2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)
Allerton Park and Retreat Center
Monticello, IL, USA, October 2-5, 2018

Linearly Solvable Mean-Field Road Traffic Games

Takashi Tanaka[1] Ehsan Nekouei[2] Karl Henrik Johansson[3]

Abstract— **We analyze the behavior of a large number of strategic drivers traveling over an urban traffic network using the mean-field game framework. We assume an incentive mechanism for congestion mitigation under which each driver selecting a particular route is charged a tax penalty that is affine in the logarithm of the number of agents selecting the same route. We show that the mean-field approximation of such a large-population dynamic game leads to the so-called linearly solvable Markov decision process, implying that an open-loop ϵ-Nash equilibrium of the original game can be found simply by solving a finite-dimensional linear system.**

I. Introduction

A reasonable approach to obtain a macroscopic model of a road traffic network is to use a game-theoretic analysis (e.g., Wardrop [1]) with assumptions that (i) the number of players involved in the game is large, (ii) each individual player's impact on the network is infinitesimal, and (iii) players' identities are indistinguishable [2]. Dynamic games with such assumptions are broadly known as *mean-field games* (MFG), and have been actively studied in recent years [3], [4]. The central result in the MFG theory shows that the game theoretic equilibria (e.g., the Markov perfect equilibria) of the original large-population game can be well-approximated by the solution (called *mean-field equilibrium*, MFE) to the pair of the backward Hamilton-Jacobi-Bellman (HJB) equation and the forward Fokker-Planck-Kolmogorov (FPK) equation. This result has a significant impact in applications where solving the HJB-FPK system is computationally more tractable than analyzing the equilibria of the original game with large number of players.

Recently, the MFG theory has been applied to the analysis of traffic systems. In [5], the authors modeled the interaction between drivers on a straight road as a non-cooperative game, and characterized the MFE of this game using an HJB equation and a mass preservation equation. In [6], the authors considered a continuous-time Markov chain to model the aggregate behavior of drivers on a traffic network. They proposed various routing schemes for balancing the traffic flow over the network. MFG has been applied to pedestrian crowd dynamics modeling in [7], [8].

In this paper, we apply the MFG framework to the analysis of an urban transportation network in which individual drivers' dynamics are decoupled from each other, but their cost functions are coupled via the tax penalty/reward mechanism imposed by the Traffic System Operator (TSO). In our context, the tax mechanism provides drivers with incentives to route themselves in such a way that their collective behavior matches the desirable traffic flow pre-calculated by the TSO. In particular, we are interested in the log-population type of the tax mechanisms (i.e., the penalty imposed on an individual driver is affine in the logarithm of the number of drivers taking the same route). Such a tax mechanism is notable, as it renders the mean-field approximation of the original large-population game *linearly solvable* using the result of [9]. This offers a tremendous computational advantage over the conventional HJB-FPK characterization of the MFE. The purpose of this paper is to delineate this observation and to demonstrate its computational advantages using a numerical simulation.

A. Mean-field game theory: Background

The MFG theory has been introduced by the authors of [3], [4] and has been applied to the analysis of large-population games appearing in engineering, economic and financial applications. The central subject in the MFG theory is the coupled HJB-FPK system, which has attracted much attention in mathematical and engineering research [4], [10]. The MFE of linear quadratic stochastic games have been extensively studied in the literature, e.g., [11], [12], [13] and references therein. MFGs with non-linear cost function and/or non-linear dynamics are also studied in, e.g., [5], [14]. The MFE of a Markov decision game with a major agent and a large number of minor agents was studied in [14], where the players are only coupled via their cost functions. These results were used in [15] to design decentralized security defense decisions in a mobile ad hoc network. The authors of [16] studied the MFE of a dynamic stochastic game wherein the dynamic of each agent is described by a (non-linear) stochastic difference equation, and agents are coupled via both dynamics and cost functions. The authors in [17] considered a stochastic dynamic game in which the dynamics and cost function of each agent is affected by its individual disturbance term. They studied the existence of a robust (minimax) MFE. In [18], the authors analyzed the MFE of a hybrid stochastic game in which the dynamics of agents are affected by continuous disturbance terms as well as random switching signals. Risk-sensitive MFG was considered in [19]. While continuous-time continuous-state models are commonly used in the references above, the MFG in discrete-time and/or discrete-state regime have been considered in, e.g., [20]–[22].

[1] Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, TX, USA. ttanaka@utexas.edu. [2] School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. nekouei@kth.se. [3] School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. kallej@kth.se.

978-1-5386-6596-1/18/$31.00 ©2018 IEEE

B. Contribution of this paper

In this paper, we apply the MFG theory to study the strategic behavior of infinitesimal drivers traveling over an urban traffic network. We consider a dynamic stochastic game wherein, at each intersection, each driver randomly selects one of the outgoing links as her next destination according to her randomized policy. The objective function of each driver consists of the cost of taking different links at different time-steps as well as a tax penalty/incentive term computed by the TSO. At a given time, the tax associated with a particular link depends on the log-population of drivers traveling on that link. Our problem set up is motivated by the *mechanism design* problem in which the TSO's objective is to keep the empirical distribution of agents over the links at each time as close as possible to a target distribution.

As the main technical contribution of this paper, we prove that the MFE of the game described above is given by the solution to a linearly solvable MDP. We emphasize that the MFE in our setting can be computed by performing a sequence of matrix multiplications backward in time *only once*, without any need of forward-in-time computations. This is a stark contrast to the standard MFG results where there is a need to solve a forward-backward HJB-FPK system, which is often a non-trivial task [10]. To highlight this computational advantage, we restrict ourselves to the discrete-time, discrete-space setting. Our result is different from [22] although the entropy-like cost considered there is similar to the Kullback-Leibler cost that appears in our analysis. The game considered in [22] involves only a fixed number of players (which can be thought of as routing policies at intersections rather than infinitesimal drivers) and no mean-field limit is considered.

