Infinite-Horizon Average-Cost Markov Decision Process Routing Games

Dan Calderone
Electrical Engineering and
Computer Science
University of California, Berkeley
Berkeley, CA 94720

S. Shankar Sastry
Electrical Engineering and
Computer Science
University of California, Berkeley
Berkeley, CA 94720

***Abstract*—We explore an extension of nonatomic routing games that we call *Markov decision process routing games* where each agent chooses a transition policy between nodes in a network rather than a path from an origin node to a destination node, i.e. each agent in the population solves a Markov decision process rather than a shortest path problem. This type of game was first introduced in [1] in the finite-horizon total-cost case. Here we present the infinite-horizon average-cost case. We present the appropriate definition of a Wardrop equilibrium as well as a potential function program for finding the equilibrium. This work can be thought of as a routing-game-based formulation of continuous population stochastic games (mean-field games or anonymous sequential games). We apply our model to ridesharing drivers competing for fares in an urban area.**

I. Introduction

Classic routing games [2]–[5] are perhaps the best studied examples of continuous population potential games [6]. The strategy choices for agents in the population of a nonatomic routing game are the various routes from their origin to their destination and each agent's goal is to find the shortest route. At the Wardrop or Nash equilibrium of the game, the population is divided up among the routes so the overall population distribution is consistent with the shortest path problem that each individual member is solving, i.e. no mass is allocated to a route with non-minimal latency.

It is well known that the shortest path problem can be formulated as a linear program. There are also similar linear programming formulations for solving Markov decision processes (MDPs) [7], [8]. This suggests that there may be a version of a continuous population potential game on a network where each agent solves an MDP as opposed to a shortest path problem. In this paper, we define such a game that we call a *Markov decision process routing game*. We define the appropriate Wardrop-type equilibrium concept and show how it can be found by minimizing a potential function in the infinite-horizon average-cost case.

MDP routing games are a specific case of continuous population stochastic games, games where each infinitesimal agent in a population solves a Markov decision process with rewards determined by the actions of the other agents. These games were first introduced as *anonymous sequential games* by Rosenthal and Jovanovic [9]. Results have focused mostly on existence and uniqueness of equilibria [10]–[12] and specific applications [13], [14]. Recently, stochastic population games have been studied in the mean-field game community starting with Lasry and Lions in 2006 [15]–[17]. Our formulation bears closest resemblance to mean-field games on graphs [18]–[20]. The standard mean-field model consists of a pair of coupled partial differential equations (PDEs): one backward time PDE that defines the value function or ”cost-to-go” for the population of agents and one forward time PDE that defines the mass evolution of the population. As in our case, when the costs agents' consider can be written as the gradient of some functional, the mean-field game is called a potential game and both PDEs can be solved by solving a single optimal control problem. A significant difference between our formulation and mean-field games on graphs is that in our formulation the potential function derivative condition is defined with respect to the mass of the population taking a specific action as opposed to the mass of the population at a given node (compare this paper with the potential game formulations in [18], [19]).

To illustrate an application of this game, we consider the problem faced by drivers who provide ridesharing services such as Uber or Lyft drivers. In order to employ more than just heuristics in their fare optimization process, drivers must consider the jobs they take over the entire time horizon, taking into account both the fare that they will receive on an individual trip as well as how the destination of that trip will position them to take advantage of the next job. Another source of complexity is that drivers are competing with each other for jobs. If an individual area becomes crowded with drivers, they will have to wait longer in order to get the job they want. This competition naturally gives rise to a game where drivers seek to optimize their profits over some period of time by choosing a transition strategy throughout the network and their rewards depend on their own transition strategy as well as the transition strategies of the other drivers in the population. We return to this example in Section IV.

The body of the paper is organized as follows. In Section II, we briefly detail the classical routing game and its relationship with the linear programming formulation of the shortest path problem. We then review a linear programming formulation of

Author contact info: {danjc,sastry}@eecs.berkeley.edu
This work is supported by NSF FORCES CNS-1239166.

978-1-5386-1526-3/17/$31.00 ©2017 IEEE

classical MDPs in the infinite-horizon average-cost case and draw connections with the shortest path program. In Section III, we detail the MDP routing game in the infinite-horizon average-cost case and make further connections with classical routing games. In Section IV, we apply our formulation to the ridesharing game. In Section V, we conclude and comment on future work.

II. Inspiration: Routing games, Shortest path LP, and Markov Decision Process LP

A. Routing games and shortest path LP

To motivate our formulation, we offer a brief, informal presentation of classical nonatomic routing games and the corresponding linear programming formulation of the shortest path problem that each individual agent in the population solves. For a thorough discussion of routing games we refer the reader to [5]. We then present a linear programming formulation of classical MDPs that serves as the motivation for the MDP routing game we present in the next section.

