2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Dynamic Routing Games: An Evolutionary Game Theoretic Approach

Hamidou Tembine and Amar Prakash Azad

***Abstract*— We consider a dynamic routing problem where the objective of each user is to obtain flow policy that minimizes its long term cost. The framework differs from other related works which consider collection of static one shot games with dynamic cost function. Instead, we motivate our problem from the two basic facts: i) the path cost may not be exactly known in advance in dynamic environment unlike static; ii) long term solution is important aspect to evaluate rather than obtaining one slot solution. Moreover, this constraint inhibits to apply traditional game theoretic approach to obtain equilibria, rather we discuss that it is not required to obtain equilibria at every slot to "cover" the dynamic environment. In this work we propose an evolutionary game theoretic approach, we intend to learn the optimal strategy exploiting the past experiences (information) instead of cost function. Further, we characterize the dynamic equilibria of the long-term game using evolutionary variational inequalities. The dynamic equilibria so obtained, optimizes the long term cost, however it need not to be an equilibrium for intermediate epochs (games). As a byproduct, this reduces drastically the computation complexity. Under strictly monotone cost function, we prove that the dynamic equilibria are also dynamic evolutionarily stable strategies.**

I. INTRODUCTION

One of the motivations to consider dynamic scenarios in evolving networks is that they seem to show up in reality more often. Routing games is widely studied in last few years for "selfish-users" in a network of limited resources. These models are thoroughly explored in existing literature ([1], [2]). They apply to networks that involve large number of selfish users such as Internet routing, peer-to-peer file sharing systems, etc. However, in most of the studies a static network model is considered which includes a game which is framed over static network cost and static user demand. As the complexity of the existing system are growing up, we need to study and explore the dynamic behaviour of such systems which involve not only the time dependent network cost but also the demand varies over time. In recent past some studies are seen to be attracted with dynamic behavior, but most of them consider time dependent cost for a game framework and study the behavior over a period of time. However in reality when the system cost varies, it gives birth of another game than the prior one and hence the resulting game solutions may not depict the actual behavior the system. This complicates the modeling problem prohibitively difficult to use the static game approach.

Thus we seek a new approach to study these problems. In this paper we propose an evolutionary game theory approach (see [3] and the references therein) which not only allows us study the behavior of the system, rather we discover that our approach turns out to be less complex to study. In dynamic networks, where the traffic and topology changes over time, it may be difficult to compute the equilibria for each time epoch. Infact this is not only impossible to obtain, rather less important because continuum. However, the important thing is to compute here is the optimal strategy to be evolved over time, which leads to an optimal cost over a long time.

This also address the major challenge in the design of wireless networks, the need for distributed routing schemes that consume a minimal amount of information and a minimal amount resources. This is important in dynamic networks, where the traffic and topology can change over time, and therefore the strategies of players must be updated frequently. It is natural to ask whether there exist less-information, less memory schemes for dynamic wireless networks that could produce and maintain efficient and fair allocation. In this paper we consider dynamically changing traffic with fixed sources and destinations. Routing in wireless networks has been a rich area of enquiry over the last decade. The two main paradigms for routing have been geographic routing and topology based routing. We formulate the problem as dynamic routing game and examine both equilibria and optimality of network.

A. Related work

One of the most prominent learning algorithm in the general setting of dynamic games is fictitious play. Fictitious play have been studied by Brown (1951). The fictitious play procedure assumes that at each time slot, players choose a best reply to the observed empirical distribution of past actions of the others players. One of the obvious problem with this algorithm is the non transparency of the other player's actions. These algorithm requires significant amount of information about others players. The assumption that players can observe the past actions of the others players is *too strong* in the context of routing games which involves many players with bounded rationality and limited observation opportunities. An alternative learning scheme has been proposed in [4]–[6] which is based on non-regret choices of players with partial information but the algorithm uses rationality assumption which is too restrictive in networking scenarios. It requires to keep track of previous history which is resource consuming as well. We will discuss on section III-A how these conditions can be weaken.

One of the motivations to consider dynamic scenarios in evolving networks is that they seem to show up in reality more often. The motivation of dynamic routing games comes from several remarks:

A. P. Azad is with UCSC, USA (amar.azad@ieee.org)
H. Tembine is with Ecole Supérieure d'Electricité (Supelec), France (tembine@ieee.org)

- A time-dependent traffic model with variable costs is needed to capture more the variability of network traffic. The demand for resource allocation evolves in time and hence, the proposed framework should take into consideration the time-dependent constraints.
- The traffic in the network can change quickly, therefore the assumption of static model studied in [2], [7], [8] is not adapted in evolving traffic networks.

