyapunov function expressed as

$$V(y_1, y_2) = y_1^T P_1 y_1 + y_2^T P_2 y_2, \quad (21)$$

where P_1 and P_2 are diagonal matrices. We use *Lyapunov functions* from control theory [30] as a standard to prove the stability of an equilibrium of a system represented via ordinary differential equations (ODEs). Let d be expressed as

$$d = \frac{2\lambda_{\min}(P_2)\psi_{\min}\lambda_{\min}(Q)}{\beta^2}, \quad (22)$$

where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of Q ,

$$\beta \geq \|P_1 A_2 + R^T[1]_{n \times n} P_2\|_2,$$

where R is a rotating matrix, and $\psi_{\min} = \min(\psi_i)$, ψ_i being the coefficient of the orthogonal vector w_i to express $VM^{\max}$ as $\sum_{i=1}^n \psi_i w_i$. We now have the following theorem characterizing stability of the dynamic market equilibrium. The proof of the theorem is in Section 3.

Theorem 2. Let A_1 be Hurwitz. Then the equilibrium (x_1^*, x_2^*) is asymptotically stable for all initial conditions in

$$\Omega_{c_{\max}} = \{(y_1, y_2) | V(y_1, y_2) \leq c_{\max}\} \text{ for } c_{\max} > 0,$$

such that

$$\Omega_{c_{\max}} \subsetneq D = \{y_2 \geq 0 | \|y_2\|_2 \leq d\}.$$

Proof. Since strong duality holds, it follows from Theorem 1 that equilibrium $(x_1^*, x_2^* \in E)$ exists. We first prove the stability of this equilibrium point and then proceed to its asymptotic stability. Differentiating the positive definite Lyapunov function $V(y_1, y_2) = y_1^T P_1 y_1 + y_2^T P_2 y_2$, with respect to time where $y_1 = x_1 - x_1^*$ and $y_2 = x_2 - x_2^*$, and by using the non-expansive property of the projection operation, we have

$$\dot{V}(y_1, y_2) \leq y_1^T (P_1 A_1 + A_1^T P_1) y_1 + y_1^T P_1 A_2 y_2 + y_2^T A_2^T P_1 y_1. \quad (23)$$

If A_1 is Hurwitz, for any $Q > 0$, there exists a positive definite matrix P_1 such that $P_1 A_1 + A_1^T P_1 = -Q$. Let $\lambda_{\min}(Q)$ denote the minimum eigenvalue of Q . Since P_2 is a symmetric positive definite matrix with a set n orthogonal, real, and non-zero eigenvectors $x_1, \dots, x_n$, can be written as

$$P_2 = \sum_{i=1}^n \lambda_i x_i x_i^T,$$

where $\lambda_i > 0$ is the eigenvalue corresponding to x_i . We can expand the vector $VM^{\max}$ using the orthogonal vector w_i as

$$VM^{\max T} [1]_{n \times n} P_2 y_2 \geq \lambda_{\min}(P_2) \psi_{\min} \|y_2\|_2, \quad (24)$$

where $\psi_{\min} = \min(\psi_i), \forall i = 1, \dots, n$. Now let

$$\beta \geq \|P_1 A_2 + R^T[1]_{n \times n} P_2\|_2.$$

Using Equation (23) and (24), we obtain

$$\begin{aligned} \dot{V}(y_1, y_2) &\leq -\lambda(Q) \left(\|y_1\|_2 - \frac{\beta}{\lambda_{\min}(Q)} \|y_2\|_2 \right)^2 \\ &\quad - \|y_2\|_2 \left(2\lambda_{\min}(P_2) \psi_{\min} - \frac{\beta^2}{\lambda_{\min}(Q)} \|y_2\|_2 \right). \end{aligned}$$

For all $\Omega_{c_{\max}} \subsetneq D$, it follows that for all solutions beginning in $\Omega_{c_{\max}}$, $V \leq 0$. Hence, the equilibrium is stable and $\Omega_{c_{\max}}$ is the region of attraction.

Since the initial conditions start in Ω_Δ and the latter is a strict subset of D_Δ , y_2 cannot be equal to $2\lambda_{\min}(P_2) \psi_{\min} \frac{\lambda_{\min}(Q)}{\beta^2}$. This in turn implies that $(\|y_1\|, \|y_2\|) = (0, 0)$ is the only invariant set. Hence, all solutions starting in Ω_Δ converge

to the equilibrium point $(x_1, x_2) = (x_1^*, x_2^*)$. Thus, we proved Theorem 2. $\square$

Theorem Implications. Intuitively, the theorem states that irrespective of any initial state the market is in, on being perturbed, it will always come back to an equilibrium state from a disequilibrium state. The Hurwitz (not the same as Hurwicz) nature of matrix A_1 is determined from the time constants in Equations (13), (14), and (15). Most real systems satisfy the Hurwitz criterion in that A_1 will be a real square matrix constructed with coefficients of a real polynomial.