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be a directed graph with nodes and edges. Let $G \in \{-1, 0, 1\}^{|\mathcal{N}|\times|\mathcal{E}|}$ be an incidence matrix for the graph

$$[G]_{ne} = \begin{cases} 1 & ; \text{ if edge } e \text{ starts at node } n \\ -1 & ; \text{ if edge } e \text{ ends at node } n \\ 0 & ; \text{ otherwise} \end{cases} \quad (1)$$

and let $s \in \{-1, 0, 1\}^{|\mathcal{N}|}$ be a source-sink vector that indicates the population's origin and destination nodes.

$$s_n = \begin{cases} 1 & ; \text{ if } n \text{ is the origin} \\ -1 & ; \text{ if } n \text{ is the destination} \\ 0 & ; \text{ otherwise} \end{cases} \quad (2)$$

We also define two other incidence matrices, $I_o, I_i \in \{0, 1\}^{|\mathcal{N}|\times|\mathcal{E}|}$, that indicate where each edge originates and terminates.

$$[I_o]_{ne} = \begin{cases} 1 & ; \text{ if } e \text{ starts at } n \\ 0 & ; \text{ otherwise} \end{cases}, \qquad [I_i]_{ne} = \begin{cases} 1 & ; \text{ if } e \text{ ends at } n \\ 0 & ; \text{ otherwise} \end{cases} \quad (3)$$

Note that $G = I_o - I_i$.

Let $m \in \mathbb{R}_+$ be the total population mass and let $x \in \mathbf{R}_+^{|\mathcal{E}|}$ be the vector of masses on each edge. Each edge has an associated *latency function*, $l_e(x_e)$, (generally positive and increasing) that describes the effect of congestion on travel time on that particular edge. The Wardrop equilibrium of the routing game is a distribution of mass across the routes of the network such that each utilized route has equal, minimal latency, i.e. no member of the population can improve their travel time by switching routes. One can find this Wardrop equilibrium by solving the following optimization problem.

$$\min_x \quad F(x) \quad (4a)$$

$$\text{s.t.} \quad Gx = sm, \quad x \geq 0 \quad (4b)$$

where $F(x)$ is the potential function

$$F(x) = \sum_e \int_0^{x_e} l_e(u)\ du \quad (5)$$

The equivalence between the minimum of Problem (4) and the Wardrop equilibrium is based on the first order optimality conditions and in particular that the gradient of the potential function $F(x)$ is equal to the vector of latencies $l(x) : x \mapsto \mathbb{R}^{|\mathcal{E}|}$. Intuitively, at equilibrium each infinitesimal agent seeks to find a route from the origin to the destination with minimal latency as defined by $l(x)$. This shortest path problem can be written as a linear program in the following way

$$\min_\xi \quad l(x)^T \xi \quad (6a)$$

$$\text{s.t.} \quad G\xi = s, \quad \xi \geq 0 \quad (6b)$$

Note that here x is fixed and the optimization variable $\xi \in [0, 1]^{|\mathcal{E}|}$ is a decision vector that represents the probability of taking each edge in the graph. This program has the same first order optimality conditions for ξ as Problem (4) has for x. Thus we can interpret the Wardrop equilibrium as the overall mass distribution whose support is consistent with individual agent solving the shortest path problem in Problem (6).

B. Markov Decision Process LP

We now present a linear programming formulation of an infinite-horizon average-cost Markov decision process. This program bears clear resemblance to Program (6) which motivates the formulation of the MDP routing game.

We assume the same graph structure $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ and define $\mathcal{N}_j$ to be the nodes that can be reached by an edge from node j for all $j \in \mathcal{N}$. Let $\mathcal{A}_j$ be a set of *actions* associated with each node j and let $P_j^a = (P_{ij}^a)_{i\in\mathcal{N}_j} \in \Delta_{|\mathcal{N}_j|}$ be a set of transition probabilities from node j to each node $i \in \mathcal{N}_j$ associated with each action a. Here, $\Delta_{|\mathcal{N}_j|}$ is the simplex of dimension $|\mathcal{N}_j|$. We will sometimes refer to the set of all actions as $\mathcal{A} = \cup_j \mathcal{A}_j$ with the understanding that each action $a \in \mathcal{A}$ is only available from a specific node.

