On the Economic Sustainability of Cloud Sharing Systems
Are Dynamic Single Resource Sharing Markets Stable?

Ranjan Pal [1], Aditya Ahuja [2], Sung-Han Lin [3], Leana Golubchik [5], Abhishek Kumar [4], Nachikethas A. Jagadeesan [6]

[1] University of Cambridge, rp631@cam.ac.uk
[2] Indian Institute of Technology Delhi, aditya.ahuja@cse.iitd.ac.in
[3] NetApp, CA, USA, SungHan.Lin@netapp.com
[4] University of Helsinki, abhishek.kumar@helsinki.fi
[5] University of Southern California, leana@usc.edu
[6] Google, CA, USA, nanantha@usc.edu

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 [29] 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 sharing of resources between competing SCs to alleviate the resource inelasticity issues that might arise. Based on this idea, a recent effort ([18]) 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 paper, 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* [9, 10] in economics, and the *gradient play* technique in game theory that allows us to iteratively converge upon efficient and stable market equilibria.

CCS CONCEPTS

• **Modeling Methodologies → Economics**;

KEYWORDS

small cloud, dynamic market, stability, disequilibrium process

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 [8]. Cloud service providers (Amazon AWS [1], Google Compute Engine [4], and Microsoft Azure [6]) allow customers to quickly deploy their services without a large initial infrastructure investment.

ACM SIGMETRICS Performance Evaluation Review, 2019
2020. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

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., Blippex [3]), thus resulting in more and more customers building their own smaller data centers [2] 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 [21]. 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 [5] 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 to help meet each others' user demand via resource sharing at low costs, thereby increasing their individual 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 [28].

1.2 Research Motivation

In this section, we briefly describe the problem setting followed by the challenges that motivate us to alleviate them.

Problem Setting. The effective sharing or borrowing resources by an SC from its peers involves mutually satisfying the interests of the stakeholders in context. In this paper, we consider three different stakeholders: (i) the SC customers, (ii) profit maximizing autonomous SCs, and (iii) a regulatory agency overseeing certain functioning aspects of the autonomous SCs (e.g., ensuring customer data privacy). The SC customers are interested in achieving certain performance measures for their jobs (e.g., low job response time, cheap storage); the SCs are interested in maximizing revenues obtained from serving customers; and a regulatory agency (e.g., the local government, a federated agency [24][23]) is interested in ensuring a proper manner by which the autonomous SCs conduct their business of lending resources to peer SCs (e.g., many European countries are now concerned about preserving data privacy and disallowing their data leaving European borders [20]). *We term the above setting as an SC market.*

The Challenges. 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). The latter option is generally more expensive than the former and also likely to be more privacy threatening. Thus, from an SC's viewpoint, its challenge is to satisfy two conflicting objectives: (i) to generate as much revenue by serving its customer demands, and (ii) to incur as low as possible, borrowing and/or buying costs from other clouds. For simplicity purposes, 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. 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 [19]. *In this regard, the authors in [18] show the existence of SC market equilibrium through numerical simulations, and do not provide a general theory for equilibrium existence.* 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 efficient equilibrium, a state of eventual disequilibrium will threaten the long-term sustainability of SC markets. 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 eventually returns to its equilibrium point(s).

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.*

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 SC customers, profit maximizing autonomous SCs, and a regulatory agency overseeing some functionality aspects of the 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).
- Using the notion of a *disequilibrium process* proposed by Arrow and Hurwicz [9, 10], we apply the *gradient play* technique in game theory [26] 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 Appendix for details on static markets, as it is not the main focus of our paper) is *asymptotically stable* in dynamic market settings. 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* [14], (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 [15], from a state of disequilibrium. More specifically, for our market setting the occurrence of (i)-(iii) is quite likely. The gradient play technique also works to achieve static market equilibrium when issues (i)-(iii) do not arise (see Section 2.3).

Differences and Drawbacks w.r.t. [18] - Related literature on cloud sharing frameworks and their economics are detailed in the very recent paper by Lin et.al., [18]. Here, we state the differences and drawbacks of our contributions in this paper with respect to the work in [18].