Case 2: We now consider stability aspects when κ_j^1, κ_j^2 does not equal zero. In this case, the equilibria of the dynamic SC market described through Equations (13a), (13b), and (13c), also lies in the set E . We define y_1, y_2 , and $V(y_1, y_2)$ as before but define d_Δ as

$$d_\Delta = d - d_{\Delta_{SC}} + d_{\Delta_C}, \quad (25)$$

where d is the same as in Equation (22), Δ_{SC} and Δ_{SC} represent the supply demand perturbation matrices, and $d_{\Delta_{SC}}$ and d_{Δ_C} are given by

$$d_{\Delta_{SC}} = \frac{4\lambda_{\min}(P_2) \psi_{\min} \|P_1\|_2 \pi_i |i \in SC|}{\beta^2}. \quad (26a)$$

$$d_{\Delta_C} = \frac{4\lambda_{\min}(P_2) \psi_{\min} \|P_1\|_2 \pi_j |j \in C|}{\beta^2}. \quad (26b)$$

We now have the following theorem characterizing market stability. The proof of the theorem is in Section 3.

Theorem 3. Let A_1 be Hurwitz, and let

$$\pi_{SC} - \pi_C < \frac{\lambda_{\min}(Q)}{2\|P_1\|_2} \quad (27)$$

Then the equilibrium (x_1^*, x_2^*) is asymptotically stable for all initial conditions in

$$\Omega_{c_{\max}} = \{(y_1, y_2) | V(y_1, y_2) \leq c_{\max}\} \text{ for } c_{\max} > 0,$$

such that $\Omega_{c_{\max}} \subsetneq D = \{y_2 \geq 0 | \|y_2\|_2 \leq d_\Delta\}$.

Proof. Differentiating the Lyapunov function $V(y_1, y_2)$ along the trajectories of 16 (see main paper), we get

$$\dot{V}(y_1, y_2) \leq -a_\Delta \left(\|y_1\| - \frac{\beta}{a_\Delta} \|y_2\| \right)^2 - \|y_2\| \left(e - \frac{\beta^2}{a_\Delta} \|y_2\| \right), \quad (28)$$

where $a_\Delta = \lambda_{\min}(Q) - 2\|P_1\| \pi_{SDC} + 2\|P_1\| \pi_C$, and $e = 2\lambda_{\min}(P_2) \psi_{\min}$.

From Equation (27) (see main paper) it follows that $a_\Delta > 0$. Therefore, (20) implies that for all $\Omega_{c_{\max}} \subsetneq D_\Delta$, for all solutions beginning in Ω_Δ , $\dot{V} \leq 0$. Hence, the market equilibrium state is stable, and Ω_Δ is the region of attraction.

The asymptotic stability of the perturbed market can be shown via the following argument: since the initial conditions start in Ω_Δ and the latter is a strict subset of D_Δ , y_2 cannot be equal to $2\lambda_{\min}(P_2) \psi_{\min} \frac{\lambda_{\min}(Q)}{\beta^2}$. This in turn implies that $(\|y_1\|, \|y_2\|) = (0, 0)$ is the only invariant set. Hence, all solutions starting in Ω_Δ converge to the equilibrium point $(x_1, x_2) = (x_1^*, x_2^*)$. Thus, we have proved Theorem 3.

TABLE 1
Parameters of Cost Functions for SCs

SC#	$VM^{\min}$	$VM^{\max}$	τ_{SC}	β	α
SC1	0	200	0.6	-0.3	90
SC2	0	200	0.2	-0.6	102
SC3	0	250	0.6	-0.25	80
SC4	0	250	0.6	-0.25	80
SC5	0	200	0.2	-0.01	20

Theorem Implications. Similar to the implications of Theorem 2, this theorem states that irrespective of any initial state the market is in, on being perturbed, it will always come back to an equilibrium state from a disequilibrium state.

5 NUMERICAL EVALUATION

In this section, we provide a numerical example of our dynamic market model to lightly illustrate system aspects not covered in the theory of this paper: (a) static market efficiency and average stakeholder utility under different Samuelson-Bergson welfare metrics, and (b) stability behavior of dynamic markets, and the corresponding speed at which markets reach stability. In the absence of a theoretical study (not our focus in this paper), (a) is important to get an approximate idea of the gaps in net stakeholder utility achieved via different welfare functions. (b) is important to characterize the speed of convergence of an SC market to go from a state of disequilibrium to a state of equilibrium. The first part of this section describes the example setting, and the second part analyzes the results.

5.1 Evaluation Setup

As an example evaluation setting, we consider five SCs and 15 customers (not including other SCs). Each SC has five customers each and they are locked in to the SCs throughout the entire duration of the experiment. Each customer requests a job and waits for it to complete before requesting another. Peer SCs are assumed to be altruistic w.r.t. VM borrowing. A reasonable set of market parameters for the SCs and the customers are shown in Tables 1 and 2 respectively. We note here that the rationale behind our parameter selection (specifically the τ values) follows directly from parameter explanations in Section 4, and also noted below. We also note that Tables 1 and 2 are just one table instance, i.e., representative example, in the set of parameter instances used for our evaluations. Our results for other table instances are very similar to those reported in the paper. We simulate a perfect price competition