The paper is structured as follows: We start with generic system model for routing games. We first discuss in routing for finite number of player in Section III and discuss about learning dynamics in dynamic environment setting. We show that players can learn the strategy to reach equilibrium and even without having similar learning scheme. We then extend the study to mean field limits and show various important properties there in. Further, we extend to non-atomic type and multi class users in sec. IV. Finally, we conclude with remarks in sec. V. We omit some proof in the paper due to space limitation (refer to detailed report at arxiv.org).

B. Our Contribution

In this paper we emphasize the following contributions

- We develop a stochastic learning procedure for dynamic atomic routing games in absence of payoff function knowledge. Based on Boltzmann-Gibbs dynamics [6], [9], [10] and some basic properties of the entropy penalty function, we show that the learning mechanism converges almost surely to equilibria for parallel links. Further more, learning need not to be restricted to Boltzman-Gibbs dynamics procedure. Based on combined fully distributed payoff and strategy reinforcement learning (CODIPAS-RL, [6], [10]), we show that even all the players need not to have same learning mechanism which is more realistic in large network scenarios.
- We then focus on the transition from microscopic to macroscopic routing game via mean field interaction. We derive non-atomic dynamic routing games, their evolutionary dynamics [3] and mean field limit dynamics.
- For non-atomic users, we analyze the dynamic routing game using evolutionary variational inequalities introduced in [11]. This the first attempt to time-dependent strategies in dynamic routing games using evolutionary variational inequalities. Moreover our model does not need any information about the others players and have less memory requirement. Note that, we do not use punishment-reputation based mechanism because the punishment mechanism is not justified in particular in the context of large networks (detection of deviants , detection of the identity of players may fail, anonymity per class etc).

II. System Model

Consider a network $(\mathcal{V},\mathcal{L})$, where $\mathcal{V}$ is a finite set of nodes and $\mathcal{L}\subseteq\mathcal{V}\times\mathcal{V}$ is a set of directed links. A set $\mathcal{I}=\{1,2,...,I\}$ of users[1] share the network $(\mathcal{V},\mathcal{L})$. Let $\mathcal{R}$ be the set of possible routes in the network. We shall assume that all users ship flow from source node "s" to a common destination "d". User i has a throughput demand that is some process with average rate d^i. For simplicity of notation and without loss of generality, we assume that at most one link exists between each pair of nodes (in each direction). For any link $l=(u,v)\in\mathcal{L}$. Considering a node $v\in\mathcal{V}$, let $\text{In}(v)=\{l: D(l)=v\}$ denote the set of its in-going links, and $\text{Out}(v)=\{l: S(l)=v\}$ the set of its out-going links. A job with a given source-destination pair arrives in the system at s and leaves it at d after visiting a series of nodes and links, which we refer to as a route or path, then it leaves the system.

[1]A node can also act as one or more users.

III. Discrete time routing games with finite number of players

Routing in discrete time with finite number of player is originally discussed in transportation problems [12]. Such settings also show up in context of networks where flow of a user is un-splittable [13]. In this section we start our discussion from learning algorithm for discrete time routing with finite number of player, latter we show that it can be studied using continuous time dynamics of eq. (2). Borrowing simple tools from [14] we show that such dynamics converges to equilibria almost surely. In other words we propose simple (Boltzmann-gibbs based) learning algorithm which converges to equilibria in *dynamic environment setting*. Further, we carrying over the discussion for large population where using mean field limits we show the convergence to equilibrium even when player do not have same learning algorithms. In large population scenario it is more practical to consider that player may not have same learning algorithm.

Policies in the dynamic routing game: We introduce a few notations here in particular related to dynamic environment setting. Denote $\mathcal{X}^i$ be the set of mixed strategies of player i of the one-shot game. In a *dynamic environment*, where long time system observation is required, we seek notion of time dependent flow which include the history of policies and payoff. Therefore, we define history-dependent policies in which we will built our learning algorithm based on own-experience and own-payoff observations. A *private history* of length t of player i is a collection $\{(r^i_{t'},C^i_{r^i_{t'}}),\ t'\leq t\}$. Denote $\mathcal{H}^i_t$, the set of histories of length t of player i and, $\mathcal{H}^i=\bigcup_{t\geq 1}\mathcal{H}^i_t$. A *behavioral strategy* of player i is a collection of maps $(\sigma^i_t)_{t\geq 1}$, where $\sigma^i_t:\ \mathcal{H}^i_t\longrightarrow\mathcal{X}^i$. A strategy σ^i is said pure if for any time t, $\sigma^i_t\in\mathcal{R}$. A mixed strategy of player i in the dynamic routing game is a probability distribution over the pure strategies. A stationary strategy is a time-independent behavioral strategy $\forall t,\ \sigma^i_t=\sigma^i\in\mathcal{X}^i$. A general strategy is a distribution over behavioral strategies. Using Aumann's generalization of Kuhn theorem [15], we restrict our attention to behavioral strategies. We believe that this class of strategies is large enough (time and experiences dependent can be explored using this class). Note that our framework of dynamic routing game differs from a standard repeated game with complete information where it is assumed that players observe the other's payoffs or the actions of the others players after each time slot. We believe that such an assumption is too strong in network routing context and may need a feedback or a central coordinator. We do not assume any information about other player's strategy. However, one's strategy in dynamic game induces a probability measure over the set of histories of infinite length. For any expectation on the time average payoff (Cesaro mean payoff), we will refer to this induced probability on the product topology.

