

How Does Driver Non-Compliance Destroy Traffic Routing Control?

Yu Tang, Li Jin and Kaan Ozbay

Abstract— Routing control is one of important traffic management strategies against urban congestion. However, it could be compromised by heterogeneous driver non-compliance with routing instructions. In this article, we model the compliance in a stochastic manner and investigate its impacts on routing control. We consider traffic routing for two parallel links. Particularly, we consider two scenarios: one ignores congestion spillback while the other one considers it. We formulate the problem as a Markov chain, given random drivers’ adherence. Then, we propose the stability and instability conditions to reveal when the routing is able or unable to stabilize the traffic. We show that for links without congestion spillback there exists a necessary and sufficient stability criterion. For links admitting congestion propagation, we present one stability condition and one instability condition. These stability conditions allow us to quantify the impacts of driver non-compliance on the two-link network in terms of throughput. Finally, we illustrate the results with a set of numerical examples.

I. INTRODUCTION

A. Motivation

Dynamic traffic routing provides drivers with route recommendations based on real-time road information. It has been used as one of promising control policies for alleviating congestion [1], [2], and is expected to find extensive applications in a connected vehicle environment [3]. Nevertheless, it is also reported that driver non-compliance with route guidance could undermine the performance of dynamic routing [4], especially social routing advice that deliberately detours part of vehicles to achieve benefits in terms of road networks [5]. Although more and more surveys have confirmed this phenomenon [5], [6], limited studies have investigated in an analytical way how drivers’ adherence influences the effect of traffic routing control.

In this paper, we focus on routing advice released by traffic system operators/agencies. We study the above problem by considering a setting with random demand and random driver non-compliance. We analyze the resulting stochastic dynamical system under routing control. Specially, we focus on a network comprised of two parallel links; see Fig. 1. Though simple, the two-link network serves as a typical

This work was in part supported by US NSF Award CMMI-1949710, USDOT Award # 69A3551747124 via the C2SMART Center, NYU Tandon School of Engineering, SJTU UM Joint Institute, and J. Wu & J. Sun Endowment Fund.

Y. Tang is with C2SMART Center, Department of Civil & Urban engineering, Tandon School of Engineering, New York University, 11201, USA. tangyu@nyu.edu

L. Jin is with UM Joint Institute and Department of Automation, Shanghai Jiao Tong University, Shanghai, 200240, China. li.jin@sjtu.edu.cn

K. Ozbay is with C2SMART Center, Department of Civil & Urban engineering, Tandon School of Engineering, New York University, 11201, USA. kaan.ozbay@nyu.edu

scenario for studying routing control [7], [8], [9]; it turns out to be an appropriate abstraction of multiple parallel links: one stands for arterials and the other denotes a set of local streets [10]. Furthermore, we adopt a Markov chain to model the compliance rate that possibly depends on traffic states. It allows us to study stability and instability criteria that determine whether the network is destabilized by the random compliance rate. We also quantify the impacts of drivers’ disobedience in terms of throughput, namely the maximum constant inflow under which the network can be stabilized.

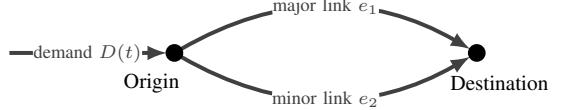


Fig. 1: The two-link network.

B. Related work

Previous work on evaluating impacts of the conformity with routing advice typically applied static or dynamic traffic assignment (STA or DTA). These methods are favored since they easily provide numerical assessment in terms of efficiency, equity and so on [5], [11], [12] and can be implemented even for large-scale networks. However, they also have disadvantages. STA finds equilibrium by solving mathematical programming. It fails to capture significant traffic dynamics, such as congestion spillback and fluctuations of drivers’ compliance rate and thus could induce unrealistic equilibrium. Though DTA can address the shortcomings of STA to some extent, it introduces a new problem. As we see later, low compliance rates could make traffic networks unstable. In that case, it could be problematic to apply DTA since we do not have guaranteed convergence in advance. Noting this, we aim at developing methods that allow stability and instability analysis, at least in some conditions, before numerical evaluation. To our best knowledge, limited studies discussed this topic for routing control subject to random compliance.

Our model belongs to discrete-time nonlinear stochastic systems. Although the general theories of stochastic stability have been studied extensively [13], how to apply them in our problem is still unclear. Typically, stability analysis can be refined for specific nonlinear systems. Besides, it is noteworthy that most of studies mainly discuss sufficient stability conditions for general nonlinear stochastic systems, while we also have interest in sufficient instability conditions.

C. Our contributions

In this paper, we address the following two questions:

- 1) How to determine whether the network can be stabilized by routing control subject to driver non-compliance?
- 2) How to evaluate efficiency losses of routing control due to the non-compliance in an analytic way?

We answer the first question for two types of networks. In the first one, the two parallel links have infinite space and there are no congestion spillback, while in the second one, the two parallel links only have finite space. We formulate discrete-time nonlinear stochastic systems for the two networks, respectively. Then we apply the Foster-Lyapunov criterion [13] to derive the stability condition and scrutinize transience of Markov chains [13] to obtain instability conditions. For the first network, we successfully obtain a sufficient and necessary stability condition (Theorem 1); for the second one, we have one stability criterion (Theorems 2) and one instability criterion (Theorems 3).