Our work is a necessarily important theoretical extension of [18] 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.* However, [18] does not consider the important problem of analyzing equilibrium stability under variations in SC resource availability, in a non-cooperative game-theoretic SC environment. *Without showing the existence of a stable SC market, one, based on the existing results showing the existence of a market equilibrium, cannot not say much regarding the sustainability of SC markets in the future.* A characterization of this scenario is an important contribution of this work. A major difference of our work with the one in [18], is the lack of a queuing-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.

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, their customers, and a regulatory agency overseeing some functionality aspects of the SCs. A Walrasian competitive market [19] 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 [7]), where the curtailment decision primarily arising from (a) the nature of VM instance prices, (b) the unnecessity of the job to continue executing beyond a certain accuracy already achieved, and (c) the unnecessity 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 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.

$$c(vm_i^r) = f_1(vm_i^r) + f_2(vm_i^r), \tag{1}$$

where $f_1(\cdot)$ (a linear function) and $f_1(\cdot)$ (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 [19]. 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.

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

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 [13]. 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* [19]. 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:

$$c(vm_i^b) = \alpha_b^i vm_i^b + \frac{\beta_b^i}{2}(vm_i^b)^2, \tag{3}$$

where α_b^i (a positive quantity) and β_b^i (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^r 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 $c(vm_i^r)$ and $c(vm_i^b)$, and express it via the following equation:

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

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.

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

where vm_j^e 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^e (a positive quantity) and β_j^e (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:

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

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

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

Here, α_j^e (a positive value) and β_j^e (a negative value) are j's utility coefficients for Type I jobs. The interpretation of vm_j^c is as follows: $\kappa_j^1 vm_j^e$ is the number of VMs required to accomplish j's curtailed task, whereas $\kappa_j^2 vm_j^e$ 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 $\kappa_j^1 vm_j^e$ 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:

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

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 $vm_j^{ag} = vm_j^e + vm_j^c + vm_j^s$ VM instances. Therefore, customer j's aggregate utility takes a similar form to his utility for a specific job type, and is given by

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

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 = $\{e, c, s\}$ 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 for his job.

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[1]. However, the presence of a regulator brings in other important benefits through (ii) and (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[2], whereas (iii) is important from an economic perspective as *maximizing* social

[1]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.

[2]Economists Laffont and Tirole have proposed *principal-agent* models in this regard [17] which will enable autonomous SCs to charge appropriate prices to customers purely out of self-interest.

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 [19], 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

$$SW = \sum_{j \in C} U_j(vm_j^{ag}) - \sum_{i \in SC} \left(c(vm_i^r) + c(vm_i^b) + c(vm_i^{pc}) \right), \quad (9)$$

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 [19] 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 [19]. 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 focusses 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 as every SC has exactly the same utility level.

3 DYNAMIC SC MARKETS

On 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 Appendix 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. *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* [10], 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.

3.1 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* [14] 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 [26] 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* [14], (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 [15]. 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 ith SC is given by

$$\tau_i^r \dot{vm}_i^r = \rho_i - \beta_i^r vm_i^r - \alpha_i^r. \quad (10a)$$

$$\tau_i^b \dot{vm}_i^b = \rho_i - \beta_i^b vm_i^b - \alpha_i^b. \quad (10b)$$

$$\tau_i^{pc} \dot{vm}_i^{pc} = \rho_i - \beta_i^{pc} vm_i^{pc} - \alpha_i^{pc}. \quad (10c)$$

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 22a-22c (see Appendix) at static market equilibrium. It can be seen that the RHSs of 22a-22c 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 22a-22c 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

$$\tau_i^{ag} \dot{vm}_i^{ag} = \beta_i^{ag} vm_i^{ag} + \alpha_i^{ag} - \rho_i. \quad (11)$$

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 22d at static market equilibrium. It can be seen that the RHS of 12 is proportional to the gradient $\nabla_{vm_i^{ag}} L$, $i \in C$, where L is the Lagrangian of OPT. Equation 12 can be solved independently by each SC customer i.