A. Learning in dynamic routing games

As pointed out earlier, we assume that payoff function is not known to players. Moreover, the variation in payoff reflects the changes in the dynamic environment. Therefore, this justifies that it is difficult to know the payoff in advance rather the players need to learn the payoff during the game, thence their strategy. In this section, we discuss a suitable payoff learning mechanism for dynamic environment.

We assume that players can have : i) an estimate of the average cost; and, ii) delay time of the alternative routes. Players make a decision based on this rough information by using a randomized decision rule to revise his strategy. The costs and time delays of the chosen alternative are then observed and is used to update the strategy for that particular route. Each player experiments only the costs and time delays of the selected route on that time slot, and uses this information to adjust his strategy for that particular route. This scheme is repeated every slot, generating a discrete time stochastic learning process. Basically we have three parameters: i) the state; ii) the rule for revision of strategies from the state to randomized actions; and, iii) a rule to update the new state. Although players observe only the costs of the specific route chosen on the given time slot, the observed values depend on the congestion levels determined by the others players choices revealing implicit information on the entire system. The natural question is whether such a learning algorithm based on a minimal piece of information may be sufficient to induce coordination [2] and make the system stabilize to an equilibrium. We show that the answer to this question is positive for dynamic routing games on:

- parallel links and a *Boltzmann-Gibbs dynamics* for route selection.
- general topology with monotone cost functions (in vectorial sense).

1) Finite number of players: We consider a finite set of players and finite set of routes. Denote by $C_r^w(k,t)$ the average $w-$weighted cost for the path r when k players chose this path at time t. The weight w simply depicts that the effective cost is the weighted sum of several cost depending on certain objective. For example, there can be a delay cost, memory cost,.., can be combined together with weight w. Again, weight w could also be different for different players due their objective. Henceforth, we omit w and work with generic cost $C_r(k,t)$ for simplicity of notation. An estimation of player i is a vector $C^i = (C^i_r)_r$ where C^i_r represents player i's estimate of the average cost of route r (the weighted cost composition). Player i update its strategy using the Boltzmann-Gibbs scheme: use route r with probability $p^i_r(C) := \frac{e^{-\frac{1}{\epsilon_i}C^i_r}}{\sum_{r'} e^{-\frac{1}{\epsilon_i}C^i_{r'}}}$, $\epsilon_i > 0$.

[2]Note that we do not assume coordination between players, there is not central authority and there is no signaling scheme to the players.

$$\begin{tikzcd}
C^i_r(k,t) \arrow[r, "Boltzman\ dynamics"] \arrow[d, "one-step"'] & p^i_r(t+1) \arrow[r, "realization"] & r^i_{t+1} \arrow[d, "random\ number"] \\
C^i_{r^i_{t+1}}(k_{r,t+1}, t+1) & & k_{r,t+1} \arrow[ll, "computation"', "estimation"]
\end{tikzcd}$$

Fig. 1. One-step of the learning algorithm.

Congestion is captured by the inequality $\forall t,\ C_r(k,t) \leq C_r(k+1,t),\ k \leq N$, where N is the total # of players. This implies, more the route is congested higher the weighted cost. Let ν_t be the step size satisfying $\sum_t \nu_t = +\infty,\ \sum_t \nu_t^2 < \infty$.
Learning Algorithm :

forall the *Players* **do**
  Initialize to some estimations $C^i_r(k_0, 0)$;
  Initial Boltzmann distribution $p^i(0)$;
end
for *t=1* **to** *max* **do** do;
foreach *Player* i **do**
  Observe its costs;
  Update via Boltzmann-Gibbs dynamics $p^i(t+1)$;
  Compute the distribution over r^i_{t+1} and $k_{r^i_{t+1}}$ from $p(t+1)$;
  Update its estimation via
  $C^i_{r,t+1} = C^i_{r,t} + \frac{\nu_{t+1}}{p^i_{r,t}} 1\!\!1_{\{r^i_{t+1}=r\}} \left(W^i_{r,t+1} - C^i_{r,t}\right)$;
  Estimate the random costs $C^i_{r^i_{t+1}}(k_{r^i_{t+1}}, t+1)$;
end