Even when the network is stable, we want to know to what extent the network performance decrease. Thus, to answer the second question, we take throughput as the metric to measure efficiency losses. However, throughput is not always available even for the two-link network. For the two links with infinite buffer sizes, we indeed derive exact values of throughput since we have a sufficient and necessary stability condition. For the two links with finite buffer sizes, we use the stability and instability conditions to yield lower and upper bounds, respectively.

The rest of the paper is organized as follows. Section II introduces our modeling framework. Section III presents the results when the two parallel links have infinite storage space, and Section IV provides the results in case of two links with limited buffer sizes. Finally, Section V summarizes our work and discusses future research.

II. MODELING AND FORMULATION

Consider the two-link network in Fig. 1: one is the major link e_1 , typically with a higher free-flow speed or capacity, and the other is the minor link e_2 . We suppose that the system operator tries to route part of flows to the minor link e_2 to reduce congestion in the major link e_1 .

We denote by $X_e(t) \in \mathbb{R}_{\geq 0}$ traffic density of link $e \in \{e_1, e_2\}$ at time t . Each link $e \in \{e_1, e_2\}$ is associated with a sending flow (demand function) $f_e(x_e) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ and a receiving flow (supply function) $r_e(x_e) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$.

Assumption 1 (Sending & receiving flows).

- 1.1 *Sending flows: For link e , $f_e(x_e)$ is Lipschitz continuous and $df_e(x_e)/dx_e \geq 0$ almost everywhere (a.e.). Moreover, $f_e(0) = 0$ and $\sup_{x_e} f_e(x_e) < \infty$.*
- 1.2 *Receiving flows: For link e with a finite buffer size $x_e^{\max} < \infty$, $r_e(x_e)$ is Lipschitz continuous and $dr_e(x_e)/dx_e \leq 0$ a.e.. Moreover, $r_e(x_e^{\max}) = 0$ and $\sup_{x_e} r_e(x_e) < \infty$. For link e with an infinite buffer size, $r_e = \infty$.*

The assumptions above follow the conventional modeling of road traffic. We also define *link capacity* as

$$Q_e := \sup_{x_e} \min\{f_e(x_e), r_e(x_e)\}, \quad (1)$$

which denotes an upper bound of sustainable discharging flow from link e .

Note that Assumption 1.2 implies that it is reasonable to only consider $X_e(t) \in [0, x_e^{\max}]$ for link e with limited storage. Compared with supposing finite buffer sizes, the assumption of infinite buffer sizes seems a little unrealistic, but it helps understand and design routing control, even on complex networks. In this paper, we discuss both of them.

For demand modeling, we consider an independent and identically distributed (i.i.d.) stochastic process $\{D(t) : t \geq 0\}$ with a distribution Γ^d , $\mathbb{E}[D(t)] = \alpha$ and $D(t) \in \mathcal{D}$ for $t \geq 0$, where \mathcal{D} is a compact set. This is based on the observation that during rush hours, of interest to traffic management, traveling demands are relatively stationary and only fluctuate within certain bounds. Obviously, we require

$$\mathbb{E}[D(t)] = \alpha < Q_{e_1} + Q_{e_2}, \quad (2)$$

otherwise the traffic densities must blow up.

Next, we introduce routing control. Let $\beta_e(x) : \mathbb{R}_{\geq 0}^2 \rightarrow [0, 1]$ denote a proportion of traffic routed to link e . We assume the routing policies to satisfy:

Assumption 2 (Routing control). *The routing proportions $\beta_{e_1}(x_{e_1}, x_{e_2})$ and $\beta_{e_2}(x_{e_1}, x_{e_2})$ are continuous and have the following monotonicity a.e.:*

- 2.1 $\frac{\partial}{\partial x_{e_1}} \beta_{e_1}(x_{e_1}, x_{e_2}) \leq 0$ and $\frac{\partial}{\partial x_{e_2}} \beta_{e_1}(x_{e_1}, x_{e_2}) \geq 0$;
- 2.2 $\frac{\partial}{\partial x_{e_1}} \beta_{e_2}(x_{e_1}, x_{e_2}) \geq 0$ and $\frac{\partial}{\partial x_{e_2}} \beta_{e_2}(x_{e_1}, x_{e_2}) \leq 0$.

The assumption above implies that the routing proportion $\beta_e(x_{e_1}, x_{e_2})$ tends to decrease (resp. increase) as link e (resp. the other link) becomes more congested. It holds true for typical routing policies, such as logit routing [?].

Recalling that routing proportions could be compromised due to heterogeneous drivers' choice behavior, we denote by $C(t) \in [0, 1]$ the compliance rate of drivers' routed to the minor link e_2 at time t . Then the compromised routing ratios, denoted by $\tilde{\beta}_e(x_{e_1}, x_{e_2}, c) : \mathbb{R}_{\geq 0}^2 \times [0, 1] \rightarrow [0, 1]$, are given by

$$\begin{aligned} \tilde{\beta}_{e_1}(X_{e_1}(t), X_{e_2}(t), C(t)) \\ = \beta_{e_1}(X_{e_1}(t), X_{e_2}(t)) + \beta_{e_2}(X_{e_1}(t), X_{e_2}(t))(1 - C(t)), \end{aligned} \quad (3a)$$

$$\begin{aligned} \tilde{\beta}_{e_2}(X_{e_1}(t), X_{e_2}(t), C(t)) \\ = \beta_{e_2}(X_{e_1}(t), X_{e_2}(t))C(t). \end{aligned} \quad (3b)$$

Note that (3a)-(3b) imply the compliance rate of drivers routed to the major link e_1 equals one. This is because in our setting drivers are assumed to prefer the major link e_1 while the system operator tries to route some of them to the minor link e_2 . The assumption is not necessary, just for simplifying the problem. In fact, we can introduce the second compliance rate, and apply our method to obtain stability and instability criteria, which are more complicated.