Define a set of losses (traditionally rewards), R_j^a, associated with taking each action $a \in \mathcal{A}_j$; a set of losses, Q_{ij}, associated with transitioning from j to i; and a set of losses S_j associated with ending up in node j. Depending on context, we will sometimes index the action losses simply by the action $a \in \mathcal{A}$ with the understanding that a is only available from node j, and we will sometimes index the transition losses by the specific edge of the graph $e \in \mathcal{E}$ with the understanding that edge e runs from j to i.

The goal of solving a Markov decision process is to find the optimal mixed strategy over the actions $\mathcal{A}_j$ to choose whenever the agent is in node j. We consider here the case where an agent optimizes their average cost over an infinite time horizon. We will denote a mixed strategy at node j as $\eta_j = (\eta_j^a)_{a\in\mathcal{A}_j} \in \Delta_{|\mathcal{A}_j|}$ where $\Delta_{|\mathcal{A}_j|}$ is the simplex of dimension $|\mathcal{A}_j|$, and we will refer to a collection of mixed strategies, $\eta = (\eta_j)_{j\in\mathcal{N}}$, as a *policy*. A policy η gives rise to a transition matrix $P(\eta) \in [0, 1]^{|\mathcal{N}|\times|\mathcal{N}|}$

$$\left[P(\eta)\right]_{ij} = \sum_{a\in\mathcal{A}_j} P_{ij}^a \eta_j^a \quad (7)$$

and the resulting stationary distribution $p(\eta) : \eta \mapsto [0,1]^{|\mathcal{N}|}$. We make the following standard assumption to guarantee that the stationary distribution exists, is unique, and describes the long term limit of the Markov chain's behavior.

Assumption 1: Assume that $P(\eta)$ is irreducible and aperiodic for every pure strategy policy η.

In the average loss infinite horizon case, we want to optimize the following program.

$$\min_{\eta} \quad \sum_j \Big(\sum_i Q_{ij} P_{ij}(\eta) + \sum_{a \in \mathcal{A}_j} R_j^a \eta_j^a + S_j \Big) p_j(\eta) \tag{8a}$$

$$\text{s.t.} \quad \eta_j \in \Delta_{|\mathcal{N}_j|}, \quad p_i(\eta) = \sum\nolimits_j P_{ij}(\eta) p_j(\eta) \quad \forall i, j \tag{8b}$$

This problem as formulated is nonlinear and difficult to solve. However, if we solve for both the policy and the stationary distribution at the same time by applying a change of variables

$$\xi_j^a = p_j(\eta) \eta_j^a \tag{9}$$

we can transform the problem into a linear program. ξ_j^a is the probability of being in node j and choosing the action a. Problem (8) can be written as

$$\min_{\xi \geq 0} \quad \sum_j \sum_{a \in \mathcal{A}_j} \left(\sum_i Q_{ij} P_{ij}^a + R_j^a + S_j \right) \xi_j^a \tag{10a}$$

$$\text{s.t.} \quad \sum_{a \in \mathcal{A}_i} \xi_i^a = \sum_j \sum_{a \in \mathcal{A}_j} P_{ij}^a \xi_j^a, \quad \sum_j \sum_{a \in \mathcal{A}_j} \xi_j^a = 1 \tag{10b}$$

Given ξ, it is straightforward to solve for η and $p(\eta)$ as

$$\eta_j^a = \frac{\xi_j^a}{\sum_{a \in \mathcal{A}_j} \xi_j^a}, \qquad p_j(\eta) = \sum_{a \in \mathcal{A}_a} \xi_j^a \tag{11}$$

For more details see [7], [8].

In order to write Problem (10) in matrix form and make connections with the shortest path problem and routing game, we can assign some ordering to the edges of the graph and as well as the action set $\mathcal{A}$. We can then define a transition matrix $\mathbf{P}_{\mathcal{A}} \in [0,1]^{|\mathcal{E}| \times |\mathcal{A}|}$ that maps ξ to the probability of taking edge e at any given transition.

$$\big[\mathbf{P}_{\mathcal{A}}\big]_{ea} = \begin{cases} P_{ij}^a & ; \text{ if } a \text{ is available in node } j \text{ and edge } e \text{ connects } j \text{ to } i \\ 0 & ; \text{ otherwise} \end{cases} \tag{12}$$

Note that $\mathbf{P}_{\mathcal{A}}$ is column stochastic. Using matrix $\mathbf{P}_{\mathcal{A}}$, we can compute the vector of probabilities of taking a specific edge in the graph at any given time, $\bar{\xi} \in [0,1]^{|\mathcal{E}|}$, as $\bar{\xi} = \mathbf{P}_{\mathcal{A}} \xi$. We also note that in the *fully deterministic* case—where an agent can choose any edge originating from a given node with probability 1—we have that $\mathbf{P}_{\mathcal{A}} = I_{|\mathcal{E}| \times |\mathcal{E}|}$ (assuming the proper ordering on the edges and actions) and $\bar{\xi} = \xi$.