Algorithm 1: Stochastic Learning Algorithm based on Boltzmann-Gibbs dynamics

Description of the learning process: At stage $t+1$ the past estimation $C^i_r(k_{r,t}, t)$ determines the transition probabilities $p^i_r(t) = p^i_r(C^i_r(k_{r,t}, t))$ which are used by player i to experiment a random route r^i_{t+1}. The action profile of all the players determines a total random number $k_{r,t+1}$ of players i such that at route $r^i_{t+1} = r$. The weighted costs of r is then $C^i_{t+1}(k_{r^i_{t+1}}, t+1) = C^i_r(k, t+1)$ if $r^i_{t+1} = r, k_{r,t+1} = k$. Finally, each player i observes only the cost of the chosen alternative r^i_{t+1} and updates his/her estimations by averaging

$$\begin{aligned} C^i_{r,t+1} &= (1 - \frac{\nu_{t+1}}{p^i_{r,t}} 1\!\!1_{\{r^i_{t+1}=r\}}) C^i_{r,t} \\ &\quad + \frac{\nu_{t+1}}{p^i_{r,t}} 1\!\!1_{\{r^i_{t+1}=r\}} C^i_{r,t+1}, \end{aligned} \tag{1}$$

if $r^i_{t+1} = r$. Otherwise the estimation is unchanged: $C^i_{r,t+1} = C^i_{r,t}$ The diagram 1 illustrates one-step of the algorithm.

The learning algorithm can be rewritten as $C_{t+1} - C_t = \frac{\nu_{t+1}}{p_{r,t}} 1\!\!1_{\{r_{t+1}=r\}} (W_{t+1} - C_t)$ where $W^i_{r,t+1} = \begin{cases} C_{r,t+1} & if\ r^i_{t+1} = r \\ C^i_{r,t} & \text{otherwise} \end{cases}$

This process has the form of a stochastic learning algorithm [4], [5] with the distribution of the random vector W_t being determined by the individual updating rules which depend upon the estimations. Assuming that the cost functions are bounded, the sequences generated the learning algorithm is also bounded. Hence the asymptotic behavior of our learning algorithm can be studied by analyzing the continuous adaptive dynamics of the drift $\mathbb{E}(W_{t+1}|C_t)$. The following holds:

Lemma III-A.2: The stochastic learning algorithm generates the continuous time dynamics given by

$$\frac{d}{dt}C_r^i(t) = \bar{C}_r^i(t) - C_r^i(t) \tag{2}$$

where $Ber_{r,t}^i$ denotes a Bernoulli random variable with the parameter $\mathbb{P}(Ber_{r,t}^i = 1) = p_r^i$, $\bar{C}_r^i(t) = \mathbb{E}\left(C_r^i(\sum_{i'} Ber_{r,t}^{i'}, t) \mid Ber_{r,t}^i = 1\right)$ represents the average cost observed by player i when he chooses route r and the others players choose it with probabilities $p_r^{i'}$.
It is easy to see that the expectation of W given C is

$$\mathbb{E}(W_r^i|C) = p_r^i(C)\bar{C}_r^i + (1-p_r^i)C_r^i.$$

Denote by $a_1 = \max\{C_r(k+1) - C_r(k),\ 1 \leq k \leq N\}$, $\epsilon = \max_i \epsilon_i > 0$, $a_2 = \frac{1}{2}\epsilon a_1$. We choose the coefficients ϵ_i such that $a_2 < \frac{1}{N-1}$.

Proposition III-A.3: The Boltzmann-Gibbs-based stochastic learning algorithm converges almost surely to equilibria.

So far, we considered finite number of players in discrete setting and showed that the proposed learning algorithm converges to equilibrium in dynamic environment setting. In the next subsection we go ahead with large number of player. We apply mean field analysis and show that players can learn based on related ODE. Further more, in the next subsection, we show that players need not have same learning algorithms to reach equilibrium.