We consider that $C(t+1)$ depends on $X_{e_1}(t) = x_{e_1}$ and $X_{e_2}(t) = x_{e_2}$ with a distribution $\Gamma_{x_{e_1}, x_{e_2}}^c$. For convenience of analysis, we assume that the distributions $\Gamma_{x_{e_1}, x_{e_2}}^c(c)$, for any x_{e_1} and x_{e_2} , have lower semi-continuous densities with the same support $\mathcal{C} \subseteq [0, 1]$. We define $\mathbb{E}_{x_{e_1}, x_{e_2}}[C] := \mathbb{E}[C(t+1)|X_{e_1}(t) = x_{e_1}, X_{e_2}(t) = x_{e_2}]$ and assume it to satisfy:

Assumption 3 (Drivers' compliance). *The expected compliance rate has the following monotonicity a.e.:*

$$\frac{\partial}{\partial x_{e_1}} \mathbb{E}_{x_{e_1}, x_{e_2}}[C] \geq 0, \text{ and } \frac{\partial}{\partial x_{e_2}} \mathbb{E}_{x_{e_1}, x_{e_2}}[C] \leq 0. \quad (4)$$

Clearly, the assumption implies that more drivers follow the routing advise to the minor link e_2 if the major link e_1 becomes more congested or the minor link e_2 becomes less congested.

The following specifies the inflows into links e_1 and e_2 . Given an upstream flow $F(t)$, we denote by $q_e^{\text{in}} : \mathbb{R}_{\geq 0}^3 \times [0, 1] \rightarrow \mathbb{R}_{\geq 0}$ the inflow into link $e \in \{e_1, e_2\}$:

$$\begin{aligned} q_e^{\text{in}}(F(t), X_{e_1}(t), X_{e_2}(t), C(t)) \\ = \min\{\tilde{\beta}_e(X_{e_1}(t), X_{e_2}(t), C(t))F(t), r_e(X_e(t))\}. \end{aligned} \quad (5)$$

Supposing $r_{e_1} = r_{e_2} = \infty$, we have the following network dynamics for any link $e \in \{e_1, e_2\}$

$$\Delta X_e(t) = \frac{\delta}{l_e} \left(q_e^{\text{in}}(D(t), X_{e_1}(t), X_{e_2}(t), C(t)) - f_e(X_e(t)) \right), \quad (6)$$

where $\Delta X_e(t) := X_e(t+1) - X_e(t)$, δ denotes the time step size, and l_e denotes length of link e .

Clearly, if links e_1 and e_2 have finite space, congestion could block the inflows. For the sake of analysis, we consider another link e_0 upstream of links e_1 and e_2 , satisfying $Q_{e_0} \geq Q_{e_1} + Q_{e_2}$ and $r_{e_0} = \infty$, to accept inflows. It leads to the network dynamics as follows:

$$\begin{aligned} \Delta X_{e_0}(t) = \frac{\delta}{l_{e_0}} \left(D(t) \right. \\ \left. - \sum_{e \in \{e_1, e_2\}} q_e^{\text{in}}(f_{e_0}(X_{e_0}(t)), X_{e_1}(t), X_{e_2}(t), C(t)) \right) \end{aligned} \quad (7a)$$

$$\begin{aligned} \Delta X_e(t) = \frac{\delta}{l_e} \left(q_e^{\text{in}}(f_{e_0}(X_{e_0}(t)), X_{e_1}(t), X_{e_2}(t), C(t)) \right. \\ \left. - f_e(X_e(t)) \right), \quad e \in \{e_1, e_2\} \end{aligned} \quad (7b)$$

For notional convenience, we assume δ/l_e to be the same for any link e and ignore them in the following analysis.

Then, (6) indicates that

$$\Phi_1 := \{(X_{e_1}(t), X_{e_2}(t), D(t), C(t)) : t \geq 0\} \quad (8)$$

is a Markov chain with a state space $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \mathcal{D} \times \mathcal{C}$, and (7a)-(7b) indicate

$$\Phi_2 := \{(X_{e_0}(t), X_{e_1}(t), X_{e_2}(t), D(t), C(t)) : t \geq 0\} \quad (9)$$

is also a Markov chain with a state space $\mathbb{R}_{\geq 0} \times \mathcal{X}_{e_1} \times \mathcal{X}_{e_2} \times \mathcal{D} \times \mathcal{C}$. Note that $\mathcal{X}_{e_1} \subseteq [0, x_{e_1}^{\max}]$ and $\mathcal{X}_{e_2} \subseteq [0, x_{e_2}^{\max}]$ are bounded sets.

We make the last assumption as follows:

Assumption 4.