Indexing Q by edges and R and ξ by actions in $\mathcal{A}$, we can write Problem (10) in matrix form as

$$\min_{\xi} \quad \big(Q^T \mathbf{P}_{\mathcal{A}} + R^T + S^T I_o \mathbf{P}_{\mathcal{A}}\big)\xi \tag{13a}$$

$$\text{s.t.} \quad G \mathbf{P}_{\mathcal{A}} \xi = 0, \quad \mathbf{1}^T \xi = 1, \quad \xi \geq 0 \tag{13b}$$

In the fully deterministic case, this becomes

$$\min_{\xi} \quad \big(Q^T + R^T + S^T I_o\big)\xi \tag{14a}$$

$$\text{s.t.} \quad G\xi = 0, \quad \mathbf{1}^T \xi = 1, \quad \xi \geq 0 \tag{14b}$$

In this form, we can see a clear connection with the shortest path linear program presented in Section II-A. ξ must live in the nullspace of the incidence matrix G which means that ξ must be some positive linear combination of the cycles of the graph $\mathcal{G}$. We comment on this more in Section III-B. We also note that this deterministic case breaks Assumption 1. We mention it here to draw connections with the routing game and we make further comments about it in the game context in Remark 3.

This linear program view of MDPs and the connections with the edge flow formulation of the routing game suggests that we can define a population potential game where the loss functions are functions of the mass of the population and each infinitesimal member of the population is seeking to optimize an MDP. We consider this game in the next section.

III. Markov Decision Process Routing Games

A. Equilibrium: Definition and Computation

We now define the *Markov decision process (MDP) routing game* in the infinite-horizon average-cost case. We consider the steady state behavior of a continuous population of anonymous agents spread out over the various states in the network each playing an optimal feedback policy. Again, we assume that Assumption 1 holds for each pure strategy policy. In our presentation, we assume a single population of agents, but as in classical routing games, the analysis extends to multiple populations.

Let x_j represent the steady state portion of the population in node j and x_j^a the subpopulation that takes action $a \in \mathcal{A}_j$.

$$x_j = \sum_{a \in \mathcal{A}_j} x_j^a \tag{15}$$

We can compute the steady state portion of the population transitioning from j to i as

$$x_{ij} = \sum_{a \in \mathcal{A}_j} P_{ij}^a x_j^a \tag{16}$$

Assuming the population is in steady state, mass conservation between nodes gives

$$x_i = \sum_{a \in \mathcal{A}_i} x_i^a = \sum_j x_{ij} = \sum_j \sum_{a \in \mathcal{A}_j} P_{ij}^a x_j^a \tag{17}$$

We also assume mass conservation over the entire network of a total population mass of m and that mass is positive everywhere.

$$\sum_j \sum_{a \in \mathcal{A}_j} x_j^a = m, \qquad x_j^a \geq 0 \tag{18}$$

We will use $x \in \mathbb{R}^{|\mathcal{A}|}$ to refer to the vector of all subpopulations x_j^a.

In the game case, the loss functions depend on the population distribution, i.e. $R_j^a(x)$, $Q_{ij}(x)$, and $S_j(x)$ respectively. As in the classical routing game, this dependence introduces competition into the scenario. We model each (infinitesimal) agent in the population as solving a Markov decision process to minimize their average expected cost given these loss functions and the overall mass distribution of the other agents. Given Assumption 1, the average cost of a given policy, η, (given the overall mass distribution) can be computed as

$$\begin{aligned}\ell(x,\eta) &= \sum_j \Big(\textstyle\sum_i Q_{ij}(x)P_{ij}(\eta)+\sum_{a\in\mathcal{A}_j} R_j^a(x)\eta_j^a+S_j(x)\Big)p_j(\eta)\\ &= \sum_j \sum_{a\in\mathcal{A}_j} \Big(\textstyle\sum_i Q_{ij}(x)P_{ij}^a+R_j^a(x)+S_j(x)\Big)\eta_j^a p_j(\eta) \quad (19)\end{aligned}$$

We now define a Wardrop equilibrium condition for the MDP routing game.

Definition 1 (Infinite-Horizon Wardrop Equilibrium): The population distribution x is an *infinite-horizon Wardrop equilibrium* for the infinite-horizon MDP routing game if for any two policies η and η' such that $\eta_j^a > 0$ only if $x_j^a > 0$,

$$\ell(x,\eta) \le \ell(x,\eta') \quad (20)$$

In other words, any policy that some portion of the population employs has minimal cost and thus no population member can improve their cost by changing their policy at any state.