TABLE 2
Parameters of Utility Functions for Customers

Customer(C)#	$VM^{\min}$	$VM^{\max}$	τ_C	β	α
C1, C2	60	100	0.1	-0.5	168
C3, C4	60	100	0.1	-0.15	140
C5, C6	70	80	0.2	-0.35	140
C7, C8	20	60	0.2	-0.2	100
C9, C10	30	60	0.2	-0.3	120
C11, C12	20	40	0.2	-0.1	125
C13, C14, C15	30	60	0.2	-0.5	135

game between the SCs, and use the *tatonnement process* (TP) [31] to converge to a static market equilibrium in practice for a distributed setting. *Tatonnement* is a trial-and-error process similar to the *hill climbing* approach in local search theory by which equilibrium is reached in competitive markets via a distributed fashion. As a measure of static market efficiency we investigate and compare the utilitarian SW function values at market equilibrium for *utilitarian*, *egalitarian*, and *Rawlsian* (see Section 2.3 in paper for more details) regulators. Note that the utilitarian SW function reflects the net stakeholder utility, and our goal is to study the net stakeholder utility at market equilibrium for regulators with different utility equitability mindsets. For the parameter values in Tables 1 and 2, we run numerical evaluations for all possible permutations (instances) of values that are applicable to SCs and their customers, and report the mean 95 percent confidence value of the results obtained (with the exception of Fig. 2c which reports (without loss of generality) on individual permutations). Note that each permutation of values can be considered as a different market setting. To experiment with dynamic markets, as a representative example, we fix τ_p to be the same for all SCs and wlog. vary it in the interval $[0, 5]$. Similarly we fix τ_{ag} to be the same for all 15 customers and vary it in the interval $[.05, .2]$. We also make κ^1 and κ^2 to be equal for all customers and vary it in the interval $[0, 0.05]$. To provide a rationale behind the values for the time constant, τ_p , first note that it represents the market time scale for the update of prices. Small values of τ_p imply a fast update in real-time price, which can introduce volatility. On the other hand, high values of τ_p contribute to reduced volatility. The time constant, τ_{ag} , represents the reciprocal of consumer demand elasticity.

5.2 Analysis of Evaluation Results

In the first part of this section, we analyze SC cost and customer utility allocations at market equilibrium under the

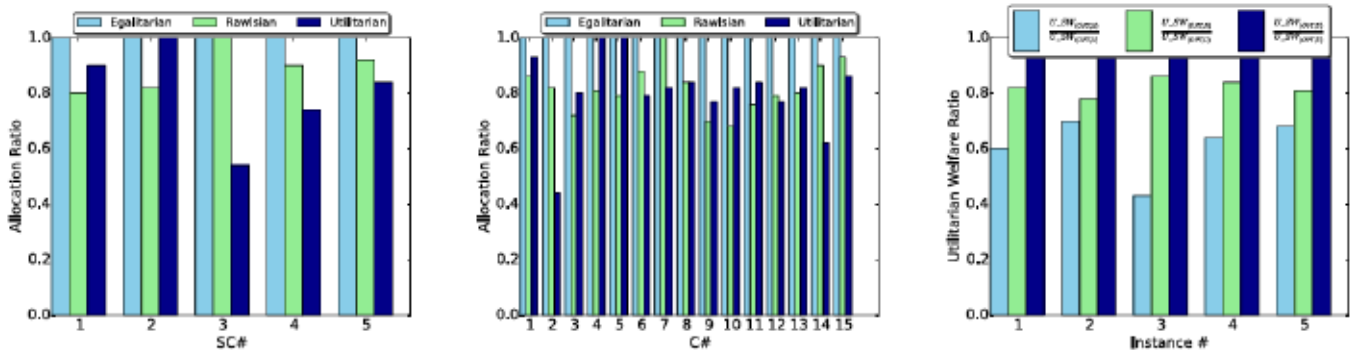


Fig. 2. ME Performance (a) SC Costs, (b) Customer Utilities, and (c) Social Welfare Ratio w.r.t to various SW metrics.

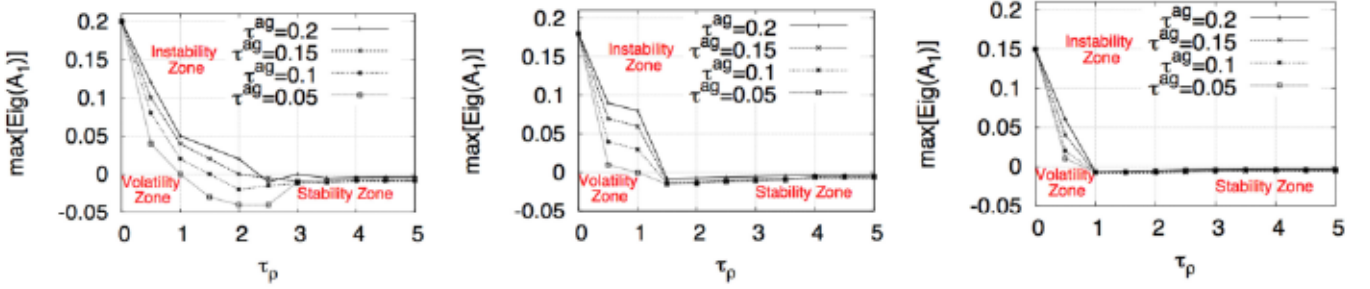


Fig. 3. Market Stability Performance when (a) $\kappa_1 = \kappa_2 = 0$, (b) $\kappa_1 = \kappa_2 = 0.02$, and (c) $\kappa_1 = \kappa_2 = 0.05$ [Utilitarian].

utilitarian, egalitarian, and Rawlsian SW paradigms. In the second part, we analyze the stability of various dynamic market settings, and also how fast a dynamic market converges to a stable equilibrium.

