Resilient Control of Transportation Networks by Using Variable Speed Limits

A. Yasin Yazıcıoğlu, Mardavij Roozbehani, and Munther A. Dahleh
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
yasiny@mit.edu, mardavij@mit.edu, dahleh@mit.edu

***Abstract*—We investigate the use of variable speed limits for resilient operation of transportation networks, which are modeled as dynamical flow networks under local routing decisions. In such systems, some external inflow is injected to the so-called origin nodes of the network. The total inflow arriving at each node is routed to its operational outgoing links based on their current densities of traffic. The density on each link has first order dynamics driven by the difference of its incoming and outgoing flows. A link fails if it reaches its jam density. Such failures may propagate in the network and cause a systemic failure. We show that larger link capacities, i.e., the maximum flows that can be sustained by the links, are not always better for preventing systemic failures under local routing. Accordingly, we propose the use of variable speed limits to operate the links below their capacities, when necessary, to compensate for the lack of global information and coordination in routing decisions. We show that systemic failures under feasible external inflows can always be averted through a proper selection of speed limits if the routing decisions are sufficiently responsive to local congestion and the network is initially uncongested. This is an attractive feature as it provides a practical alternative to building more physical capacity or altering routing decisions that are determined by social behavior.**

I. Introduction

Resilience is a critical aspect in the design and operation of infrastructure systems such as transportation, power, water, and communication networks. In many applications, the network faces various disturbances and operates under the actions taken by users with limited information about the system. Such a distributed operation of the network often results in a suboptimal global performance (e.g., [1]), and it may even lead to a cascading failure with severe systemic consequences (e.g., [2], [3]). Accordingly, the resilience of the network can be quantified as the smallest amount of disturbance that leads to such a failure. There has been a significant research interest in modeling cascading failures on networks (e.g., [3], [4], [5]), investigating the influence of network topology on failure propagation (e.g., [6], [7]), and designing networks (e.g., [8], [9]) and control policies (e.g, [2], [3], [10]) for resilient operation.

The main focus of this paper is on transportation networks under local routing decisions, which are modeled as dynamical flow networks as in [2], [3]. In such networks, some external inflow is injected to the so-called origin nodes. The total inflow arriving at each node is routed to its operational outgoing links, i.e., the outgoing links that have not reached their jam densities, based on their current densities of traffic. The density on each link has first order dynamics driven by the difference of its incoming and outgoing flows. A link fails if it reaches its jam density. Such failures may propagate in the network since the outflow of a link drops to zero when there are no operational links in its immediate downstream. A systemic failure is observed if the long-term average outflow of the network (throughput) is less than the total external inflow.

Although links with larger capacities can sustain larger flows, the link capacities do not have a monotonic influence on the throughput of locally routed network flows [11]. Such a counterintiutive influence of the link capacities mainly stems from the lack of global information rather than the lack of cooperative decisions captured by the Braess' paradox [12]. More specifically, a link with a larger capacity may attract a larger demand under local routing decisions, and the resulting flow may lead to a systemic failure by overloading some critical links further in the downstream. Such failures can be avoided if the dynamics yield a sufficient back-propagation of congestion effects through the densities [13]. Otherwise, additional control and communication mechanisms are needed to mitigate the fragility arising from local decisions. One approach in this direction is to make the global network state accessible via communications. For instance, GPS-based route guidance systems are used in road traffic networks to provide the drivers with real-time traffic data. However, in distributed systems, the availability of more information does not guarantee the desired global performance in the absence of a properly coordinated response (e.g., [14], [15], [16]). Alternatively, some system objectives can be achieved via a controlled operation of the network, without requiring any global information and coordination in the user decisions (e.g., [11], [17]).

In this paper, we investigate the use of variable speed limits for avoiding systemic failures in transportation networks under local routing decisions. Some preliminary results of this work appeared in [18]. The main contributions of this paper are as follows:

- We show that for a large family of network topologies, a constant use of maximum speed limits can lead to systemic failures under feasible external inflows for any local routing policy.
- We show that the systemic failures due to local routing decisions can always be averted through a proper use of variable speed limits as a flow control mechanism if the routing decisions are sufficiently responsive to local congestion and the initial densities on the links are sufficiently small.

- We provide a tractable method for generating such speed limits based on some limited information regarding the aggregate routing behavior of drivers. Accordingly, throughput optimality is ensured without needing any global information and coordination in the routing decisions, even when there is not an inherent sufficient back-propagation of congestion effects in the network.

The organization of this paper is as follows: Section II presents some related work. Section III provides some preliminaries. Section IV defines the dynamical model of network flows under local routing. Section V provides some motivation and examples. Section VI presents our main results. Finally, Section VII concludes the paper.

II. Related Work

Recent advances in sensing, communication, and computation technologies have created tremendous opportunities for improving the resilience of transportation systems by properly responding to real-time data. Accordingly, there has been a growing interest in dynamical (macroscopic) models of transportation networks (e.g., [2], [3], [19], [20], [21], [22], [23]). The dynamical model in this paper is closely related to the model in [2], [3] with the following main differences: We consider flow functions that are non-monotonic in densities due to congestion collapse, and we incorporate the speed limits as external control inputs in the flow functions. Such flow-density relationships are indicated by many empirical studies on transportation networks (e.g., see [24] and the references therein) and violate the monotonicity of the dynamical flow model, which was a key system property for the derivations in [2], [3]. In the absence of montonicity, our analysis shows how variable speed limits can be used to induce some positively invariant sets of density profiles, which ensure throughput optimality, as long as the routing choices are sufficiently responsive to local congestion. In this regard, similar to the various routing (e.g., back-pressure [25]), flow control, and congestion control policies (e.g., congestion pricing [26], [27]) in the literature on queuing theory and communication networks, the variable speed limits serve as a feedback mechanism for facilitating a proper response in the upstream based on the condition of the downstream. Accordingly, throughput optimality is achieved without requiring any alteration of the external inflow or the routing behavior.

Some earlier works in the literature have also investigated the use of variable speed limits for mitigating congestion in transportation systems (e.g., [28], [29], [30]). The studies in [28], [29], [30] consider discrete-time second-order flow models (e.g., [20]) for freeways, which are divided into discrete cells, and investigate the design of speed limit policies for reducing congestion. In [28], the authors utilize integrator backstepping for obtaining the speed limits that stabilize the density of traffic in each cell of a freeway to some desired value. Another prevalent approach is to formulate the design of variable speed limits as a (stochastic) optimal control problem (e.g., minimize the average delay). Such formulations typically lead to numerical solutions or approximate methods due to the complexity of the flow dynamics. A stochastic finite-horizon optimal control problem is posed in [29] and a suboptimal solution is obtained by pairing an extended Kalman filter with a parametrized controller that is tuned based on Monte Carlo simulations. Model predictive control is employed in [30] for optimally coordinating variable speed limits with the aim of suppressing shock waves. Similar to variable speed limits, traffic lights can also be used as a flow control mechanism for mitigating congestion in transportation systems. In this regard, one of the prevalent methods is ramp metering (e.g., [19], [31], [32], [33]). In [19], [31], the authors consider some first-order models similar to the one in [22] and present ramp metering strategies for optimizing the throughput or the average delay. In [32], the authors consider a second-order model and present a heuristic that is based on the coordination of metered ramps, each of which executes a local feed-back policy. The co-design of ramp metering and variable speed limits is formulated in [33] as a finite-horizon optimal control problem, which is then solved numerically via a feasible directions algorithm. In this paper, we consider a continuous-time first-order flow model and present how variable speed limits can be used as a robust way of ensuring throughput optimality under rich families of routing behavior that are sufficiently responsive to local congestion.

III. Preliminaries

A. Notation

For any finite set A with cardinality $|A|$, we use $\mathbb{R}^A$ ($\mathbb{R}^A_+$, $\mathbb{R}^A_{++}$) to denote the space of real-valued (nonnegative-real-valued, positive-real-valued) $|A|-$dimensional vectors whose components are indexed by the elements of A. Accordingly, for any $a \in A$ and $x \in \mathbb{R}^A$, $x_a \in \mathbb{R}$ denotes the corresponding entry of x. Similarly, for any $A' \subseteq A$, we use $x_{A'} \in \mathbb{R}^{A'}$ to denote the $|A'|$-dimensional vector consisting only of the components of x whose indices are in A'. For any pair of vectors $\underline{x}, \bar{x} \in \mathbb{R}^A$, we use $[\underline{x}, \bar{x})$ to denote the set of vectors $x \in \mathbb{R}^A$ such that $\underline{x} \leq x < \bar{x}$, i.e., $\underline{x}_a \leq x_a < \bar{x}_a$ for all $a \in A$. The all-zero vector, its size being clear from the context, will be denoted by $\mathbf{0}$.

B. Graph basics

A directed graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, consists of a set of nodes, $\mathcal{V}$, and a set of edges, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, given by ordered pairs of nodes. A graph is a multi-graph if multiple edges are allowed between the nodes, i.e., if $\mathcal{E}$ is a multi-set. Each $(v, w) \in \mathcal{E}$ denotes a link from v (the tail) to w (the head). A path is a sequence of nodes such that $(v, w) \in \mathcal{E}$ for any pair of nodes $v, w \in \mathcal{V}$ that are consecutive in the sequence. For each $v \in \mathcal{V}$, we use $\mathcal{E}^-_v$ and $\mathcal{E}^+_v$ to denote the corresponding sets of incoming and outgoing links, respectively. Similarly, for any $\mathcal{U} \subseteq \mathcal{V}$, $\mathcal{E}^-_{\mathcal{U}} = \{(v, w) \in \mathcal{E} \mid v \notin \mathcal{U}, w \in \mathcal{U}\}$ and $\mathcal{E}^+_{\mathcal{U}} = \{(v, w) \in \mathcal{E} \mid v \in \mathcal{U}, w \notin \mathcal{U}\}$.

IV. System Model

In this section, we provide some definitions and the dynamical model of locally routed flows.