- 4.1 For the system (6), there exists $c \in \mathcal{C}$ and $d \in \mathcal{D}$ such that $\lim_{t \rightarrow \infty} X_e(t) = x_e^* < \infty$, $e \in \{e_1, e_2\}$, given $C(t) \equiv c$ and $D(t) \equiv d$;
- 4.2 For the system (7a)-(7b), there exists $c \in \mathcal{C}$ and $d \in \mathcal{D}$ such that $\lim_{t \rightarrow \infty} X_e(t) = x_e^* < \infty$, $e \in \{e_0, e_1, e_2\}$, given $C(t) \equiv c$ and $D(t) \equiv d$. Moreover,

$$\tilde{\beta}_{e_1}((x_{e_1}^*, x_{e_2}^*), c) f_{e_0}(x_{e_0}^*) < r_{e_1}(x_{e_1}^*), \quad (10a)$$

$$\tilde{\beta}_{e_2}((x_{e_1}^*, x_{e_2}^*), c) f_{e_0}(x_{e_0}^*) < r_{e_2}(x_{e_2}^*). \quad (10b)$$

The above assumption essentially states that there exists c and d such that the systems (6) and (7a)-(7b) are stable. Note that (10a)-(10b) are mild technical assumptions. The system (6) does not require them due to $r_{e_1} = r_{e_2} = \infty$. The equations (10a)-(10b) imply that the inflows into links e_1 and e_2 are strictly fewer than the corresponding receiving flows. That is, the inflow can smoothly pass links e_1 and e_2 when there is no congestion. By noting (2), (10a)-(10b) are easy to achieve for appropriate routing polices.

We have the following lemma by Proposition 7.1.4 and Theorem 7.2.6 in [13]:

Lemma 1. *Given Assumption 4.1, the Markov chain (8) is φ -irreducible; and given Assumption 4.2, the Markov chain (9) is φ -irreducible.*

Here φ is a certain measure. The φ -irreducibility means that any set with positive measure can be reached by the Markov chain given any initial state. It implies that any large set can be reached from any initial condition and thus the state space is indecomposable. It is a prerequisite of discussing stability of Markov chains.

Finally, we define the stability of interest below:

Definition 1 (Stability & Instability). *A stochastic process $\{Y(t) : t \geq 0\}$ with a state space \mathcal{Y} is stable if there exists a scalar $Z < \infty$ such that for any initial condition $y \in \mathcal{Y}$*

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^t \mathbb{E}[|Y(\tau)| | Y(0) = y] \leq Z, \quad (11)$$

where $|Y(\tau)|$ denotes 1-norm of $Y(\tau)$. The network is unstable if there does not exist $Z < \infty$ such that (11) holds for any initial condition $y \in \mathcal{Y}$.

The notion of stability follows a classical definition [14] and is widely used in studying traffic control [15].

III. STABILITY ANALYSIS OF THE NETWORK WITHOUT CONGESTION PROPAGATION

We state the main result as follows:

Theorem 1. *The Markov chain (8) with the state space $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \mathcal{D} \times \mathcal{C}$ is stable if and only if there exists a vector $\theta := [\theta_{e_1}, \theta_{e_2}]^T \in \mathbb{R}_{\geq 0}^2$ such that*

$$\left(\beta_{e_1}(\theta) + \beta_{e_2}(\theta) \mathbb{E}_\theta[1 - C] \right) \alpha - f_{e_1}(\theta_{e_1}) < 0, \quad (12a)$$

$$\beta_{e_2}(\theta) \mathbb{E}_\theta[C] \alpha - f_{e_2}(\theta_{e_2}) < 0. \quad (12b)$$

In the following sections, we first present a numerical example and then prove Theorem 1.

A. Numerical example

First, we set $\delta = 0.1$ and $l_{e_1} = l_{e_2} = 1$. We consider the sending flows $f_e(x_e) = \min\{v_{ex_e}, Q_e\}$ for $e \in \{e_1, e_2\}$, with $v_{e_1} = 1$, $v_{e_2} = 0.8$, $Q_{e_1} = 0.6$, $Q_{e_2} = 0.4$. Besides, we adopt the classical logit routing as follows:

$$\beta_e(x) = \frac{e^{-\nu_e x_e}}{e^{-\nu_{e_1} x_{e_1}} + e^{-\nu_{e_2} x_{e_2}}}, \quad e \in \{e_1, e_2\}, \quad (13)$$

where $\nu_{e_1} = 1$ and $\nu_{e_2} = 2$ are routing parameters.

We assume the demands $D(t) \in [\underline{d}, 1.2]$, $t \geq 0$, are independent and identically distributed (i.i.d.) uniform random variables. It follows $\mathbb{E}[D(t)] = \underline{d}/2 + 0.6$. We also assume the routing compliance rates $C(t) \in [0, \bar{c}]$, $t \geq 0$, are i.i.d. uniform random variables, along with $\mathbb{E}[C(t)] = \bar{c}/2$. It indicates that in our numerical example the compliance rates are independent of traffic states. It should be noted that this independence is not necessary for our approach. Here we assume it just for simplification. However, we still have non-trivial observations in this case.

We first analyze the stability and instability of scenarios with different \underline{d} and compliance rate \bar{c} . Fig. 2a shows the time-average traffic densities after 5×10^5 steps and reveals the stability and instability regions. We observe a non-linear boundary: given moderate traffic demands, improvements of compliance rates can stabilize the network; but given a high demand close to the network capacity, we hardly see the effect of improving compliance rate.

Then we compute the throughput, the maximum expected demand under which the network can be stabilized. It is interesting to find that we can achieve a relatively high throughput (around 0.987) when $\mathbb{E}[C(t)] = 0.395$. Further improvement is marginal when $\mathbb{E}[C(t)]$ exceed 0.395.

