Mean Field Equilibria for Competitive Exploration in Resource Sharing Settings

Pu Yang,* Krishnamurthy Iyer,† and Peter Frazier‡

School of Operations Research and Information Engineering, Cornell University

Abstract

We consider a model of nomadic agents exploring and competing for time-varying location-specific resources, arising in crowdsourced transportation services, online communities, and in traditional location-based economic activity. This model comprises a group of agents, and a set of locations each endowed with a dynamic stochastic resource process. Each agent derives a periodic reward determined by the overall resource level at her location, and the number of other agents there. Each agent is strategic and free to move between locations, and at each time decides whether to stay at the same node or switch to another one. We study the equilibrium behavior of the agents as a function of dynamics of the stochastic resource process and the nature of the externality each agent imposes on others at the same location. In the asymptotic limit with the number of agents and locations increasing proportionally, we show that an equilibrium exists and has a threshold structure, where each agent decides to switch to a different location based only on their current location's resource level and the number of other agents at that location. This result provides insight into how system structure affects the agents' collective ability to explore their domain to find and effectively utilize resource-rich areas. It also allows assessing the impact of changing the reward structure through penalties or subsidies.

1 Introduction

We consider a model of nomadic agents exploring and competing for time-varying stochastic location-specific resources. Such multi-agent systems arise in crowdsourced transportation services like Uber and Lyft where drivers position themselves to be close to demand; in online communities like Twitch and Reddit where webizens choose in which subcommunities to participate; and in location-specific activity in the traditional economy, such as food trucks choosing where to position themselves, fisherman choosing where to fish, and pastoralists choosing where to graze their animals. In each of these examples, overall social welfare is determined both by agents' willingness to explore their domain to find and exploit resource-rich locations, the level of antagonism or synergy inherent to having multiple agents at the same location, and the equilibrium distribution of agents across locations that these factors induce.

The model we study comprises a set of locations and a group of agents. Each location has a resource level that varies randomly with time. Each agent periodically derives resource from the location at which she currently resides, whose amount is determined by the number of other agents currently residing there, and the location's current overall resource level. Based on these quantities, the agent then decides whether to stay at the same location or switch to another. The agents are fully strategic and seek to maximize their total rewards over their lifetime.

We study the equilibrium behavior of the agents in this system as a function of dynamics of the spatio-temporal resource process and the level of synergy or antagonism in the agents' sharing of resources. We analyze the system under the limit where the number of agents and locations both increase proportionally, using the methodology of a mean field equilibrium. We show that an equilibrium exists, and the agents'

---
*py75@cornell.edu

†kriyer@cornell.edu

‡pf98@cornell.edu

optimal strategy has a simple threshold structure, in which it is optimal to leave a location when the number of other agents exceeds a threshold that depends on the resource level at that location. In the limit as the system grows large, this induces a joint probability distribution over the number of agents and level of resource at each location.

Our results allow us to obtain economic insights into how the nature of the externality agents impose on others at the same location affects the exploration of the locations for resources, and consequently the overall welfare of the economy. In particular, our methodology allows us to analyze settings where the overall reward at a location either increases or decreases with the number of agents at the location, and how these two settings affect the equilibrium exploration. Furthermore, our methodology allows us to evaluate engineering interventions, such as providing subsidies to or imposing costs on agents to promote or discourage exploration to improve welfare.

Examples The model we study is a simplified version of systems appearing in real-world settings. It arises in the shared economy, in crowdsourced transportation services such as Uber and Lyft, in which drivers choose neighborhoods, and then earn money (reward) based on the number of riders requesting service within that neighborhood (the overall resource level), and the number of other drivers working there. This overall resource level varies stochastically over time in a neighborhood-specific way as demand rises and falls, and the resource derived by a driver decreases with the number of other drivers working in her neighborhood.

This model also arises in the internet economy, in online communities such as Reddit and Twitch, in which participants choose sub-communities, and then derive enjoyment depending both on some underlying but transitory societal interest in the sub-community's topic of focus (the overall resource) and the number of other participants in the sub-community. When the number of other participants is too small, lack of social interaction prevents enjoyment; when the number of other participants is too large, crowding diminishes the sense of community.

This model also arises in the traditional economy, for example in food trucks deciding in which neighborhoods to locate, in pastoralists deciding where to graze their livestock, and in fishermen deciding where to fish. In these examples, the level of resource derived by each agent from their location (whether profit from hungry passers by; or food for livestock provided by the range-land; or profit from the catch) depends both on the number of other agents at the location, and on the location's overall and stochastically varying resource level.

This model even arises among scientific researchers, who must choose a research area in which to work, and derive value from this choice based both on the underlying level of societal interest and funding in their chosen area, and in the number of other researchers working in it. As with online communities, the number of other researchers should be neither too large nor too small too maximize the value derived.

Related Work Our paper adds to the growing literature on mean field equilibrium [1, 16, 19, 21, 26], that studies complex systems under a large system limit and obtains insights about agent behavior that are hard to obtain from analyzing finite models. The main insight behind this line of literature, that in the large system limit, agents' behavior are characterized by their (private) state and an aggregate distribution of the rest of system, has been used to study settings that include industry dynamics and oligopoly models [15, 26, 25], repeated dynamics auctions [5, 18], online labor markets [2], and queueing systems [27].

Our model can be seen as an extension of the Kolkata Paise Restaurant Problem [7]. In this game, each agent chooses (simultaneously) a restaurant to visit, and earns a reward that depends both on the restaurant's rank, which is common across agents, and the number of other agents at that restaurant. This reward is inversely proportional to the number of agents visiting the restaurant.

The Kolkata Paise Restaurant Problem is itself a generalization of the El Farol bar problem [3, 8]. The Kolkata Paise Restaurant Problem is studied both in the one-shot and repeated settings, with results on the limiting behavior of myopic [7] and other strategies [12], although we are not aware of existing results on mean-field equilibria in this model. Our model is both more general, in that we allow general reward functions and allow location's resource to vary stochastically, and more specific, in that our locations are homogeneous. Our model also differs in that our agents' decisions are made asynchronously.

Our model is also related to congestion games [22, 24], in which agents choose paths on which to travel, and then incur costs that depend on the number of other agents that have chosen the same path. One may view paths as being synonymous with locations in our model, and observe that in both cases the utility/cost

derived from a path/location depends on the number of other agents using that path, or portion thereof. The main difference between our model and congestion games is the stochastic time-varying nature of our overall level of resource (making our model more complex), and the lack of interaction between locations contrasting with the interaction between paths (making our model simpler).

Our model has within it an exploration vs. exploitation tradeoff, in which an agent faces the decision of whether to stay at his current node, exploiting its resource and obtaining a known reward, or to leave and go to another randomly chosen location with unknown reward. Visiting this new location provides information upon which future decisions may be based. Similar tradeoffs between exploration and exploitation appear widely, and have been studied extensively in the single-agent setting [4, 13, 20, 23]. Exploration and learning in multi-agent settings has been considered by [9, 17].

2 Model

We consider a setting with N agents, each situated at each time $t \geq 0$ in one of K locations. Each location k has a stochastic dynamic resource process, denoted by $\{Z_t^{(k)} : t \geq 0\}$, that determines the reward obtained by each agent at that location, as we describe below. We assume that each process $\{Z_t^{(k)} : t \geq 0\}$ is a finite state continuous time Markov chain, that is distributed identically and independently from the rest of the system. For the purpose of analysis, we assume that each $Z_t^{(k)}$ takes values in $\{0,1\}$, with holding time at state $z \in \{0,1\}$ distributed as $\mathsf{Exp}(\mu_{z,1-z})$ for some fixed $\mu_{z,1-z} > 0$.

Agents may switch between locations to explore for resources. Formally, each agent i has an associated independent Poisson process X_t^i with rate $\lambda > 0$, at whose jump times $\{T_\ell^i : \ell \geq 0\}$ the agent makes the decision to either stay in her current location or switch instantaneously to a different location that is chosen uniformly at random. Let k_t^i denote the location of agent i at time $t \geq 0$ and let $N_t^{(k)}$ denote the number of agents at location k at time t.

The agents in the model are short-lived, and at each jump time T_ℓ^i, the agent i departs the system with probability $1 - \gamma > 0$. Thus, each agent i lives in the system for a random time τ^i which is distributed according to $\mathsf{Exp}(\lambda(1-\gamma))$. For each agent i that leaves the system, a new agent (with the same label i) arrives at a location chosen uniformly at random. We make this modeling assumption to ensure that the number of agents in the system is always positive; this assumption can be relaxed to allow for random arrivals and departures, with the arrival rate equal to the departure rate.