We now define the notion of a potential game for the MDP routing game as follows.

Definition 2 (Infinite-horizon potential game): We say the infinite-horizon MDP routing game is a *potential game* if there exists a C^1 function $F : x \mapsto \mathbb{R}$ such that

$$\frac{\partial F}{\partial x_j^a}(x) = R_j^a(x) + \sum_i Q_{ij}(x)P_{ij}^a + S_j(x) \quad (21)$$

for each element x_j^a of x.
Intuitively, the derivative of the potential function with respect to the mass taking a particular action captures the immediate payoff of that action. We note that if we write $F(x)$ not only as a function of x_j^a but also explicitly as a function of x_{ij} and x_j and $F(\cdot)$ satisfies

$$\frac{\partial F}{\partial x_j^a} = R_j^a(x), \quad \frac{\partial F}{\partial x_{ij}} = Q_{ij}(x), \quad \frac{\partial F}{\partial x_j} = S_j(x) \quad (22)$$

then Condition (21) is satisfied by applying the chain rule and Equations (15) and (16).

Remark 1: In the special case where $R_j^a(\cdot)$ is simply a function of x_j^a, $Q_{ij}(\cdot)$ is simply a function of x_{ij}, and $S_j(\cdot)$ is simply a function of x_j, we can use the potential

$$F(x) = \sum_j \sum_{a\in\mathcal{A}_j} \int_0^{x_j^a} R_j^a(u)\,du \;+ \sum_{ij}\int_0^{x_{ij}} Q_{ij}(u)\,du + \sum_j \int_0^{x_j} S_j(u)\,du$$

Note the similarities with the potential function for the classical routing game.

Remark 2: This definition of a potential function is a substantial deviation from mean-field games on graphs where the potential function differentiation condition is defined with respect to the mass on the nodes as opposed to the mass taking a particular action. (See [18], [19] for details.)

We now show that we can find the Wardrop equilibrium by minimizing the potential function.

Theorem 1: Given a potential function F for the infinite-horizon MDP routing game, if x satisfies the KKT first order necessary conditions for minimizing F, then x is a Wardrop equilibrium.

Proof 1: The optimization problem and corresponding Lagrangian are given by

$$\min_{x\ge 0} \quad F(x) \quad (23a)$$

$$\text{s.t.} \quad \sum_{a\in\mathcal{A}_i} x_i^a = \sum_j \sum_{a\in\mathcal{A}_j} P_{ij}^a x_j^a, \quad \sum_{j,a\in\mathcal{A}_j} x_j^a = m \quad (23b)$$

$$\begin{aligned}\mathcal{L}(x,\pi,\lambda,\mu) &= F(x) - \lambda\Bigg(\sum_j \sum_{a\in\mathcal{A}_j} x_j^a - m\Bigg)\\ &- \sum_i \pi_i \Bigg(\sum_{a\in\mathcal{A}_i} x_i^a - \sum_j\sum_{a\in\mathcal{A}_j} P_{ij}^a x_j^a\Bigg) - \sum_{j,a}\mu_j^a x_j^a \quad (24)\end{aligned}$$

with Lagrange multipliers $\pi \in \mathbb{R}^{|\mathcal{N}|}$, $\lambda \in \mathbb{R}$, and $\mu_j \in \mathbb{R}_+^{|\mathcal{A}_j|}$. Given that $F(x)$ is a potential function, the first order necessary conditions are

$$\sum_i Q_{ij}(x)P_{ij}^a + R_j^a(x) + S_j(x) - \pi_j + \sum_i \pi_i P_{ij}^a - \lambda - \mu_j^a = 0 \quad (25)$$

$\mu_j^a \ge 0$, and $\mu_j^a x_j^a = 0$. Using (25), we can compute the cost of any individual's policy η as follows. (When it is obvious, we assume the appropriate action set $\mathcal{A}_j$ from context.)

$$\begin{aligned}\ell(x,\eta) &= \sum_j\sum_{a\in\mathcal{A}_j}\Bigg(\sum_i Q_{ij}(x)P_{ij}^a + R_j^a(x) + S_j(x)\Bigg)\eta_j^a p_j(\eta) \quad (26a)\\ &= \sum_j\sum_{a\in\mathcal{A}_j}\Bigg(\pi_j - \sum_i \pi_i P_{ij}^a + \lambda + \mu_j^a\Bigg)\eta_j^a p_j(\eta) \quad (26b)\\ &= \sum_j \pi_j p_j(\eta) - \sum_i \pi_i p_i(\eta) + \lambda + \sum_{j,a}\eta_j^a\mu_j^a p_j(\eta) \quad (26c)\\ &= \lambda + \sum_j\sum_{a\in\mathcal{A}_j}\eta_j^a\mu_j^a p_j(\eta) \quad (26d)\end{aligned}$$