II. Problem Formulation

Let the directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ (referred to as the *traffic graph*) represent the topology of the underlying traffic network, where $\mathcal{V} = \{1, 2, ..., V\}$ is the set of nodes (intersections) and $\mathcal{E} = \{1, 2, ..., E\}$ is the set of directed edges (links). For each $i \in \mathcal{V}$, denote by $\mathcal{V}(i) \subseteq \mathcal{V}$ the set of intersections to which there is a directed link from the intersection i. Let $\mathcal{N} = \{1, 2, ..., N\}$ be the set of players (drivers) on the graph $\mathcal{G}$. At any given time step $t \in \mathcal{T} = \{0, 1, ..., T-1\}$, each player is located at an intersection. The node at which the n-th player is located at time step t is denoted by $i_{n,t} \in \mathcal{V}$. At every time step, player n selects an *action* $j_{n,t} \in \mathcal{V}(i_{n,t})$, which represents her next destination. By selecting $j_{n,t}$ at time t, the player n moves to the node $j_{n,t}$ at time $t+1$ deterministically (i.e., $i_{n,t+1} = j_{n,t}$). Let $P_0 = \{P_0^i\}_{i \in \mathcal{V}}$ be a probability mass function over $\mathcal{V}$. At $t = 0$, we assume $i_{n,0}$ for each $n \in \mathcal{N}$ is realized independently with probability distribution P_0.

A. Players' strategies

Each player traverses on $\mathcal{G}$ according to her individual randomized policy. More precisely, for each $n \in \mathcal{N}$, $i \in \mathcal{V}$ and $t \in \mathcal{T}$, let $Q_{n,t}^i = \{Q_{n,t}^{ij}\}_{j \in \mathcal{V}(i)}$ be the decision policy of player n at the intersection i at time t, representing a probability distribution according to which she selects the next destination $j \in \mathcal{V}(i)$. We consider the collection $Q_{n,t} = \{Q_{n,t}^i\}_{i \in \mathcal{V}}$ of such probability distributions as the strategy of player n at time t. Namely, for each $n \in \mathcal{N}$ and $t \in \mathcal{T}$, we have $Q_{n,t} \in \mathcal{Q}$, where

$$\mathcal{Q} = \left\{ \{Q^{ij}\}_{i \in \mathcal{V}, j \in \mathcal{V}(i)} : \begin{array}{l} Q^{ij} \geq 0 \;\; \forall i \in \mathcal{V}, j \in \mathcal{V}(i) \\ \sum_{j \in \mathcal{V}(i)} Q^{ij} = 1 \;\; \forall i \in \mathcal{V} \end{array} \right\}$$

is the space of strategies. As clarified below, in this paper we consider a game with the open-loop information pattern [23].

If the strategy $\{Q_{n,t}\}_{t \in \mathcal{T}}$ of player n is fixed, then the probability distribution $P_{n,t} = \{P_{n,t}^i\}_{i \in \mathcal{V}}$ of her location at time t is recursively computed by

$$P_{n,t+1}^j = \sum_i P_{n,t}^i Q_{n,t}^{ij} \;\; \forall t \in \mathcal{T}, j \in \mathcal{V} \tag{1}$$

with the initial condition $P_{n,0} = P_0$. If $(i_{n,t}, j_{n,t})$ is the location-action pair of player n at time t, it has a joint distribution $P_{n,t}^i Q_{n,t}^{ij}$, and it is statistically independent of $(i_{m,t}, j_{m,t})$ with $m \neq n$.

B. Action costs

For each $i \in \mathcal{V}$, $j \in \mathcal{V}(i)$ and $t \in \mathcal{T}$, let C_t^{ij} be a given constant that represents the cost (e.g., fuel cost) incurred to each agent who takes action j at location i at time t. We will also introduce the terminal cost C_T^i, $i \in \mathcal{V}$ for each player arriving at state i at the final time step $t = T$.

C. Tax mechanisms and incentives

We assume that players are also subject to individual and time-varying tax penalties calculated by the TSO. Individual tax values depend not only on the players' locations and actions, but also on how the entire population is distributed over the traffic graph $\mathcal{G}$. Specifically, we consider the following log-population tax mechanism, where the tax charged to player n taking action j at location i at time t is

$$\pi_{N,t,n}^{ij} = \alpha \left(\log \frac{K_{N,t}^{ij}}{K_{N,t}^i} - \log R_t^{ij} \right). \tag{2}$$

In (2), $\alpha > 0$ is a fixed constant. The parameter $R_t^{ij} > 0$ is also a fixed constant satisfying $\sum_j R_t^{ij} = 1$ for each i. R_t^{ij} can be interpreted as the reference policy (state transition probability) designated by the TSO in advance. $K_{N,t}^i$ is the number of agents (including player n) who are located at the intersection i at time t, and $K_{N,t}^{ij}$ is the number of agents (including player n) who takes the action j at the intersection i at time t. The tax rule (2) indicates that agent n receives a positive payment by taking action j at location i at time t if $K_{N,t}^{ij}/K_{N,t}^i < R_t^{ij}$, while she is penalized by doing so if $K_{N,t}^{ij}/K_{N,t}^i > R_t^{ij}$. Since $K_{N,t}^i$ and $K_{N,t}^{ij}$ are random variables, $\pi_{N,t,n}^{ij}$ is also a random variable. We assume that the TSO is able to observe $K_{N,t}^i$ and $K_{N,t}^{ij}$ at every time step and hence $\pi_{N,t,n}^{ij}$ is computable.[1]

[1] Whenever $\pi_{N,t,n}^{ij}$ is computed, we have both $K_{N,t}^{ij} \geq 1$ and $K_{N,t}^i \geq 1$ since at least player n herself is counted. Hence (2) is well-defined.