We now describe the decision problem faced by an agent i in more detail. At each jump time $t = T_\ell^i$, an agent i at location $k = k_t^i$ receives a reward $R_t^i = F(Z_t^{(k)}, N_t^{(k)})$ that depends on the state of the resource process $Z_t^{(k)}$ and the overall number of agents $N_t^{(k)}$ at the location k. In the following, we assume that the function F governing the reward at each location is given by $F(z,n) = zf(n)$ for some function $f : \mathbb{N} \to \mathbb{R}^+$ that is non-increasing in n, with $\lim_{n \to +\infty} f(n) = 0$. Essentially, this implies that the reward at a location at any time is zero if the resource process is in state 0, and it is equal to $f(n)$ if the resource process is in state 1, if there are n agents at that location. Furthermore, this reward $f(n)$ decreases to zero as the number of agents at a location increases. Given this setting, each agent i at any time prefers to be at a location with resource process state equal to 1, and where the number of other agents is small.

Within this setting, we will focus on three cases: (1) for each $n \in \mathbb{N}$, we have $nf(n) = 1$. In this case the (unit) resource, if available at a location, is shared equally among the agents at that location; (2) $nf(n)$ is non-decreasing in n. In this case adding agents to a location increases the total reward earned, either through synergy, or because a small number of agents cannot fully utilize a location's resource; and (3) $nf(n)$ is non-increasing in the number of agents n at that location. In this case antagonism or overutilization causes the total reward earned to decrease as agents are added.

Next, we discuss the information each agent i has while making their decision to stay or switch. We assume that an agent i has access to the states of the resource process $Z_t^{(k)}$ and the number of agents $N_t^{(k)}$ of a location k during the time she is present at the location k, i.e., when $k_t^i = k$. We further assume that agents have perfect recall, and hence, at a jump time $t = T_\ell^i$, each agent i bases her decision to switch or stay on the entire history h_t^i she has observed until time t, namely the resource process states and the number of

3

agents at each location she has visited during the time period she visited that location:

$$h_t^i = \left\{ (Z_s^{(k_s^i)}, N_s^{(k_s^i)}) : s \leq t \right\}.$$

Thus, a strategy ξ^i for an agent i, specifies a (mixed) action between stay and switch at each jump time $t = T_\ell^i$ of her associated Poisson process X_t^i, based on her history h_t^i.

Given this informational assumption, each agent i seeks to maximize the total expected reward accrued over her lifetime, given by

$$\mathbf{E} \left[\sum_{\ell=0}^{\infty} R_{T_\ell^i}^i \mathbf{I}\{\tau^i \geq T_\ell^i\} \right].$$

Observe that since the agent departs the system with probability $(1 - \gamma)$ at each jump time independently, the total expected reward can be equivalently written as

$$\mathbf{E} \left[\sum_{\ell=0}^{\infty} \gamma^\ell R_{T_\ell^i}^i \right].$$

Thus, each agent's decision problem is equivalent to maximizing her total discounted expected reward assuming she persists in the system.

Since the reward at any location is determined by the number of agents at that location, each agent's decision to stay in her current location or to switch to a new one depends on all the other agents' behavior. Consequently, the interaction among the agents is a dynamic game, and analyzing the agents' behavior requires an equilibrium analysis.

The standard equilibrium concept to analyze the induced dynamic game is a perfect Bayesian equilibrium (PBE). A PBE consists of a strategy ξ^i and a belief system μ^i for each player i. A belief system μ^i for agent i specifies a belief $\mu^i(h_t^i)$ after any history h_t^i over all aspects of the system that she is uncertain of and that influence her expected payoff. A PBE then requires two conditions to hold: (1) each agent i's strategy ξ^i is a best response after any history h_t^i, given their belief system and given all other agents' strategies; and (2) each agent i's beliefs $\mu^i(h_t^i)$ are updated via Bayes' rule whenever possible (see [10, 11] for more details).

Observe that a PBE supposes a complex model of agent behavior. It requires each player i to keep track of her entire history, and maintain complex beliefs about the rest of the system. While this may be plausible in small settings, this behavioral model seems implausible for large systems. On the contrary, in such settings, it is more plausible that each agent would base her decision to stay or switch solely on the current state of the location she is in — specifically on its level of resource, and the number of other agents there— and on the aggregate features of the entire system. Moreover, we expect that if an agent were to base her decision only on this information, then she would pursue a "threshold" strategy: she would stay in her current location if the number of agents at that location is low, and switch to a different location if that number is high, with the threshold used depending on that location's level of resource.

Below, we seek to uncover this intuitive behavioral model as an equilibrium in large systems by letting the number of agents and the number of location both increase proportionally to infinity, and studying the limiting infinite system.

3 Limiting infinite system

In this section, we consider an infinite system that is obtained as the limit of the finite system as the number of location K and the number of agents N both tend to infinity, with $N = \beta K$, for some fixed $\beta > 0$. In the limiting system, there are infinite number of locations and agents, with the expected number of agents per location fixed at $\beta > 0$. In such a limit, given certain consistency conditions that bind the mean dynamics of all the locations, the dynamics of each location essentially decouples from the rest of the system. Under such a decoupling, instead of focusing on the entire limiting system, it suffices to focus on the dynamics of a single location, as well as the empirical distribution of the states of all the locations. We begin with the description of the dynamics of a single location in such an infinite system.

3.1 Location dynamics

To analyze the agents' behavior in the infinite system, we fix a location k and focus on the decision problem faced by an agent i at location k about when to switch to a different location. Let X_t^i denote the Poisson process with rate $\lambda > 0$ associated with agent i, with jump times T_ℓ^i for $l \geq 0$. As before, let $Z_t^{(k)}$ denote the state of the resource process at location k and let $N_t^{(k)}$ denote the number of agents at location k at time t. We assume that agents arrive at location k according to a Poisson arrival process with rate $\kappa > 0$. Note that these arriving include new agents arriving to the system (following a departure), as well as existing agents who have chosen to switch from their current location.

Inspired by the discussion at the end of the preceding section, we focus on a family of threshold strategies for the agents. A threshold strategy is characterized by a pair $(n_0, n_1) \in [0, +\infty)^2$. In a threshold strategy (n_0, n_1), an agent at a location with resource level $z \in \{0, 1\}$ chooses to stay at her current location if the number of agents is strictly below $\lfloor n_z \rfloor$; chooses to switch her location if the number of agents at her current location is strictly above n_z; and stays with probability $n_z - \lfloor n_z \rfloor$ and switches with the remaining probability if the number of agents is equal to $\lfloor n_z \rfloor$. Our eventual goal is to show that there exists an equilibrium for agents' behavior where all agents follow (the same) threshold strategy. For now, we assume that all agents except agent i adopt a threshold strategy (n_0, n_1), and seek an optimal strategy over the class of all history-dependent strategies (not just the class of threshold strategies) for agent i.

Note that given the arrival rate $\kappa > 0$, and the threshold policy (n_0, n_1), the process $(Z_t^{(k)}, N_t^{(k)})$ evolves as a continuous time Markov chain on the state space $\mathcal{S} = \{0, 1\} \times \mathbb{N}$ with the following transition rate matrix: for each $z \in \{0, 1\}$, and for all $n \in \mathbb{N}$, we have

$$\mathsf{Q}((z, n) \to (1 - z, n)) = \mu_{z, 1-z},$$
$$\mathsf{Q}((z, n) \to (z, n + 1)) = \kappa,$$
$$\mathsf{Q}((z, n) \to (z, n - 1)) = \lambda(n - 1)\,(1 - \gamma +$$
$$\gamma\,(\mathbf{1}\{n > n_z\} + (n + 1 - n_z)\mathbf{1}\{n = \lfloor n_z \rfloor\}))\,.$$

Here, the first equation represents the transitions in $Z_t^{(k)}$, which is an independent Markov chain on $\{0, 1\}$ with holding times μ_{01} and μ_{10}. The second equation follows from the assumption that agents arrive at location k according to a Poisson process with rate $\kappa > 0$. The third equation represents a transition where an agent at location k leaves. This transition can occur in two ways: first, the agent could leave the system with probability $1 - \gamma$; second, the agent could survive, with probability γ, but choose to switch to a different location, which happens with probability 1 if $n > n_z$, with probability $(n + 1 - n_z)$ if $n = \lfloor n_z \rfloor$, and zero otherwise. Since there are $(n - 1)$ other agents that make this decision to stay or switch at rate λ, these transitions occur at rate $\lambda(n - 1)$. We denote this continuous time Markov chain describing the dynamics of a single location, where all agents adopt the threshold policy (n_0, n_1) and the rate of arrival of agents is κ, by $\mathcal{MC}(n_0, n_1, \kappa)$.