It follows that

$$\ell(x,\eta) - \sum_{j,a}\eta_j^a\mu_j^a p_j(\eta) = \ell(x,\eta') - \sum_{j,a}{\eta'}_j^a\mu_j^a p_j(\eta') \quad (27)$$

for any two policies η, η'. If $\eta_j^a > 0$ only if $x_j^a > 0$ for all j, then by complementary slackness we have that $\eta_j^a\mu_j^a = 0$. It follows that

$$\ell(x,\eta) \le \ell(x,\eta') \quad (28)$$

since ${\eta'}_j^a \mu_j^a p_j(\eta') \ge 0$ for all j and $a \in \mathcal{A}_j$.

Remark 3: Many interesting cases break the irreducible, aperiodic assumption (Assumption 1), the deterministic transition case being the most obvious. We note that the steady state equilibrium concept may be valid even in these cases. A reducible Markov chain will have several irreducible subsets of recurrent states. Problem (10) will assign a certain amount of mass to each of these irreducible subsets in order to minimize the expected loss. If the initial probability distribution of the agent is fixed, there may not be a policy that divides up the probability mass between the irreducible subsets in this optimal way. If, however, we think of the agent as choosing their initial probability distribution as well as a policy, they would be able to assign the appropriate amount of mass to each recurrent subset. In the case of a periodic Markov chain assuming strictly increasing losses, we note that oscillating solutions cause some members of the population to experience more congestion than others. If agents have the option of remaining in a node at any time, it will be advantageous for them to damp out these oscillations. We could recover the full aperiodic assumption by adding a small probability of remaining in the same node to each action; however, from this argument, assuming that agents have the option of remaining in any node at any time seems to be enough for the steady state equilibrium concept to make sense. This argument is particularly applicable in the deterministic case where we are really thinking of agents as playing a routing game where they choose between cycles as opposed to routes.

B. Parallels with traditional routing games

Writing Problem (23) in matrix form (and indexing x to match the columns of $\mathbf{P}_{\mathcal{A}}$) we get

$$\min_{x\geq 0} \quad F(x) \tag{29a}$$

$$\text{s.t.} \quad G\mathbf{P}_{\mathcal{A}}x = 0, \quad \mathbf{1}^T x = m \tag{29b}$$

which parallels Problem (4) in the fully deterministic case where routes from the origin to destination have been replaced by cycles of the graph. Indeed, when $\mathbf{P}_{\mathcal{A}} = I_{|\mathcal{E}|\times|\mathcal{E}|}$, x is a positive vector contained in the nullspace of G, it must be a positive linear combination of indicator vectors for the cycles in the graph . We could write the optimization problem in this deterministic case in a way that parallels the path formulation of the traditional routing game by enumerating the simple cycles of the graph and solving directly for the mass on each cycle. Let $\mathcal{C}$ be the set of cycles and let $\mathcal{E}_c$ be the set of edges in a cycle $c \in \mathcal{C}$. Define an indicator matrix for the cycles, $\mathbf{C} \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{C}|}$, as

$$[\mathbf{C}]_{ec} = \begin{cases} \frac{1}{|\mathcal{E}_c|} & ; \text{ if } e \in \mathcal{E}_c \\ 0 & ; \text{ otherwise} \end{cases} \tag{30}$$

Using this indicator matrix, we can solve directly for the masses on the cycles $z \in \mathbb{R}_+^{|\mathcal{C}|}$.

$$\min_{z\geq 0} \quad F(\mathbf{C}z) \tag{31a}$$

$$\text{s.t.} \quad \mathbf{1}^T z = m \tag{31b}$$

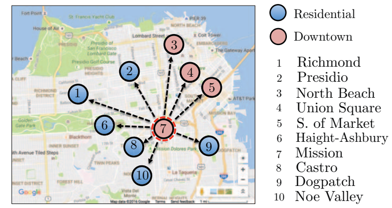

Fig. 1. (a) Illustration of SF neighborhoods with their types.

IV. SIMULATION: RIDESHARING GAME

In this section, we simulate the game ridesharing drivers might play in downtown San Francisco seeking to optimize their fares. We abstract the city as a set of neighborhoods (nodes) drivers travel between (shown in Figure 1). We assume that when drivers pick up a rider, that rider is selected at random, i.e. drivers do not get to select the trip they want. (We could also model situations where drivers get to select their riders specifically, in which case the transitions would be deterministic.) We assume that at each node a given percentage of riders want to travel to each of the other nodes. This model would be useful for ridesharing services where drivers are simply assigned riders or taxi drivers who queue at transportation hubs.