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

$$\tau_{\rho_i} \dot{\rho}_i = \sum_{j \in C_i} vm_j^{ag}(1 - \kappa_j^1 - \kappa_j^2) - (vm_i^r + vm_i^b + vm_i^{pc}), \quad (12)$$

where the goal is to drive the solution ρ_i, $\forall i \in SC$ to ρ_i^*, the solution of 22e at static market equilibrium. Here, τ_{ρ_i} is the free parameter denoting the speed with which ρ_i can be adjusted. Equations 10-12 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^{type}, type $\in \{e, c, s\}$, 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 of the Appendix. This representation paves the way for analytically analyzing the stability of such markets.*

3.1.1 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 3.2). Using Equations 10-12, our proposed dynamic market mechanism can be compactly represented in the matrix form via the following equation:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} A_1 + \Delta A_1 & A_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} \bar{\alpha} \\ f_2(x_1 x_2). \end{bmatrix} \quad (13)$$

Definiton of Equation Parameters. We now describe the parameters of Equation 13. 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: $M_1 = Diag(\frac{1}{\tau_i^{type}} \beta_i^{type})$, type $\in \{r, b, pc\}$. We assume that all for a given type, τ_i^{type}'s are equal for all $i \in SC$. $M_2 = Diag(\frac{1}{\tau_i^{type}} A_{SC}^T)$, type $\in \{r, b, pc\}$, where $A_{SC} = Diag(1)$. $M_3 = Diag(\frac{1}{\tau_i^{type}} \beta_i^{type})$, type $\in \{e, c, s\}$.

$M_4 = Diag(\frac{1}{\tau_i^{type}} A_C^T)$, type $\in \{e, c, s\}$, where $A_C = Diag(1)$. $M_5 = Diag(A'^T BA)$, 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. $M_6 = Diag(\frac{1}{\tau_{\rho_i}^{type}} A_{SC})$, type $\in \{r, b, pc\}$. $M_7 = Diag(\frac{1}{\tau_{\rho_i}^{type}} A_C)$, type $\in \{e, c, s\}$. $M_8 = Diag(\frac{1}{\tau_{\rho_i}^{type}} A^T BA')$, 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

$$f_2(x_1, x_2) = [cx_1 - VM^{\max}]^+_{x_2}, \tag{14}$$

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 nth row of the projection $[cx_1 - VM^{\max}]^+_{x_2}$ is denoted as

$$\left[[cx_1 - VM^{\max}]^+_{x_2}\right]_n = \begin{cases} \max(0, [cx_1]_n - VM_n^{\max}, & \text{if } [x_2]_n = 0 \\ [cx_1]_n - VM_n^{\max}, & \text{if} [x_2]_n > 0 \end{cases} \tag{15}$$

ΔA_1 in Equation 13 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

$$E = \{\Delta_A = \Delta_{SC} - \Delta_C | \Delta_{SC} \in E_{SC}; \Delta_C \in E_C.\} \tag{16}$$

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

$$E_{SC} = \{\Delta_{SC} | ||\Delta_{SC}|| = \sqrt{\lambda_{\max}(\Delta_{SC}^T \Delta_{SC})} \le \pi_{SC}\},$$

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 $Diag\left(\frac{1}{\tau_{\rho_i}^{type}} A_C^T (I - \kappa_j^1 - \kappa_j^2)\right)$. We also have

$$E_C = \{\Delta_C | ||\Delta_C|| = \sqrt{\lambda_{\max}(\Delta_C^T \Delta_C)} \le \pi_C\},$$

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

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

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.

3.2 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.

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 22a - 22c (via the use of the gradient play technique), lies in the set

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

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 22a - 22e. The proof of the theorem is in the Appendix.

Theorem 3.1. *The equilibrium (x_1^*, x_2^*) is identical to the unique static market equilibrium obtained from the solution of OPT.*

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

$$V(y_1, y_2) = y_1^T y_1 + y_2^T P_2 y_2, \tag{17}$$

where P_1 and P_2 are diagonal matrices. We use *Lyapunov functions* from control theory [11] as a standard to prove the stability of an equilibrium of a system represented via ordinary differential equations (ODEs). Let d be expressed as

$$d = \frac{2\lambda_{min}(P_2)\psi_{\min}\lambda_{\min}(Q)}{\beta^2}, \tag{18}$$

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

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

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 the Appendix.