3.2 Agent's decision problem

We are now ready to describe the decision problem faced by the agent i regarding when to switch from her current location. At each jump time $t = T_\ell^i$ of X_t^i, the agent i receives an immediate payoff of $Z_t^{(k)} f(N_t^{(k)})$ and may leave the system with probability $1 - \gamma$. If she does not leave the system, then she has to decide between two actions "stay" or "switch". On choosing "stay" continues until the next jump time $T_{\ell+1}^i$; on choosing "switch", the decision problem terminates with an immediate payoff of $C > 0$, that does not depend on the state of the location k.

Before proceeding, we provide a brief interpretation of the termination payoff C. Observe that in a finite system, an agent on switching from a location, moves on to a different location that is chosen uniformly at random, and continues to accrue payoffs until she leaves the system. This suggests that one may interpret the termination payoff C as capturing the notion of a continuation payoff on switching in the finite system in the context of the limiting infinite system. Subsequently, we impose conditions on our equilibrium notion that ensure that indeed C denotes the continuation payoff in the infinite system.

Given these payoffs and actions for agent i, it follows that the decision problem facing agent i is an optimal stopping problem, which we denote by $\mathcal{OS}(n_0, n_1, \kappa, C)$. We next specify the dynamic programming formulation of $\mathcal{OS}(n_0, n_1, \kappa, C)$.

Note that in the decision problem, when the Markov chain $\mathcal{MC}(n_0, n_1, \kappa)$ is in state (z, n), events occur at rate $\kappa + \mu_{z,1-z} + \lambda n$: with rate κ a new agent arrives, with rate $\mu_{z,1-z}$ the resource level changes, with rate $\lambda(n-1)$ one of the other agents either leaves the system or survives and makes the decision to stay or switch, and finally with rate λ, agent i arrives at a jump time to make a decision herself. Thus, we define the following transition probabilities for the state transition:

$$\mathsf{P}_{dec}(z, n) = \frac{\lambda}{n\lambda + \mu_{z,1-z} + \kappa},$$
$$\mathsf{P}_{exit}(z, n) = \frac{(n-1)\lambda(1-\gamma)}{n\lambda + \mu_{z,1-z} + \kappa},$$
$$\mathsf{P}_{sur}(z, n) = \frac{(n-1)\lambda\gamma}{n\lambda + \mu_{z,1-z} + \kappa},$$
$$\mathsf{P}_{res}(z, n) = \frac{\mu_{z,1-z}}{n\lambda + \mu_{z,1-z} + \kappa},$$
$$\mathsf{P}_{arr}(z, n) = \frac{\kappa}{n\lambda + \mu_{z,1-z} + \kappa}. \tag{1}$$

Here, $\mathsf{P}_{dec}(z, n)$ denotes the probability that the next event corresponds to agent i's decision epoch, P_{exit} denotes the probability the next event corresponds to one of the other agents exiting the system, P_{sur} denotes the probability the next event corresponds to one of the other agents persisting in the system, P_{res} denotes the probability the next event corresponds to change in the resource level, and finally, P_{arr} denotes the probability that the next event corresponds to a new arrival.

Given $(Z_t^{(k)}, N_t^{(k)}) = (z, n)$ with $T_\ell^i = t$, let $V(z, n)$ denote the optimal expected total reward of agent i just after her associated Poisson process X_t^i has undergone a jump, but prior to her making a decision or receiving any reward. Similarly, let $\hat{V}(z, n)$ denote the optimal expected total reward of agent i after a jump time, conditional on the decision problem not terminating either due to the agent leaving the system or choosing to switch to a different location. Then, we have the following Bellman equation:

$$V(z, n) = zf(n) + \gamma \max\{\hat{V}(z, n), C\},$$
$$\hat{V}(z, n) = \mathsf{P}_{dec}(z, n)V(z, n) + \mathsf{P}_{exit}(z, n)\hat{V}(z, n-1)$$
$$+ \mathsf{P}_{sur}(z, n)\left(\mathbf{1}\{n > n_z\} + (n+1-n_z)\mathbf{1}\{n = \lfloor n_z \rfloor\}\right)\hat{V}(z, n-1)$$
$$+ \mathsf{P}_{sur}(z, n)\left(\mathbf{1}\{n < \lfloor n_z \rfloor\} + (n_z-n)\mathbf{1}\{n = \lfloor n_z \rfloor\}\right)\hat{V}(z, n)$$
$$+ \mathsf{P}_{res}(z, n)\hat{V}(1-z, n) + \mathsf{P}_{arr}(z, n)\hat{V}(z, n+1). \tag{2}$$

Here, the first equation follows from the fact that subsequent to a jump time, the agent receives an immediate payoff equal to $zf(n)$. Following this, she continues with survival probability γ, and has to make a decision to stay, which gets her expected payoff equal to $\hat{V}(z, n)$ or switch, which gets her an expected payoff equal to C. The second equation relates $\hat{V}(z, n)$ to the agent's expected payoff subsequent to various events that can occur at the next transition. For a solution V and $\hat{V}$ to the Bellman's equation, an optimal strategy ξ^i for the agent i requires agent i to stay if $\hat{V}(z, n) > C$, and to switch if $\hat{V}(z, n) < C$ (any mixed action is optimal if $\hat{V}(z, n) = C$). Let $\mathcal{OPT}(n_0, n_1, \kappa, C)$ denote the set of all optimal strategies (not necessarily threshold strategies) for the agent's decision problem $\mathcal{OS}(n_0, n_1, \kappa, C)$.

3.3 Mean field equilibrium

Given the Markov chain $\mathcal{MC}(n_0, n_1, \kappa)$ and an agent's decision problem $\mathcal{OS}(n_0, n_1, \kappa, C)$, we are now ready to state the equilibrium conditions on the limiting system. First, in the infinite system, we require the agents' strategies to be in equilibrium. Since $\mathcal{MC}(n_0, n_1, \kappa)$ describes the dynamics of a location where all agents other than agent i use the threshold policy (n_0, n_1), for equilibrium we must impose the condition

that (n_0, n_1) is an optimal strategy for the agent's decision problem. This leads to the following condition:

$$(n_0, n_1) \in \mathcal{OPT}(n_0, n_1, \kappa, C). \tag{3}$$

If all agents at location k, including agent i, follow the threshold policy (n_0, n_1), then the transitions in $(Z_t^{(k)}, N_t^{(k)}) = (z, n)$ follow a Markov chain with transition rate matrix $\overline{\mathsf{Q}}$ that is equal to Q except for the transition $\overline{\mathsf{Q}}((z, n) \to (z, n-1))$ which is equal to $\overline{\mathsf{Q}}((z, n) \to (z, n-1)) = \lambda n(1 - \gamma + \gamma(\mathbf{1}\{n > n_z\} + (n + 1 - n_z)\mathbf{1}\{n = \lfloor n_z \rfloor\}))$. This is because the arrival and the changes in the resource level occur at the same rate, but now any one of the agents at location k might choose to leave the location, as opposed to any one of the agents other than agent i as defined in Q. Denote this Markov chain by $\overline{\mathcal{MC}}(n_0, n_1, \kappa)$ and let π denote an invariant distribution of this chain:

$$\pi^{\mathsf{T}}\overline{\mathsf{Q}} = 0. \tag{4}$$

In a large system, a natural requirement to impose is that the invariant distribution of a single location k equals the steady state empirical distribution of the resource level and the number of agents across all locations. Requiring this condition to hold leads to two consequences. First, because in the infinite system the "agent density", i.e., mean number of agent across all locations, is equal to β, this implies that the expected number of agents at location k must equal β:

$$\sum_{(z,n) \in \mathcal{S}} n\pi(z, n) = \beta. \tag{5}$$

Note that this equation imposes a restriction on the arrival rate κ of the Markov chain $\mathcal{MC}(n_0, n_1, \kappa)$. In particular, it requires the arrival rate to be such that in steady state the expected number of agents at each location is equal to β.