Definition (*Flow Network*): A flow network is a directed multi-graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ such that there exists a path from any node to some destination node $v \in \mathcal{V}_D$, where

$$\mathcal{V}_D = \{v \in \mathcal{V} \mid \mathcal{E}_v^+ = \emptyset\}.$$

Furthermore, the set of origin nodes $\mathcal{V}_O$ is defined as

$$\mathcal{V}_O = \{v \in \mathcal{V} \mid \mathcal{E}_v^- = \emptyset\}.$$

The nodes that are neither an origin nor a destination constitute the set of intermediate nodes $\mathcal{V}_I = \mathcal{V} \setminus \mathcal{V}_O \setminus \mathcal{V}_D$.

Tranportation networks are often studied by using single-commodity flow models (e.g., [13], [19], [20], [22], [28]-[33]). Such models provide a natural representation for single-destination networks, and they can also be used for multi-destination networks when routing is assumed to be independent of the state (e.g., see [22]). While such an assumption does not introduce a major drawback when analyzing a network at some nominal operating conditions, the feedback in routing decisions is an important aspect of the system in the face of disturbances. In this regard, we present a single-commodity dynamical flow model where the routing decisions are responsive to the local densities of traffic. Accordingly, we restrict our focus to single-destination networks. Analyzing multi-destination networks, where each vehicle is routed based on the observed congestion and must arrive at a specific destination, requires a multi-commodity model (e.g., [34]) since drivers with different destinations would respond differently to the same observations.

In this single-commodity dynamical flow model, each link $e \in \mathcal{E}$ carries some flow $f_e \in \mathbb{R}_+$ that is related to the average number of vehicles per unit length (density) $\rho_e \in \mathbb{R}_+$ and their average speed $s_e \in \mathbb{R}_+$ as

$$f_e = \rho_e s_e. \tag{1}$$

Furthermore, the vehicles are assumed to move at the maximum admissible speed, i.e.,

$$s_e = \min(\bar{s}_e(\rho_e), u_e), \tag{2}$$

where $\bar{s}_e(\rho_e)$ denotes the maximum speed at which the vehicles can travel along e when the density is ρ_e, and $u_e \in [0, \bar{u}_e]$ is the speed limit imposed on the link. In transportation networks, $\bar{s}_e(\rho_e)$ is a monotonically decreasing function since a higher density corresponds to a shorter following distance, for which the maximum safe speed is lower. Furthermore, each $\bar{u}_e \in \mathbb{R}_{++}$ is typically set based on some safety, efficiency, and environmental considerations.

In light of (1) and (2), the flow on each link is a function of its density and speed limit, i.e.,

$$f_e : [0, \bar{\rho}_e] \times [0, \bar{u}_e] \mapsto [0, \bar{f}_e],$$

where $\bar{f}_e \in \mathbb{R}_{++}$ denotes the finite capacity of the link, that is, the maximum flow that can be sustained by the link, and $\bar{\rho}_e \in \mathbb{R}_{++}$ is the finite jam density, for which $\bar{s}_e(\bar{\rho}_e) = 0$. Accordingly,

$$f_e(0, u_e) = f_e(\bar{\rho}_e, u_e) = 0, \quad \forall u_e \in [0, \bar{u}_e]. \tag{3}$$

Note that the flow becomes a function of only the density when the speed limit is constant. For instance, $f_e(\rho_e, \bar{u}_e)$ represents how the flow depends on the density when the link is operated constantly with the maximum speed limit.

Assumption 1 Each $f_e(\rho_e, \bar{u}_e)$ is continuous in ρ_e, and there exists a congestion threshold $\rho_e^c \in (0, \bar{\rho}_e)$ such that $f_e(\rho_e, \bar{u}_e)$ is strictly increasing on the interval $(0, \rho_e^c)$ and strictly decreasing on the interval $(\rho_e^c, \bar{\rho}_e)$.

The structure in Assumption 1 is typical for the flow functions in transportation networks (e.g., see [24]). Accordingly, the capacity of link e is $\bar{f}_e = f_e(\rho_e^c, \bar{u}_e)$, and the link is said to be congested when $\rho_e > \rho_e^c$.

In this paper, we focus on networks where the total inflow to each node is routed locally to its operational outgoing links based on their current densities. Such routing behaviors are modeled as local routing policies.

Definition (*Local Routing Policy*): For any flow network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the finite jam densities $\bar{\rho} \in \mathbb{R}_{++}^{\mathcal{E}}$, a local routing policy $\mathcal{R}$ is a family of functions

$$\mathcal{R}^v : [\mathbf{0}, \bar{\rho}_{\mathcal{E}_v^+}] \times \mathbb{R}_+ \mapsto \mathbb{R}_+^{\mathcal{E}_v^+}, \; v \in \mathcal{V} \setminus \mathcal{V}_D,$$

such that, for any inflow $\mu_v \in \mathbb{R}_+$,

$$\sum_{e \in \mathcal{E}_v^+} \mathcal{R}_e^v(\rho_{\mathcal{E}_v^+}, \mu_v) = \mu_v, \; \forall \rho_{\mathcal{E}_v^+} \neq \bar{\rho}_{\mathcal{E}_v^+}, \tag{4}$$

$$\mathcal{R}_e^v(\rho_{\mathcal{E}_v^+}, \mu_v) = 0, \; \forall e \in \mathcal{E}_v^+ : \rho_e = \bar{\rho}_e, \tag{5}$$

where $\mathcal{R}_e^v(\rho_{\mathcal{E}_v^+}, \mu_v)$ is the flow routed to the link $e \in \mathcal{E}_v^+$.

As per the definition above, local routing policies satisfy two realistic constraints: 1) For every non-destination node, the total inflow equals the total outflow as long as the node has at least one operational outgoing link, i.e., (4), and 2) jammed links cannot receive any inflow since they are fully occupied and have no outflow, i.e., (5). Accordingly, each vehicle arriving at a non-destination node takes one of the operational outgoing links since jammed links cannot recover and the destination is reachable from every other node in the network.

Definition (*Locally Routed Flow*): For any flow network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, flow functions f, local routing policy $\mathcal{R}$, and external inflow $\lambda \in \mathbb{R}_+^{\mathcal{V}_O}$, the locally routed flow $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is a dynamical system such that, for every $e = (v, w) \in \mathcal{E}$,

$$\dot{\rho}_e = \begin{cases} \mathcal{R}_e^v(\rho_{\mathcal{E}_v^+}, \mu_v) - f_e(\rho_e, u_e), & \text{if } \rho_{\mathcal{E}_w^+} \neq \bar{\rho}_{\mathcal{E}_w^+}, \\ \mathcal{R}_e^v(\rho_{\mathcal{E}_v^+}, \mu_v), & \text{o.w.} \end{cases}, \tag{6}$$

where μ_v denotes the total inflow to v, i.e.,

$$\mu_v = \begin{cases} \lambda_v, & \text{if } v \in \mathcal{V}_O, \\ \sum\limits_{\epsilon \in \mathcal{E}_v^-} f_\epsilon(\rho_\epsilon, u_\epsilon), & \text{o.w.} \end{cases}, \tag{7}$$

and $u_e \in [0, \bar{u}_e]$ is the speed limit imposed on e.

In light of (3) and (5), each $\bar{\rho}_e$ is an equilibrium point under (6). As such, a link stops carrying flow (fails) if it reaches its jam density. Furthermore, since (6) implies $\dot{\rho}_e \geq 0$ for any $e = (v, w) \in \mathcal{E}$ such that $\rho_{\mathcal{E}_w^+} = \bar{\rho}_{\mathcal{E}_w^+}$, link failures can propagate in the network. While some failures may not have a severe impact on the overall performance, others can cause a cascading failure of some critical links and lead to a non-transferring system.

Definition (*Transferring System*): A locally routed flow is transferring (throughput optimal) if the long-term average inflow to the destination nodes (throughput) is equal to the total external inflow, i.e.,

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t \sum_{v\in\mathcal{V}_D} \mu_v(\tau)d\tau = \sum_{v\in\mathcal{V}_O} \lambda_v.$$

In transferring systems, the amount of vehicles that are unable to reach a destination node remains bounded as time goes to infinity. In contrast, non-transferring systems experience a systemic failure, i.e., at least one origin node, which has a non-zero external inflow, becomes disconnected from the network in finite time due to link failures (see Lemma 6.2). As such, eventually it becomes impossible to transfer the corresponding portion of the external inflow through the network. Of course in reality, a very small portion of this external inflow can still enter the network since links might have a very small amount of outflow even in bumper-to-bumper traffic. However, as long as the demand persists, the outgoing links of an origin can not recover once they are jammed since the vehicles at that origin node would try to flow into the network as much as they can, i.e., each of those jammed links would receive an inflow equal to its outflow. Hence, a severe throughput loss would be observed.

V. Motivation

Consider any locally routed network flow $(\mathcal{G}, f, \mathcal{R}, \lambda)$, where all the links are operated constantly under their respective maximum admissible speed limits, i.e., $u(t) = \bar{u}$, so that they all can sustain flows up to their respective capacities. If such a system is non-transferring despite the initial densities being sufficiently small (e.g., $\rho(0) = \mathbf{0}$), this failure is due to one of the two possible reasons:

- The external inflow λ is too large (infeasible) and cannot be transferred under any routing policy.
- The external inflow λ is feasible, yet the routing decisions lead to failure.