Since player n's probability of taking action j at location i at time step t is given by $P^i_{n,t}Q^{ij}_{n,t}$, the total number $K^{ij}_{N,t}$ of such players follows the Poisson binomial distribution

$$\Pr(K^{ij}_{N,t}=k)=\sum_{A\in F_k}\prod_{n\in A}P^i_{n,t}Q^{ij}_{n,t}\prod_{n^c\in A^c}(1-P^i_{n^c,t}Q^{ij}_{n^c,t}).$$

Here, F_k is the set of all subsets of size k that can be selected from $\mathcal{N}=\{1,2,...,N\}$, and $A^c=F_k\backslash A$. Similarly, the distribution of $K^i_{N,t}$ is given by

$$\Pr(K^i_{N,t}=k)=\sum_{A\in F_k}\prod_{n\in A}P^i_{n,t}\prod_{n^c\in A^c}(1-P^i_{n^c,t}).$$

Notice also that the conditional probability distribution of $K^{ij}_{N,t}$ given player n's location-action pair $(i_{n,t},j_{n,t})=(i,j)$ is

$$\begin{aligned}&\Pr(K^{ij}_{N,t}=k+1\mid i_{n,t}=i,j_{n,t}=j)\\&=\sum_{A\in F_k^{-n}}\prod_{m\in A}P^i_{m,t}Q^{ij}_{m,t}\prod_{m^c\in A^{-c}}(1-P^i_{m^c,t}Q^{ij}_{m^c,t}).\end{aligned}\tag{3}$$

Here, F_k^{-n} is the set of all subsets of size k that can be selected from $\mathcal{N}\backslash\{n\}$, and $A^{-c}=F_k^{-n}\backslash A$. Similarly, the conditional probability distribution of $K^i_{N,t}$ given $i_{n,t}=i$ is

$$\begin{aligned}&\Pr(K^i_{N,t}=k+1\mid i_{n,t}=i)\\&=\sum_{A\in F_k^{-n}}\prod_{m\in A}P^i_{m,t}\prod_{m^c\in A^{-c}}(1-P^i_{m^c,t}).\end{aligned}\tag{4}$$

Therefore, given the prior knowledge that the player n's location-action pair at time t is (i,j), the expectation of her tax penalty $\pi^{ij}_{N,n,t}$ is

$$\begin{aligned}&\Pi^{ij}_{N,n,t}\triangleq\mathbb{E}\left[\pi^{ij}_{N,n,t}\mid i_{n,t}=i,j_{n,t}=j\right]\\&=\sum_{k=0}^{N-1}\alpha\log\frac{k+1}{N}\sum_{A\in F_k^{-n}}\prod_{m\in A}P^i_{m,t}Q^{ij}_{m,t}\prod_{m^c\in A^c}(1-P^i_{m^c,t}Q^{ij}_{m^c,t})\\&-\sum_{k=0}^{N-1}\alpha\log\frac{k+1}{N}\sum_{A\in F_k^{-n}}\prod_{m\in A}P^i_{m,t}\prod_{m^c\in A^c}(1-P^i_{m^c,t})\\&-\alpha\log R^{ij}_t.\end{aligned}\tag{5}$$

Notice that the quantity (5) depends on the strategies $Q_{-n}\triangleq\{Q_m\}_{m\neq n}$, but not on Q_n. In other words, $\pi^{ij}_{N,n,t}$ is a random variable whose distribution does not depend on player n's own strategy. This fact will be used in Section IV.

D. Road traffic game

The N-player dynamic game considered in this paper is now formulated as follows.

1) State dynamics: We consider the probability distribution $P_{n,t}$ as the state of player n at time t, and $Q_{n,t}$ as her control input. Each individual's state dynamics is governed by (1). Notice that different players' dynamics are decoupled.

2) Cost functionals: The n-th player's cost functional is given by

$$\begin{aligned}&J\left(\{Q_{n,t}\}_{t\in\mathcal{T}},\{Q_{-n,t}\}_{t\in\mathcal{T}}\right)\\&=\sum_{t=0}^{T-1}\sum_{i,j}P^i_{n,t}Q^{ij}_{n,t}\left(C^{ij}_t+\Pi^{ij}_{N,n,t}\right)+\sum_i P^i_{n,T}C^i_T.\end{aligned}$$

Notice that this quantity depends not only on the n-th player's own strategy $\{Q_{n,t}\}_{t\in\mathcal{T}}$ but also on the other players' strategies $\{Q_{-n,t}\}_{t\in\mathcal{T}}$ through the term $\Pi^{ij}_{N,n,t}$, whose precise expression is given by (5).

3) Information pattern: Throughout this paper, we restrict our analysis to the open-loop information pattern [23]. More precisely, each player n must fix a sequence of strategies $\{Q_{n,t}\}_{t\in\mathcal{T}}$ in advance based only on the public knowledge $\mathcal{G},\alpha,\mathcal{N},\mathcal{T},R^{ij}_t,C^{ij}_t,C^i_T$ and P_0. Players are not allowed to update their strategies in real-time based on the observations $\{K^{ij}_{N,t}\}_{t\in\mathcal{T}}$.[2]

4) Solution concept: We introduce the following equilibrium concepts for the game described above.

Definition 1: The N-tuple of strategies $\{Q^{\text{NE}}_{n,t}\}_{n\in\mathcal{N},t\in\mathcal{T}}$ is said to be an (open-loop) *Nash equilibrium* if the following inequality holds for each $n\in\mathcal{N}$ and $\{Q_{n,t}\}_{t\in\mathcal{T}},Q_{n,t}\in\mathcal{Q}$:

$$J\left(\{Q_{n,t}\}_{t\in\mathcal{T}},\{Q^{\text{NE}}_{-n,t}\}_{t\in\mathcal{T}}\right)\geq J\left(\{Q^{\text{NE}}_{n,t}\}_{t\in\mathcal{T}},\{Q^{\text{NE}}_{-n,t}\}_{t\in\mathcal{T}}\right).$$

Definition 2: A Nash equilibrium $\{Q^{\text{NE}}_{n,t}\}_{n\in\mathcal{N},t\in\mathcal{T}}$ is said to be *symmetric* if

$$\{Q^{\text{NE}}_{1,t}\}_{t\in\mathcal{T}}=\{Q^{\text{NE}}_{2,t}\}_{t\in\mathcal{T}}=\cdots=\{Q^{\text{NE}}_{N,t}\}_{t\in\mathcal{T}}.$$

Remark 1: The N-player game described above is a *symmetric game* in the sense of [24]. Thus, [24, Theorem 3] is applicable to show that it has a symmetric equilibrium.