At each node, the driver can choose from several actions. The first action, a_r is to wait for a random rider and transition to whatever node that rider wants to go to. The transition probabilities of this action are determined by the percentages of riders at that node that want to make specific trips derived from the following table.

From ea. / To ea.	Resident	Downtown
Resident	0.06	0.083
Downtown	0.175	0.167

The driver can also choose to transition without a rider. We will refer to the action of transitioning to node i without a rider as a_i. In general, this would result in the driver paying the travel costs without receiving a fare; however, there is a small possibility that the driver will find a customer along the way.

In order to model this scenario, we define a graph with two sets of edges going between each node. The first set model transitioning with a rider. We will denote these edges $\mathcal{E}_{\text{rider}}$. The second set model transitioning without a rider which we will denote $\mathcal{E}_{\text{norider}}$. This allows us to differentiate between the rewards received for taking a rider and for driving without a rider. Drivers who take a_r at each node travel on edges in $\mathcal{E}_{\text{rider}}$. If a driver transitions without a rider, they take the appropriate edge in $\mathcal{E}_{\text{norider}}$ with probability 0.82 and they take each of the edges in $\mathcal{E}_{\text{rider}}$ (coming from that node) with probability 0.02. This is meant to represent the small chance that they might pick up a rider along the way.

The costs on the actions are

$$R_j^{a_r} = (C_j)_{\text{wait}} x_j^{a_r}, \qquad R_j^{a_i} = 0 \tag{32}$$

where the waiting cost coefficient is

$$(C_j)_{\text{wait}} = \tau \cdot \Big(\underbrace{\text{Customer Demand Rate}}_{\text{rides/hr}}\Big)^{-1} \tag{33}$$

and τ is a time-money tradeoff parameter which we take to be \$27/hr. We take the driver demands as 20 rides per hr from residential neighborhoods and 50 rides per hr from downtown neighborhoods.

The transition costs differ depending on whether the drivers take a rider or not. The transition costs are given by

$$Q_e(\mathrm{x}) = \begin{cases} -M_e + (C_e)_{\text{trav}} & ;\ \text{if } e \in \mathcal{E}_{\text{rider}} \\ (C_e)_{\text{trav}} & ;\ \text{if } e \in \mathcal{E}_{\text{norider}} \end{cases} \tag{34}$$

where M_e is the fare for a trip on edge e and $(C_e)_{\text{trav}}$ is the cost of travel on that edge which we take as

$$M_e = \big(\text{Rate}\big) \cdot \big(\text{Dist}\big) \tag{35}$$

$$(C_e)_{\text{trav}} = \tau \underbrace{\big(\text{Dist}\big)}_{\text{mi}} \underbrace{\big(\text{Vel}\big)^{-1}}_{\text{hr/mi}} + \underbrace{\big(\text{Fuel Price}\big)}_{\$/\text{gal}} \underbrace{\big(\text{Fuel Eff}\big)^{-1}}_{\text{gal/mi}} \underbrace{\big(\text{Dist}\big)}_{\text{mi}} \tag{36}$$

The values that are not specifically edge dependent are listed in the table below.

Rate	Velocity	Fuel Price	Fuel Eff
\$6 /mi	8 mph	\$2.5/gal	20 mi/gal

We compute both the equilibrium and the socially optimal flows. The social optimum is the mass distribution that minimizes the overall cost the population experiences. Figure 2 shows the steady state distribution of drivers at the nodes in both cases including the portion that take riders and the portion that do not as well as the transitions that drivers make without riders.