In this regard, an external inflow is feasible if the corresponding set of feasible equilibrium flows, i.e.,

$$\begin{aligned}\mathcal{F}(\bar{f}, \lambda) = \{f \in \mathbb{R}_+^{\mathcal{E}} \mid f_e \le \bar{f}_e,\ \forall e \in \mathcal{E}\\ \lambda_v = \sum_{e\in\mathcal{E}_v^+} f_e,\ \forall v \in \mathcal{V}_O\\ \sum_{e\in\mathcal{E}_v^-} f_e = \sum_{e\in\mathcal{E}_v^+} f_e,\ \forall v \in \mathcal{V}_I\}\end{aligned} \tag{8}$$

is nonempty. The feasibility of an external inflow can be checked based on the max-flow min-cut theorem [35], which implies that $\mathcal{F}(\bar{f}, \lambda) \neq \emptyset$ if and only if

$$\sum_{e\in\mathcal{E}_\mathcal{U}^+} \bar{f}_e - \sum_{v\in\mathcal{U}} \lambda_v \ge 0,\ \forall \mathcal{U} \subseteq \mathcal{V} \setminus \mathcal{V}_D. \tag{9}$$

Accordingly, for any infeasible external inflow, there exists some $\mathcal{U} \subseteq \mathcal{V} \setminus \mathcal{V}_D$, which receives a total external inflow larger than the total maximum flow sustainable by its outgoing links. For such a set of nodes $\mathcal{U}$, the total density on the outgoing links of the nodes in $\mathcal{U}$ keeps increasing as long as all the origin nodes in $\mathcal{U}$ remain connected to the network. Accordingly, at least one of those origin nodes must eventually become disconnected from the network since the jam densities are finite. Hence, infeasible external inflows lead to a non-transferring system, regardless of the routing behavior.

In light of (9), systemic failures due to infeasible external inflows cannot be avoided without reducing the external inflow or building more capacity in the network. However, this is not the case for failures due to routing decisions. In fact, we will show that such failures can be avoided by using the speed limits to intentionally operate some of the links below their capacity, provided that the routing policy is sufficiently congestion-aware.

Definition (*Congestion Awareness*): For any flow network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and flow functions f, a local routing policy $\mathcal{R}$ is congestion-aware at a density profile $\rho^* \in [\mathbf{0}, \bar{\rho}]$ if, for every $e = (v, w) \in \mathcal{E}$,

$$\rho_e \ge \rho_e^* \Rightarrow \mathcal{R}_e^v(\rho_{\mathcal{E}_v^+}, \mu_v) \le \phi_e(\rho_e),\ \forall \mu_v \le \sum_{j\in\mathcal{E}_v^+} \phi_j(\rho_j), \tag{10}$$

where $\phi_e(\rho_e)$ denotes the maximum sustainable inflow, i.e.,

$$\phi_e(\rho_e) = \begin{cases} \bar{f}_e, & \text{if } \rho_e \le \rho_e^c, \\ f_e(\rho_e, \bar{u}_e), & \text{o.w.} \end{cases} \tag{11}$$

Note that every routing policy is trivially congestion-aware at $\bar{\rho}$ due to (5). Furthermore, if a policy is congestion-aware at some ρ^*, then by definition it is also congestion-aware at any density profile in $[\rho^*, \bar{\rho}]$. In the remainder of this paper, we will refer to a routing policy simply as a congestion-aware policy if it satisfies (10) for every $\rho^* \in [\mathbf{0}, \bar{\rho}]$.

For any link $e = (v, w)$ and $\rho_e \in [0, \bar{\rho}_e]$, $\phi_e(\rho_e)$ is the maximum constant inflow that can be sustained by the link when its current density is ρ_e. Indeed, if $\mathcal{R}_e^v(\rho_{\mathcal{E}_v^+}, \mu_v)$ in (6) is replaced by a constant value larger than $\phi_e(\rho_e)$, then ρ_e is guaranteed to converge to $\bar{\rho}_e$ under the resulting dynamics. As such, (10) indicates some local failure-avoidance behavior. In particular, when the density of a link $e = (v, w)$ is equal to or greater than the corresponding ρ_e^*, the flow routed to e does not exceed its maximum sustainable inflow as long as node v is receiving a sufficiently small inflow μ_v. It is worth mentioning that having the inflow to link e upper bounded by $\phi_e(\rho_e)$ is not sufficient by itself to avoid the failure of e. For example, if all the links in its immediate downstream, $\mathcal{E}_w^+$, become jammed (i.e., $\rho_{\mathcal{E}_w^+} = \bar{\rho}_{\mathcal{E}_w^+}$) or if the speed limit u_e is kept very low, then link e may still be driven to failure. However, as we will show later in the paper, a proper choice of speed limits ensures a transferring system under feasible external inflows when the local routing decisions are sufficiently congestion-aware.

A. Motivating Example

Consider the flow network in Fig. 1a. Let all the links have similar characteristics except for their cross-sectional dimensions. Accordingly, let $f_0(\rho_0, u_0)$ denote a normalized flow function such that

$$f_e(\rho_e, u_e) = c_e f_0\Big(\frac{\rho_e}{c_e}, u_e\Big), \quad \forall e \in \mathcal{E}, \tag{12}$$

where $c_e \in \mathbb{R}_{++}$ is determined by the cross-sectional dimension of the link e. For instance, each c_e may correspond to the number of lanes in e, whereas f_0 denotes the flow function for a single lane. Let all the links be constantly operated under the maximum speed limit $\bar{u}_0 \in \mathbb{R}_{++}$, and let $f_0(\rho_0, \bar{u}_0)$ be as illustrated in Fig. 1b. Accordingly, the capacity of each link is $\bar{f}_e = c_e \bar{f}_0$ and provided next to itself in Fig. 1a. Suppose that initially there are no vehicles in the network, i.e., $\rho(0) = \mathbf{0}$, and an external inflow of $\lambda = 5\bar{f}_0$ is injected. Note that this is a feasible external inflow for this network since the minimum value for the left side of (9) is $2\bar{f}_0$, which is obtained for $\mathcal{U} = \{1, 2\}$.

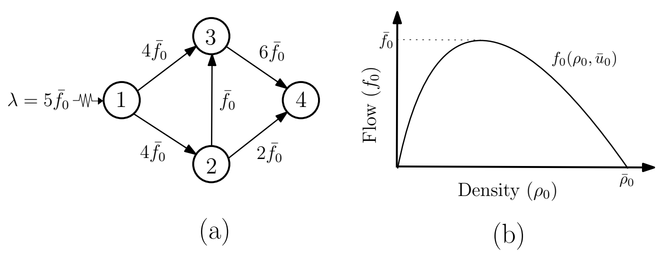


Fig. 1. A flow network along with the external inflow and link capacities are illustrated in (a), and the normalized flow function is shown in (b).

Since the external inflow is feasible and $\rho(0) = \mathbf{0}$, whether the system is transferring depends on the routing policy. Consider a local routing policy $\mathcal{R}$ such that the flow routed to each edge is proportional to its maximum sustainable inflow, that is, for every $e = (v, w) \in \mathcal{E}$,

$$\mathcal{R}_e^v(\rho_{\mathcal{E}_v^+}, \mu_v) = \mu_v \frac{\phi_e(\rho_e)}{\sum\limits_{j \in \mathcal{E}_v^+} \phi_j(\rho_j)}, \forall \rho_{\mathcal{E}_v^+} \neq \bar{\rho}_{\mathcal{E}_v^+}. \tag{13}$$

Such a routing policy models the aggregate behavior that emerges under some intuitive driver choices. At each junction, suppose that the drivers randomly route to the outgoing links with probabilities proportional to their capacities when they are all uncongested. For the flow functions satisfying (12), this is equivalent to routing with probabilities proportional to the number of lanes on the roads. For example, if a junction has two outgoing links such that the first link has two lanes and the second link has three lanes, then the drivers route to the first link with probability 0.4 if both links are uncongested. When some of the outgoing links are congested, then allocated probabilities are adjusted in accordance with the maximum inflows those link can support. When all drivers make such randomized routing decisions, the resulting local outflows would approach their expected values as given in (13). It is also worth mentioning that the proportional routing is provided here just as one illustrative example of local routing behavior and our technical analysis will not be limited to this specific policy.

In light of (10), the proportional routing policy in (13) is congestion-aware. Furthermore, it can be shown that when $\rho(0) \in [\mathbf{0}, \rho^c]$ and the maximum speed limits are constantly applied as $u(t) = \bar{u}$, the tail of the first link that exceeds its congestion threshold under this policy must have received an inflow larger than the total capacity of its outgoing links. For the network in Fig. 1a, it can be easily verified that node 1 (since λ is less than its total outgoing capacity) and node 3 (since its total incoming capacity is less than its total outgoing capacity) never receive such excessive inflows. That is also true for node 2 since $\mathcal{R}^1$ routes λ evenly to the links in $\mathcal{E}_1^+$ as long as they are both uncongested. Accordingly, μ_2 monotonically increases to $2.5\bar{f}_0$, which is less than the total capacity of the links in $\mathcal{E}_2^+$. Hence, none of the links exceed their congestion thresholds in this example, and the resulting system $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is transferring.

As a second example, consider the same network when $c_{(2,4)}$ is reduced by one due to some disturbance. For instance, such a reduction can be due to one of the lanes in the link $(2, 4)$ being closed due to maintenance or an accident. Note that the external inflow of $\lambda = 5\bar{f}_0$ is still feasible under the resulting link capacities as shown in Fig. 2a. In this case, the minimum value for the left side of (9) is $\bar{f}_0$, which is obtained for $\mathcal{U} = \{1, 2\}$. Same as the previous case, consider the routing policy $\mathcal{R}$ that satisfies (13) for the perturbed flow functions in this example, and let the system have $\rho(0) = \mathbf{0}$ and $u(t) = \bar{u}$. Under the resulting dynamics, the densities of the links in $\mathcal{E}_1^+$ initially follow the same trajectory as before. However, once μ_2 exceeds $2\bar{f}_0$ (the total capacity of the links in $\mathcal{E}_2^+$), the links in $\mathcal{E}_2^+$ will be driven to their jam densities. Note that λ is not feasible for the remaining network once those two links fail, and the system inevitably reaches the state illustrated in Fig. 2d. As such, $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is non-transferring for this example.