The second condition imposes a restriction on the immediate termination reward on switching C. Recall that we interpret C as modeling the optimal continuation payoff on switching in the finite system. Since the empirical distribution of the states of other locations is given by π, the optimal expected reward an agent can obtain on switching is given by $\sum_{(z,n) \in \mathcal{S}} \pi(z, n)\hat{V}(z, n + 1)$. This is because, after the agent moves to a location in state (z, n), which happens with probability $\pi(z, n)$, the number of agents at that location becomes $n + 1$, and the expected payoff to that agent is $\hat{V}(z, n + 1)$. We require that this quantity equals the immediate reward C:

$$C = \sum_{(z,n) \in \mathcal{S}} \pi(z, n)\hat{V}(z, n + 1). \tag{6}$$

Given these consistency conditions, we are now ready to define a mean field equilibrium for the infinite system:

Definition 1 (Mean Field Equilibrium). *A mean field equilibrium is characterized by a threshold strategy (n_0, n_1), an arrival rate $\kappa > 0$, an distribution π over $\mathcal{S}$, and an immediate reward $C > 0$, such that the set of equations* (3),(4),(5), *and* (6) *hold.*

Note that as opposed to a PBE, a mean field equilibrium adopts a fairly natural and simple model of agent behavior, where each agent needs to keep track only of current state and the number of agents at the location she is in, along with the immediate payoff for switching.

4   Main results

Having defined the equilibrium concept, we now consider the problem of existence of a mean field equilibrium in the infinite system.

We begin with the following lemma that shows that for any level of resource at a location, the value function $\hat{V}(z, n)$ is non-increasing with the number of agents at that location. The proof may be found in Appendix A.

Lemma 4.1. *For each $z \in \{0, 1\}$, the value function $\hat{V}(z, n)$ is non-increasing in n.*

Using this lemma, it is straightforward to show that for any level of resource $z \in \{0, 1\}$, if it is optimal to switch when the number of agents at the location is n, then it is still optimal to switch when the number of agents is greater than n. From this, we obtain the first result of this section.

Theorem 4.1. *For any $(n_0, n_1) \in \mathbb{R}_+^2$, $\kappa > 0$ and $C > 0$, there always exists an optimal strategy with a threshold structure for the decision problem $\mathcal{OS}(n_0, n_1, \kappa, C)$.*

The preceding theorem suggests that set of threshold strategies is closed under best-responses: if all agents other than agent i adopt the same threshold strategy, then it is optimal for agent i to also follow a threshold strategy. Thus, it suffices to focus on the set of optimal threshold strategies for the agent decision problem, which we denote by $\mathcal{T}(n_0, n_1, \kappa, C)$. Note that $\mathcal{T}(n_0, n_1, \kappa, C)$ can be characterized as a subset of $\mathbb{R}_+^2$ corresponding to the values of the thresholds of the optimal threshold strategies. In Lemma A.1 in the Appendix A, we show that $\mathcal{T}(n_0, n_1, \kappa, C)$ is a convex set.

Building on this result, we obtain the main theorem of our paper.

Theorem 4.2. *For any $\lambda > 0$, $\beta > 0$ and $\mu_{01}, \mu_{10} > 0$, there exists a mean field equilibrium for the infinite system.*

We emphasize that the existence of a mean field equilibrium is obtained under very general conditions, requiring only that the reward function $f(n)$ is non-increasing in the number of agents n at a location, and converging to zero as n tends to infinity. Our proof technique can be extended to cases where the resource level $Z_t^{(k)}$ at any location can take values in any finite subset of $\mathbb{R}_+$.

The full proof of Theorem 4.2 is technical and is omitted due to space constraints. Instead, in the next section, we sketch the main ideas behind the proof and provide a brief outline. Selected portions of the proof may be found in Appendix A, and a full proof will appear later in a longer version of the paper.

5   Proof outline

The proof of Theorem 4.2 follows by applying Kakutani's fixed point theorem on carefully defined map $\mathscr{T}$, whose fixed points correspond to the mean field equilibria of the infinite system. To define the map $\mathscr{T}$ requires a number of intermediate steps, which we outline below.

The first step of the proof involves showing that given any $(n_0, n_1) \in \mathbb{R}_+^2$ and $C > 0$, there exists a unique κ and distribution π such that π is the unique invariant distribution of the Markov chain $\overline{\mathcal{MC}}(n_0, n_1, \kappa)$ (i.e., π satisfies equation (4)), and for which condition (5) holds. This step itself involves first showing that for any $\kappa > 0$, the Markov chain $\overline{\mathcal{MC}}(n_0, n_1, \kappa)$ is irreducible and positive recurrent, and hence has a unique stationary distribution π_κ. To show this, we use coupling arguments to bound the Markov chain $\overline{\mathcal{MC}}(n_0, n_1, \kappa)$ between two $M/M/\infty$ queues. Then, we show that the quantity $\sum_{(z,n) \in \mathcal{S}} n \pi_\kappa(z, n)$ is strictly increasing and continuous over an interval of values of κ, which suffices to show that there exists a value of κ satisfying (5).

We then compute the value function $\hat{V}$ satisfying the Bellman's equation (2) for the decision problem $\mathcal{OS}(n_0, n_1, \kappa, C)$, where κ is the value of the arrival rate obtained in the first step. Using this value function, we identify the set $\mathcal{T}(n_0, n_1, \kappa, C)$ of optimal threshold strategies.

Finally, using the value function $\hat{V}$ and the invariant distribution π from the first step, we compute the total expected payoff of an agent subsequent to switching to a different location chosen uniformly at random, defined as $\tilde{C} \triangleq \sum_{(z,n) \in \mathcal{S}} \pi(z, n) \hat{V}(z, n + 1)$.

The map $\mathscr{T}$ is then defined as follows: for each $(n_0, n_1) \in \mathbb{R}_+^2$ and $C > 0$, we define $\mathscr{T}(n_0, n_1, C) = \mathcal{T}(n_0, n_1, \kappa, C) \times \{\tilde{C}\}$. We depict the map pictorially in Fig. 1.

By definition, any fixed point (n_0, n_1, C) of the map $\mathscr{T}$ must satisfy $(n_0, n_1) \in \mathcal{T}(n_0, n_1, \kappa, C)$, and $C = \tilde{C}$. This implies that (n_0, n_1) is an optimal threshold strategy for the decision problem $\mathcal{OS}(n_0, n_1, \kappa, C)$, and hence is also an optimal strategy and equation (3) holds. Recall that the arrival rate κ and the invariant distribution π satisfy equations (4) and (5). Finally, $\tilde{C} = C$ implies that equation (6) holds. From this we conclude that (n_0, n_1), C and resulting κ and the invariant distribution π together constitute a mean field equilibrium. Thus each fixed point of the map $\mathscr{T}$ corresponds to a mean field equilibrium.
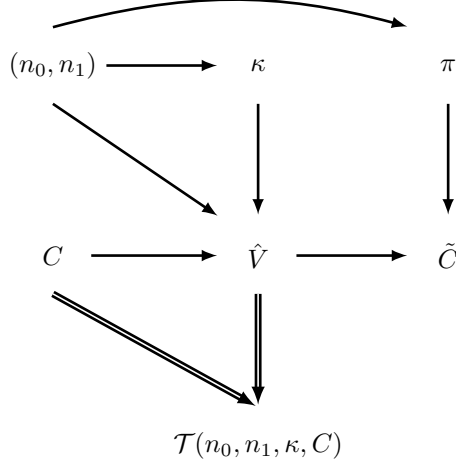
Figure 1: Illustration of the correspondence $\mathscr{T}(n_0, n_1, C) = \mathcal{T}(n_0, n_1, \kappa, C) \times \{\tilde{C}\}$.

To show that the map $\mathscr{T}$ has a fixed point, we apply Kakutani's fixed point theorem. For this, we first identify a compact set $\mathcal{X} \triangleq [0, M]^2 \times [\underline{C}, \bar{C}]$ of $\mathbb{R}_0^3$, with $0 < \underline{C} < \bar{C}$ and $M > 0$, and show that $\mathscr{T}(\mathcal{X}) \subseteq \mathcal{X}$. We include the proof of this step in Appendix A.2. Second, we show that the map is upper hemicontinuous under the Euclidean topology. To show this, we need to show that each of the intermediate maps in Fig. 1 is continuous (or upper hemicontinuous if the map is a correspondence) under appropriate topologies. We apply Berge's Maximum Theorem [6] to show the map from (n_0, n_1, C) to (κ, π) is continuous, and the continuity of fixed points of continuous contraction mappings [14] to show the map from (n_0, n_1, κ, C) to the value function $\hat{V}$ is continuous. Finally, we show that the image of $\mathscr{T}$ is convex and non-empty for all values of (n_0, n_1, C). Then, it follows by a direct application of Kakutani's fixed point theorem that the map $\mathscr{T}$ has a fixed point, and consequently a mean field equilibrium exist in the infinite system.