B. From micro to macro: Scaling and Mean Field Limits

Define the scaled cost functions as $C_r^N(\frac{k_r}{N}, t)$ and the mean profile $X^N = \frac{1}{N}\sum_{i'} \delta_{\{r_t^{i'}=r\}}$. Assuming that the second moment of the number of players that use the same route is finite[3], the mean process converges weakly to a mean field limit, solution of system ordinary differential equation given by the drift $\frac{1}{\Delta_N} f^N(x(t))$ where $f^N(x(t)) = \mathbb{E}\left(X^N(t+\Delta_N) - X^N(t)|X^N(t) = x(t)\right)$ is the expected change in the system in one-time slot with duration Δ_N. This can be directly inferred from [16].

For example the Boltzmann-Gibbs dynamics (also called logit dynamics or smooth best response dynamics) is given by

$$\frac{d}{dt}x_r^i(t) = \sum_{r'} x_{r'}^i(t) p_{r'r}^i(x(t), t) - x_r^i(t)$$

Where $p_{r'r}^i(x(t)) = \frac{e^{\frac{-C_r^i(x(t),t)}{\epsilon_i}}}{\sum_{\bar{r}} e^{\frac{-C_{\bar{r}}^i(x(t),t)}{\epsilon_i}}}$. Players from class i can learn $x(t)$ via the ordinary differential equation (ODE) and can use a route r with probability p_r^i. Note that our study can be extended to the case where one has both atomic and non-atomic players by considering the weighted population profile: $\tilde{X}^N(t)_{(i,r)} = \frac{1}{\sum_j \gamma_j^N}\sum_j \gamma_j^N \delta_{R_j^N(t)=(i,r)}$ where γ_j^N is the weight ("power") of j in the population of size N.

[3]Notice that the number of players that can interact can be very large

C. Players need not to use the same learning scheme

We now study how to combine various learning schemes based on mean field limit dynamics. In the previous studies the players have to follow the same rule of learning, they have to learn in the same way. We ask the following question: what happens if players have different learning schemes? The motivation to study different learning scheme comes from the fact that in real scenario it is not practical to enforce a learning scheme to player, rather players learning may depend on various factors e.g. their capability. On the other hand we are interested in a class of learning scheme in which player use less information on the others players, less memory on the history and need not to use the same learning scheme [10]. Thus, in this section we study and characterize the system behavior when different learning schemes work together.

Consider a population in which the players can adopt different learning schemes in $\{\eta^1, \eta^2, \eta^3, \ldots, \eta^\kappa\}$, $\kappa < \infty$. Then, based on the composition of population and the use of each learning scheme we build a spatial hybrid game dynamics. The intra-incoming and the intra-outgoing flow as well as the inter-neighborhood flow are expressed in term of the weighted combination of different learning schemes picked from the set $\{\eta^1, \eta^2, \eta^3, \ldots\}$.

Definition 3.1 (Property): **WES**: Every rest point of the mean field limit routing game dynamics generated by the weighted cost is a weighted equilibrium and every constrained weighted equilibrium is a rest point of the dynamics. Note that this property is not satisfied by the well-known *replicator dynamics* as it is known that the replicator dynamics may not lead to equilibria. We have the following results:

Proposition III-C.1: If all the learning schemes contained in the support of $\lambda = (\lambda^1, \ldots, \lambda^\kappa) \in \mathbb{R}_+^\kappa$ satisfy the property (WES) then, the hybrid mean field limit routing game dynamics generated by these learning schemes satisfies also the weighted equilibrium stationarity property.

Example: The family of learning scheme generated by $\eta_{r',r}^\theta = \max(0, -C_r(x(t)) + C_{r'}(x(t)))^\theta$, $\theta \geq 1$ satisfies (WES). Note that $\theta = 1$ is well known smith dynamics. For $\theta \geq 1$, one can refer [17].

IV. Non-atomic multi-class dynamic routing

In this section we move on to non-atomic class of routing games which simply means that change of strategy of a single player has negligible impact on the system behavior. We extend the traditional variational calculus approach to dynamic routing setting and define notion of dynamic equilibrium. We establish conditions under which uniqueness of equilibrium is sustained. We further discuss the learning schemes which appeared correlated with evolutionary dynamics.

The network is crossed through by infinitely many jobs that have to choose their routes (collection of consecutive edges, no cycle). Jobs are classified into different types or classes. Denote by $\mathcal{E}$ the set of classes. For example, in the context of road traffic, a type may represent the set of a given type of vehicles, such as busses, car or bicycle. In the context of telecommunications a type may represent the jobs or packets sent by all the players of a given operator. We consider that the packets of a given type e may arrive in the system at some different possible points, and leave the system at some different possible points. Each individual packet of type e with source-destination pair, chooses its route trough the system, by means of the choice of a path. Denote by SD the set of source-destination pairs, $\mathcal{R}_{sd}$ the set of routes from s to d. The current traffic generated between s and d by class e is $m_{sd}^e(t)$, also represent the rate at which the jobs of type e with source-destination $(s,d) \in \mathcal{N}^2$ arrives in the network. $m_{sd}(t)$ is the total arrival rate of data with source-destination sd at time t. This can be viewed as concentrated mass on sd.