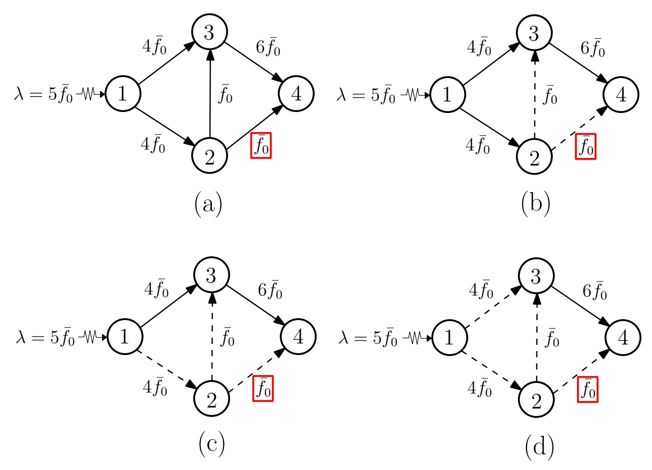


Fig. 2. A reduction in the capacity of the link $(2, 4)$ as shown in (a) leads to a cascade of failures as illustrated in (b)-(c)-(d).

To avoid the failures in the second example, a larger portion of λ should be routed to the link $(1, 3)$ before the links in $\mathcal{E}_2^+$ become overloaded. However, local routing policies only respond to the densities on the immediate links. Hence, even a congestion-aware routing policy such as (13) may fall short in circumventing systemic failures when the links can sustain flows that may overload some critical links further in their downstream. In this regard, large capacity of the link $(1,2)$ was actually a liability in the second example. As we will show later in the paper, the systemic failure in this example could actually be avoided by reducing the capacity of the link $(1,2)$ so that its outflow never exceeds $2\bar{f}_0$.

It should also be emphasized that the issue highlighted in this example and the proposed solution are not limited to the case where the outflow of each link has a discontinuous dependence on the congestion in its immediate downstream as in (6). Systemic failures due to local routing can occur even if the outflow of each link degrades gradually as its downstream becomes congested, unless such a degradation facilitates a sufficiently fast back-propagation of congestion. An example and some further discussion on this matter is provided in the Appendix.

While such a non-monotonic influence of link capacities on throughput is reminiscent of the Braess' paradox [12], there are major differences between the two phenomena. The analysis in [12] is based on a static flow model, where each driver is assumed to observe the whole network and take a route with minimum delay (time it takes to reach the destination). Under such a model, it was shown that adding capacity to a network can sometimes reduce the efficiency of equilibrium and cause larger delays for everyone due to selfish routing. In contrast, this work is focused on dynamical flows under local routing decisions and shows that having more capacity can sometimes lead to a systemic failure, which is a form of instability, due to local routing decisions.

B. Capacity Allocation via Speed Limits

As illustrated in the previous section, having larger link capacities may lead to systemic failures when the routing decisions are local. In this section, we present how speed limits can be used to address this issue by intentionally operating certain links below their capacities when necessary.

Given a flow function $f_e(\rho_e, \bar{u}_e)$ and some $\bar{f}_e^* \in [0, \bar{f}_e]$, suppose that we would like to design the speed limit u_e such that the flow on the link never exceeds $\bar{f}_e^*$. In order to avoid unnecessary delays, it would also be desired to keep u_e as large as possible. The solution to this problem depends on the available flexibility in the design. For instance, if a constant speed limit is required, then (1) and (2) together with Assumption 1 imply that the solution is

$$u_e = \frac{\bar{f}_e^*}{\hat{\rho}_e(\bar{f}_e^*)}, \tag{14}$$

where

$$\hat{\rho}_e(\bar{f}_e^*) = \max\{\rho_e \in [0, \bar{\rho}_e] \mid f_e(\rho_e, \bar{u}_e) = \bar{f}_e^*\}. \tag{15}$$

Such a constant speed limit may be too conservative when the density on the link is low. Alternatively, a feedback law can induce the desired capacity more efficiently. For instance, the following policy achieves this task while allowing for the maximum possible speed at all times:

$$u_e(\rho_e) = \begin{cases} \bar{u}_e, & \text{if } f_e(\rho_e, \bar{u}_e) \le \bar{f}_e^*, \\ \dfrac{\bar{f}_e^*}{\rho_e}, & \text{o.w.} \end{cases} \tag{16}$$

Fig. 3 illustrates how the speed limits in (14) and (16) influence the flow on a link for a typical flow function. Accordingly, if $\bar{f}_e^*$ is chosen properly, such an operation of the link ensures that the link accumulates a higher density itself rather than overloading some critical links further in its downstream. This way, potential failures due to the lack of global information can be avoided if the local routing decisions properly respond to the increasing density on the link. In the next section, we provide our formal results regarding such use of speed limits to avert systemic failures in locally routed flows.

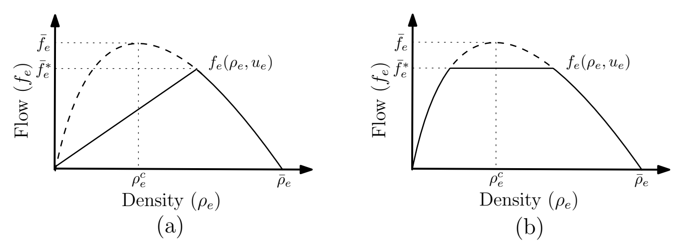


Fig. 3. Capacity allocation via the speed limit policies in (14) and (16) are illustrated in (a) and (b), respectively. The solid curves show the resulting flow functions, the dashed curves illustrate the deviation from $f_e(\rho_e, \bar{u}_e)$.

VI. Main Results

We start our analysis with our first main result, which formalizes the message illustrated through the examples in Section V-A, i.e., constantly using the maximum speed limits (utilizing the full capacity of all links) may lead to systemic failures due to local routing decisions. In fact, for a rich family of flow networks, it is possible to have perturbed flow functions and feasible external inflows such that a systemic failure is inevitable under the constant speed limits $u(t) = \bar{u}$ for any initial condition and local routing policy.

Theorem 6.1. *Consider any flow network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the flow functions $f_e(\rho_e, \bar{u}_e)$. If $\mathcal{G}$ has an origin whose immediate downstream, $\mathcal{W} \subset \mathcal{V}$, consists of multiple nodes, each of which has a link to a node outside $\mathcal{W}$, i.e.,*

$$\mathcal{E}_w^+ \cap \mathcal{E}_{\mathcal{W}}^+ \neq \emptyset, \forall w \in \mathcal{W}, \tag{17}$$

then, for any local routing policy $\mathcal{R}$ and initial condition $\rho(0) \in [\mathbf{0}, \bar{\rho}]$, there exist perturbed flow functions $f'_e(\rho_e, \bar{u}_e) \le f_e(\rho_e, \bar{u}_e)$ and feasible (under the perturbed flow functions) external inflows $\lambda \in \mathbb{R}_+^{\mathcal{V}_O}$ such that $(\mathcal{G}, f', \mathcal{R}, \lambda)$ is non-transferring under the constant speed limits $u(t) = \bar{u}$.

Proof. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a flow network with the flow functions $f_e(\rho_e, \bar{u}_e)$, and let $v \in \mathcal{V}_O$ be an origin node such that its immediate downstream, $\mathcal{W} \subset \mathcal{V}$, satisfies $|\mathcal{W}| \ge 2$ and (17). For the sake of simplicity in notation, we will assume that $\mathcal{G}$ contains no parallel edges, and the proof can easily be extended to multi-graphs. Consider some external inflow

$\lambda \in \mathbb{R}_+^{\mathcal{V}_O}$ such that

$$\lambda_v \in (0, \min_{e \in \mathcal{E}} \bar{f}_e],$$

$$\lambda_o = 0, \forall o \in \mathcal{V}_O \setminus \{v\}.$$

Consider some flow functions $f'_e(\rho_e, \bar{u}_e) \leq f_e(\rho_e, \bar{u}_e)$ such that only the outgoing links of the nodes in W, i.e.,

$$\mathcal{E}' = \bigcup_{w \in \mathcal{W}} \mathcal{E}_w^+,$$

are perturbed and $f'_e(\rho_e, \bar{u}_e) = f_e(\rho_e, \bar{u}_e)$ for every other link $e \in \mathcal{E} \setminus \mathcal{E}'$. Let the resulting link capacities satisfy

$$\bar{f}'_e = \alpha_e \lambda_v, \forall e \in \mathcal{E}', \tag{18}$$

for some $\alpha \in (0,1)^{\mathcal{E}'}$ satisfying

$$\sum_{e \in \mathcal{E}_{\mathcal{W}}^+} \alpha_e = 1, \tag{19}$$

Note that λ is a feasible inflow for any such f', i.e.,

$$\sum_{e \in \mathcal{E}_{\mathcal{U}}^+} \bar{f}'_e - \sum_{o \in \mathcal{U}} \lambda_o \geq 0, \ \forall \mathcal{U} \subseteq \mathcal{V} \setminus \mathcal{V}_D.$$

Furthermore,

$$\sum_{e \in \mathcal{E}_{\mathcal{W}}^+} \bar{f}'_e - \lambda_v = 0.$$

Hence, if all the outgoing links of any $w \in \mathcal{W}$ fail or if any outgoing link of v fails, then $(\mathcal{G}, f', \mathcal{R}, \lambda)$ is non-transferring since λ becomes infeasible for the remaining network. In the remainder of the proof, we will show that, for any local routing policy $\mathcal{R}$ and initial condition $\rho(0) \in [\mathbf{0}, \bar{\rho}]$, there exists $\alpha \in (0,1)^{\mathcal{E}'}$ that satisfies (19) and ensures such a failure under the constant speed limits $u(t) = \bar{u}$.