Static Market Equilibrium Performance. Using TP and in the presence of a regulator with different social welfare (SW) mindsets, we arrive at a different single market equilibrium (ME) maximizing SW. Note here that ME might not be unique, and in this case TP will converge locally to a ME in a distributed manner. The regulator will then have the option to work upon the ME to maximize SW. We observe (as a mean of multiple instances) from Figs. 2a and 2b that with respect to SC and customer allocation ratio equitability, Egalitarian MEs are the best as they ensure nearly identical cost and utility allocation ratios across all autonomous SCs and customers respectively, followed closely by Rawlsian MEs, and utilitarian MEs that are not very fair (equitable) in the utility allocation sense. Here, we define *allocation ratio* as the ratio of the cost (utility) of SC (customer) i to the maximum cost (utility) of any SC (customer) at ME, for each given market type. On the other hand, we see that market equilibrium in utilitarian markets, MEs lead to a considerably greater additive stakeholder satisfaction (utility) (see Fig. 2c.) when compared to egalitarian and Rawlsian markets, i.e., the utilitarian SW metric is highest in utilitarian markets. This is true from theory as marginal stakeholder utility at utilitarian ME is equal across all stakeholders. In addition, from theory, SW maximizing ME in utilitarian competitive markets are *always* Pareto optimal. In Fig. 2c, $U - SW_{OPT,t}, t \in \{U, E, R\}$, denotes the utilitarian social welfare value at the optimal market situation of type t , and $\frac{U - SW_{OPT,t}}{U - SW_{OPT,U}}$ is the ratio of the utilitarian social welfare value at the optimal market situation of type t to the optimal utilitarian social welfare at utilitarian ME.

Dynamic Market Stability Performance. Through Figs. 3, 4, and 5, we study dynamic markets for three different instances of (κ_1, κ_2) pairs, for utilitarian, Rawlsian, and egalitarian market types, respectively. For each instance, and for any market type, we observe that low values of τ_ρ for a given

instance correspond to market instability, i.e., a state of disequilibrium, because they imply a fast update in SC prices charged to customers, indicating market volatility in supply and demand as well. Here, stability is indicated through the maximum of the eigenvalues of Hurwitz matrix A_1 (see Section 3) formed from the market instance, which are negative in the stable zone, and positive in the unstable zone. It is logical to expect that market instability can be reduced if the price update is slower, i.e., if τ_ρ is larger. In this regard, we observe that the reduction in market instability takes place the slowest for egalitarian market types because in such markets a stable condition needs to satisfy a strict requirement of *absolute* stakeholder utilities to be equal at an efficient ME (see Section 2.3), and this condition is quite hard for the system to re-satisfy once market parameters are subject to variations - hence the slow pace of converging to market stability. On a similar note, the Rawlsian market type is relatively faster to market stability in events of system parameter variations because at an efficient ME a max/min utility allocation is the indicator of stability, and this intuitively is faster to achieve than a strict absolute equality of stakeholder utilities as in the egalitarian case. Finally, the reduction in market instability is the fastest for utilitarian market types, as the system needs to only guarantee equal marginal (and not absolute) utilities at an efficient ME, and this state is comparatively faster to attain in events of system parameter variations compared to the egalitarian and Rawlsian paradigms.

We also observe from Figs. 3, 4, and 5 that market volatility is increased due to a decrease in τ^{ag} values because the latter trend corresponds to the increase in demand elasticity (and hence price variations) that contributes to a market being volatile. We infer from the plots that it is possible to design an SC market where volatility (arising due to either low τ_ρ or low τ^{ag} values) can be contained by increasing market latency, i.e., increasing τ_ρ values. An increase in the market latency reflects the situation of a market of relatively low demand elasticity on variations in systems parameters

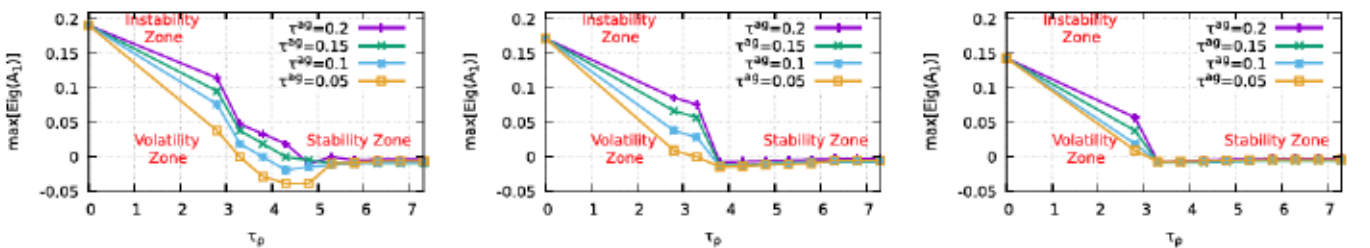


Fig. 4. Market Stability Performance when (a) $\kappa_1 = \kappa_2 = 0$, (b) $\kappa_1 = \kappa_2 = 0.02$, and (c) $\kappa_1 = \kappa_2 = 0.05$ [Rawlsian].

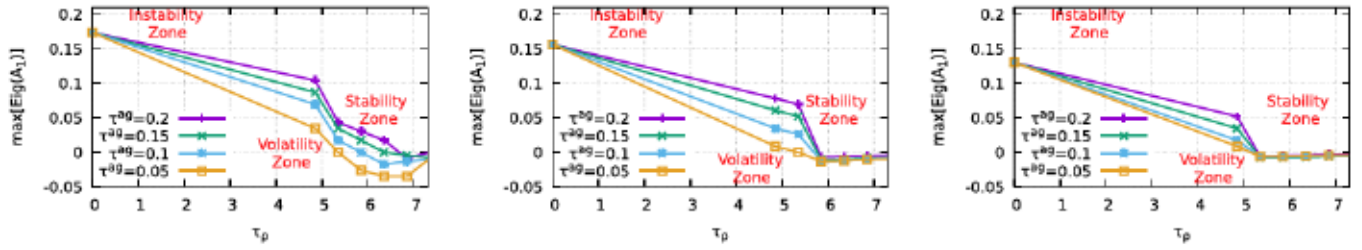


Fig. 5. Market Stability Performance when (a) $\kappa_1 = \kappa_2 = 0$, (b) $\kappa_1 = \kappa_2 = 0.02$, and (c) $\kappa_1 = \kappa_2 = 0.05$ [Egalitarian].