6   Comparative statics

Having shown the existence of a mean field equilibrium for the infinite system, we now study how the model parameters and the reward function f affect the equilibrium agent behavior. We perform this investigation numerically by computing the mean field equilibrium for a range of parameter values, and studying how the equilibrium thresholds (n_0, n_1) and the stationary distribution vary.

As our model is invariant to the proportional scaling of the decision epoch rates λ, and the holding rates $\mu_{0,1}$ and $\mu_{1,0}$, we assume for our computations that $\lambda = 1$ and vary the holding rates. Further, we restrict our attention to the symmetric case where the holding rates are equal: $\mu_{0,1} = \mu_{1,0} = \mu$. In all our computation, we set the agent density $\beta = 20$, and the survival probability $\gamma = 0.95$. For this, we study the mean field equilibrium in three reward settings: (1) $f(n) = 1/n$ for all n; (2) $f(n) = 1/n^2$ for all n; and (3) $f(n) = 1/\sqrt{n}$ for all n.

Before we discuss the results of our numerical investigation, we briefly describe our approach to compute an (approximate) mean field equilibrium of the infinite system.

6.1   Computation of MFE

Recall that the mean field equilibria of our model are the fixed points of the correspondence $\mathscr{T}$. We thus seek to find (approximate) fixed points of this map. To do this, we adopt a brute-force approach. We first truncate the state space $\mathcal{S}$ of the agent decision problem to $\mathcal{S}_{200} = \{0, 1\} \times \{0, 1, \cdots, 199\}$. We restrict the thresholds (n_0, n_1) to grid of values in $[0, 50]^2$, where we set the grid resolution r adaptively over different runs. We do a similar adaptive meshing of the set of values of the immediate payoff C. Having restricted the set of values of (n_0, n_1) and C thus, we search over all values to find lie in the image of $\mathscr{T}$, within some pre-specified tolerance ϵ. We describe this process in detail:

9

μ	$f(n) = 1/\sqrt{n}$	$f(n) = 1/n$	$f(n) = 1/n^2$
0.1	2.80, (1.0, 43.9)	1.98, (1.0, 4.0)	0.14, (1.0, 7.7)
0.5	2.39, (4.3, 43.8)	1.72, (1.0, 4.0)	0.25, (1.0, 4.7)
1.0	2.63 , (1.0, 27.2)	0.96, (1.0, 10.4)	0.18, (1.0, 5.0)
10.0	2.37, (11.0, 18.2)	0.93, (4.0, 7.1)	0.16, (3.0, 5.0)
100.0	2.46, (11.0, 12.0)	0.97, (4.0, 4.0)	0.80, (1.0, 8.0)

Table 1: Computationally determined approximate equilibrium values of the payoff for switching C, and the thresholds n_0 and n_1 used to decide whether to switch or not. Values are reported in the table as $C, (n_0, n_1)$, for each value for reward function f and the rate μ at which the resource level changes.

1. For each value of (n_0, n_1), we perform a binary search on κ to find a value for which the stationary distribution π of the Markov chain $\mathcal{MC}(n_0, n_1, \kappa)$ restricted to $\mathcal{S}_{200}$ satisfies equation (5) with a tolerance ϵ. (The stationary distribution π is obtained by solving the set of linear equations (4).)

2. For this value of κ, (n_0, n_1) and each value of C, we perform value iteration to compute the value function $\hat{V}$, again with a tolerance of ϵ, and compute the set of optimal thresholds $\mathcal{T}(n_0, n_1, \kappa, C)$. Let $\mathsf{dist}(n_0, n_1)$ denote the distance between (n_0, n_1) and the $\mathcal{T}(n_0, n_1, \kappa, C)$ under the Euclidean norm. (Note that the latter set is convex and the distance is well defined.)

3. Next, using the stationary distribution π and the value function $\hat{V}$, we compute the immediate payoff $\tilde{C}$.

4. For each value of n_0, n_1 and C, we compute $d(n_0, n_1, C) \triangleq \|C - \tilde{C}\| + \mathsf{dist}(n_0, n_1)$. We output the value of (n_0, n_1, C) that minimizes $d(n_0, n_1, C)$ over all values.

To make the brute-force search efficient, we run this algorithm sequentially and adaptively by first identifying candidate regions where equilibria might exist, and restricting the search to those regions with lower tolerance and finer grid values.

6.2 Numerical results

In Table 1 we report computationally determined values for three different reward functions, obtained over five different rates for the underlying resource process.

For large values of μ, we see n_0 and n_1 are close. This is natural because large values of μ imply that the resource level is changing very quickly, and so the level of resource at the time of the decision has little impact on what the resource level will be at the next time the agent receives a reward. Thus, the current level of resource z has little impact on the threshold n_z. On the other hand, for small values of μ, the thresholds differ significantly with the resource level.

We also observe that, when comparing reward functions $f(n) = 1/\sqrt{n}$, $1/n$, and $1/n^2$, when f decreases more quickly, agents are more willing to switch (the threshold for switching is lower), and the payoff for switching is also lower.

7 Conclusion

We have studied a multi-agent location-specific resource-sharing game, and have established the existence of an equilibrium in this game, and have characterized each agent's policy in this equilibrium as being a threshold policy. This result provides economic insight into such multi-agent resource-sharing games, and also allows evaluating the effects of designing and modifying these games, through subsidies or penalties added to natural occurring rewards and costs.

This work also sets the stage for studying information sharing in multi-agent resource-sharing systems. Our current analysis assumes that agents observe only the number of other agents and resource level at their current location, and the locations they have visited in the past. One may extend this model to allow for information sharing among the agents as well as between the agents and a central planner who has access to

the current state(s) of the location(s) that the agent mights switch to. A first question of interest would then be whether such information sharing necessarily improves social welfare, or whether it can in fact degrade it. A second question is how this information sharing mechanism should be designed to maximize social welfare.

References

[1] S. Adlakha, R. Johari, and G. Y. Weintraub. Equilibria of dynamic games with many players: Existence, approximation, and market structure. *Journal of Economic Theory*, 156:269–316, March 2015.

[2] N. Arnosti, R. Johari, and Y. Kanoria. Managing congestion in decentralized matching markets. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, pages 451–451, New York, NY, USA, 2014. ACM.

[3] W. B. Arthur. Inductive reasoning and bounded rationality. *The American economic review*, pages 406–411, 1994.

[4] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

[5] S. R. Balseiro, O. Besbes, and G. Y. Weintraub. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science*, 61(4):864–884, 2015.

[6] C. Berge. *Topological Spaces: Including a treatment of multi-valued functions, vector spaces, and convexity, translated by E. M. Patterson*. Dover, 1963.

[7] A. S. Chakrabarti, B. K. Chakrabarti, A. Chatterjee, and M. Mitra. The kolkata paise restaurant problem and resource utilization. *Physica A: Statistical Mechanics and its Applications*, 388(12):2420–2426, 2009.

[8] B. K. Chakrabarti. Kolkata restaurant problem as a generalised el farol bar problem. In *Econophysics of Markets and Business Networks*, pages 239–246. Springer, 2007.

[9] P. I. Frazier, D. Kempe, J. Kleinberg, and R. Kleinberg. Incentivizing exploration. In *Proceedings of the 15th ACM Conference on Economics and Computation (EC'14)*, pages 5–22, New York, NY, USA, 2014. Association for Computing Machinery.

[10] D. Fudenberg and J. Tirole. *Game theory*. The MIT Press, Cambridge, Massachusetts, 1991.

[11] D. Fudenberg and J. Tirole. Perfect bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, 53(2):236–260, 1991.

[12] A. Ghosh, A. Chatterjee, M. Mitra, and B. K. Chakrabarti. Statistics of the kolkata paise restaurant problem. *New Journal of Physics*, 12(7):075033, 2010.

[13] J. Gittins. *Multi-Armed Bandit Allocation Indices*. John Wiley and Sons, New York, 1989.

[14] A. Granas and J. Dugundji. *Fixed point theory*. Springer Monographs in Mathematics. Springer-Verlag, New York, 2003.

[15] H. A. Hopenhayn. Entry, exit and firm dynamics in long run equilibrium. *Econometrica*, 60(5):1127 – 1150, 1992.

[16] M. Huang, P. E. Caines, and R. P. Malhamé. Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ϵ-Nash equilibria. *IEEE Transactions on Automatic Control*, 52(9):1560–1571, 2007.

[17] J. R. Ilan Lobel, Ankur Mani. Learning via external sales networks. in preparation, 2014.

[18] K. Iyer, R. Johari, and M. Sundararajan. Mean field equilibria of dynamic auctions with learning. *Management Science*, 60(12):2949–2970, 2014.