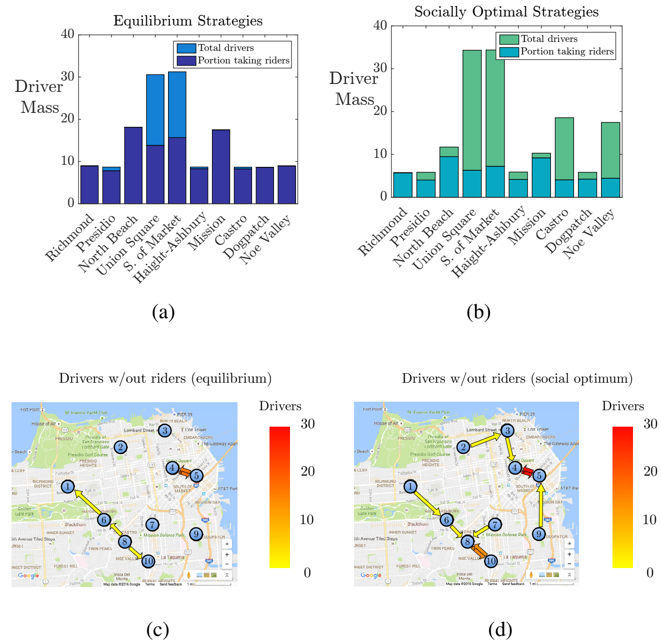

Fig. 2. Steady state distribution of drivers at each node under the (a) equilibrium strategies and (b) socially optimal strategies showing the portion of drivers that take riders and the portion that do not take riders. Drivers transitioning between nodes without riders in the (c) equilibrium case and (d) under the socially optimal strategies.

V. Conclusion

We have presented a continuous population game analogous to nonatomic routing games where individuals solve an infinite-horizon average-cost MDP as opposed to a shortest path problem. Future work includes considering the infinite-horizon discounted cost case.

References

[1] D. Calderone and S. S. Sastry, "Markov decision process routing games," in *Proceedings of the 8th International Conference on Cyber-Physical Systems*, ser. ICCPS '17. New York, NY, USA: ACM, 2017, pp. 273–279. [Online]. Available: http://doi.acm.org/10.1145/3055004.3055026

[2] J. G. Wardrop, "Road paper. some theoretical aspects of road traffic research." in *ICE Proceedings: Engineering Divisions*, vol. 1, no. 3. Thomas Telford, 1952, pp. 325–362.

[3] M. Beckmann, C. McGuire, and C. B. Winsten, "Studies in the economics of transportation," Tech. Rep., 1956.

[4] S. C. Dafermos and F. T. Sparrow, "The traffic assignment problem for a general network," *Journal of Research of the National Bureau of Standards, Series B*, vol. 73, no. 2, pp. 91–118, 1969.

[5] M. Patriksson, *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015.

[6] W. H. Sandholm, "Potential games with continuous player sets," *Journal of Economic Theory*, vol. 97, no. 1, pp. 81–108, 2001.

[7] A. S. Manne, "Linear programming and sequential decisions," *Management Science*, vol. 6, no. 3, pp. 259–267, 1960.

[8] J. L. Nazareth and R. B. Kulkarni, "Linear programming formulations of markov decision processes," *Operations research letters*, vol. 5, no. 1, pp. 13–16, 1986.

[9] B. Jovanovic and R. W. Rosenthal, "Anonymous sequential games," *Journal of Mathematical Economics*, vol. 17, no. 1, pp. 77–87, 1988.

[10] J. Bergin, D. Bernhardt *et al.*, "Anonymous sequential games with general state space," Tech. Rep., 1991.

[11] J. Bergin and D. Bernhardt, "Anonymous sequential games: existence and characterization of equilibria," *Economic Theory*, vol. 5, no. 3, pp. 461–489, 1995.

[12] P. Wiecek and E. Altman, "Stationary anonymous sequential games with undiscounted rewards," *Journal of optimization theory and applications*, vol. 166, no. 2, pp. 686–710, 2015.

[13] P. Wiecek, E. Altman, and Y. Hayel, "An anonymous sequential game approach for battery state dependent power control," in *International Conference on Network Control and Optimization*. Springer, 2009, pp. 121–136.

[14] P. Wiecek, E. Altman, and Y. Hayel, "Stochastic state dependent population games in wireless communication," *IEEE Transactions on Automatic Control*, vol. 56, no. 3, pp. 492–505, 2011.

[15] J.-M. Lasry and P.-L. Lions, "Jeux à champ moyen. i–le cas stationnaire," *Comptes Rendus Mathématique*, vol. 343, no. 9, pp. 619–625, 2006.

[16] ——, "Jeux à champ moyen. ii–horizon fini et contrôle optimal," *Comptes Rendus Mathématique*, vol. 343, no. 10, pp. 679–684, 2006.

[17] ——, "Mean field games," *Japanese Journal of Mathematics*, vol. 2, no. 1, pp. 229–260, 2007.

[18] D. A. Gomes, J. Mohr, and R. R. Souza, "Discrete time, finite state space mean field games," *Journal de mathématiques pures et appliquées*, vol. 93, no. 3, pp. 308–328, 2010.

[19] O. Guéant, "From infinity to one: The reduction of some mean field games to a global control problem," *arXiv preprint arXiv:1110.3441*, 2011.

[20] ——, "Existence and uniqueness result for mean field games with congestion effect on graphs," *Applied Mathematics & Optimization*, vol. 72, no. 2, pp. 291–303, 2015.