For the sake of contradiction, suppose that, for some local routing policy $\mathcal{R}$ and initial condition $\rho(0) \in [\mathbf{0}, \bar{\rho}]$, $(\mathcal{G}, f', \mathcal{R}, \lambda)$ is transferring under the constant speed limits $u(t) = \bar{u}$ for any $\alpha \in (0,1)^{\mathcal{E}'}$ satisfying (19). Then, since every $w \in \mathcal{W}$ must always have at least one operational outgoing link, (6) implies

$$\dot{\rho}_e = \mathcal{R}_e^v(\rho_{\mathcal{E}_v^+}, \lambda_v) - f'_e(\rho_e, \bar{u}_e), \forall e \in \mathcal{E}_v^+.$$

In that case, the densities on the outgoing links of v, $\rho_{\mathcal{E}_v^+}(t)$, evolve independent of the rest of the system and their trajectory is determined by the local initial state $\rho_{\mathcal{E}_v^+}(0)$. For each $e \in \mathcal{E}_v^+$, let $\beta_e \in [0,1]$ denote the ratio of its long-term average outflow to λ_v, i.e.,

$$\beta_e = \frac{\liminf_{t \to \infty} \frac{1}{t} \int_0^t f'_e(\tau) d\tau}{\lambda_v}. \tag{20}$$

Since the total inflow to any node $w \in \mathcal{W}$ equals the total outflow of its incoming edges, (18) and (20) together imply that, for every $e = (v,w) \in \mathcal{E}_v^+$,

$$\sum_{\epsilon \in \mathcal{E}_w^+} \alpha_\epsilon < \beta_e \Rightarrow \liminf_{t \to \infty} \frac{1}{t} \int_0^t \mu_w(\tau) d\tau > \sum_{\epsilon \in \mathcal{E}_w^+} \bar{f}'_\epsilon.$$

As such, if $\sum_{\epsilon \in \mathcal{E}_w^+} \alpha_\epsilon < \beta_e$, then all the outgoing links of w are guaranteed to reach their jam densities since the long-term average inflow to w is larger than its total outgoing capacity. Since all the links in $\mathcal{E}_v^+$ must remain operational for the system to be transferring,

$$\sum_{e \in \mathcal{E}_v^+} \beta_e = 1. \tag{21}$$

Due to (21), there is at least one link $e = (v,w) \in \mathcal{E}_v^+$ such that $\beta_e > 0$. Then, for any $\alpha \in (0,1)^{\mathcal{E}'}$ such that (19) is satisfied and $\sum_{\epsilon \in \mathcal{E}_w^+} \alpha_\epsilon < \beta_e$, we reach a contradiction since all the links in $\mathcal{E}_w^+$ are bound to fail, which implies that $(\mathcal{G}, f', \mathcal{R}, \lambda)$ is non-transferring.

□

Theorem 6.1 shows that the constant use of maximal speed limits $u(t) = \bar{u}$ introduces a fragility under any local routing policy for a large family of flow networks. For example, the network in Fig. 1a satisfies the premise of Theorem 6.1 since each node in the immediate downstream of the origin, i.e., $\mathcal{W} = \{2,3\}$, has a link to node 4. Hence, no matter what the local routing policy and initial densities are, there exist perturbed flow functions and feasible external inflows such that a systemic failure is inevitable on this network under the constant speed limits $u(t) = \bar{u}$. In this regard, the significant structural property is a node with links to multiple intermediate nodes. At such junctions, local routing decisions cannot ensure global failure-avoidance without a proper flow control mechanism since the conditions of the links further in the downstream of those intermediate nodes are unknown.

Our second main result, Theorem 6.3, shows that the systemic failures due to the lack of global information in routing decisions can always be prevented through a proper choice of variable speed limits. In particular, if the local routing policy is congestion-aware at a density profile $\rho^* \in [\mathbf{0}, \bar{\rho}]$ that is compatible with the external inflow, then there exist speed limits $u(t) \in [\mathbf{0}, \bar{u}]$ that ensure a transferring system for any initial condition $\rho(0) \in [\mathbf{0}, \rho^*]$. First, we present Lemma 6.2, which will be used in proving Theorem 6.3.

Lemma 6.2. *If a locally routed flow $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is non-transferring, then there exists an origin node $v \in \mathcal{V}_O$ such that its external inflow is non-zero and all its outgoing links reach their jam densities in some finite time τ, i.e.,*

$$\lambda_v > 0, \quad \rho_{\mathcal{E}_v^+}(t) = \bar{\rho}_{\mathcal{E}_v^+}, \forall t \geq \tau.$$

Proof. We will first show that for every non-transferring system there exists an origin node $v \in \mathcal{V}_O$ such that

$$\lambda_v > 0, \quad \lim_{t \to \infty} \rho_{\mathcal{E}_v^+}(t) = \bar{\rho}_{\mathcal{E}_v^+}. \tag{22}$$

Since every local routing policy $\mathcal{R}$ satisfies (4) and (5), under the dynamics in (6) and (7),

$$\sum_{e \in \mathcal{E}} \dot{\rho}_e = \sum_{v \in \mathcal{V}_O : \rho_{\mathcal{E}_v^+} \neq \bar{\rho}_{\mathcal{E}_v^+}} \lambda_v - \sum_{v \in \mathcal{V}_D} \mu_v. \tag{23}$$

For the sake of contradiction, suppose that $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is non-transferring, i.e.,

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t \sum_{v\in\mathcal{V}_D} \mu_v(\tau)d\tau < \sum_{v\in\mathcal{V}_O} \lambda_v, \tag{24}$$

and there is no $v \in \mathcal{V}_O$ satisfying (22). In that case, (23) and (24) together imply

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t \sum_{e\in\mathcal{E}} \dot{\rho}_e(\tau)d\tau > 0,$$

which leads to

$$\liminf_{t\to\infty} \sum_{e\in\mathcal{E}} \rho_e(t) = +\infty. \tag{25}$$

Note that (25) contradicts with every $\bar{\rho}_e$ being finite. Consequently, if $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is non-transferring, then there exists $v \in \mathcal{V}_O$ satisfying (22).

Since each flow function $f_e(\rho_e, \bar{u}_e)$ is a continuous in ρ_e and satisfies $f_e(\bar{\rho}_e, \bar{u}_e) = 0$ and $0 \leq f_e(\rho_e, u_e) \leq f_e(\rho_e, \bar{u}_e)$ for every $u_e \in [0, \bar{u}_e]$, (22) implies that the total flow on the outgoing links of v converges to zero, i.e.,

$$\lim_{t\to\infty} \sum_{e\in\mathcal{E}_v^+} f_e(t) = 0.$$

Hence, for any $\epsilon \in (0, \lambda_v)$ there exists a finite τ' such that

$$\sum_{e\in\mathcal{E}_v^+} f_e(t) < \lambda_v - \epsilon,\ \forall t > \tau'.$$

Accordingly, under (6), the total density on the outgoing links of v increases with a positive rate bounded away from zero for $t > \tau'$ until all of those links fail, i.e.,

$$\sum_{e\in\mathcal{E}_v^+} \dot{\rho}_e(t) = \lambda_v - \sum_{e\in\mathcal{E}_v^+} f_e(t) > \epsilon,\ \forall t > \tau' : \rho_{\mathcal{E}_v^+}(t) \neq \bar{\rho}_{\mathcal{E}_v^+}.$$

Consequently, since all the jam densities are finite,

$$\rho_{\mathcal{E}_v^+}(t) = \bar{\rho}_{\mathcal{E}_v^+}, \forall t \geq \tau,$$

for some τ satisfying

$$\tau \leq \tau' + \frac{1}{\epsilon}\sum_{e\in\mathcal{E}_v^+} \bar{\rho}_e - \rho_e(\tau').$$

□

Theorem 6.3. *For any locally routed flow $(\mathcal{G}, f, \mathcal{R}, \lambda)$, if $\mathcal{R}$ is congestion-aware at some $\rho^* \in [\mathbf{0}, \bar{\rho}]$ such that the corresponding maximum sustainable inflows in* (11) *satisfy*

$$\sum_{e\in\mathcal{E}_\mathcal{U}^+} \phi_e(\rho_e^*) - \sum_{v\in\mathcal{U}} \lambda_v \geq 0,\ \forall \mathcal{U} \subseteq \mathcal{V}\setminus\mathcal{V}_D, \tag{26}$$

then there exist speed limits $u(t) \in [\mathbf{0}, \bar{u}]$ such that $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is transferring for any initial condition $\rho(0) \in [\mathbf{0}, \rho^*]$.

Proof. Let $(\mathcal{G}, f, \mathcal{R}, \lambda)$ be any locally routed flow such that $\mathcal{R}$ is congestion-aware at some $\rho^* \in [\mathbf{0}, \bar{\rho}]$ satisfying (26). Let $\bar{\mathcal{F}}^*(\phi(\rho^*), \lambda)$ be the set of feasible capacity allocations that are upper bounded by $\phi(\rho^*)$ and ensure that the total outgoing capacity is greater than the external inflow for each origin node and greater than the total incoming capacity for each intermediate node, i.e.,

$$\begin{aligned}\bar{\mathcal{F}}^*(\phi(\rho^*), \lambda) = \{\bar{f}^* \in \mathbb{R}_+^{\mathcal{E}} \mid \bar{f}_e^* \leq \phi_e(\rho_e^*),\ \forall e \in \mathcal{E}\\ \lambda_v \leq \sum_{e\in\mathcal{E}_v^+} \bar{f}_e^*,\ \forall v \in \mathcal{V}_O\\ \sum_{e\in\mathcal{E}_v^-} \bar{f}_e^* \leq \sum_{e\in\mathcal{E}_v^+} \bar{f}_e^*,\ \forall v \in \mathcal{V}_I\}\end{aligned} \tag{27}$$

Based on (8), $\mathcal{F}(\phi(\rho^*), \lambda) \subseteq \bar{\mathcal{F}}^*(\phi(\rho^*), \lambda)$. Furthermore, due to (26), $\mathcal{F}(\phi(\rho^*), \lambda) \neq \emptyset$. Hence, $\bar{\mathcal{F}}^*(\phi(\rho^*), \lambda) \neq \emptyset$. Let the speed limits $u(t) \in [\mathbf{0}, \bar{u}]$ be assigned in accordance with (14) or (16) for any $\bar{f}^* \in \bar{\mathcal{F}}^*(\phi(\rho^*), \lambda)$. Such speed limits ensure that the flow on each link e is upper bounded by $\bar{f}_e^*$. Accordingly, (7) and (27) together imply

$$\mu_v(t) \leq \sum_{e\in\mathcal{E}_v^+} \bar{f}_e^*,\ \forall v \in \mathcal{V},\ \forall t \geq 0. \tag{28}$$