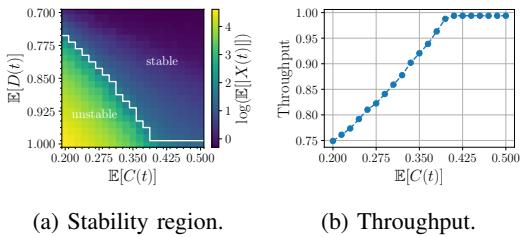


Fig. 2: Analysis given links e_1 and e_2 with infinite buffer sizes.

B. Proof of Theorem 1

We first prove the sufficiency by the Foster-Lyapunov criterion [13]:

Foster-Lyapunov criterion. Consider a φ -irreducible Markov chain $\{Y(t); t \geq 0\}$ with a state space \mathcal{Y} , an infinitesimal generator \mathcal{L} , and a Lyapunov function $V : \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$. If there exist constants $m > 0$, $n < \infty$, a function $g : \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$ and a compact set \mathcal{E} such that for any $y \in \mathcal{Y}$

$$\mathbb{E}[V(Y(t+1))|Y(t) = y] - V(y) \leq -mg(y) + n\mathbf{1}_{\mathcal{E}}(y),$$

where $\mathbf{1}_{\mathcal{E}}(y)$ is an indicator function, then, for each initial condition $y(0) \in \mathcal{Y}$,

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^t \mathbb{E}[g(Y(\tau))] \leq m/n.$$

To proceed, we consider the following Lyapunov function

$$V(x) = \begin{cases} 0 & x \in \mathcal{X}^1, \\ \frac{1}{2}(x_{e_1} - \theta_{e_1})_+^2 & x \in \mathcal{X}^2, \\ \frac{1}{2}(x_{e_2} - \theta_{e_2})_+^2 & x \in \mathcal{X}^3, \\ \frac{1}{2}((x_{e_1} - \theta_{e_1})_+ + (x_{e_2} - \theta_{e_2})_+)^2 & x \in \mathcal{X}^4, \end{cases} \quad (14)$$

where $(\cdot)_+ := \max\{\cdot, 0\}$, $\mathcal{X}_{e_1} := [0, \theta_{e_1}] \times [0, \theta_{e_2}]$, $\mathcal{X}_{e_2} := (\theta_{e_1}, \infty) \times [0, \theta_{e_2}]$, $\mathcal{X}_{e_3} := [0, \theta_{e_1}] \times (\theta_{e_2}, \infty)$ and $\mathcal{X}_{e_4} := (\theta_{e_1}, \infty) \times (\theta_{e_2}, \infty)$.

The rest is devoted to show that there exist constants $m' > 0$ and $n' < \infty$ such that for every $x \in \mathbb{R}_{\geq 0}^2$

$$\begin{aligned} & \mathbb{E}[V(X(t+1))|X(t) = x] - V(x) \\ & \leq -m'((x_{e_1} - \theta_{e_1})_+ + (x_{e_2} - \theta_{e_2})_+) + n'. \end{aligned} \quad (15)$$

If (15) holds, we must have $0 < m < m'$, $n < \infty$, and a compact set $\mathcal{E} = [0, M] \times [0, M]$ such that

$$\begin{aligned} & \mathbb{E}[V(X(t+1))|X(t) = x] - V(x) \\ & \leq -m((x_{e_1} - \theta_{e_1})_+ + (x_{e_2} - \theta_{e_2})_+) + n\mathbf{1}_{\mathcal{E}}(x), \end{aligned} \quad (16)$$

which indicates that $X_{e_1}(t)$ and $X_{e_2}(t)$ and thus concludes the stability.

To show (15), we need to discuss whether $X_{e_1}(t)$ is larger than θ_{e_1} and whether $X_{e_2}(t)$ is larger than θ_{e_2} , up to four cases. Here we present the proofs for the typical cases and the remaining can be proved in a similar way.

When $X_{e_1}(t) \leq \theta_{e_1}$ and $X_{e_2}(t) \leq \theta_{e_2}$, the proof is trivial by noting that $X_{e_1}(t+1)$ and $X_{e_2}(t+1)$ must be bounded a sufficiently large number.

Now we assume $X_{e_1}(t) > \theta_{e_1}$ and $X_{e_2}(t) \leq \theta_{e_2}$. It follows

$$\begin{aligned} & \mathbb{E}[V(X(t+1))|X(t) = x] - V(x) \\ & \leq \left((\beta_{e_1}(x) + \beta_{e_2}(x)\mathbb{E}_x[1 - C])\alpha - f_{e_1}(x_{e_1}) \right) x_{e_1} + n, \end{aligned}$$

where n is a sufficiently large number. Note that we omit δ/l_{e_1} . By Assumptions 1-3,

$$(\beta_{e_1}(x_{e_1}, x_{e_2}) + \beta_{e_2}(x_{e_1}, x_{e_2})\mathbb{E}_x[1 - C])\alpha - f_{e_1}(x_{e_1})$$

is non-increasing in x_{e_1} and non-decreasing in x_{e_2} . Thus (12a) indicates there exists $m'_1 > 0$ such that for any $x_{e_1} > \theta_{e_1}, x_{e_2} \leq \theta_{e_2}$, we have

$$\mathbb{E}[V(X(t+1))|X(t) = x] - V(x) \leq -m'_1 x_{e_1} + n.$$

Finally, we consider $X_{e_1}(t) > \theta_{e_1}$ and $X_{e_2}(t) > \theta_{e_2}$. It turns out that we obtain

$$\begin{aligned} & \mathbb{E}[V(X(t+1))|X(t) = x] - V(x) \\ & \leq (\alpha - f_{e_1}(x_{e_1}) - f_{e_2}(x_{e_2}))(x_{e_1} + x_{e_2}) + n. \end{aligned}$$