[19] B. Jovanovic and R. W. Rosenthal. Anonymous sequential games. *Journal of Mathematical Economics*, 17:77–87, 1988.

[20] L. Kaelbling. *Learning in embedded systems*. MIT Press, Cambridge, MA, 1993.

[21] J. M. Lasry and P. L. Lions. Mean field games. *Japanese Journal of Mathematics*, 2:229–260, 2007.

[22] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic game theory*. Cambridge University Press Cambridge, 2007.

[23] W. Powell and I. Ryzhov. *Optimal Learning*. Wiley, 2012.

[24] R. W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

[25] G. Y. Weintraub, C. L. Benkard, and B. van Roy. Industry dynamics: Foundations for models with an infinite number of firms. *Journal of Economic Theory*, 146(5):1965 – 1994, 2011.

[26] G. Y. Weintraub, C. L. Benkard, and B. VanRoy. Markov perfect industry dynamics with many firms. *Econometrica*, 76(6):1375–1411, 2008.

[27] J. Xu, B. Hajek, et al. The supermarket game. *Stochastic Systems*, 3(2):405–441, 2013.

Appendix A Proofs

In this appendix we provide proofs of selected results discussed in the main text. A full proof of the main theorem is technical, and is omitted due to space constraints. This full proof will appear in a future version of the paper.

A.1 Structure of optimal strategies

of Lemma 4.1. By Markov property we may assume at $t = 0$ the agent makes his current decision, and let t' be the time his next decision epoch starts. We denote $\nu_{(z,n)}$ as the distribution of $(Z_{t'}, N_{t'})$ given $(Z_0, N_0) = (z, n)$, then we can write

$$\hat{V}(z,n) = \mathbf{E}_{\nu_{(z,n)}} V(Z_{t'}, N_{t'}).$$

Let $\hat{V}^{(0)}(z,n) = 0$ for all z and n. Compute $V^{(m)}(z,n)$ and $\hat{V}^{(m+1)}(z,n)$ using value iteration for the Bellman's equation (2).

$$V^{(m)}(z,n) = zf(n) + \gamma \max\{\hat{V}^{(m)}(z,n), C\},$$
$$\hat{V}^{(m+1)}(z,n) = \mathbf{E}_{\nu_{(z,n)}} V^{(m)}(Z_{t'}, N_{t'}).$$

By convergence of value iteration we have $\|V^{(m)} - V\|_\infty \to 0$ and $\|\hat{V}^{(m)} - \hat{V}\|_\infty \to 0$ as $m \to \infty$. Therefore, it suffices to show $\forall m \in \mathbb{N}_0$, for $z \in \{0,1\}$, $V^{(m)}(z,n)$ and $\hat{V}^{(m)}(z,n)$ are non-increasing in n. We can prove this by induction on m.

The base case follows trivially. Now assume $\hat{V}^{(m)}(z,n)$ is non-increasing in n for $z \in \{0,1\}$, for some $m \in \mathbb{N}_0$. From this it is straightforward to conclude that $V^{(m)}(z,n) = zf(n) + \gamma \max\{\hat{V}^{(m)}(z,n), C\}$ must be non-increasing in n since both $f(n)$ and $\hat{V}^{(m)}(z,n)$ are non-increasing in n. Thus, we only need to show that $\hat{V}^{(m+1)}(z,n)$ is also non-increasing in n.

Observe showing $\hat{V}^{(m+1)}(z,n) \geq \hat{V}^{(m+1)}(z, n+1)$ is equivalent to showing $\mathbf{E}_{\nu_{(z,n)}} V^{(m)}(Z_{t'}, N_{t'}) \geq \mathbf{E}_{\nu_{(z,n+1)}} V^{(m)}(Z_{t'}, N_{t'})$. We show this using a sample path argument by considering two processes. Let (Z_t^1, N_t^1) be a copy of $\mathcal{MC}(n_0, n_1, \kappa)$ that starts at $(Z_0^1, N_0^1) = (z, n)$ and let (Z_t^2, N_t^2) be a copy of $\mathcal{MC}(n_0, n_1, \kappa)$ that starts at $(Z_0^2, N_0^2) = (z, n+1)$. By carefully coupling the two processes, the details of which we omit due to space restrictions, it can be shown that for all $t \geq 0$, $Z_t^1 = Z_t^2$ and $N_t^1 \leq N_t^2$. Since $(Z_{t'}^1, N_{t'}^1) \sim \nu_{(z,n)}$ and $(Z_{t'}^2, N_{t'}^2) \sim \nu_{(z,n+1)}$, we have

$$\mathbf{E}_{\nu_{(z,n)}} V^{(m)}(Z_{t'}, N_{t'}) = \mathbf{E}[V^{(m)}(Z_t^1, N_t^1)],$$

and

$$\mathbf{E}_{\nu_{(z,n+1)}} V^{(m)}(Z_{t'}, N_{t'}) = \mathbf{E}[V^{(m)}(Z_t^2, N_t^2)].$$

Since $V^{(m)}(z,n)$ is non-increasing in n for both $z = 0, 1$, $V^{(m)}(Z_t^2, N_t^2) \leq V^{(m)}(Z_t^1, N_t^1)$ in all sample paths, and therefore $\mathbf{E}[V^{(m)}(Z_t^1, N_t^1)] \geq \mathbf{E}[V^{(m)}(Z_t^2, N_t^2)]$, which completes the proof. □

We now consider the set of optimal thresholds $\mathcal{T}(n_0, n_1, \kappa, C)$.

Lemma A.1. *For each (n_0, n_1), $\kappa > 0$, and $C > 0$, the set of optimal thresholds $\mathcal{T}(n_0, n_1, \kappa, C)$, as a subset of $\mathbb{R}_+^2$, is convex.*

Proof. From Lemma 4.1, we know that the value function $\hat{V}(z,n)$ is non-increasing in n for each $z \in \{0,1\}$. Now for $z \in \{0,1\}$, denote $\underline{n}_z = \max\{n : \hat{V}(z,n) > C\}$ and $\bar{n}_z = \min\{n : \hat{V}(z,n) < C\}$. For all $n \leq \underline{n}_z$, the optimal action is to stay at the current location and for all $n \geq \bar{n}_z$, the optimal action is to switch to a different location. For any integer $n \in [\underline{n}_z, \bar{n}_z]$, we have $\hat{V}(z,n) = C$, meaning the agent is indifferent between staying in the current location or switching to a different location when the state at the current location is (z, n).

Let (n_0^1, n_1^1) and (n_0^2, n_1^2) be two threshold strategies that are both optimal: $(n_0^\ell, n_1^\ell) \in \mathcal{T}(n_0, n_1, \kappa, C)$ for $\ell \in \{1, 2\}$. By the definition of the threshold strategy, this implies that $\lfloor n_z^\ell \rfloor \geq \underline{n}_z$, and $\lceil n_z^\ell \rceil \leq \bar{n}_z$ for each

ℓ. This is because in the threshold strategy (n_0^ℓ, n_1^ℓ), at state (z, n), the agent stays at her current location for all $n < \lfloor n_z^\ell \rfloor$, switches to a different location for all $n > n_z^\ell$ and and stays with probability $n_z^\ell - \lfloor n_z^\ell \rfloor$ and switches with the remaining probability if the number of agents n is equal to $\lfloor n_z^\ell \rfloor$. Since this is true for each ℓ, we have for any $\alpha \in (0, 1)$, $\lfloor \alpha n_z^1 + (1 - \alpha) n_z^2 \rfloor \geq \underline{n}_z$, and $\lceil \alpha n_z^1 + (1 - \alpha) n_z^2 \rceil \leq \bar{n}_z$. This implies that a threshold strategy that at state (z, n) stays in the current location if $n < \lfloor \alpha n_z^1 + (1 - \alpha) n_z^2 \rfloor$, switches to a different location if $n > \lfloor \alpha n_z^1 + (1 - \alpha) n_z^2 \rfloor$, and stays with probability $(n + 1 - \lfloor \alpha n_z^1 + (1 - \alpha) n_z^2 \rfloor)$ and switches otherwise if $n = \lfloor \alpha n_z^1 + (1 - \alpha) n_z^2 \rfloor$ is also optimal. This implies that $(\alpha n_0^1 + (1 - \alpha) n_0^2, \alpha n_1^1 + (1 - \alpha) n_1^2)$ also lies in the set $\mathcal{T}(n_0, n_1, \kappa, C)$, and hence the latter set is convex. □

A.2 Restriction of $\mathcal{T}$ to a compact set

In this section, we show that the map $\mathcal{T}$ can be restricted to a compact subset $\mathcal{X}$ of $\mathbb{R}_+^3$, whose image $\mathcal{T}(\mathcal{X})$ is a subset of $\mathcal{X}$.