Theorem 3.2. *Let A_1 be Hurwitz. Then the equilibrium (x_1^*, x_2^*) is asymptotically stable for all initial conditions in*

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

such that

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

.

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 10-12. 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 10a - 10c, 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

$$d_\Delta = d - d_{\Delta_{SC}} + d_{\Delta_C}, \tag{19}$$

where d is the same as in Equation 18, Δ_{SC} and Δ_{SC} represent the supply demand perturbation matrices, and $d_{\Delta_{SC}}$ and d_{Δ_C} are given by

$$d_{\Delta_{SC}} = \frac{4\lambda_{min}(P_2)\psi_{\min}||P_1||_2\pi_i|i \in SC}{\beta^2}. \tag{20a}$$

$$d_{\Delta_C} = \frac{4\lambda_{min}(P_2)\psi_{\min}||P_1||_2\pi_j|j \in C}{\beta^2}. \tag{20b}$$

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

THEOREM 3.3. *Let A_1 be Hurwitz, and let*

$$\pi_{SC} - \pi_C < \frac{\lambda_{\min}(Q)}{2||P_1||_2} \tag{21}$$

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

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

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

Theorem Implications. Similar to the implications of Theorem 3.2, 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.

4 CONCLUSION AND FUTURE WORK

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. 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 using *general equilibrium theory*[19], (ii) under heterogeneous VM profiles, and (iii) a coalitional market setting where SCs have the capability to collude with one another.