Note the $\alpha - f_{e_1}(x_{e_1}) - f_{e_2}(x_{e_2})$ is non-increasing in both x_{e_1} and x_{e_2} . Combing (12a)-(12b) indicates there exists $m'_2 > 0$ such that for any $x_{e_1} > \theta_{e_1}, x_{e_2} > \theta_{e_2}$, we have

$$\mathbb{E}[V(X(t+1))|X(t)=x] - V(x) < -m'_2(x_{e_1} + x_{e_2}) + n.$$

Thus we finish proving the sufficient condition.

The following proves the necessary condition. We prove it by showing the Markov chain (8) is transient [13]. We consider the following transience criterion [13]:

Transience criterion. Consider a φ -irreducible Markov chain $\{Y(t) : t \geq 0\}$ with a state space \mathcal{Y} . Then $\{Y(t) : t \geq 0\}$ is transient if there exists a bounded function $V : \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$ and a sublevel set of V , denoted by S , such that

- (i) $\varphi(S) > 0$ and $\varphi(\mathcal{Y} \setminus S) > 0$;
- (ii) $\mathbb{E}[V(Y(t+1))|Y(t)=y] - V(y) \geq 0$, $\forall y \in \mathcal{Y} \setminus S$.

Note that our Markov chain (8) is φ -irreducible, stated by Lemma 1. To proceed, we first assume that for any $\theta \in \mathbb{R}_{\geq 0}^2$

$$\left(\beta_{e_1}(\theta) + \beta_{e_2}(\theta) \mathbb{E}_\theta[1 - C] \right) \alpha - f_{e_1}(\theta_{e_1}) \geq 0. \quad (17)$$

We consider a bounded test function $W : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$:

$$W(x_{e_1}) = \xi_1 - \frac{1}{x_{e_1} + \xi_2} \quad (18)$$

where ξ_1 and ξ_2 are sufficiently large numbers. It turns out that given (17), we obtain

$$\mathbb{E}[W(X_{e_1}(t+1))|X_{e_1}(t)=x_{e_1}] - f_{e_1}(x_{e_1}) \geq 0.$$

Note that the above inequality holds over $\mathbb{R}_{\geq 0}^2$. It indicates that the conditions (i)-(ii) above are satisfied. Thus we conclude the Markov chain (8) is transient given (17).

Then we assume that for any $\theta \in \mathbb{R}_{\geq 0}^2$

$$\beta_{e_2}(\theta) \mathbb{E}_\theta[C] \alpha - f_{e_2}(\theta_{e_2}) \geq 0. \quad (19)$$

We can prove in a similar way that the Markov chain (8) is transient in this case. Finally, we conclude that if there does not exist a vector θ satisfying (12a)-(12b), the system is unstable.

IV. STABILITY ANALYSIS OF THE NETWORK WITH CONGESTION PROPAGATION

We state the main results as follows:

Theorem 2. Given Assumptions 1-4, the Markov chain (9) with the state space $\mathbb{R}_{\geq 0} \times \mathcal{X}_{e_1} \times \mathcal{X}_{e_2} \times \mathcal{D} \times \mathcal{C}$ is stable if there exists a vector $\theta := [\theta_{e_1}, \theta_{e_2}]^\top \in [0, 1]^2$ and a positive scalar $\gamma > 0$ such that

$$\begin{aligned} \alpha - \sum_{e \in \{e_1, e_2\}} (1 - \theta_e) \mathbb{E}_{x_{e_1}, x_{e_2}} [q_e^{\text{in}}(f_{e_0}(x_{e_0}^c), x_{e_1}, x_{e_2}, C)] - \\ - \sum_{e \in \{e_1, e_2\}} \theta_e f_e(x_e) < -\gamma, \quad \forall (x_{e_1}, x_{e_2}) \in \mathcal{X}_{e_1} \times \mathcal{X}_{e_2}, \end{aligned} \quad (20)$$

where $x_{e_0}^c := \inf\{x_{e_0} | f_{e_0}(x_{e_0}) = Q_{e_0}\}$ and

$$\begin{aligned} & \mathbb{E}_{x_{e_1}, x_{e_2}} [q_e^{\text{in}}(f_{e_0}(x_{e_0}), x_{e_1}, x_{e_2}, C)] \\ & := \int_{\mathcal{C}} q_e^{\text{in}}(f_{e_0}(x_{e_0}), x_{e_1}, x_{e_2}, c) \Gamma_{x_{e_1}, e_2}(\mathrm{d}c). \end{aligned} \quad (21)$$

Theorem 3. Given Assumptions 1-4, the Markov chain (9) with the state space $\mathbb{R}_{\geq 0} \times \mathcal{X}_{e_1} \times \mathcal{X}_{e_2} \times \mathcal{D} \times \mathcal{C}$ is unstable if there exists a vector $\theta := [\theta_{e_1}, \theta_{e_2}]^\top \in [0, 1]^2$ and a non-negative scalar $\gamma \geq 0$ such that

$$\begin{aligned} \alpha - \sum_{e \in \{e_1, e_2\}} (1 - \theta_e) \mathbb{E}_{x_{e_1}, x_{e_2}} [q_e^{\text{in}}(f_{e_0}(\bar{x}_{e_0}), x_{e_1}, x_{e_2}, C)] - \\ - \sum_{e \in \{e_1, e_2\}} \theta_e f_e(x_e) \geq \gamma, \quad \forall (x_{e_1}, x_{e_2}) \in \mathcal{X}_{e_1} \times \mathcal{X}_{e_2}, \end{aligned} \quad (22)$$

where $\bar{x}_{e_0} := \infty$.