(e.g., number of available VMs) thereby leading to the system spending more time in a ‘stable’ state, and being less volatile in terms of exiting and entering such ‘stable’ states. With respect to the speed of convergence, from Figs. 3, 4, and 5, we observe in general that SC markets converge fast (approximately exponentially, based on our model) to the stable zone, i.e., even at low values of τ_{tho} - this irrespective of the social welfare paradigm type, and the speed of convergence increases with increasing κ_1, κ_2 values. The latter is because increasing κ values indicate more demand curtailment by SC customers, thereby leading to system states where there are lesser chances of scarcity of resources. This lack of scarcity sustains an efficient static market equilibrium for comparatively longer periods of time when compared to the case when there is a lesser resource demand curtailment by the SC customers - consequently leading to increasing non-volatility in resource supply and prices.

6 RELATED WORK

We give an overview of efforts related to ours and highlight the relevant differences. Works on hybrid clouds [32], [33] are related as they allow private (or smaller-scale) clouds to outsource their requests to large-scale public providers. However, since that can potentially be costly for a small-scale provider, our work differs in that it focuses on a sharing framework, while minimizing cost of using public clouds.

Earlier efforts also study the competition and cooperation within a federated cloud. For instance, authors in [34], [35] characterize the cloud federation to help cloud providers maximize their profits via dynamic pricing models. Earlier efforts [36], [37], [38], [39] also study the competition and cooperation among cloud providers, but assume that each cloud provider has sufficient resources to serve all users’ requests, while [37] incorporates a penalty function to address the service delay penalty. Authors in [40] propose a hierarchical cooperative game theoretic model for better resources integration and achieving a higher profit in the federation. Earlier efforts [28], [41] proposed coalitional games to form a federation among highly reputed providers to achieve high individual profits and high reputation while maintaining the level of committed QoS. To completely remove costly financial transactions, [42] prioritizes the most urgent requests in a new multi-agent based Cloud Resource Bartering System (CRBS) to maintain consistent service delivery. Similarly to our work, [43] studies a federation formation game but assumes that cloud providers share everything with others, while [44] adopts cooperative game theoretic approaches to model a cloud federation and study the motivation for cloud providers to participate in a federation. Another line of work focuses on designing sharing policies in

the federation to obtain higher profit. For instance, [45] proposes a decentralized cloud platform *SpotCloud* [46], a real-world system allowing customers or SCs to sell idle compute resources at specified prices, and presents a resource pricing scheme (resulting from a repeated seller game) plus an optimal resource provisioning algorithm. [47] employs various cooperation strategies under varying workloads, to reduce the request rejection rate (i.e., the efficiency metric in [47]). Another effort [48] combines resource outsourcing and rejection of less profitable requests in order to increase resource utilization and profit. [49], [50] proposes to efficiently deploy distributed applications on federated clouds by considering security requirements, the cost of computing power, data storage and inter-cloud communication. [51] groups resources of various SCs into computational units, in order to serve customers’ requests. [52] proposes to incorporate both historical and expected future revenue into VM sharing decisions in order to maximize an SC’s profit. [53] designs a reinsurance-emulated collaboration mechanism to maximize the shared resource utilization.

Differences and Drawbacks. Our work is a necessarily important theoretical extension of a very recent analytical work in [2] that was the first of its kind in the analysis of small cloud markets. There, the authors considered consequences of performance (i.e., queueing theory) driven non-cooperative game-theoretic (with no SC willing to share its utility and capacity information with others, i.e., an incomplete information game-theoretic setting) resource sharing on the resulting performance delivered to customers at *static market equilibrium*, something not considered by any of the above-mentioned efforts. In this regard, the authors in [2] show the existence of SC market equilibrium for their game setting through numerical simulations, and do not provide a general theory for equilibrium existence. In addition, [2] does not consider the important problem of analyzing equilibrium stability under variations in SC resource availability, in a non-cooperative game-theoretic SC environment. A major drawback of our previous work is that Without showing the existence of a stable SC market (conservatively assuming that the resource allocation competition has at least one market equilibria, and that one would locally or globally encounter such equilibria via simulations), cannot not say much regarding the sustainability of SC markets in the future. Without certain guarantees on the sustainability of such markets, cloud firms will not be encouraged to be in the resource sharing business. A characterization of this scenario is an important contribution of this work. A major difference of our work with the one in [2], is the lack of a queueing-driven performance model to reduce the equilibrium search space. However, our work is orthogonal in the sense that, given the existence of (efficient) market equilibria, we investigate whether such a state is sustainable in the long run. In contrast

to the work in [2], we also mathematically characterize equilibrium in market competition. In a very recent effort [28], the authors study stability of a market coalitional equilibrium in a strategically cooperative federated setting. With respect to stability, our work is different from theirs on two counts: (i) we study stability of staying at an efficient static market equilibrium in the presence of resource variations, whereas the authors in [28] study stability of a coalition formation between the members of a cloud sharing federation, and (ii) our notion of a federation is a strategically non-cooperative setting where autonomous and selfish cloud providers do business by being competitors to other cloud providers but at the same time think of the overarching interests of the federation in sharing their resources 'non-altruistically' when feasible. With respect to the work in [28], while stable coalition formation is definitely a relevant area for research in federated cloud settings, we are not quite certain about cooperative behavior between profit maximizing SCs in practice, and would argue in favor of a co-opetitive resource sharing setting when compared to a cooperative one.