Towards that goal, we define $\bar{C}$, $\underline{C}$ as follows:

$$\bar{C} \triangleq \frac{f(1)}{1 - \gamma},$$

$$\underline{C} \triangleq \frac{\lambda}{\lambda + \mu_{1,0} + \beta\lambda} \frac{\mu_{0,1}}{\lambda + \mu_{0,1} + \beta\lambda} e^{-\frac{\beta}{1-\gamma}} f(1).$$

It is straightforward that $0 < \underline{C} < f(1) < \bar{C}$. Also, for $n \in \mathbb{N}_0$, define

$$g(n) \triangleq \frac{1}{1 - \gamma} \left[f(\frac{\sqrt{n}}{2}) + \exp(-\frac{\sqrt{n}}{8}) + \frac{2}{\sqrt{\log n}} \right]$$
$$+ \frac{\gamma^{\lfloor (\log n)^{1/2} \rfloor}}{1 - \gamma} f(1). \tag{7}$$

Note $g(n)$ is decreasing in n and $g(n) \to 0$ as $n \to +\infty$. We pick M such that

$$M \triangleq \min\{n : g(n) < (1 - \gamma)\underline{C}\}.$$

Define $\mathcal{X} = [0, M]^2 \times [\underline{C}, \bar{C}]$. The following theorem shows that the image of map $\mathcal{T}$ is a subset of $\mathcal{X}$ for all values therein.

Theorem A.1. *For all $(n_0, n_1, C) \in \mathcal{X}$, $\mathcal{T}(n_0, n_1, C) \subseteq \mathcal{X}$.*

We prove this theorem in two steps. First, in Lemma A.2, we show that for all values of (n_0, n_1, C) in $\mathcal{X}$, the value of $\tilde{C}$ lies in $[\underline{C}, \bar{C}]$. Then, we show in Lemma A.3, that for all values of (n_0, n_1, C) in $\mathcal{X}$, the optimal thresholds must always be less than M. We begin with the first lemma.

Lemma A.2. *For any $(n_0, n_1, C) \in [0, +\infty)^2 \times [\underline{C}, \bar{C}]$ and $\kappa \in [\beta\lambda(1 - \gamma), \beta\lambda]$, let $\hat{V}$ be the solution of the Bellman equation (2) with given parameters (n_0, n_1, C, κ). Let π be the unique stationary distribution of $\overline{\mathcal{MC}}(n_0, n_1, \kappa)$. Let $\tilde{C} \triangleq \sum_{(z,n) \in \mathcal{S}} \pi(z, n) \hat{V}(z, n + 1)$. Then $\tilde{C} \in [\underline{C}, \bar{C}]$.*

Proof. We first show $\tilde{C} \leq \bar{C}$. We have

$$\tilde{C} = \sum_{(z,n) \in \mathcal{S}} \pi(z, n) \hat{V}(z, n + 1)$$
$$\leq \max_{(z,n) \in \mathcal{S}} \hat{V}(z, n + 1)$$
$$= \max_{z \in \{0,1\}} \hat{V}(z, 1), \tag{8}$$

where the last equality is implied by Lemma 4.1. Also note $\mathsf{P}_{exit}(z, 1) = \mathsf{P}_{sur}(z, 1) = 0$ for $z \in \{0, 1\}$, therefore

$$\hat{V}(z, 1) = \mathsf{P}_{dec}(z, 1)V(z, 1) + \mathsf{P}_{res}(z, 1)\hat{V}(1 - z, 1)$$
$$+ \mathsf{P}_{arr}(z, 1)\hat{V}(z, 2), \quad \text{for } z \in \{0, 1\}. \tag{9}$$

14

From Lemma 4.1 we have $\hat{V}(z,2) \le \hat{V}(z,1)$ for $z \in \{0,1\}$. Further, assume $z^* \in \{0,1\}$ attains $\max_{z \in \{0,1\}} \hat{V}(z,1)$, then $\hat{V}(1-z^*,1) \le \hat{V}(z^*,1)$, and (9) becomes

$$\hat{V}(z^*,1) \le \mathsf{P}_{dec}(z^*,1)V(z^*,1) + \mathsf{P}_{res}(z^*,1)\hat{V}(z^*,1)$$
$$+ \mathsf{P}_{arr}(z^*,1)\hat{V}(z^*,1). \tag{10}$$

We have $1 - \mathsf{P}_{res}(z^*,1) - \mathsf{P}_{arr}(z^*,1) = \mathsf{P}_{dec}(z^*,1) = \lambda/(\lambda + \mu_{z^*,1-z^*} + \kappa) > 0$, along with (10) this gives $\hat{V}(z^*,1) \le V(z^*,1)$. Thus, we have

$$\hat{V}(z^*,1) \le V(z^*,1)$$
$$= z^* f(1) + \gamma \max\{C, \hat{V}(z^*,1)\}$$
$$\le z^* f(1) + \gamma \max\{C, \hat{V}(z^*,1)\}. \tag{11}$$

If $\hat{V}(z^*,1) < C$, then by (8) we have

$$\tilde{C} \le \max_{z \in \{0,1\}} \hat{V}(z,1) = \hat{V}(z^*,1) < C \le \bar{C}.$$

Otherwise $C \le \hat{V}(z^*,1)$ and (11) becomes

$$\hat{V}(z^*,1) \le z^* f(1) + \gamma \hat{V}(z^*,1),$$

which gives us

$$\hat{V}(z^*,1) \le \frac{z^* f(1)}{1-\gamma} \le \frac{f(1)}{1-\gamma} = \bar{C}.$$

Hence in both cases we have $\tilde{C} \le \bar{C}$.

Next we show $\tilde{C} \ge C$. We have

$$\tilde{C} = \sum_{(z,n) \in \mathcal{S}} \pi(z,n)\hat{V}(z,n+1) \ge \sum_{z=0,1} \pi(z,0)\hat{V}(z,1). \tag{12}$$

Note that

$$\hat{V}(1,1) = \mathsf{P}_{dec}(1,1)V(1,1) + \mathsf{P}_{res}(1,1)\hat{V}(0,1)$$
$$+ \mathsf{P}_{arr}(1,1)\hat{V}(1,2)$$
$$\ge \mathsf{P}_{dec}(1,1)V(1,1)$$
$$= \mathsf{P}_{dec}(1,1)[f(1) + \gamma \max\{\hat{V}(1,1), C\}]$$
$$\ge \mathsf{P}_{dec}(1,1)f(1),$$

and

$$\hat{V}(0,1) = \mathsf{P}_{dec}(0,1)V(0,1) + \mathsf{P}_{res}(0,1)\hat{V}(1,1)$$
$$+ \mathsf{P}_{arr}(0,1)\hat{V}(0,2)$$
$$\ge \mathsf{P}_{res}(0,1)\hat{V}(1,1)$$
$$\ge \mathsf{P}_{res}(0,1)\mathsf{P}_{dec}(1,1)f(1).$$

Therefore, with (12) we have

$$\tilde{C} \ge \mathsf{P}_{res}(0,1)\mathsf{P}_{dec}(1,1)f(1) \sum_{z=0,1} \pi(z,0).$$

Let N_t^1 be the number of agents in our system at time t, and N_t^2 be that of an $M/M/\infty$ queue with arrival rate κ and service rate $\lambda(1-\gamma)$. Assume $N_0^1 = N_0^2$. Using a coupling argument, which we omit due

to space limitations, it can be shown that N_t^2 stochastically dominates N_t^1 for all $t \geq 0$. Therefore,

$$
\begin{aligned}
\sum_{z=0,1} \pi(z,0) &= \sum_{z=0,1} \lim_{t \to +\infty} \mathbf{P}(N_t^1 = 0, Z_t = z) \\
&= \lim_{t \to +\infty} \mathbf{P}(N_t^1 = 0) \\
&\geq \lim_{t \to +\infty} \mathbf{P}(N_t^2 = 0) \\
&= e^{-\frac{\kappa}{\lambda(1-\gamma)}},
\end{aligned}
$$

where the last equality follows from the steady state distribution of the $M/M/\infty$ queue.