Note that $x_{e_0}^c$ defined in Theorem 2 is usually interpreted as *critical density* since link e_0 with $x_{e_0} > x_{e_0}^c$ is considered as “congested” in practice. Theorem 2 indicates that though link e_0 could be congested with extremely high traffic densities, we only need to check the critical density. Besides, Theorem 3 says that we need to check $x_{e_0} = \infty$, namely to consider $\sup f_{e_0}$.

One can implement Theorem 2 by solving the following semi-infinite programming (SIP [16]):

$$(P_1) \min_{\theta_1, \theta_2, \gamma} \gamma \text{ s.t. (20)}, \quad (23)$$

and Theorem 3 by solving the SIP:

$$(P_2) \max_{\theta_1, \theta_2, \gamma} \gamma \text{ s.t. (22)}. \quad (24)$$

The programmings P_1 and P_2 belong to SIPs because they have infinite constraints over the continuous set $\mathcal{X}_{e_1} \times \mathcal{X}_{e_2}$. But noting \mathcal{X}_{e_1} and \mathcal{X}_{e_2} are bounded, we have efficient algorithms to solve P_1 and P_2 [16].

Note that Theorem 2 is proved based on the Lyapunov function $\tilde{V} : \mathbb{R}_{\geq 0}^3 \rightarrow \mathbb{R}_{\geq 0}$:

$$\tilde{V}(x_{e_0}, x_{e_1}, x_{e_2}) = x_{e_0} \left(\frac{1}{2} x_{e_0} + \theta_{e_1} x_{e_1} + \theta_{e_2} x_{e_2} \right). \quad (25)$$

and Theorem 3 is based on the test function $\tilde{W} : \mathbb{R}_{\geq 0}^3 \rightarrow \mathbb{R}_{\geq 0}$:

$$\tilde{W}(x_{e_0}, x_{e_1}, x_{e_2}) = \xi_1 - \frac{1}{x_{e_0} + \theta_{e_1} x_{e_1} + \theta_{e_2} x_{e_2} + \xi_2}, \quad (26)$$

where ξ_1 and ξ_2 are sufficiently large numbers. Clearly, we can improve the stability and instability conditions by considering more sophisticated Lyapunov/test functions, but with more computational costs.

The following section presents a numerical example. The proofs of Theorems 2 and 3 are omitted since they are similar to those of Theorem 1, except for different Lyapunov/test functions.

A. Numerical example

Besides the setting in Section III-A, we introduce the receiving flows $r_e(x_e) = R_e - w_e x_e$, with $R_{e_1} = 1.2$, $R_{e_2} = 0.8$, $w_{e_1} = 0.5$ and $w_{e_2} = 0.4$. It follows $x_{e_1}^{\max} = 2.4$ and $x_{e_2}^{\max} = 2$. For the upstream link e_0 , we suppose that its sending flow with $v_{e_0} = 1$ and $Q_{e_0} = 1$, which indicates the critical density $x_{e_0}^c = 1$.

Analyzing the Markov chain (9) is more difficult since traffic dynamics involving congestion spillback is more complicated. We consider the technique of invariant sets [17] and focus on our analysis of the Markov chain (9) on the state space

$$[\underline{d}, \infty) \times [\underline{x}_{e_1}, \bar{x}_{e_1}] \times [0, \bar{x}_{e_2}] \times [\underline{d}, 1.2] \times [0, \bar{c}], \quad (27)$$

where the boundaries \underline{x}_{e_1} , \bar{x}_{e_1} and \bar{x}_{e_2} satisfy

$$\beta_{e_1}(\underline{x}_{e_1}, 0)x_{e_1}\underline{d} + \beta_{e_2}(\underline{x}_{e_1}, 0)(1 - \bar{c})\underline{d} = f_{e_1}(\underline{x}_{e_1}), \quad (28a)$$

$$r_{e_1}(\bar{x}_{e_1}) = f_{e_1}(\bar{x}_{e_1}), \quad (28b)$$

$$\beta_2(\bar{x}_{e_1}, \bar{x}_{e_2})f_{e_0}(x_{e_0}^c)\bar{c} = f_{e_2}(\bar{x}_{e_2}). \quad (28c)$$

Note that restricting analysis on the state space (27) does not lose any generality. In fact, we can prove that the set $\tilde{\mathcal{X}} := [\underline{d}, \infty) \times [\underline{x}_{e_1}, \bar{x}_{e_1}] \times [0, \bar{x}_{e_2}]$ is positively invariant and globally attracting. That is, for any initial condition $(X_{e_0}(0), X_{e_1}(0), X_{e_2}(0)) \in \tilde{\mathcal{X}}$, $(X_{e_0}(t), X_{e_1}(t), X_{e_2}(t)) \in \tilde{\mathcal{X}}$ for $t \geq 0$ given any $(D(t), C(t)) \in [\underline{d}, 1.2] \times [0, \bar{c}]$. Besides, for any initial condition $(X_{e_0}(0), X_{e_1}(0), X_{e_2}(0)) \in \mathbb{R}_{\geq 0} \times [0, x_{e_1}^{\max}] \times [0, x_{e_2}^{\max}]$, $(X_{e_0}(t), X_{e_1}(t), X_{e_2}(t))$ enters $\tilde{\mathcal{X}}$ almost surely.