REFERENCES

[1] [n. d.]. Amazon AWS. http://aws.amazon.com. ([n. d.]).
[2] [n. d.]. As Private Cloud Grows, Rackspace Expands Options Inside Equinix. https://blog.equinix.com/blog/2016/05/16/as-private-cloud-grows-rackspace-expands-options-inside-equinix/. ([n. d.]).
[3] [n. d.]. Blippex - Why we moved away from AWS. http://blippex.github.io/updates/2013/09/23/why-we-moved-away-from-aws.html. ([n. d.]).
[4] [n. d.]. Google Compute Engine. https://cloud.google.com/compute. ([n. d.]).
[5] [n. d.]. It's clear to Linode: There's a market to bring cloud services to small companies. http://www.njbiz.com/article/20131104/NJBIZ01/311019994/It. ([n. d.]).
[6] [n. d.]. Microsoft Azure. https://azure.microsoft.com. ([n. d.]).
[7] G. Ananthnarayanan, M.C. Hung, X. Ren, I. Stoica, A. Wierman, and M. Yu. 2014. GRASS: Trimming Stragglers in Approximation Analytics. In *NSDI*.
[8] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. 2010. A View of Cloud Computing. *Commun. ACM* 53 (2010).
[9] K. J. Arrow, H. D. Block, and L. Hurwicz. 1959. On the Stability of the Competitive Equilibrium II. *Econometrica* 27, 1 (1959).
[10] K. J. Arrow and L. Hurwicz. 1958. On the Stability of the Competitive Equilibrium I. *Econometrica* 26, 4 (1958).
[11] A. Bacciotti and L. Rosier. [n. d.]. *Liapunov Functions and Stability in Control Theory*. Springer, 2005.
[12] J. Bulow, J. Geneakolos, and P. Klemperer. 1985. Multi-market Oligopoly: Strategic Substitutes and Complements. *Journal of Political Economy* 93 (1985).
[13] O. Candogan, K. Bimpikis, and A. Ozdaglar. 2012. Optimal Pricing in Networks with Externalities. *Operations Research* 60, 4 (2012).
[14] D.Fudenberg and J.Tirole. [n. d.]. *Game Theory*. MIT Press, 1991.
[15] S. D. Flam. 2002. Equilibrium, Evolutionary Stability, and Gradient Dynamics. *International Game Theory Review* 4, 4 (2002).
[16] J. J. Laffont and J.Tirole. [n. d.]. *A Theory of Incentives in Procurement and Regulation*. MIT Press, 1993.
[17] J-J. Laffont and J. Tirole. [n. d.]. *A Theory of Incentives in Procurement and Regulation*. MIT Press.
[18] Sung-Han Lin, Ranjan Pal, Marco Paolieri, and Leana Golubchik. 2017. Performance Driven Resource Sharing Markets for the Small Cloud. In *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*. IEEE, 241–251.
[19] A. Mas-Collel, M. D. Whinston, and J. R. Green. [n. d.]. *Microeconomic Theory*. Oxford University Press, 1995.
[20] Mayer and Brown. [n. d.]. *European and German Privacy Laws*. http://www.martindale.com.
[21] Jeffrey C Mogul and Lucian Popa. 2012. What we talk about when we talk about cloud network performance. *ACM SIGCOMM Computer Communication Review* (2012).
[22] Y. Nesterov and A. Nemirovski. [n. d.]. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics, 1994.
[23] B. Rochwerger, D. Breitgand, A. Epstein, D. Hadas, I. Loy, K. Nagin, J. Tordsson, C. Ragusa, M. Villari, S. Clayman, E. Levy, A. Maraschini, P. Massonet, H. Munoz, and G. Toffetti. 2011. Reservoir: When One Cloud is Not Enough. *Computer* 44, 3 (2011).
[24] N. Samaan. 2014. A Novel Economic Sharing Model in a Federation of Selfish Cloud Providers. *IEEE Transactions on Parallel and Distributed Systems* 25, 1 (2014).
[25] S.Boyd and L.Vanderberghe. [n. d.]. *Convex Optimization*. Cambridge University Press, 2005.
[26] J. S. Shamma and G. Arslan. 2005. Dynamic Fictitious Play, Dynamic Gradient Play, and Distributed Convergence to Nash Equilibria. *IEEE Trans. Automat. Control* 50, 3 (2005).
[27] G. Wang, M. Negrete, A. Kowli, E. Shafieepoorfard, S. Meyn, and U. V. Shanbhag. [n. d.]. *Dynamic Competitive Equilibria in Electricity Markets*. Springer-Verlag, 2011.
[28] Wikipedia. [n. d.]. *Business Cluster*. http://en.wikipedia.org.
[29] H. Zhuang, R. Rahman, and K. Aberer. 2014. Decentralizing the Cloud: How Can Small Data Centers Cooperate. In *IEEE International Conference on Peer-to-Peer Computing*.

5 APPENDIX

5.1 Static Market Analysis

In this section 5.1 we derive and analyze perfectly competitive 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 [12]), 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 [12]. 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 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 [16] 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:

$$\textbf{OPT:} \quad \max SW$$

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 above) 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 [25]. The advantage of using the primal-dual approach is that the dual optimization problem of the primal is always convex [25], 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:

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

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

$$\textbf{DOPT:} \quad \max \inf_{t:=\{vm^e, vm^c, vm^s, vm^r, vm^b, vm^{pc}, \rho\}} L,$$

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 [25] that are stated through equations 22a-22g. 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 (22a)-(22e) as follows.

$$\frac{d(c(vm_i^r))}{dvm_i^r}|vm_i^{r*} - \rho_i^* = 0,\ \forall i \in SC. \tag{22a}$$

$$\frac{d(c(vm_i^b))}{dvm_i^b}|vm_i^{b*} - \rho_i^* = 0,\ \forall i \in SC. \tag{22b}$$

$$\frac{d(c(vm_i^{pc}))}{dvm_i^{pc}}|vm_i^{pc*} - \rho_i^* = 0,\ \forall i \in SC. \tag{22c}$$

$$\rho_i^* - \frac{\partial(U_i(vm_i^e))}{\partial vm_i^e}|vm_i^{e*} = 0,\ \forall i \in C. \tag{22d}$$

$$\rho_i^* - \frac{\partial(U_i(vm_i^c))}{\partial vm_i^c}|vm_i^{c*} = 0,\ \forall i \in C. \tag{22e}$$

$$\rho_i^* - \frac{\partial(U_i(vm_i^s))}{\partial vm_i^s}|vm_i^{s*} = 0,\ \forall i \in C. \tag{22f}$$

$$\sum_{j\in C_i} vm_j^{ag}(1 - \kappa_j^1 - \kappa_j^2) = (vm_i^r + vm_i^b + vm_i^{pc}),\ \forall i \in SC. \tag{22g}$$