Therefore we have

$$
\begin{aligned}
\tilde{C} &\geq \mathsf{P}_{dec}(1,1)\mathsf{P}_{res}(0,1)f(1) \sum_{z=0,1} \pi(z,0) \\
&= \left(\frac{\lambda}{\lambda + \mu_{1,0} + \kappa} \right) \left(\frac{\mu_{0,1}}{\lambda + \mu_{0,1} + \kappa} \right) e^{-\frac{\kappa}{\lambda(1-\gamma)}} f(1) \\
&\geq \left(\frac{\lambda}{\lambda + \mu_{1,0} + \beta\lambda} \right) \left(\frac{\mu_{0,1}}{\lambda + \mu_{0,1} + \beta\lambda} \right) e^{-\frac{\beta}{1-\gamma}} f(1) = C,
\end{aligned}
$$

where the last inequality is achieved on picking $\kappa = \beta\lambda$. $\qquad\square$

Next we restrict our choice of thresholds (n_0, n_1) to a compact set.

Lemma A.3. *Given* $(n_0, n_1) \in [0, +\infty)^2$, $C \in [\underline{C}, \bar{C}]$ *and* $\kappa \in [\beta\lambda(1 - \gamma), \beta\lambda]$, *for any* $(\tilde{n}_0, \tilde{n}_1) \in \mathcal{T}(n_0, n_1, \kappa, C)$, *we have* $\tilde{n}_0 \leq M$ *and* $\tilde{n}_1 \leq M$.

Proof. Suppose at $t = 0$ the 0th decision epoch the agent starts, and there are n agents at the node. We show for large enough n, the total expected payoff the agent receives if he chooses to stay at $t = 0$ will be less than what he would receive if he chooses to switch to a different location.

Let $\tau \in \mathbb{N}_0$ be the first decision epoch the agent chooses to leave. We seek to show that $\tau = 0$. Suppose, for the sake of arriving at a contradiction, $\tau \geq 1$. Let R_i be the immediate reward the agent receives at his i^{th} decision epoch if he is still at the current node at the time he makes his i^{th} decision, and R be the total reward he receives on choosing to stay at $t = 0$. We have:

$$
\begin{aligned}
\mathbf{E}[R] &= \sum_{i=0}^{+\infty} \gamma^i \mathbf{E}[R_i \mathbf{1}\{\tau > i\}] + \mathbf{E}[\gamma^\tau C] \\
&\leq \sum_{i=0}^{+\infty} \gamma^i \mathbf{E}[R_i \mathbf{1}\{\tau > i\}] + \gamma C. \tag{13}
\end{aligned}
$$

The first term in the first equation of (13) is the expected total reward until the agent chooses to leave, and the second term is the aggregated expected payoff on leaving. The inequality follows from the assumption that $\tau \geq 1$. Our goal is to show $\mathbf{E}[R_i \mathbf{1}\{\tau > i\}]$ vanishes as $n \to +\infty$, hence $\mathbf{E}[R] \to \gamma C < C$ as $n \to +\infty$, and since C is the aggregated expected payoff on leaving at $t = 0$, we would have $\tau = 0$.

Let T_i be the time the ith decision epoch of the agent starts. We have $T_i \sim \text{Gamma}(i, \lambda)$ since the interval between two consecutive decision epochs are i.i.d. $\exp(\lambda)$ and T_i is the sum of i such intervals.

Let $n' = \frac{1}{2}\sqrt{n}$. We have for all $i \in \mathbb{N}^+$,

$$
\begin{aligned}
\mathbf{E}[R_i \mathbf{1}\{\tau \geq i\}] &= \mathbf{E}[R_i \mathbf{1}\{\tau \geq i\}|N_{T_i} \geq n']\mathbf{P}(N_{T_i} \geq n') \\
&\quad + \mathbf{E}[R_i \mathbf{1}\{\tau \geq i\}|N_{T_i} < n']\mathbf{P}(N_{T_i} < n') \\
&\leq f(n')\mathbf{P}(N_{T_i} \geq n') + \mathbf{P}(N_{T_i} < n')f(1) \\
&\leq f(n') + \mathbf{P}(N_{T_i} < n')f(1). \tag{14}
\end{aligned}
$$

We first show $\mathbf{P}(N_{T_i} < n')$ vanishes as $n \to +\infty$. Consider an alternative system with n agents at $t = 0$ where each agent stays an $\exp(\lambda)$ time and then leaves the system. Let N_t' be the number of agents in the

16

system at time t. For any agent in this alternative system, the probability he is still in the system at time T_i is $e^{-\lambda T_i}$, therefore we have $N'_{T_i} \sim \text{Bin}(n, e^{-\lambda T_i})$.

Using a similar argument as in the proof of Lemma 4.1, we can show for all $t \geq 0$, N_t is no less than N'_t in all sample paths, i.e., N_t stochastically dominates N'_t in the first order. Let $k = \lfloor (\log n)^{\frac{1}{2}} \rfloor$. Note $N'_{T_i} \geq N'_{T_k}$, for all $i \leq k$. Thus, pick $t_k = \frac{1}{2\lambda} \log n$. We have

$$
\begin{aligned}
\mathbf{P}(N_{T_i} < n') &\leq \mathbf{P}(N'_{T_i} < n') \\
&\leq \mathbf{P}(N'_{T_k} < n') \\
&\leq \mathbf{P}(N'_{T_k} < n' | T_k < t_k)\mathbf{P}(T_k < t_k) + P(T_k \geq t_k) \\
&\leq \mathbf{P}(N'_{t_k} < n' | T_k < t_k)\mathbf{P}(T_k < t_k) + \mathbf{P}(T_k \geq t_k) \\
&\leq \mathbf{P}(N'_{t_k} < n') + \mathbf{P}(T_k \geq t_k).
\end{aligned}
$$

We have $N'_{t_k} \sim \text{Bin}(n, e^{-\lambda t_k})$. Given $ne^{-\lambda t_k} = \sqrt{n} > \frac{1}{2}\sqrt{n} = n'$, we can apply the Chernoff bound to obtain,

$$
\begin{aligned}
\mathbf{P}(N'_{t_k} < n') &\leq \exp(-\frac{(ne^{-\lambda t_k} - n')^2}{2ne^{-\lambda t_k}}) \\
&= \exp(-\frac{n'}{4}) \\
&= \exp(-\frac{\sqrt{n}}{8}).
\end{aligned}
$$

Also, by Markov's inequality we have

$$
\mathbf{P}\left(T_k \geq t_k\right) \leq \frac{\mathbf{E}[T_k]}{t_k} = \frac{k}{\lambda t_k} = \frac{2\lfloor (\log n)^{\frac{1}{2}} \rfloor}{\log n} \leq \frac{2}{\sqrt{\log n}}.
$$

Therefore,

$$
\mathbf{P}(N_{T_i} < n') \leq \exp(-\frac{\sqrt{n}}{8}) + \frac{2}{\sqrt{\log n}}.
$$

By (14) and using the fact that $n' = \frac{1}{2}\sqrt{n}$, we have

$$
\mathbf{E}[R_i \mathbf{1}\{\tau > i\}] \leq f(\frac{\sqrt{n}}{2}) + \exp(-\frac{\sqrt{n}}{8}) + \frac{2}{\sqrt{\log n}},
$$

for all $i = 1, \ldots, k$. Therefore, we have

$$
\begin{aligned}
\sum_{i=0}^{\infty} \gamma^i \mathbf{E}[R_i \mathbf{1}\{\tau > i\}] &= \sum_{i=0}^{k-1} \gamma^i \mathbf{E}[R_i \mathbf{1}\{\tau > i\}] + \sum_{i=k}^{\infty} \gamma^i \mathbf{E}[R_i \mathbf{1}\{\tau > i\}] \\
&\leq \sum_{i=0}^{k-1} \gamma^i \left[f(\frac{\sqrt{n}}{2}) + \exp(-\frac{\sqrt{n}}{8}) + \frac{2}{\sqrt{\log n}} \right] + \sum_{i=k}^{\infty} \gamma^i f(1) \\
&\leq \frac{1}{1-\gamma} \left[f(\frac{\sqrt{n}}{2}) + \exp(-\frac{\sqrt{n}}{8}) + \frac{2}{\sqrt{\log n}} \right] + \frac{\gamma^k}{1-\gamma} f(1).
\end{aligned}
$$

The righthand side is denoted as $g(n)$ in (7). As $n \to +\infty$, $g(n)$ goes to 0. Hence by picking M such that

$$
M \triangleq \min\{n : \frac{1}{1-\gamma} g(n) + \frac{\gamma^k}{1-\gamma} f(1) < (1-\gamma)C\},
$$

we have

$$
\mathbf{E}[R] < (1-\gamma)C + \gamma C < C,
$$

for all $n \geq M$. Thus we have proved choosing to stay at $t = 0$ when there are more than M agents at the location is suboptimal no matter what the resource level is. This contradicts the assumption that $\tau \geq 1$. $\square$