Let $\hat{\rho}$ denote the vector of densities such that each $\hat{\rho}_e$ satisfies (15). Due to Assumption 1, $\hat{\rho} \in [\rho^*, \bar{\rho}]$ and

$$\phi_e(\rho_e) \geq \phi_e(\hat{\rho}_e) = \bar{f}_e^*,\ \forall \rho_e \in [0, \hat{\rho}_e],\ \forall e \in \mathcal{E}. \tag{29}$$

As such, (10), (28) and (29) together imply that, under (6),

$$\rho_e = \hat{\rho}_e \Rightarrow \dot{\rho}_e \leq 0,\ \forall \rho \in [\mathbf{0}, \hat{\rho}],\ \forall e \in \mathcal{E}.$$

Hence, $[\mathbf{0}, \hat{\rho}]$ is positively invariant. Also, due to (27),

$$\sum_{e\in\mathcal{E}_v^+} \bar{f}_e^* > 0,\ \forall v \in \mathcal{V}_O : \lambda_v > 0,$$

which implies

$$\hat{\rho}_{\mathcal{E}_v^+} \neq \bar{\rho}_{\mathcal{E}_v^+},\ \forall v \in \mathcal{V}_O : \lambda_v > 0. \tag{30}$$

Since $[\mathbf{0}, \hat{\rho}]$ is positively invariant, Lemma 6.2 and (30) together imply that $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is transferring for any $\rho(0) \in [\mathbf{0}, \hat{\rho}]$. Since $\hat{\rho} \geq \rho^*$, we conclude that $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is transferring for any $\rho(0) \in [\mathbf{0}, \rho^*]$.

□

Theorem 6.3 implies that the systemic failures under any feasible external inflow can be averted through a proper selection of speed limits if the local routing policy is sufficiently congestion-aware and the initial densities on the links are sufficiently small. For instance, since $\phi(\rho^c) = \bar{f}$ and any feasible external inflow satisfies (9), for any $(\mathcal{G}, f, \mathcal{R}, \lambda)$ such that λ is feasible and $\mathcal{R}$ is congestion-aware at the congestion threshold ρ^c, there exist speed limits $u(t) \in [\mathbf{0}, \bar{u}]$ such that $(\mathcal{G}, f, \mathcal{R}, \lambda)$ is transferring for any $\rho(0) \in [\mathbf{0}, \rho^c]$. Furthermore, it follows from the proof of Theorem 6.3 that one way to ensure the throughput optimality of such a system is using the speed limits for capacity allocation such that the total outgoing capacity is at least equal to the total incoming capacity (external inflow) for all the intermediate nodes (origin nodes). Accordingly, a proper choice of speed limits ensures a transferring system by compensating for the lack of global information in routing decisions. We will conclude this section by providing a tractable method of generating such speed limits.

A. Optimal Capacity Allocation via Speed Limits

The main results of this paper pave the way for investigating the design of optimal speed limit policies that are guaranteed to prevent systemic failures under feasible external inflows. Based on the proof of Theorem 6.3, for any locally routed flow $(\mathcal{G}, f, \mathcal{R}, \lambda)$ with the initial condition $\rho(0)$, if the routing policy is congestion-aware at some $\rho^* \in [\rho(0), \bar{\rho}]$ satisfying the premise, then one way to ensure that the system is transferring is to use the speed limits for realizing any of the feasible capacity allocations in $\bar{\mathcal{F}}^*(\phi(\rho^*), \lambda)$. Regarding the routing behavior of users, this approach only requires some knowledge of the congestion awareness in the system, i.e., a suitable ρ^*. This is particularly useful since the routing policy, which emerges through the behavior of drivers, may not be exactly known to the network operator in most cases. Accordingly, an optimal capacity allocation problem can be formulated as

$$\bar{f}^* = \underset{\bar{f}' \in \bar{\mathcal{F}}^*(\phi(\rho^*), \lambda)}{\operatorname{argmin}} \; J(\bar{f}'), \tag{31}$$

where $J : \mathbb{R}_+^{\mathcal{E}} \mapsto \mathbb{R}$ is the cost function that captures the impact of the allocated capacities on some local (e.g., the cut-through traffic in residential areas) or global (e.g., the average delay) performance measures of interest. Furthermore, since $\bar{\mathcal{F}}^*(\phi(\rho^*), \lambda)$ is defined by the linear inequalities in (27), such a capacity allocation problem is convex for any convex cost function. In that case, the problem can be efficiently solved via a centralized algorithm or a distributed method that requires only some local communications among the controllers of the links (e.g., see [36], [37] and the references therein). Once $\bar{f}^*$ is obtained, each controller can impose $\bar{f}_e^*$ on the corresponding link e by assigning the speed limit as a function of the density of traffic on that link as per (16).

Example: One of the typical performance measures in flow networks is the average amount of time each vehicle spends in the network, i.e., the average delay. For any transferring network, in light of Little's law [38], the average delay is equal to the long-term average total density in the system divided by the total inflow rate, i.e.,

$$\text{Average Delay} = \frac{\limsup\limits_{t\to\infty} \dfrac{1}{t}\displaystyle\int_0^t \sum_{e\in\mathcal{E}} \rho_e(\tau) d\tau}{\displaystyle\sum_{v\in\mathcal{V}_O} \lambda_v}.$$

It is possible to obtain an upper bound for the average delay under the proposed capacity allocation method by using the maximum densities corresponding to the allocated capacities, i.e., $\hat{\rho}$ with the entries as given in (15). As shown in the proof of Theorem 6.3, using the speed limit policy in (16) for inducing some capacities $\bar{f}^* \in \bar{\mathcal{F}}^*(\phi(\rho^*), \lambda)$ renders $[\mathbf{0}, \hat{\rho}]$ positively invariant. Hence, this method comes with the following upper bound on the average delay:

$$\text{Average Delay} \leq \frac{\displaystyle\sum_{e\in\mathcal{E}} \hat{\rho}_e(\bar{f}_e^*)}{\displaystyle\sum_{v\in\mathcal{V}_O} \lambda_v}. \tag{32}$$

Accordingly, one approach to optimal capacity allocation is to minimize the upper bound in (32), i.e., to solve (31) for

$$J(\bar{f}') = \sum_{e\in\mathcal{E}} \hat{\rho}_e(\bar{f}'_e). \tag{33}$$

For example, many models in transportation networks (e.g., [21], [22]) consider flow functions $f_e(\rho_e, \bar{u}_e)$ that decrease linearly beyond the congestion threshold. In that case,

$$\hat{\rho}_e(\bar{f}'_e) = \rho_e^c + \alpha_e(\bar{f}_e - \bar{f}'_e),$$

for some $\alpha_e \in \mathbb{R}_{++}$. Accordingly, solving (31) with the cost function in (33) becomes a linear program as

$$\bar{f}^* = \underset{\bar{f}' \in \bar{\mathcal{F}}^*(\phi(\rho^*), \lambda)}{\operatorname{argmin}} \; -\sum_{e\in\mathcal{E}} \alpha_e \bar{f}'_e, \tag{34}$$

which implements the proposed capacity allocation method through some minimal reduction in the link capacities. Since allocating smaller capacities requires imposing lower speed limits, such a minimal reduction avoids unnecessary delays.

As an application of this approach, let us revisit the perturbed system in Section V-A. Given the link capacities $\bar{f}$ as shown in Fig. 4a, suppose that we would like to assign the speed limits to ensure a transferring system for any feasible inflow, i.e., $\lambda \in [0, 6\bar{f}_0]$, local routing policy $\mathcal{R}$ that is congestion-aware at the congestion threshold ρ^c, and uncongested initial condition $\rho(0) \in [\mathbf{0}, \rho^c]$. Note that, without such a flow control mechanism, congestion-awareness is not enough by itself to ensure throughput optimality. For example, the system could fail as shown in Section V-A, even before the demand reaches the network capacity (e.g., $\lambda = 5\bar{f}_0$), under some congestion-aware policies such as the proportional routing. Since it follows from (27) that $\bar{\mathcal{F}}^*(\bar{f}, \lambda) \subseteq \bar{\mathcal{F}}^*(\bar{f}, \lambda')$ for any $\lambda' \leq \lambda$, one way to achieve the desired performance guarantee in this example is to solve (34) with the feasible set $\bar{\mathcal{F}}^*(\bar{f}, 6\bar{f}_0)$. For this example, the resulting optimization problem has a unique solution for any $\alpha \in \mathbb{R}_{++}^{\mathcal{E}}$, i.e.,

$$\bar{f}^*_{(1,2)} = 2\bar{f}_0, \quad \bar{f}^*_e = \bar{f}_e, \forall e \in \mathcal{E} \setminus \{(1,2)\},$$

which can be realized by using the maximal speed limits in (16). Such speed limits would prevent the systemic failure illustrated in Fig. 2, and a transferring system would be guaranteed under any feasible external inflow and sufficiently congestion-aware routing policy. Furthermore, the resulting average delay would be upper bounded as given in (32). Such robust performance guarantees, which hold for a family of local routing policies, are particularly useful in practice since the actual routing policy is determined by the aggregate behavior of drivers and may not be exactly known to the network operator.