Unlike previous section the flow is splittable. A flow on route r at time t of class e is assumed non-negative and is denoted by $x_r^e(t)$. A flow configuration which corresponds to the population profile of the network $\mathcal{X}$ follows from the choices of each of the infinitely many packets. We have the conservation flow equation

$$\forall (s,d) \in SD,\ \forall e,\ \sum_{r \in \mathcal{R}_{sd}} x_r^e(t) = m_{sd}^e(t)$$

that is, the demand associated with (s,d) pair and class e must be equal to the sum of the flows of that class on the routes that connect s to d. There is a capacity constraint per class $\forall e, \forall r,\ 0 \leq x_r^e(t) \leq \beta_r^e(t)$, $\beta_r^e(t)$ is the capacity on route r at time t.

The link flow is denoted by $f(t)$ and satisfies the following relation: $f_l^e(t) = \sum_{r \in \mathcal{R}} x_r^e(t)\delta_{l\in r}$ $\delta_{a \in r}$ is equal to one if the link l is contained in the route r and zero otherwise. Then the costs on route r can be written as $C_r^e(x(t),t) = \sum_{l \in \mathcal{L}} C_l^e(x(t),t)\delta_{l \in r}$. Note that, we focus on continuous cost functions.

An important generalization of variational inequalities are quasi-variational inequalities and evolutionary variational inequalities. We model and study a dynamic traffic network problem with multi-class of traffic and with feasible path flows which have to satisfy time-dependent capacity constraints and demands. We construct a unified definition of equilibrium and the constraint set that arises in the applications of time-dependent traffic network, and in non-atomic games. We formulate a dynamic equilibrium which can be expressed as an evolutionary variational inequality. A dynamic equilibrium in the evolving network is a time-dependent trajectory equilibrium that satisfies an evolutionary (time-dependent) variational inequalities.

A. Dynamic Equilibria : Evolutionary Variational Solutions

Consider traffic networks in which the demand varies over the time horizon as well as the capacities on the flows on the paths connecting the origins to the destinations. We are interested how traffic network equilibria evolve in the presence of such variations.

Consider the Hilbert space $\mathbb{H} := L^2([0,T], \mathbb{R}^{(\sharp\mathcal{E}\times\sharp\mathcal{R})})$, (the set Lebesgue-measurable function $m(.)$ such that $\int_0^T m^2(t)\ dt < +\infty$).

$$\mathcal{M} = \{x \in \mathbb{H},\ \Phi x(t) = m(t)\ a.e,\ 0 \leq x(t) \leq \beta(t) \text{ a.e}\}$$

where $a.e$ stands for "almost everywhere" for the Lebesgue measure restricted to the interval $[0,T]$ and Φ denotes the traffic matrix.

The set $\mathcal{M}$ is closed, convex and bounded in $\mathbb{H}$. Define the incidence matrix χ with $\chi_{r,sd} = 1$ if $r \in \mathcal{R}_{sd}$ and zero otherwise. We assume that $0 \leq m(t) \leq \chi\beta(t)$. Then, the feasible set $\mathcal{M}$ is non-empty. Denote by $\mathbb{H}^*$ be the dual of $\mathbb{H}$. Define the cost function as a mapping from $\mathcal{M}$ to $\mathbb{H}^*$. The bilinear form on $\mathbb{H} \times \mathbb{H}^*$ is given by $\langle C, x\rangle := \int_0^T \langle C(t), x(t)\rangle\ dt$. where $\langle f,g\rangle = \int_0^T f(t)g(t)dt$.

Dynamic equilibrium: We say $x \in \mathcal{M}$ is a dynamic equilibrium if $\forall y \in \mathcal{M},\ \int_0^T \langle y(t)-x(t), C(x(t),t)\rangle\ dt \geq 0$. This concept generalizes the standard Wardrop first principle developed in static traffic scenarios.