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 [19], market equilibria in a perfectly competitive autonomous setting of firms is known as *Walrasian equilibria.* It turns out from general equilibrium results in [19] 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 [19], (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 equilbria 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* [25]. 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 [22]. 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:

$$v(t+\epsilon) = v(t) - k_v \nabla_x L \cdot \epsilon. \tag{23a}$$

$$\rho(t+\epsilon) = \rho(t) + k_\rho \nabla_x L \cdot \epsilon. \tag{23b}$$

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

$$\tau_v \dot{v}(t) = -\nabla_v L, \tag{24a}$$

$$\tau_v \dot{\rho}(t) = -\nabla_\rho L, \tag{24b}$$

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 [25].

5.2 Theorem Proofs

Proof of Theorem 3.1. The equilibrium (x_1^*, x_2^*) when setting κ_j^1, κ_j^2 to zero, is a solution of the following.

$$\rho_i^* - \beta_i^r vm_i^{r*} - \alpha_i^r = 0, \ \forall i \in SDC. \tag{25a}$$

$$\rho_i^* - \beta_i^b vm_i^{b*} - \alpha_i^b = 0, \ \forall i \in SDC. \tag{25b}$$

$$\rho_i^* - \beta_i^{pc} vm_i^{pc*} - \alpha_i^{pc} = 0, \ \forall i \in SDC. \tag{25c}$$

$$\beta_i^{type} vm_i^{type} + \alpha_i^{type} - \rho_i^* = 0, \ \forall i \in C, \ type \in \{e, c, s, ag\}. \tag{25d}$$

$$\sum_{j \in C_i} vm_j^{ag}(1 - \kappa_j^1 - \kappa_j^2) = (vm_i^r + vm_i^b + vm_i^{pc}), \ \forall i \in SDC. \tag{25e}$$

Using Theorem 3.3 in [27], strong duality implies that equilibrium $(x_1^*, x_2^*$ exists is identical to the solution of the KKT conditions in 22a-22e. It can be seen that (25a) follows by replacing the cost function for SDCs in (2)-(4) in (22a). Similarly, (25b) follows by replacing the utility function of SDC customers in (5)-(8) in (22d). Furthermore (25c) is identical to (22e). Thus, $(x_1^*, x_2^*$ is identical to the equilibrium in (22a)-(22e). Thus, we proved Theorem 3.1. ■

Proof of Theorem 3.2. Since strong duality holds, it follows from Theorem 3.1 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

$$V(\bar{y_1}, y_2) \leq y_1^T(P_1A_1 + A_1^T P_1)y_1 + y_1^T P_1 A_2 y_2 + y_2 A_2^T P_1 y_1 \tag{26}$$

If A_1 is Hurwitz, for any $Q > 0$, there exists a positive definite matrix P_1 such that $P_1A_1 + A_1^T P_1 = -Q$. 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 $x_1, ..., x_n$, 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

$$VM^{max^T}[1]_{n\times n} P_2 y_2 \geq \lambda_{\min}(P_2)\psi_{\min}||y_2||_2, \tag{27}$$

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

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

Using (26) and (27), we obtain

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

For all $\Omega_{\max} \subsetneq D$, it follows that for all solutions beginning in Ω_{max}, $V \leq 0$. Hence, the equilibrium is stable and $\Omega_{\max}$ 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 3.2. ■

Proof of Theorem 3.3. Differentiating the Lyapunov function $V(y_1, y_2)$ along the trajectories of (13), we get

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

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 (21) it follows that $a_\Delta > 0$. Therefore, (25) implies that for all $\Omega_{c_{max}} \subsetneq D_\Delta$, for all solutions beginning in Ω_Δ, $\bar{V} \leq 0$. 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 3.3. ■