Definition 3: A set of strategies $\{Q^{\text{MFE}}_{n,t}\}_{n\in\mathcal{N},t\in\mathcal{T}}$ is said to be an *MFE* if the following conditions are satisfied.

(a) It is symmetric, i.e.,

$$\{Q^{\text{MFE}}_{1,t}\}_{t\in\mathcal{T}}=\{Q^{\text{MFE}}_{2,t}\}_{t\in\mathcal{T}}=\cdots=\{Q^{\text{MFE}}_{N,t}\}_{t\in\mathcal{T}}.$$

(b) There exists a sequence $\{\epsilon_N\}$ satisfying $\epsilon_N\searrow 0$ as $N\to\infty$ such that for each $n\in\mathcal{N}=\{1,2,...,N\}$ and for each $\{Q_{n,t}\}_{t\in\mathcal{T}}$ with $Q_{n,t}\in\mathcal{Q}$,

$$\begin{aligned}&J\left(\{Q_{n,t}\}_{t\in\mathcal{T}},\{Q^{\text{MFE}}_{-n,t}\}_{t\in\mathcal{T}}\right)+\epsilon_N\\&\quad\geq J\left(\{Q^{\text{MFE}}_{n,t}\}_{t\in\mathcal{T}},\{Q^{\text{MFE}}_{-n,t}\}_{t\in\mathcal{T}}\right).\end{aligned}$$

III. Linearly Solvable MDPs

In this section, we introduce an auxiliary optimal control problem that is closely related to the road traffic game introduced in the previous section. The result in this section will serve as a tool to find an MFE in the road traffic game described above. The main emphasis in this section is that the introduced auxiliary optimal control problem belongs to the class of *linearly-solvable MDPs* [9], [25]. This fact provides

[2] In the future, we will consider a closed-loop implementation in which the open-loop optimization is performed repeatedly over the receding horizon.

a tremendous advantage in the computation of mean-field equilibria in the road traffic game.

For each $t = 0, ..., T$, let P_t be the probability distribution over the vertices $\mathcal{V}$ that evolves according to

$$P_{t+1}^{j} = \sum_i P_t^i Q_t^{ij} \ \ \forall j \in \mathcal{V}$$

with the initial state P_0. We consider P_t as the state of the dynamics and Q_t as the control action in the optimal control problem below. We assume C_t^{ij}, R_t^{ij} for each $t \in \mathcal{T}, i \in \mathcal{I}, j \in \mathcal{I}$, C_T^i for $i \in \mathcal{I}$, and α are given positive constants. The T-step optimal control problem of our interest is:

$$\min_{\{Q_t\}_{t\in\mathcal{T}}} \sum_{t=0}^{T-1} \sum_{i,j} P_t^i Q_t^{ij} \left(C_t^{ij} + \alpha \log \frac{Q_t^{ij}}{R_t^{ij}} \right) + \sum_i P_T^i C_T^i. \tag{6}$$

Notice that the logarithmic term in (6) can be written as the Kullback–Leibler divergence from the reference policy R_t^{ij} to the selected policy Q_t^{ij}. For each $t = 0, 1, ..., T$, introduce the value function:

$$V_t(P_t) \triangleq \min_{\{Q_\tau\}_{\tau=t}^{T-1}} \sum_{\tau=t}^{T-1} \sum_{i,j} P_\tau^i Q_\tau^{ij} \left(C_\tau^{ij} + \alpha \log \frac{Q_\tau^{ij}}{R_\tau^{ij}} \right) + \sum_i P_T^i C_T^i$$

and the associated Bellman equation

$$V_t(P_t) = \min_{Q_t} \left\{ \sum_{i,j} P_\tau^i Q_\tau^{ij} \left(C_\tau^{ij} + \alpha \log \frac{Q_\tau^{ij}}{R_\tau^{ij}} \right) + V_{t+1}(P_{t+1}) \right\}. \tag{7}$$

The next theorem states that the optimal control problem (6) is *linearly* solvable [9].

Theorem 1: Let $\{\phi_t\}_{t\in\mathcal{T}}$ be the sequence of V-dimensional vectors defined by the backward recursion

$$\phi_t^i = \sum_j R_t^{ij} \exp\left(-\frac{C_t^{ij}}{\alpha} \right) \phi_{t+1}^j \ \ \forall i \in \mathcal{V} = \{1, 2, ..., V\} \tag{8}$$

with the terminal condition $\phi_T^i = \exp(-C_T^i/\alpha) \ \forall i$. Then, for each $t \in \mathcal{T}$ and P_t, the value function can be written as

$$V_t(P_t) = -\alpha \sum_i P_t^i \log \phi_t^i. \tag{9}$$

Moreover, the optimal policy for (6) is given by

$$Q_t^{ij*} = \frac{\phi_{t+1}^j}{\phi_t^i} R_t^{ij} \exp\left(-\frac{C_t^{ij}}{\alpha} \right). \tag{10}$$

Proof: Due to the choice of the terminal condition $\phi_T^i = \exp(-C_T^i/\alpha)$, notice that

$$V_T(P_T) = \sum_i P_T^i C_T^i = -\alpha \sum_i P_T^i \log \phi_T^i.$$

Thus, (9) holds for $t = T$. To complete the proof by backward induction, assume that

$$V_{t+1}(P_{t+1}) = -\alpha \sum_i P_{t+1}^i \log \phi_{t+1}^i$$

holds for some $0 \le t \le T-1$. Then, due to the Bellman equation (7), we have

$$V_t(P_t) = \min_{Q_t} \left\{ \sum_{i,j} P_\tau^i Q_\tau^{ij} \left(\rho_\tau^{ij} + \alpha \log \frac{Q_\tau^{ij}}{R_\tau^{ij}} \right) \right\}$$

where $\rho_t^{ij} = C_t^{ij} - \alpha \log \phi_{t+1}^j$ is a constant. It is elementary to show that the minimum is attained by

$$Q_t^{ij*} = \frac{R_t^{ij}}{\phi_t^i} \exp\left(-\frac{\rho_t^{ij}}{\alpha} \right)$$

from which (10) follows. By substitution, the optimal value is shown to be

$$V_t(P_t) = -\alpha \sum_i P_t^i \log \phi_t^i.$$

This completes the induction proof. ■

We stress that (8) is linear in ϕ and can be computed by matrix multiplications backward in time.