Proposition IV-A.1: $x \in \mathcal{M}$ is a dynamic equilibrium if and only if for every pair $(s,d) \in \mathcal{N}^2 \cap SD$, every route $r \in \mathcal{R}_{sd}$, every class $e \in \mathcal{E}$, the following holds on $[0,T]$ almost everywhere (a.e)

(a) $x_r^e(t) = \beta_r^e(t) \Longrightarrow C_r^e(x(t)) \leq \min_{r' \in \mathcal{R}_{sd}} C_{r'}^e(x(t))$,

(b) $0 < x_r^e(t) < \beta_r^e(t) \Longrightarrow C_r^e(x(t)) = \min_{r' \in \mathcal{R}_{sd}} C_{r'}^e(x(t))$,

(c) $x_r^e(t) = 0 \Longrightarrow C_r^e(x(t)) \geq \min_{r' \in \mathcal{R}_{sd}} C_{r'}^e(x(t))$.

For infinite horizon we consider the following definition: $\limsup_T \frac{1}{T}\int_0^T\ \langle x(t)-y(t), C(x(t),t)\rangle\ dt \leq 0$. If x is a *static*, (Cournot/Nash/Wardrop) equilibrium for network traffic with parameters α, β then the constant trajectory $z\ :\ t \longrightarrow z(t) = x$ is a dynamic equilibrium. That is if $\forall y(t),\ \langle z(t)-y(t), C(x(t),t)\rangle \leq 0$ then, $\int_0^T\ \langle x(t) - y(t), C(x(t),t)\rangle\ dt \leq 0$. This says that the "standard" Wardrop first principle (user equilibrium) is a particular case of dynamic equilibrium.

Stable game: We say that the game is a *stable game* if the (weighted) cost function satisfies

$$\forall y,y' \in \mathbb{H},\ \int_0^T \langle C(y(t),t) - C(y'(t),t), y(t)-y'(t)\rangle\ dt \geq 0$$

The game is said strictly stable if this inequality is strict. If the game is stable then the set of dynamic equilibria is a convex set. If the game is strictly stable and the payoff continuous then, there is at most one dynamic equilibrium (for Lebesgue mesaure).

Proposition IV-A.2: Under the feasibility condition on the flow and continuity of the cost function, the routing game has a least one dynamic equilibrium. Moreover if the costs define a constrained strictly stable population game. Then, the game has a unique dynamic equilibrium (for the Lebesgue measure).

B. Sensitivity and Stability of Dynamic Equilibria

We study the sensitivity of equilibria in the routing game. We study how the equilibrium responds to $\epsilon-$perturbation of the flow profile. We say that the "static" equilibrium is a global evolutionarily stable strategy (GESS) if for all flow configuration $y \neq x$, there exists a threshold $\epsilon_y > 0$ such that $\langle x-y, C(\epsilon y+(1-\epsilon)x)\rangle < 0$ for all $\epsilon \in (0, \epsilon_y)$. We say that a trajectory $x \in \mathcal{M}$ is a time-dependent global evolutionarily stable strategy if for all trajectory $y \in \mathcal{M}$ such that the set $EE := \{t,\ x(t) \neq y(t)\}$ is of non-zero Lebesgue measure: $\int_{[0,T]}\ \mathbb{1}_{EE}(t)\ dt > 0$, there is $\epsilon_y > 0$ such that

$$\int_0^T \langle x(t)-y(t), C(\epsilon y(t)+(1-\epsilon)x(t), t)\rangle\ dt < 0,$$

for all $\epsilon \in (0, \epsilon_y)$.

Proposition IV-B.1: If the game is a strict stable game then, the unique dynamic equilibrium is a time-dependent GESS.

Proposition IV-B.2: Assume that the cost function are autonomous and stable, and consider different learning schemes $\eta^1, \eta^2, \ldots, \eta^\kappa$ are adopted by the players where $\eta^\theta_{r,r'}(x(t)) = (\max[0, -C_{r'}(x(t)) + C_r(x(t))])^\theta$ with the fractions $\lambda = (\lambda^\theta)_\theta$. Then, these revision of strategies lead to an evolutionary game dynamics which converges to the set of equilibria.

3) Control of demand : Given a total demand $\bar{m}$ the operator wishes to split over $[0,T]$ such that $\int_0^T m(t) = \bar{m}$. The operator problem then to minimize to social cost $\int_0^T \langle x(m(t)), C(x(m(t))\rangle\ dt$ over $m(.)$. In the Braess graph this minimization problem can be written as:

$$\begin{cases} \inf_{m(.)} \quad \frac{11}{2}\int_0^T m^2(t)\ dt + 50\int_0^T m(t)dt \quad \text{such that} \\ \qquad\qquad \int_0^T m(t) = \bar{m}, \\ \qquad\qquad m(t) \geq 0,\ \forall t \in [0,T] \end{cases}$$

It is easy to see that this problem can be reduced to the constrained minimization of the L^2-norm in $\mathbb{H}$.

$$\begin{cases} \inf_{m(.)} \qquad \int_0^T m^2(t)\ dt \qquad \text{such that} \\ \qquad\qquad \int_0^T m(t) = \bar{m}, \\ \qquad\qquad m(t) \geq 0,\ \forall t \in [0,T] \end{cases}$$

The set of constraints is convex and the objective function $f \longrightarrow \int_0^T f^2(t)\ dt$ is also convex. The solution is given by $m(t) = \frac{\bar{m}}{T}$.