7 DISCUSSION AND SUMMARY

Discussion. The model presented in this paper is quite general for resource allocation settings where (a) resources influencing the QoS of an engineering system (e.g., renewable energy in the smart grid) are not owned in enough quantities at all times by any of the service providing firms and (b) the firms form a competitive market, and need to mutually rely on one another at times of resource deficiency, rather than resorting to expensive third party (e.g., commercial public cloud, energy utility, etc.) buyouts. On top of this, our model is also extensible to a graphical setting where a particular firm can only have access to resources from a certain set of firms based on geographical proximity. Another important thing to note about our model is the single resource economics, which in practice can be multi-resourced (e.g., #VMs, energy consumption amounts with an eye towards a greener environment). In order to extend our model to such settings, we need to resort to *general equilibrium theory* in microeconomics (unlike our analysis that is based on *partial equilibrium theory* of a single commodity type, i.e., VMs in our case) to capture (a) the tradeoffs between allocating different commodities, using indifference curves, and (b) optimize social welfare over a multi-variate function of individual firm utilities. Consequently, one needs to study the existential nature of resource sharing market equilibria using Arrow-Debreu theorems [14], according to which at an equilibrium point, the marginal rate of substitution (MRS) [14] between different commodities for the different firms are equal.

Summary. In this paper, we addressed the problem of effective resource sharing between small clouds (SCs). We modeled the problem as an efficient supply-demand market design task consisting of (i) autonomous SCs, (ii) their customers, and (iii) a regulator, as the market stakeholders. We first showed that a welfare allocation policy for the stakeholders by the regulator maximizes utilitarian social welfare at the static market equilibrium and results in the best/most efficient state at which the SC markets could operate. Fortunately, courtesy Arrow-Debreu welfare theorems in welfare economics, this unique optimal operating point is also achieved in a

distributed manner by the autonomous SCs in perfect price competition with one another, thereby guaranteeing no efficiency loss in a non-centralized market setting. The optimal market equilibrium point is prone to perturbations due to the dynamic nature of the SC market, thereby potentially leading to market disequilibrium. In this context, we designed a dynamic market mechanism based on Arrow and Hurwicz's disequilibrium process that uses the gradient play technique in game theory to converge upon the optimal static market efficient equilibrium from a disequilibrium state caused due to supply-demand perturbations, and results in market stability. As part of future work, we plan to design provably fast distributed algorithms to allow markets to roll back to efficient equilibria when perturbed from an equilibrium state, and study dynamic SC markets under (i) a setting of imperfect (multi-resource) competition between SCs and (ii) under heterogeneous VM profiles.

ACKNOWLEDGMENTS

The authors would like to thank the multiple reviewers for their valuable suggestions to improve this article. This work was supported in part by the NSF CNS-1816887 award. Sung-Han Lin and Aditya Ahuja contributed equally.

REFERENCES

- [1] H. Zhuang, R. Rahman, and K. Aberer, "Decentralizing the cloud: How can small data centers cooperate," in *Proc. IEEE Int. Conf. Peer-to-Peer Comput.*, 2014, pp. 1–10.
- [2] S.-H. Lin, R. Pal, M. Paolieri, and L. Golubchik, "Performance driven resource sharing markets for the small cloud," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst.*, 2017, pp. 241–251.
- [3] K. J. Arrow and L. Hurwicz, "On the stability of the competitive equilibrium I," *Econometrica*, vol. 26, no. 4, pp. 522–552, 1958.
- [4] K. J. Arrow, H. D. Block, and L. Hurwicz, "On the stability of the competitive equilibrium II," *Econometrica*, vol. 27, no. 1, pp. 82–109, 1959.
- [5] M. Armbrust *et al.*, "A view of cloud computing," *Commun. ACM*, vol. 53, pp. 50–58, 2010.
- [6] "Amazon AWS," Accessed: 2018. [Online]. Available: <http://aws.amazon.com>
- [7] "Google Compute Engine," Accessed: 2018. [Online]. Available: <https://cloud.google.com/compute>
- [8] "Microsoft Azure," Accessed: 2018. [Online]. Available: <https://azure.microsoft.com>
- [9] "Blippex - Why we moved away from AWS," Accessed: 2018. [Online]. Available: <http://blippex.github.io/updates/2013/09/23/why-we-moved-away-from-aws.html>
- [10] "As Private Cloud Grows, Rackspace Expands Options Inside Equinix," Accessed: 2018. [Online]. Available: <https://blog.equinix.com/blog/2016/05/16/as-private-cloud-grows-rackspace-expands-options-inside-equinix/>
- [11] J. C. Mogul and L. Popa, "What we talk about when we talk about cloud network performance," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 42, pp. 44–48, 2012.
- [12] "It's clear to Linode: There's a market to bring cloud services to small companies," *New Jersey Business Journal*, Nov. 2013.
- [13] Wikipedia, *Business Cluster*, Accessed: 2018. [Online]. Available: <http://en.wikipedia.org>
- [14] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic Theory*. London, U.K.: Oxford Univ. Press, 1995.
- [15] N. Samaan, "A novel economic sharing model in a federation of selfish cloud providers," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 12–21, Jan. 2014.
- [16] B. Rochwerger *et al.*, "Reservoir: When one cloud is not enough," *Comput.*, vol. 44, no. 3, pp. 44–51, Mar. 2011.
- [17] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to nash equilibria," *IEEE Trans. Autom. Control*, vol. 50, no. 3, pp. 312–327, Mar. 2005.