IV. Mean Field Equilibrium

Let $\{Q_t^*\}_{t\in\mathcal{T}}$ be the control policy obtained in (10), and let $\{P_t^*\}_{t\in\mathcal{T}}$ be defined recursively by

$$P_{t+1}^{j*} = \sum_i P_t^{i*} Q_t^{ij*} \ \ \forall j \in \mathcal{V}.$$

In this section, we consider the situation in which all players other than n adopt the strategy $\{Q_t^*\}_{t\in\mathcal{T}}$, and then analyze player n's best response. For each $t \in \mathcal{T}$, the probability that the m-th player ($m \neq n$) is located at i is P_t^{i*}. As before, define $\Pi_{N,n,t}^{ij}$ as the expected value of the tax penalty charged on player n at time t when she takes action j at location i. Since players' dynamics over the traffic graph are decoupled, and $\Pi_{N,n,t}^{ij}$ is computed by the population excluding player n, $\Pi_{N,n,t}^{ij}$ does not depend on player n's strategy. Therefore, the best response by player n is characterized by the solution to the following optimal control problem:

$$\min_{\{Q_{n,t}\}_{t\in\mathcal{T}}} \sum_{t=0}^{T-1} \sum_{i,j} P_{n,t}^i Q_{n,t}^{ij} \left(C_t^{ij} + \Pi_{N,n,t}^{ij} \right) + \sum_i P_{n,T}^i C_T^i, \tag{11}$$

where $\Pi_{N,n,t}^{ij}$ can be considered as a fixed constant. To evaluate $\Pi_{N,n,t}^{ij}$ when all players other than player n takes the same strategy (i.e., $Q_{m,t} = Q_t^*$ for $m \neq n$), notice that the conditional distributions of $K_{N,t}^{ij}$ and $K_{N,t}^i$ given $(i_{n,t}, j_{n,t}) = (i, j)$, provided by (3) and (4), simplify to the binomial distributions

$$\begin{aligned}
&\Pr(K_{N,t}^{ij} = k+1 | i_{n,t} = i, j_{n,t} = j) \\
&\quad = \binom{N-1}{k} (P_t^{i*} Q_t^{ij*})^k (1 - P_t^{i*} Q_t^{ij*})^{N-1-k} \\
&\Pr(K_{N,t}^{i} = k+1 | i_{n,t} = i) \\
&\quad = \binom{N-1}{k} (P_t^{i*})^k (1 - P_t^{i*})^{N-1-k}.
\end{aligned}$$

Thus, the expression (5) simplifies to

$$\begin{aligned}\Pi^{ij}_{N,n,t} &= \mathbb{E}\left[\pi^{ij}_{N,n,t} \mid i_n = i, j_n = j\right] \\ &= \sum_{k=0}^{N-1} \alpha \log \frac{k+1}{N} \binom{N-1}{k} (P_t^{i*} Q_t^{ij*})^k (1-P_t^{i*} Q_t^{ij*})^{N-1-k} \\ &\quad - \sum_{k=0}^{N-1} \alpha \log \frac{k+1}{N} \binom{N-1}{k} (P_t^{i*})^k (1-P_t^{i*})^{N-1-k} \\ &\quad - \alpha \log R_t^{ij} \end{aligned} \tag{12}$$

A. *Optimal solution to* (11) *when* $N \to \infty$

Next, we study the asymptotic limit of $\Pi^{ij}_{N,n,t}$ as $N \to \infty$.

Lemma 1: Let $\Pi^{ij}_{N,n,t}$ be defined by (12). If $P_t^{i*} Q_t^{ij*} > 0$, then

$$\lim_{N \to \infty} \Pi^{ij}_{N,n,t} = \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}}.$$

Proof: See Appendix A. ■

Lemma 1 implies that in the limit $N \to \infty$, the optimal control problem (11) becomes

$$\min_{\{Q_{n,t}\}_{t \in \mathcal{T}}} \sum_{t=0}^{T-1} \sum_{i,j} P^i_{n,t} Q^{ij}_{n,t} \left(C_t^{ij} + \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \right) + \sum_i P^i_{n,T} C^i_T. \tag{13}$$

Notice that (13) is different from the auxiliary optimal control problem (6) studied in Section III in that the logarithmic term in (13) is a fixed constant that does not depend on the control policy. Nevertheless, these two optimal control problems are closely related as we show below. To solve (13), we once again apply dynamic programming. For each $P_{n,t}$, define the value function

$$V_{n,t}(P_{n,t}) \triangleq \min_{\{Q_{n,\tau}\}_{\tau=t}^{T}} \sum_{\tau=t}^{T-1} \sum_{i,j} P^i_{n,\tau} Q^{ij}_{n,\tau} \left(C_\tau^{ij} + \alpha \log \frac{Q_\tau^{ij*}}{R_\tau^{ij}} \right) + \sum_i P^i_{n,T} C^i_T$$

The value function satisfies the Bellman equation:

$$V_{n,t}(P_{n,t}) = \min_{Q_{n,t}} \left\{ \sum_{i,j} P^i_{n,t} Q^{ij}_{n,t} \left(C_t^{ij} + \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \right) + V_{n,t+1}(P_{n,t+1}) \right\}. \tag{14}$$

The next key lemma shows that $V_{n,t}(\cdot)$ coincide with the value function $V_t(\cdot)$ for (6). It also shows an interesting property of the optimal control problem (13) that *any feasible control policy is an optimal control policy.*

Lemma 2: Let $\{\phi_t\}_{t \in \mathcal{T}}$ be the sequence defined by (8).

(a) For each $t \in \mathcal{T}$ and $P_{n,t}$, we have

$$V_{n,t}(P_{n,t}) = -\alpha \sum_i P^i_{n,t} \log \phi^i_t.$$

(b) An arbitrary sequence of control actions $\{Q_{n,t}\}_{t \in \mathcal{T}}$ with $Q_{n,t} \in \mathcal{Q}$ is an optimal solution to (13).

