52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Throughput Optimal Distributed Routing in Dynamical Flow Networks

Giacomo Como, Enrico Lovisari, and Ketan Savla

***Abstract*— A class of distributed routing policies is shown to be throughput optimal for single-commodity dynamical flow networks. The latter are modeled as systems of ODEs based on mass conservation laws on directed graphs with maximum flow capacities on links and constant external inflow at some origin nodes. Distributed routing regulates the flow splitting at each node, as a function of information on the densities of the local links around the nodes. Under monotonicity properties of routing, it is proven that, if no cut capacity constraint is violated by the external inflow, then a globally asymptotically stable equilibrium exists and the network achieves maximal throughput. This holds for finite or infinite buffer capacities for the densities. The overload behavior, if any cut capacity constraint is violated, is also characterized: there exists a cut on which the link densities grow linearly in time for infinite buffer capacities, while they simultaneously reach their respective buffer capacities, when these are finite. Numerical simulations illustrate and confirm the theoretical contributions.**

I. Introduction

Rapid advancements in technologies are facilitating real-time control of infrastructure networks, such as transportation, in order to achieve the efficient utilization of these networks. Static network flows, e.g., see [1], have traditionally dominated the modeling framework for infrastructure networks. However, in order to realize the true potential of the emerging technologies, one needs to develop control design within a dynamical framework. In this paper, we study single-commodity dynamical flow networks, modeled as systems of ordinary differential equations derived from mass conservation laws on weighted directed graphs, possibly with cycles, and having constant external inflow at each of possibly multiple origins. The weights on the links are their maximum flow capacities. The flow of particles is regulated from a link to links downstream to it by deterministic rules, or routing policies, which depend on the state of the network. The particles leave the network when they arrive at any of the possibly multiple destination nodes. Our first objective is to characterize routing policies that allow the network to achieve maximum throughput, i.e., the maximum possible external inflow at the origin nodes under which the link densities remain within the buffer capacities. Our secondary objective is the detailed characterization of the overload behavior of the network, when the external inflow at the origin nodes is greater than the maximum throughput.

We focus on routing policies that are distributed: the routing at each link depends only on the local information consisting of density of itself and the links downstream to it. We propose a novel class of distributed routing policies, called *monotone distributed routing policies*, that are characterized by general monotonicity assumptions on the sensitivity of their action with respect to local information. We then establish throughput optimality of this routing policy, and give a detailed characterization of the overload behavior of the network operating under monotone distributed routing policies. Our main result is in the form of a dichotomy. If the external inflow at the origin nodes does not violate any cut capacity constraints, then there exists a globally asymptotically stable equilibrium, and thus the network achieves maximal throughput. When the external inflow at the origin nodes violates some cut capacity constraint, then the resulting overload behavior of the network exhibits the following feature: if the buffer capacities are infinite, then there exists a constraint-violating cut, independent of the initial condition, such that the particle densities on the origin side of the cut grow at most linearly in time; if the buffer capacities are finite, then there exists a constraint-violating cut, in general dependent on the initial condition, such that the links constituting the cut hit their buffer capacities simultaneously. The last case implies that a link reaches its buffer capacity only at the very moment at which it is unavoidable. We emphasize again that such an efficient utilization of the network is induced by a *distributed* routing policy relying only on local information. Our results rely on the ability of the routing policy to implicitly propagate congestion effects upstream, allowing the flow to be routed through the less congested parts of the network in a timely fashion. For the proofs and a thorough discussion of the results we refer to [2].

While algorithms for distributed computation of maximum network flow have long been known (e.g., see [3]) the novelty of our contribution consists in proving throughput optimality for flow dynamics naturally arising in physical networks.

The distributed routing architecture of this paper and the ensuing result on throughput optimality are reminiscent of the backpressure routing algorithm for data networks, first proposed in [4], and the maxweight-α policies for switched networks, e.g., see [5], [6]. Indeed, modulo the discrete-time setting of [4], [5], [6], it can be shown [2] that backpressure and maxweight-α policies satisfy the monotonicity properties of the policies proposed in this paper.