- [18] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA, USA: MIT Press, 1991.
- [19] S. D. Flam, "Equilibrium, evolutionary stability, and gradient dynamics," *Int. Game Theory Rev.*, vol. 4, no. 4, 2002.
- [20] G. Ananthanarayanan, M. Hung, X. Ren, I. Stoica, A. Wierman, and M. Yu, "GRASS: Trimming stragglers in approximation analytics," in *Proc. 11th USENIX Conf. Netw. Syst. Design Implementation*, 2014, pp. 289–302.
- [21] O. Candogan, K. Bimpikis, and A. Ozdaglar, "Optimal pricing in networks with externalities," *Operations Res.*, vol. 60, no. 4, pp. 883–905, 2012.
- [22] J.-J. Laffont and J. Tirole, *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA, USA: MIT Press.
- [23] A. Sen et al., *On Economic Inequality*. London, U.K.: Oxford Univ. Press, 1997.
- [24] J. Bulow, J. Geneakolos, and P. Klemperer, "Multi-market oligopoly: Strategic substitutes and complements," *J. Political Economy*, vol. 93, pp. 488–511, 1985.
- [25] J. J. Laffont and J. Tirole, *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA, USA: MIT Press, 1993.
- [26] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [27] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA, USA: SIAM Studies in Applied Mathematics, 1994.
- [28] B. Ray, A. Saha, S. Khatua, and S. Roy, "Quality and profit assured trusted cloud federation formation: Game theory based approach," *IEEE Trans. Services Comput.*, to be published, doi: 10.1109/TSC.2018.2833854.
- [29] G. Wang, M. Negrete, A. Kowli, E. Shafiepoorfar, S. Meyn, and U. V. Shanbhag, *Dynamic Competitive Equilibria in Electricity Markets*. Berlin, Germany: Springer, 2011.
- [30] A. Bacciotti and L. Rosier, *Liapunov Functions and Stability in Control Theory*. Berlin, Germany: Springer, 2005.
- [31] D. P. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1992.
- [32] H. Zhang, G. Jiang, K. Yoshihira, and H. Chen, "Proactive workload management in hybrid cloud computing," *IEEE Trans. Netw. Service Manage.*, vol. 11, no. 1, pp. 90–100, Mar. 2014.
- [33] M. Shifrin, R. Atar, and I. Cidon, "Optimal scheduling in the hybrid-cloud," in *Proc. IFIP/IEEE Int. Sympo. Integr. Netw. Manage.*, 2013, pp. 51–59.
- [34] I. Gori, J. Guitart, and J. Torres, "Economic model of a cloud provider operating in a federated cloud," *Inf. Syst. Frontiers*, vol. 14, pp. 827–843, 2012.
- [35] M. Hadji and D. Zeglache, "Mathematical programming approach for revenue maximization in cloud federations," *IEEE Trans. Cloud Comput.*, vol. 5, no. 1, pp. 99–111, Jan.–Mar. 2015.
- [36] T. Truong-Huu and C.-K. Tham, "A novel model for competition and cooperation among cloud providers," *IEEE Trans. Cloud Comput.*, vol. 2, no. 3, pp. 251–265, Jul.–Sep. 2014.
- [37] H. Chen, B. An, D. Niyato, Y. Soh, and C. Miao, "Workload factoring and resource sharing via joint vertical and horizontal cloud federation networks," *IEEE J. Selected Areas Commun.*, vol. 35, no. 3, pp. 557–570, Mar. 2017.
- [38] J. Cohen and L. Echabbi, "Pricing composite cloud services: The cooperative perspective," in *Proc. 5th Int. Conf. Next Gener. Netw. Services*, 2014, pp. 335–342.
- [39] H. J. Moon, Y. Chi, and H. Hacigümüs, "SLA-aware profit optimization in cloud services via resource scheduling," in *Proc. 6th World Congress Services*, 2010, pp. 152–153.
- [40] D. Niyato, A. V. Vasilakos, and Z. Kun, "Resource and revenue sharing with coalition formation of cloud providers: Game theoretic approach," in *Proc. 11th IEEE/ACM Int. Symp. Cluster Cloud Grid Comput.*, 2011, pp. 215–224.
- [41] L. Mashayekhy, M. M. Nejad, and D. Grosu, "A trust-aware mechanism for cloud federation formation," *IEEE Trans. Cloud Comput.*, to be published, doi: 10.1109/TCC.2019.2911831.
- [42] S. Z. Goher, P. Bloodsworth, R. U. Rasool, and R. McClatchey, "Cloud provider capacity augmentation through automated resource bartering," *Future Gener. Comput. Syst.*, vol. 81, pp. 203–218, 2018.
- [43] L. Mashayekhy, M. M. Nejad, and D. Grosu, "Cloud federations in the sky: Formation game and mechanism," *IEEE Trans. Cloud Comput.*, vol. 3, no. 1, pp. 14–27, Jan.–Mar. 2015.
- [44] M. M. Hassan, M. S. Hossain, A. J. Sarkar, and E.-N. Huh, "Cooperative game-based distributed resource allocation in horizontal dynamic cloud federation platform," *Inf. Syst. Frontiers*, vol. 4, pp. 523–542, 2014.
- [45] H. Wang, F. Wang, J. Liu, D. Wang, and J. Groen, "Enabling customer-provided resources for cloud computing: Potentials, challenges, and implementation," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 7, pp. 1874–1886, Jul. 2015.
- [46] "SpotCloud," Accessed: 2018. [Online]. Available: <http://www.spotcloud.com/>
- [47] H. Zhuang, R. Rahman, and K. Aberer, "Decentralizing the cloud: How can small data centers cooperate?," in *Proc. 14th IEEE Int. Conf. Peer-to-Peer Comput.*, 2014, pp. 1–10.
- [48] A. N. Toosi, R. N. Calheiros, R. K. Thulasiram, and R. Buyya, "Resource provisioning policies to increase iaaS provider's profit in a federated cloud environment," in *Proc. IEEE 13th Int. Conf. High Perform. Comput. Commun.*, 2011, pp. 279–287.
- [49] Z. Wen, J. Cala, P. Watson, and A. Romanovsky, "Cost effective, reliable and secure workflow deployment over federated clouds," *IEEE Trans. Services Comput.*, vol. 10, no. 6, pp. 929–941, Nov./Dec. 2016.
- [50] T. Gouasmi, W. Louati, and A. H. Kacem, "Cost-efficient distributed mapreduce job scheduling across cloud federation," in *Proc. IEEE Int. Conf. Services Comput.*, 2017, pp. 289–296.
- [51] O. Babaoglu, M. Marzolla, and M. Tamburini, "Design and implementation of a P2P Cloud system," in *Proc. 27th Annu. ACM Symp. Appl. Comput.*, 2012, pp. 412–417.
- [52] N. Samaan, "A novel economic sharing model in a federation of selfish cloud providers," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 12–21, Jan. 2014.
- [53] S. Ye, H. Liu, Y.-W. Leung, and X. Chu, "Reinsurance-emulated collaboration mechanism in cloud federation," in *Proc. IEEE 10th Int. Conf. Cloud Comput.*, 2017, pp. 727–732.