Proof: (a). Proof is by backward induction. If $t = T$, the claim trivially holds due to the definition $V_{n,T}(P_T) = \sum_i P^i_{n,T} C^i_T$ and the fact that the terminal condition for (8) is given by $\phi^i_T = \exp(-C^i_T/\alpha)$. Thus, for $0 \le t \le T-1$, assume that

$$V_{n,t+1}(P_{n,t+1}) = -\alpha \sum_j P^j_{n,t+1} \log \phi^j_{t+1}$$

holds. Using $\rho_t^{ij} = C_t^{ij} - \alpha \log \phi^j_{t+1}$, the Bellman equation (14) can be written as

$$V_{n,t}(P_{n,t}) = \min_{Q_{n,t}} \sum_{i,j} P^i_{n,t} Q^{ij}_{n,t} \left(\rho_t^{ij} + \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \right). \tag{15}$$

Substituting Q_t^{ij*} obtained by (10) into (15), we have

$$\begin{aligned} V_{n,t}(P_{n,t}) &= \min_{Q_{n,t}} \sum_{i,j} P^i_{n,t} Q^{ij}_{n,t} \left(-\alpha \log \phi^i_t \right) \\ &= \min_{Q_{n,t}} \sum_i P^i_{n,t} \left(-\alpha \log \phi^i_t \right) \underbrace{\textstyle\sum_j Q^{ij}_{n,t}}_{=1} \qquad \text{(16a)} \\ &= -\alpha \sum_i P^i_{n,t} \log \phi^i_t. \qquad \text{(16b)} \end{aligned}$$

This completes the proof.

(b). Since the final expression (16b) does not depend on $Q_{n,t}$, any control action $Q_{n,t} \in \mathcal{Q}$ is a minimizer of the right hand side of the Bellman equation (14). ■

Lemma 2 (b) shows that an arbitrary policy is optimal in the optimal control problem (13). This result is a reminiscent of the *Wardrop's principle* [1] (see also [26] and references therein), which states that travel costs are equal on all used routes at the game-theoretic equilibrium among strategic and infinitesimal travelers.

B. *Mean field equilibrium*

We are now ready to state the main result of this paper. The next theorem, together with Theorem 1, provides a numerical method to compute an MFE of the road traffic game presented in Section II.

Theorem 2: A symmetric strategy profile $Q^{ij}_{n,t} = Q_t^{ij*}$ for each $n \in \mathcal{N}, t \in \mathcal{T}$ and $i, j \in \mathcal{V}$, where Q_t^{ij*} is obtained by (8)–(10), is an MFE of the road traffic game.

Proof: Let the policies $Q^{ij}_{m,t} = Q_t^{ij*}$ for $m \neq n$ be fixed. It is sufficient to show that there exists a sequence $\epsilon_N \searrow 0$ such that the cost of adopting a strategy $Q^{ij}_{n,t} = Q_t^{ij*}$ for player n is no greater than ϵ_N plus the cost of adopting any other policy. Since

$$\Pi^{ij}_{N,n,t} \to \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \text{ as } N \to \infty,$$

there exists a sequence $\delta_N \searrow 0$ such that

$$\Pi^{ij}_{N,n,t} + \delta_N > \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \;\; \forall i, j, t.$$

Now, for all policy $\{Q_{n,t}\}_{t\in\mathcal{T}}$ of player n and the induced distributions $P^j_{n,t} = \sum_i P^i_{n,t}Q^{ij}_{n,t}$, we have

$$\begin{aligned}
&\sum_{t=0}^{T-1}\sum_{i,j} P^i_{n,t}Q^{ij}_{n,t}\left(C^{ij}_t + \Pi^{ij}_{N,n,t}\right)\\
&> \sum_{t=0}^{T-1}\sum_{i,j} P^i_{n,t}Q^{ij}_{n,t}\left(C^{ij}_t + \alpha\log\frac{Q^{ij*}_t}{R^{ij}_t} - \delta_N\right)\\
&= \sum_{t=0}^{T-1}\sum_{i,j} P^i_{n,t}Q^{ij}_{n,t}\left(C^{ij}_t + \alpha\log\frac{Q^{ij*}_t}{R^{ij}_t}\right) - TV^2\delta_N\\
&\geq \min_{\{Q_{n,t}\}_{t\in\mathcal{T}}}\sum_{t=0}^{T-1}\sum_{i,j} P^i_{n,t}Q^{ij}_{n,t}\left(C^{ij}_t + \alpha\log\frac{Q^{ij*}_t}{R^{ij}_t}\right)\\
&\quad - TV^2\delta_N.
\end{aligned}$$

Notice that the minimization in the last line is attained by adopting $Q^{ij}_{n,t} = Q^{ij*}_t$. Since $\epsilon_N \triangleq TV^2\delta_N \searrow 0$, this completes the proof. ■

V. Numerical Illustration

In this section, we illustrate the result of Theorem 2 applied to a simple mean-field road traffic game over a traffic graph with 100 nodes (a grid world with obstacles) shown in Fig. 1 and over time horizon $T = 70$. At $t = 0$, the population is concentrated in the origin cell (indicated by "O"). For each player n, the terminal cost is given by $C^i_T = 10\sqrt{\text{dist}(i, \text{D})}$, where $\text{dist}(i, \text{D})$ is the Manhattan distance between the player's final location i and the destination cell (indicated by "D"). For each time step t, the action cost for each player is given by

$$C^{ij}_t = \begin{cases} 0 & \text{if } j = i\\ 1 & \text{if } j \in \mathcal{V}(i)\\ 100000 & \text{if } j \notin \mathcal{V}(i) \text{ or } j \text{ is an obstacle.}\end{cases}$$

where $\mathcal{V}(i)$ contains the north, east, south, and west neighborhood of the cell #i. As the reference distribution, we use $R^{ij}_t = 1/|\mathcal{V}(i)|$ (uniform distribution) for each $i \in \mathcal{V}$ and $t \in \mathcal{T}$ to incentivize players to spread over the traffic graph.