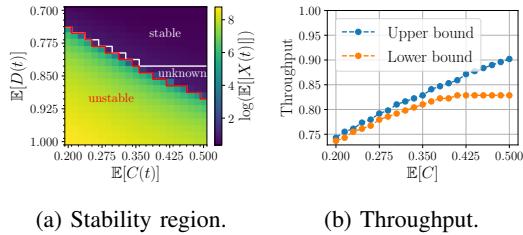


Fig. 3: Analysis given links e_1 and e_2 with finite buffer sizes.

Fig. 3a presents the time-average of traffic densities after 5×10^5 steps and discloses the stability and instability regions. We have two observations. First, there exists a gap between the stability and instability regions. This is because our stability and instability criteria are only sufficient. Second, the stability region in Fig. 3a shrinks significantly, compared with that in Fig. 2a. It indicates that congestion spillback may not be neglected in analyzing real-world scenarios.

Fig. 3b shows upper and lower bounds of throughput. We note that the gap tends to be enlarged as the compliance rate increases. Per our discussion on Theorems 2 and 3, it is possible to narrow down the gap by considering more advanced Lyapunov/test functions.

V. CONCLUDING REMARKS

This paper considered the traffic stability and throughput of a parallel-link network subject to non-compliant traffic

flows. We formulated a Markov chain that captures the traffic evolution under a dynamic routing strategy and in the face of a state-dependent non-compliance rate of drivers. Using Lyapunov methods, we derived stability conditions for typical settings with or without congestion spillback. We also used the results to analytically quantify the impact of driver non-compliance on network throughput. Possible future directions include extension of the results to general networks with cyclic structures and multi-commodity scenarios.

REFERENCES

- [1] P. Kachroo and K. Ozbay, "Solution to the user equilibrium dynamic traffic routing problem using feedback linearization," *Transportation Research Part B: Methodological*, vol. 32, no. 5, pp. 343–360, 1998.
- [2] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, "Robust distributed routing in dynamical networks with cascading failures," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 7413–7418.
- [3] R. Han, Q. Guan, F. R. Yu, J. Shi, and F. Ji, "Congestion and position aware dynamic routing for the internet of vehicles," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 16082–16094, 2020.
- [4] W. B. Powell, M. T. Towns, and A. Marar, "On the value of optimal myopic solutions for dynamic routing and scheduling problems in the presence of user noncompliance," *Transportation Science*, vol. 34, no. 1, pp. 67–85, 2000.
- [5] M. Van Essen, O. Eikenbroek, T. Thomas, and E. Van Berkum, "Travelers' compliance with social routing advice: Impacts on road network performance and equity," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1180–1190, 2019.
- [6] M. van Essen, T. Thomas, E. van Berkum, and C. Chorus, "Travelers' compliance with social routing advice: evidence from sp and rp experiments," *Transportation*, vol. 47, pp. 1047–1070, 2020.
- [7] A. Ephremides, P. Varaiya, and J. Walrand, "A simple dynamic routing problem," *IEEE Transactions on Automatic Control*, vol. 25, no. 4, pp. 690–693, 1980.
- [8] Y. Tang, Y. Wen, and L. Jin, "Security risk analysis of the shorter-queue routing policy for two symmetric servers," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 5090–5095.
- [9] Q. Xie and L. Jin, "Resilience of dynamic routing in the face of recurrent and random sensing faults," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 1173–1178.
- [10] X. Pi and Z. S. Qian, "A stochastic optimal control approach for real-time traffic routing considering demand uncertainties and travelers' choice heterogeneity," *Transportation Research Part B: Methodological*, vol. 104, pp. 710–732, 2017.
- [11] Z.-Z. Zhang, T.-Q. Tang, and H.-J. Huang, "Modeling the social-influence-based route choice behavior in a two-route network," *Physica A: Statistical Mechanics and its Applications*, vol. 531, p. 121744, 2019.
- [12] Y. Ning and L. Du, "Robust and resilient equilibrium routing mechanism for traffic congestion mitigation built upon correlated equilibrium and distributed optimization," *Transportation Research Part B: Methodological*, vol. 168, pp. 170–205, 2023.
- [13] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [14] J. G. Dai and S. P. Meyn, "Stability and convergence of moments for multiclass queueing networks via fluid limit models," *IEEE Transactions on Automatic Control*, vol. 40, no. 11, pp. 1889–1904, 1995.
- [15] S. Barman and M. W. Levin, "Throughput properties and optimal locations for limited deployment of max-pressure controls," *Transportation Research Part C: Emerging Technologies*, vol. 150, p. 104105, 2023.
- [16] O. Stein, "How to solve a semi-infinite optimization problem," *European Journal of Operational Research*, vol. 223, no. 2, pp. 312–320, 2012.
- [17] F. Blanchini, "Set invariance in control," *Automatica*, vol. 35, no. 11, pp. 1747–1767, 1999.