Ranjan Pal received the PhD degree in computer science from the University of Southern California, in 2014. He is currently a faculty member of ECE at the University of Michigan Ann Arbor and was a visiting research fellow with the Computer Laboratory at the University of Cambridge. He specializes in the application of mathematical decision sciences to engineer welfare improving solutions to problems in cyber-security and privacy. Prior to this, he was a research fellow a Provost PhD Fellow at the University of Southern California.

During different points in his career, he has held research appointments at Princeton, MIT, Tsinghua University, NUS, IIT Delhi, IIM Ahmedabad, IIM Calcutta, the University of Helsinki, and Aalborg University. He is a member of the IEEE, the ACM, the American Mathematical Society (AMS), INFORMS, SIAM, and the Game Theory Society.



Sung-Han Lin received the PhD degree in computer science in quantitative evaluation & design group (QED) from the University of Southern California (USC), Los Angeles, CA, in 2017. He has been a performance analyst at NetApp, USA, since October 2017. His main research interests include the performance modeling, analysis, and design of large-scale distributed systems, including cloud computing systems, peer-to-peer networking, and multimedia systems. He is a member of the IEEE and ACM.



Aditya Ahuja received the bachelor's degree in computer science from the Manipal Institute of Technology, India, and the master's degree in computer science from VIT University, India. He is currently working toward the PhD degree in computer science at the Indian Institute of Technology Delhi, India, and has been a visiting research scholar with the University of Southern California. His primary research interests lie in network and distributed systems security, cryptography, and game theory. He is a student member of the IEEE.



Nachikethas A. Jagadeesan received the BTech degree in engineering physics from the Indian Institute of Technology Madras, in 2011, and the first MS degree in electrical engineering, the second MS degree in statistics, and the PhD degree in electrical engineering from the University of Southern California, in 2015, 2017, and 2018, respectively. He was an Annenberg Fellow during his PhD studies. He is a researcher and software engineer at Google, USA. His research interests broadly lie in wireless networks, statistics, and machine learning. He is a member of the IEEE.



Abhishek Kumar received the BTech degree in computer science & engineering from the National Institute of Technology Patna, India, in 2014, and the MS degree in knowledge service engineering from the Korea Advanced Institute of Science and Technology (KAIST), in 2016. He is currently working toward the PhD degree in computer science at the University of Helsinki. His research interests broadly lie in security and privacy in Apps and IoT space, human computer interaction, and applied machine learning. He is a student member of the IEEE.



Leana Golubchik is a Stephen and Etta Varra professor of computer science and ECE systems with the University of Southern California, Los Angeles, CA. Prior to that, she was with the University of Maryland and Columbia University. Her research interests are broadly in the design and evaluation of large scale distributed systems including hybrid clouds and their applications in data analytics, QoS-based design of P2P and multimedia systems, and reliability of software architectures. She is a past chair of ACM SIG-METRICS and a member of the IFIP WG 7.3. She was the recipient of several awards, including the NSF CAREER award, the Okawa Foundation award, and the IBM and NSF doctoral fellowships. She is a senior member of the IEEE.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.