For various values of $\alpha > 0$, the backward formula (8) is solved and the optimal policy is calculated by (10). If α is small (e.g., $\alpha = 0.1$), it is expected that players will take the shortest path since the action cost is dominant compared to the tax cost (2). Numerical results confirm this intuition; three figures in the top row of Fig. 1 show snapshots of the population distribution at time steps $t = 20, 35$ and 50, assuming all the players take the MFE policy obtained by (10). In the bottom row, similar plots are generated with a larger α ($\alpha = 1$). In this case, it can be seen that the equilibrium strategy will choose longer paths with higher probability to reduce congestion.

VI. Conclusion and Future Work

In this paper, we showed that the mean-field approximation of a large-population road traffic game under the log-population tax mechanism can be obtained by the linearly solvable MDP. This result will serve as the basis for further research in the future. For instance, the close-loop implementation of the considered game (similar to the receding horizon implementation of model predictive control) and the corresponding feedback Nash equilibria are worthwhile to study. How the obtained results in this paper can be used in the mechanism design problems should also be investigated in the future. For instance, in this paper we have not discussed how the reference policy R^{ij}_t should be chosen by the TSO.

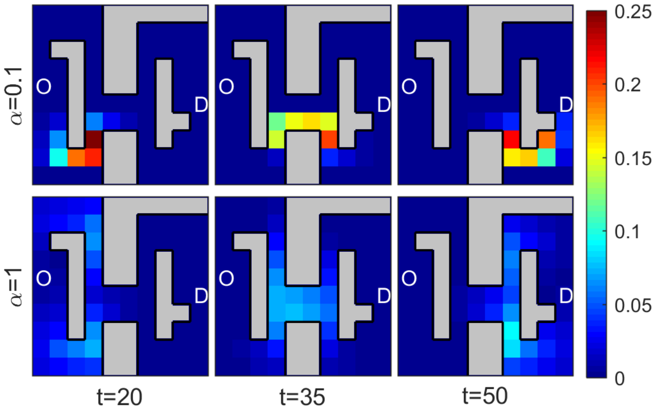

Fig. 1. Simulation results for road traffic game at $t = 20, 35, 50$ and for $\alpha = 0.1$ and 1.

Acknowledgement

The authors would like to thank Mr. Matthew T. Morris at the University of Texas at Austin for his contributions to the numerical study in Section V.

Appendix

A. Proof of Lemma 1

Let $K^i_{N,-n,t}$ denote the number of agents, except agent n, which are located at intersection i at time t and let $K^{ij}_{N,-n,t}$ denote the number of agents, except agent n, which are located at intersection i at time t and select intersection j as their next destination. Thus, we have $K^i_{N,-n,t} = \sum_{l\neq n} 1\{i_{l,t} = i\}$ and $K^{ij}_{N,-n,t} = \sum_{l\neq n} 1\{i_{l,t} = i, j_{l,t} = j\}$, where $1\{\cdot\}$ is the indicator function. Then, $\Pi^{ij*}_{N,n,t}$ can be written as

$$\begin{aligned}
\Pi^{ij*}_{N,n,t} =& \mathbb{E}\left[\log\frac{1+K^{ij}_{N,-n,t}}{1+K^{i}_{N,-n,t}}\right] - \log R^{ij}_t\\
=& \mathbb{E}\left[\log\left(\frac{1+K^{ij}_{N,-n,t}}{N}\right)\right] -\\
& \mathbb{E}\left[\log\left(\frac{1+K^{i}_{N,-n,t}}{N}\right)\right] - \log R^{ij}_t
\end{aligned}$$

Using Jensen inequality, we have

$$\begin{aligned}
\mathbb{E}\left[\log\left(\frac{1+K^{ij}_{N,-n,t}}{N}\right)\right] &\leq \log\left(\frac{1}{N} + \mathbb{E}\left[\frac{K^{ij}_{N,-n,t}}{N}\right]\right)\\
&\overset{(a)}{=} \log\left(\frac{1}{N} + \frac{N-1}{N}P^{i*}_t Q^{ij*}_t\right)
\end{aligned}$$

where (a) follows from $\mathbb{E}\left[\frac{K^{ij}_{N,-n,t}}{N}\right] = \frac{N-1}{N} P_t^{i*} Q_t^{ij*}$ and the fact that all the agents employ the policy $\{Q_t^*\}$. Thus, we have

$$\limsup_{N\to\infty} \mathbb{E}\left[\log \frac{1+K^{ij}_{N,-n,t}}{N}\right] \leq \log P_t^{i*} Q_t^{ij*} \qquad (17)$$

Next we show the other direction. For $\epsilon \in (0,\ \frac{P_t^{i*}Q_t^{ij*}}{2}]$, we can write $\mathbb{E}\left[\log \frac{1+K^{ij}_{N,-n,t}}{N}\right]$ as

$$\begin{aligned}\mathbb{E}\left[\log \frac{1+K^{ij}_{N,-n,t}}{N}\right] =& \mathbb{E}\left[\log \frac{1+K^{ij}_{N,-n,t}}{N} 1\left\{\frac{K^{ij}_{N,-n,t}}{N} > \epsilon\right\}\right] + \\ & \mathbb{E}\left[\log \frac{1+K^{ij}_{N,-n,t}}{N} 1\left\{\frac{K^{ij}_{N,-n,t}}{N} \leq \epsilon\right\}\right]\end{aligned}$$

Using the Hoeffding inequality, it follows that $\frac{K^{ij}_{N,-n,t}}{N}$ converges to $P_t^{i*}Q_t^{ij*}$ in probability as N becomes large. From continues mapping theorem, we have the convergence of $\log \frac{K^{ij}_{N,-n,t}}{N}$ in probability to $\log P_t^{i*}Q_t^{ij*}$ for $P_t^{i*}Q_t^{ij*} > 0$. Similarly, $1\left\{\frac{K^{ij}_{N,-n,t}}{N} > \epsilon\right\}$ converges to 1 in probability. Thus, from Slutsky's Theorem, we have $\log \frac{1+K^{ij}_{N,-n,t}}{N} 1\left\{\frac{K^{ij}_{N,-n,t}}{N} > \epsilon\right\}$ converges to $P_t^{i*}Q_t^{ij*}$ in distribution. Using Fatou's lemma and the fact that $\log \frac{1+K^{ij}_{N,-n,t}}{N} 1\left\{\frac{K^{ij}_{N,-n,t}}{N} > \epsilon\right\} \geq \log \epsilon$, we have

$$\liminf_{N\to\infty} \mathbb{E}\left[\log \frac{1+K^{ij}_{N,-n,t}}{N} 1\left\{\frac{K^{ij}_{N,-n,t}}{N} > \epsilon\right\}\right] \geq \log P_t^{i*}Q_t^{ij*}$$