VII. Conclusions

In this paper, we investigated the use of variable speed limits for avoiding systemic failures in transportation networks, which were modeled as dynamical flow networks under local routing decisions. We showed that for a rich family of network topologies, operating the links constantly at their full capacities may lead to systemic failures due to the lack of global information in routing decisions. In particular, constantly utilizing the

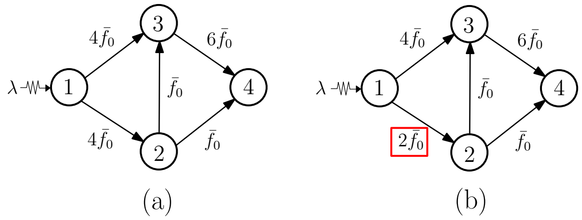


Fig. 4. For the network in (a), the solution to the minimal capacity reduction problem in (34) for $\rho^* = \rho^c$, $\lambda = 6\bar{f}_0$, and any $\alpha \in \mathbb{R}^{\mathcal{E}}_{++}$ is shown in (b).

full capacity of a link has such a negative effect on the overall network when the link sustains a large flow that overloads some critical links further in its downstream. In this regard, variable speed limits can be employed to operate some of the links below their full capacities when necessary, to compensate for the lack of global information. We showed that the systemic failures under feasible external inflows can always be averted through a proper selection of speed limits if the local routing decisions are sufficiently congestion-aware and the initial densities on the links are sufficiently low. Accordingly, throughput optimality can be ensured without requiring any global information and coordination in the routing decisions.

The technical results in this paper pave the way for designing speed limit policies which not only ensure a transferring system but also optimize some additional performance measures. In this regard, we presented a tractable approach and provided an upper bound on the resulting average amount of time each vehicle spends in the network, i.e., the average delay. The proposed approach requires only some knowledge regarding the congestion awareness in the system. This is particularly useful in practice since the actual routing policy, which emerges through the behavior of drivers, may not be exactly known to the network operator.

As a future direction, we plan to further investigate the design of optimal speed limit policies under limited knowledge of the routing behavior and develop strategies that also facilitate the recovery of congested links. We also plan to analyze how the availability of more information (e.g., multi-hop densities) in routing affects the overall performance under such network operation. Furthermore, we plan to extend our work to networks with multiple destination nodes, where each driver aims to arrive at a specific destination (multi-commodity flows).

Appendix

We will show that neither the fragility due to local routing decisions nor the proposed solution hinges on the absence of a continuous back-propagation of congestion effects as in (6). To this end, we will revisit the failure scenario in Section 5A under a modified dynamical model where the outflow of each link degrades gradually as its downstream becomes congested. For every link $e = (v, w) \in \mathcal{E}$, let

$$\dot{\rho}_e = \mathcal{R}^v_e(\rho_{\mathcal{E}^+_v}, \mu_v) - \min(g_e(\rho_{\mathcal{E}^+_w}), f_e(\rho_e, u_e)),$$

where $g_e(\rho_{\mathcal{E}^+_w}) \in [0, \bar{f}_e]$ denotes how much the outflow of the link is obstructed due to the density of traffic in its immediate downstream and μ_v denotes the total inflow to v, i.e.,

$$\mu_v = \begin{cases} \lambda_v, & \text{if } v \in \mathcal{V}_O, \\ \sum\limits_{\epsilon \in \mathcal{E}^-_v} \min(g_\epsilon(\rho_{\mathcal{E}^+_v}), f_\epsilon(\rho_\epsilon, u_\epsilon)), & \text{o.w.} \end{cases},$$

Note that (6) and (7) indicate a special case where

$$g_e(\rho_{\mathcal{E}^+_w}) = \bar{f}_e, \ \forall \rho_{\mathcal{E}^+_w} \neq \bar{\rho}_{\mathcal{E}^+_w}, \quad \text{and} \quad g_e(\bar{\rho}_{\mathcal{E}^+_w}) = 0.$$

As an example of a continuous back-propagation term, let

$$g_e(\rho_{\mathcal{E}^+_w}) = \bar{f}_e \frac{\sum\limits_{j \in \mathcal{E}^+_w} \phi_j(\rho_j)}{\sum\limits_{j \in \mathcal{E}^+_w} \bar{f}_j},$$

where $\phi_e(\rho_e)$ is as given in (11). In that case, the flow on a link can be disrupted once any of the links in its downstream enters the congested regime. However, we will show that the second example in Section 5A would still be non-transferring under the resulting dynamics. In this analysis, we will omit the arguments of several functions for brevity, e.g., $\phi_e(\rho_e)$ will be denoted as ϕ_e.

Since links $(2,4)$ and $(3,4)$ are connected to the destination, the modification in the model does not affect their outflow. Furthermore, since $\bar{f}_{(3,4)} > \bar{f}_{(1,3)} + \bar{f}_{(2,3)}$, link $(3,4)$ can never enter the congested regime. Hence, links $(1,3)$ and $(2,3)$ will not experience a limitation on their outflows either. The only link whose outflow will degrade due to downstream congestion is $(1,2)$. However, link $(1,2)$ can still allow for flows larger than links $(2,3)$ and $(2,4)$ can support since the link capacities in Fig. 2a imply that

$$g_{(1,2)} = \bar{f}_{(1,2)} \frac{\phi_{(2,3)} + \phi_{(2,4)}}{\bar{f}_{(2,3)} + \bar{f}_{(2,4)}} = 2(\phi_{(2,3)} + \phi_{(2,4)}). \tag{35}$$

Since all the links start uncongested, initially $\phi_e = \bar{f}_e$ for every link. Hence, the external inflow $\lambda = 5\bar{f}_0$ is initially split evenly as $2.5\bar{f}_0$ over the links $(1,2)$ and $(1,3)$. Accordingly, those links will start accumulating density and their outflows will keep increasing to $2.5\bar{f}_0$. When the outflow of $(1,2)$ exceeds $\bar{f}_{(2,3)} + \bar{f}_{(2,4)} = 2\bar{f}_0$, the links $(2,3)$ and $(2,4)$ will eventually become congested and $g_{(1,2)}$ will start decreasing. When those two links become sufficiently congested so that $\phi_{(2,3)} + \phi_{(2,4)} < 1.25\bar{f}_0$ (hence $g_{(1,2)} < 2.5\bar{f}_0$), link $(1,2)$ will be driven to the congested regime as well since its outflow will always be less than $2.5\bar{f}_0$ and it will keep receiving an inflow of $2.5\bar{f}_0$ under the proportional routing until it enters the congested regime. However, as long as $g_{(1,2)} \geq \bar{f}_0$, link $(1,3)$ will remain uncongested and $\rho_{(1,2)}$ cannot go beyond some $\rho'_{(1,2)} \in [\rho^c_{(1,2)}, \bar{\rho}_{(1,2)}]$, which is actually a monotonic continuous function of $g_{(1,2)}$ and approaches $\bar{\rho}_{(1,2)}$ as $g_{(1,2)}$ approaches zero, such that

$$f_{(1,2)}(\rho'_{(1,2)}, \bar{u}_{(1,2)}) = g_{(1,2)}. \tag{36}$$

This is simply because for every $\rho_{(1,3)} \in [0, \rho^c_{(1,3)}]$ the proportional routing satisfies

$$\mathcal{R}^1_{(1,2)}(\rho'_{(1,2)}, \rho_{(1,3)}, 5\bar{f}_0) = 5\bar{f}_0 \frac{g_{(1,2)}}{g_{(1,2)} + 4\bar{f}_0},$$

Hence, as long as $g_{(1,2)} \geq \bar{f}_0$ and $\rho_{(1,3)} \in [0, \rho^c_{(1,3)}]$,

$$\bar{f}_0 \leq \mathcal{R}^1_{(1,2)}(\rho'_{(1,2)}, \rho_{(1,3)}, 5\bar{f}_0) \leq g_{(1,2)}. \quad (37)$$

Since for every $\rho_{(1,2)} \leq \rho'_{(1,2)}$

$$\mathcal{R}^1_{(1,2)}(\rho_{(1,2)}, \rho_{(1,3)}, 5\bar{f}_0) \geq \mathcal{R}^1_{(1,2)}(\rho'_{(1,2)}, \rho_{(1,3)}, 5\bar{f}_0),$$

the inflow of $(1,3)$ will never exceed $4\bar{f}_0$ and $(1,3)$ will remain uncongested as long as $g_{(1,2)} \geq \bar{f}_0$ and (37) holds. Also, as long as $(1,3)$ remains uncongested, $\rho_{(1,2)}$ will not go beyond $\rho'_{(1,2)}$ since (36) and (37) together imply that the inflow of $(1,2)$ is upper bounded by its outflow when $\rho_{(1,2)} = \rho'_{(1,2)}$. As such, we have shown that $\rho_{(1,3)}$ will remain in $[0, \rho^c_{(1,3)}]$ as long as $g_{(1,2)} \geq \bar{f}_0$. Furthermore, when $g_{(1,2)} \in [\bar{f}_0, 2.5\bar{f}_0)$, $\rho_{(1,2)}$ will be driven to $[\rho^c_{(1,2)}, \rho'_{(1,2)}]$ and result in $\mu_2 = g_{(1,2)}$. Hence, as long as $g_{(1,2)} \geq \bar{f}_0$, $(2,3)$ and $(2,4)$ will keep getting more congested due to (35). Eventually, their congestion will make $g_{(1,2)}$ smaller than $\bar{f}_0$. When that happens, $\rho_{(1,2)}$ can go beyond the corresponding $\rho'_{(1,2)}$ and possibly reach a value such that μ_2 is not larger than $\phi_{(2,3)} + \phi_{(2,4)}$ anymore. However, for $g_{(1,2)} < \bar{f}_0$, it can be shown that links $(1,2)$ and $(1,3)$ will be driven to failure once $\rho_{(1,2)} \geq \rho'_{(1,2)}$. Note that $\phi_{(1,2)} + \phi_{(1,3)} < \bar{f}_0 + \bar{f}_{(1,3)} = 5\bar{f}_0$ for any such $\rho_{(1,2)}$. Accordingly, under the proportional routing, both of those links will keep receiving inflows larger than the maximum amounts they can sustain, i.e., $\phi_{(1,2)}$ and $\phi_{(1,3)}$ respectively, and they will eventually fail. Consequently, $(2,3)$ and $(2,4)$ will not stop approaching their jam densities unless the system reaches a state where the failures of $(1,2)$ and $(1,3)$ become inevitable. Hence, the origin is guaranteed to eventually become disconnected from the network and the system is non-transferring.