V. Discussion and Concluding Remarks

We give existence and uniqueness conditions for dynamic equilibria as well as sufficiency condition for evolutionary stability. We have shown that the set of dynamic equilibria is not only from the collection of equilibria of stage-games. We have proposed a learning algorithm based on Boltzmann-Gibbs dynamics and proved its almost sure convergence for the games in parallel link networks with monotone cost functions, our result is also true for any congestion-dependent resource selection problems with finite number of choices with imperfect observation but the qualitative properties for *general network topology* remains open. When the mean field limit is non-deterministic, the resulting dynamics leads to stochastic mean field limit dynamics

$$dx_r(t) = V^r_\eta(x(t))dt + \sigma(t)dW_t,\ \text{supp}\ \eta \subseteq \mathcal{L}$$

where V_η is the drift generated by the learning scheme η and W is a noise. The stochastic learning scenario where players have a noise in their stage-cost and dynamic routing in random networks are interesting directions of future investigation. Since equilibrium costs in selfish routing games can be inefficient, we plan to examine alternative solutions such as Stackelberg-based solution and team solution in future.

References

[1] Tim Roughgarden. Stackelberg scheduling strategies. In *SIAM Journal on Computing*, volume 33, pages 332–350, 2004.

[2] Ariel Orda, Raphael Rom, and Nahum Shimkin. Competitive routing in multi-user communication networks. *IEEE/ACM Transactions on Networking*, 1:510–521, 1993.

[3] H. Tembine, E. Altman, R. ElAzouzi, and Y. Hayel. Evolutionary games in wireless networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40:634–646, 2010.

[4] Michel Benaïm, Josef Hofbauer, and Sylvain Sorin. Stochastic approximations and differential inclusions, part ii: Applications. *Math. Oper. Res.*, 31(4):673–695, 2006.

[5] Hart S. and Mas-Colell A. Stochastic uncoupled dynamics and nash equilibrium. *Games and Economic Behavior*, 52(2):286–303, November 2006.

[6] H. Tembine. Distributed strategic learning for wireless engineers. *Notes, 300 pages*, 2010.

[7] Yannis A. Korilis, Aurel A. Lazar, and Ariel Orda. Achieving network optima using stackelberg routing strategies. *IEEE/ACM Transactions on Networking*, 5:161–173, 1997.

[8] H. Kameda, E. Altman, T. Kozawa, and Y. Hosokawa. Braess-like paradoxes in distributed computer systems. *IEEE Transaction on Automatic control*, 45(9):1687–1691, 2001.

[9] J. Willard Gibbs. *Elementary Principles in Statistical Mechanics*. Ox Bow Press, 1901.

[10] H. Tembine. Dynamic robust games in mimo systems. *IEEE Transactions on Systems, Man, Cybernetics, Part B, 99*, 41:990 – 1002, August 2011.

[11] H. Brezis. Inequations d'evolution abstraites. *Comptes Rendue d'Academie des Sciences*, 1967.

[12] Eitan Altman and Laura Wynter. Equilibrium, games, and pricing in transportation and telecommunications networks. *Networks and Spatial Economics*, 4(1), March 2004.

[13] Dimitris Fotakis, Spyros Kontogiannis, and Paul Spirakis. Selfish unsplittable flows. *Theoretical Computer Science*, 348(2-3):226 – 239, 2005.

[14] S. Sorin R. Cominetti, E. Meloy. A simple model of learning and adaptive dynamics in traffic network games. 2007.

[15] R. J. Aumann. Mixed and behavioural strategies in infinite extensive games. pages 627–650, 1964.

[16] Tembine H., Le Boudec J. Y., El-Azouzi R., and Altman E. Mean field asymptotics of markov decision evolutionary games and teams. In *IEEE Gamenets*, May 2009.

[17] H. Tembine, E. Altman, R. ElAzouzi, and W.H. Sandholm. Evolutionary game dynamics with migration for hybrid power control in wireless communications. In *47th CDC 2008.*, pages 4479–4484, Dec. 2008.