We also have

$$\begin{aligned}&\left|\mathbb{E}\left[\log \frac{1+K^{ij}_{N,-n,t}}{N} 1\left\{\frac{K^{ij}_{N,-n,t}}{N} \leq \epsilon\right\}\right]\right| \\ &\leq \log(N)\, \mathsf{Pr}\left(\frac{K^{ij}_{N,-n,t}}{N} \leq \epsilon\right)\end{aligned}$$

Using the Hoeffding inequality, it is straightforward to show that $\mathsf{Pr}\left(\frac{K^{ij}_{N,-n,t}}{N} \leq \epsilon\right)$ decays to zero exponentially in N which implies that

$$\lim_{N\to\infty} \mathbb{E}\left[\log \frac{1+K^{ij}_{N,-n,t}}{N} 1\left\{\frac{K^{ij}_{N,-n,t}}{N} \leq \epsilon\right\}\right] = 0$$

Thus, we have

$$\liminf_{N\to\infty} \mathbb{E}\left[\log \frac{1+K^{ij}_{N,-n,t}}{N}\right] \geq \log P_t^{i*}Q_t^{ij*} \qquad (18)$$

which implies that $\lim_{N\to\infty} \mathbb{E}\left[\log\left(\frac{1+K^{ij}_{N,-n,t}}{N}\right)\right] = \log P_t^{i*}Q_t^{ij*}$. Following similar steps, it is straightforward to show that $\lim_{N\to\infty} \mathbb{E}\left[\log\left(\frac{1+K^{i}_{N,-n,t}}{N}\right)\right] = \log P_t^{i*}$ which completes the proof.

References

[1] J. G. Wardrop, "Some theoretical aspects of road traffic research," in *Inst Civil Engineers Proc London/UK/*, 1952.

[2] M. Patriksson, *The Traffic Assignment Problem: Models and Methods*. Dover Publications, 2015.

[3] P. E. Caines, M. Huang, and R. Malhamé, "Mean field games," *in Handbook of Dynamic Game Theory (T. Başar, G. Zaccour, Eds.)*, 2018.

[4] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese journal of mathematics*, vol. 2, no. 1, pp. 229–260, 2007.

[5] G. Chevalier, J. L. Ny, and R. Malhame, "A micro-macro traffic model based on mean-field games," *2015 American Control Conference (ACC)*, pp. 1983–1988, July 2015.

[6] D. Bauso, X. Zhang, and A. Papachristodoulou, "Density flow in dynamical networks via mean-field games," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1342–1355, March 2017.

[7] A. Lachapelle and M.-T. Wolfram, "On a mean field game approach modeling congestion and aversion in pedestrian crowds," *Transportation research part B: methodological*, vol. 45, no. 10, pp. 1572–1589, 2011.

[8] C. Dogbé, "Modeling crowd dynamics by the mean-field limit approach," *Mathematical and Computer Modelling*, vol. 52, no. 9-10, pp. 1506–1520, 2010.

[9] E. Todorov, "Linearly-solvable markov decision problems," in *Advances in neural information processing systems*, 2007, pp. 1369–1376.

[10] Y. Achdou and I. Capuzzo-Dolcetta, "Mean field games: numerical methods," *SIAM Journal on Numerical Analysis*, vol. 48, no. 3, pp. 1136–1162, 2010.

[11] M. Huang, "Large-population LQG games involving a major player: The nash certainty equivalence principle," *SIAM Journal on Control and Optimization*, vol. 48, no. 5, pp. 3318–3353, 2010.

[12] J. Huang, X. Li, and T. Wang, "Mean-field linear-quadratic-gaussian (lqg) games for stochastic integral systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2670–2675, Sept 2016.

[13] J. Moon and T. Basar, "Linear quadratic risk-sensitive and robust mean field games," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1062–1077, March 2017.

[14] M. Huang, "Mean field stochastic games with discrete states and mixed players," *International Conference on Game Theory for Networks*, pp. 138–151, 2012.

[15] Y. Wang, F. R. Yu, H. Tang, and M. Huang, "A mean field game theoretic approach for security enhancements in mobile ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1616–1627, March 2014.

[16] H. Tembine and M. Huang, "Mean field difference games: Mckean-vlasov dynamics," *The 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 1006–1011, Dec 2011.

[17] D. Bauso, H. Tembine, and T. Başar, "Robust mean field games," *Dynamic Games and Applications*, vol. 6, no. 3, pp. 277–303, Sep 2016.

[18] Q. Zhu, H. Tembine, and T. Basar, "Hybrid risk-sensitive mean-field stochastic differential games with application to molecular biology," *The 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 4491–4497, Dec 2011.

[19] H. Tembine, Q. Zhu, and T. Başar, "Risk-sensitive mean-field games," *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 835–850, 2014.

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

[21] G. Y. Weintraub, L. Benkard, and B. Van Roy, "Oblivious equilibrium: A mean field approximation for large-scale dynamic games," *Advances in neural information processing systems*, pp. 1489–1496, 2006.

[22] 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.

[23] T. Basar and G. Olsder, *Dynamic Noncooperative Game Theory*. Society for Industrial and Applied Mathematics, 1999.

[24] S.-F. Cheng, D. M. Reeves, Y. Vorobeychik, and M. P. Wellman, "Notes on equilibria in symmetric games," 2004.

[25] K. Dvijotham and E. Todorov, "A unified theory of linearly solvable optimal control," *Proceedings of Uncertainty in Artificial Intelligence (UAI)*, 2011.

[26] J. R. Correa and N. E. Stier-Moses, "Wardrop equilibria," *Wiley encyclopedia of operations research and management science*, 2011.