This example shows that a continuous degradation of outflow due to downstream congestion can facilitate a smoother spread of congestion towards the upstream. Furthermore, the accumulation of congestion on the downstream links can slow down, or even stop at some point, since the inflow of any intermediate node approaches zero as its outgoing links approach their jam densities. However, the long term consequences are equally severe as long as such a gradual degradation does not facilitate a sufficiently fast propagation of congestion towards the upstream. It is also worth mentioning that the proposed solution in Section 6A, i.e., using variable speed limits to keep the outflow of $(1,2)$ in $[0, 2\bar{f}_0]$, would save the system in this example under the modified dynamics as well since links $(2,3)$ and $(2,4)$ would never become congested in the first place.

References

[1] T. Roughgarden and É. Tardos, "How bad is selfish routing?," *Journal of the ACM (JACM)*, vol. 49, no. 2, pp. 236–259, 2002.

[2] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, "Robust distributed routing in dynamical networks–part I: Locally responsive policies and weak resilience," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 317–332, 2013.

[3] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, "Robust distributed routing in dynamical networks–part II: Strong resilience, equilibrium selection and cascaded failures," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 333–348, 2013.

[4] J. Adler, "Bootstrap percolation," *Physica A: Statistical Mechanics and its Applications*, vol. 171, no. 3, pp. 453–470, 1991.

[5] I. Dobson, B. A. Carreras, and D. E. Newman, "A loading-dependent model of probabilistic cascading failure," *Probability in the Engineering and Informational Sciences*, vol. 19, no. 01, pp. 15–32, 2005.

[6] L. Blume, D. Easley, J. Kleinberg, R. Kleinberg, and É. Tardos, "Which networks are least susceptible to cascading failures?," in *IEEE Annual Symposium on Foundations of Computer Science*, pp. 393–402, 2011.

[7] D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, "The network origins of aggregate fluctuations," *Econometrica*, vol. 80, no. 5, pp. 1977–2016, 2012.

[8] J. Ash and D. Newth, "Optimizing complex networks for resilience against cascading failure," *Physica A: Statistical Mechanics and its Applications*, vol. 380, pp. 673–683, 2007.

[9] A. Y. Yazıcıoğlu, M. Egerstedt, and J. S. Shamma, "Formation of robust multi-agent networks through self-organizing random regular graphs," *IEEE Transactions on Network Science and Engineering*, vol. 2, no. 4, pp. 139–151, 2015.

[10] A. E. Motter, "Cascade control and defense in complex networks," *Physical Review Letters*, vol. 93, no. 9, p. 098701, 2004.

[11] A. Y. Yazıcıoğlu, M. Roozbehani, and M. A. Dahleh, "Resilience of locally routed network flows: More capacity is not always better," in *IEEE Conference on Decision and Control*, pp. 111–116, 2016.

[12] D. Braess, A. Nagurney, and T. Wakolbinger, "On a paradox of traffic planning," *Transportation science*, vol. 39, no. 4, pp. 446–450, 2005.

[13] G. Como, E. Lovisari, and K. Savla, "Throughput optimality and overload behavior of dynamical flow networks under monotone distributed routing," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 1, pp. 57–67, 2015.

[14] R. Arnott, A. De Palma, and R. Lindsey, "Does providing information to drivers reduce traffic congestion?," *Transportation Research Part A: General*, vol. 25, no. 5, pp. 309–318, 1991.

[15] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Informational Braess' paradox: The effect of information on traffic congestion," *arXiv preprint arXiv:1601.02039*, 2016.

[16] M. Roozbehani, M. A. Dahleh, and S. K. Mitter, "Volatility of power grids under real-time pricing," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1926–1940, 2012.

[17] S. Coogan, E. A. Gol, M. Arcak, and C. Belta, "Traffic network control from temporal logic specifications," *IEEE Transactions on Control of Network Systems*, vol. 3, no. 2, pp. 162–172, 2016.

[18] A. Y. Yazıcıoğlu, M. Roozbehani, and M. A. Dahleh, "Resilient operation of transportation networks via variable speed limits," in *American Control Conference*, pp. 5623–5628, 2017.

[19] S. Coogan and M. Arcak, "A compartmental model for traffic networks and its dynamical behavior," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2698–2703, 2015.

[20] A. Messner and M. Papageorgiou, "Metanet: A macroscopic simulation program for motorway networks," *Traffic Engineering & Control*, vol. 31, no. 8-9, pp. 466–470, 1990.

[21] G. F. Newell, "A simplified theory of kinematic waves in highway traffic, part i: General theory," *Transportation Research Part B: Methodological*, vol. 27, no. 4, pp. 281–287, 1993.

[22] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[23] C. F. Daganzo, V. V. Gayah, and E. J. Gonzales, "Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 278–288, 2011.

[24] C. F. Daganzo, *Fundamentals of transportation and traffic operations*. Emerald Group Publishing Limited, 1997.

[25] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, 1992.

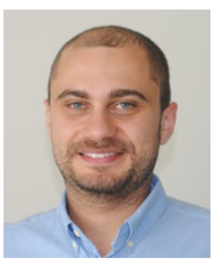
[26] F. P. Kelly, A. K. Maulloo, and D. K. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research society*, pp. 237–252, 1998.

[27] S. H. Low, F. Paganini, and J. C. Doyle, "Internet congestion control," *IEEE Control Systems*, vol. 22, no. 1, pp. 28–43, 2002.

[28] C.-C. Chien, Y. Zhang, and P. A. Ioannou, "Traffic density control for automated highway systems," *Automatica*, vol. 33, no. 7, pp. 1273–1285, 1997.

[29] A. Alessandri, A. Di Febbraro, A. Ferrara, and E. Punta, "Nonlinear optimization for freeway control using variable-speed signaling," *IEEE*

Transactions on Vehicular Technology, vol. 48, no. 6, pp. 2042–2052, 1999.
[30] A. Hegyi, B. De Schutter, and J. Hellendoorn, "Optimal coordination of variable speed limits to suppress shock waves," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 1, pp. 102–112, 2005.
[31] G. Gomes and R. Horowitz, "Optimal freeway ramp metering using the asymmetric cell transmission model," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 4, pp. 244–262, 2006.
[32] I. Papamichail and M. Papageorgiou, "Traffic-responsive linked ramp-metering control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 111–121, 2008.
[33] R. C. Carlson, I. Papamichail, M. Papageorgiou, and A. Messmer, "Optimal motorway traffic flow control involving variable speed limits and ramp metering," *Transportation Science*, vol. 44, no. 2, pp. 238–253, 2010.
[34] G. Nilsson, G. Como, and E. Lovisari, "On resilience of multicommodity dynamical flow networks," in *IEEE Conference on Decision and Control*, pp. 5125–5130, 2014.
[35] P. Elias, A. Feinstein, and C. Shannon, "A note on the maximum flow through a network," *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 117–119, 1956.
[36] D. P. Bertsekas, "Incremental gradient, subgradient, and proximal methods for convex optimization: A survey," *arXiv preprint arXiv:1507.01030*, 2015.
[37] A. Nedic, "Asynchronous broadcast-based convex optimization over a network," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1337–1351, 2011.
[38] J. D. Little, "A proof for the queuing formula: L= λ w," *Operations research*, vol. 9, no. 3, pp. 383–387, 1961.



A. Yasin Yazıcıoğlu received the B.S. and M.S. degrees in Mechatronics Engineering from Sabancı University, Turkey, in 2007 and 2009 respectively, and the Ph.D. degree in Electrical and Computer Engineering from the Georgia Institute of Technology in 2014. He is currently a postdoctoral associate at the Laboratory for Information and Decision Systems in the Massachusetts Institute of Technology. His recent research is primarily focused on distributed decision making and control, multi-agent systems, game theory, and networks, with applications to cyber-physical and societal systems. He is particularly interested in designing distributed algorithms for achieving properties such as controllability, efficiency, and robustness of networked dynamical systems in a provably-correct manner.

Mardavij Roozbehani received the BSc degree from Sharif University of Technology in 2000, the MSc degree from University of Virginia at Charlottesville, VA in 2003, and the PhD degree in Aeronautic and Astronautics from MIT in 2008. His PhD research focused on generating automated proofs for safety and performance guarantees for software systems, finite-state systems, and embedded systems, and bridged a gap between computer science and control theory. His earlier work also included finite-state approximation of dynamical systems, quantized control, and Delta-Sigma modulation. He is currently a Principal Research Scientist at the Laboratory for Information and Decision Systems (LIDS) and a principal investigator at the Institute for Data, Systems, and Society (IDSS) at MIT. His main research focus is on engineering and economic applications of control theory and optimization, analysis of robustness and efficiency in distributed and networked control systems, and dynamics and economics of power systems with an emphasis on robustness and risk. He has various publications in the IEEE Transactions on Automatic Control, IEEE Transactions on Power Systems, and the related conferences. Dr Roozbehani is a recipient of the 2007 AIAA graduate award for safety verification of real-time software systems, and a co-recipient of the 2015 SIAM/SICON best paper award.

Munther Dahleh received the Ph.D. degree from Rice University, Houston, TX, USA, in 1987. Since then, he has been with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, where he is now the William A. Coolidge Professor of electrical engineering and computer science, and the founding director of the MIT Institute for Data, Systems, and Society (IDSS). Previously, he held the position of Associate Department Head of EECS. He is interested in networked systems with applications to social and economic networks, transportation networks, and the power grid. Specifically, he focuses on the development of foundational theory necessary to understand, monitor, and control systemic risk in interconnected systems, drawing from various fields including game theory, optimal control, distributed optimization, information theory, and distributed learning. He is the co-author (with I. Diaz-Bobillo) of the book Control of Uncertain Systems: A Linear Programming Approach (Englewood Cliffs, NJ, USA: Prentice-Hall, 1995) and the co-author (with N. Elia) of the book Computational Methods for Controller Design (New York, NY, USA: Springer, 1998). Dr. Dahleh is the four times recipient of the George Axelby outstanding paper award for best paper in the IEEE Transactions on Automatic Control. He is also the recipient of the Donald P. Eckman award from the American Control Council in 1993 for the best control engineer under 35. .