Throughput optimality for the finite buffer capacity under

The first two authors were partially supported by the Swedish Research Council through the Junior Research Grant Information Dynamics in Large Scale Networks and the Linnaeus Center LCCC. G. Como and E. Lovisari are with the Department of Automatic Control, Lund University, SE-221 00 Lund, Sweden giacomo.como, enrico.lovisari@control.lth.se. K. Savla is with the Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA 90089-2531 ksavla@usc.edu

distributed routing has to be contrasted with existing work in [7], [8], where either throughput optimality is obtained under a centralized routing policy or there is a trade-off between throughput and buffer capacity under distributed routing policies. The dynamical formulation of this paper is also reminiscent of models of dynamic traffic flow over networks, e.g., see [9], [10], [11]. Here the key difference is that, unlike these existing works, routing policies in our framework depend on the local state of the network.

It is imperative to compare this paper with our previous work [12], [13]. There we proposed a class of *locally responsive policies* able to control the splitting of incoming flow at a node among outgoing links only, we established conditions for existence and stability of equilibrium for directed acyclic network topologies with infinite buffer capacities, and we studied the resilience properties of the network.

In the present paper we extend and modify the framework from [12], [13] to allow for finite buffer capacities and cyclic network topologies. Also, the routing policies are allowed to to control the outflow from each link. In the extended framework of the present paper we are able to provide conditions for the existence of a global asymptotic stability of equilibrium. Moreover, unlike [12], [13], we characterize in detail the overload behavior of the network.

The paper is organized as follows: in section II, we propose a general model for dynamical flow in networks and we formulate the problem. In section III, we state our main results, while section IV provides numerical simulations confirming the developed theory. Finally, section V states conclusions and possible directions for future research.

We conclude this section by introducing some notational conventions to be used throughout the paper. Let $\mathbb{R}$ be the set of real numbers and $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$ be the set of nonnegative real numbers. Let $\mathcal{A}$ and $\mathcal{B}$ be finite sets. Then $|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$, $\mathbb{R}^{\mathcal{A}}$ (respectively, $\mathbb{R}_+^{\mathcal{A}}$) the space of real-valued (nonnegative-real-valued) vectors whose components are indexed by elements of $\mathcal{A}$, and $\mathbb{R}^{\mathcal{A}\times\mathcal{B}}$ the space of matrices whose real entries are indexed by pairs in $\mathcal{A}\times\mathcal{B}$. If $\mathcal{B} \subseteq \mathcal{A}$ and $x \in \mathbb{R}^{\mathcal{A}}$, then $x_{\mathcal{B}} \in \mathbb{R}^{\mathcal{B}}$ stands for the projection of x on $\mathcal{B}$. The transpose of a matrix $M \in \mathbb{R}^{\mathcal{A}\times\mathcal{B}}$ is denoted by $M' \in \mathbb{R}^{\mathcal{B}\times\mathcal{A}}$, while $\mathbb{1}$ stands for an all-one vector of suitable dimension. The natural partial ordering of $\mathbb{R}^{\mathcal{A}}$ will be denoted by $x \preceq y$ for two vectors $x, y \in \mathbb{R}^{\mathcal{A}}$ such that $x_a \leq y_a$ for all $a \in \mathcal{A}$.

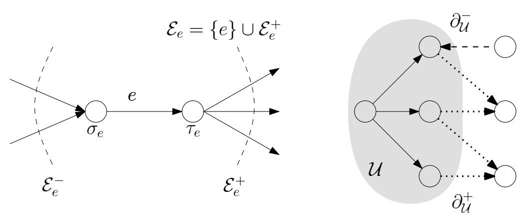

Fig. 1. Graphical depiction of some key notations. In the right, links comprising $\partial_{\mathcal{U}}^-$ and $\partial_{\mathcal{U}}^+$ are shown by dashed and dotted arrows, respectively; links comprising $\mathcal{E}_{\mathcal{U}}^+ \setminus \partial_{\mathcal{U}}^+$ are shown in solid arrows.

A weighted directed multi-graph is a triple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, C)$,

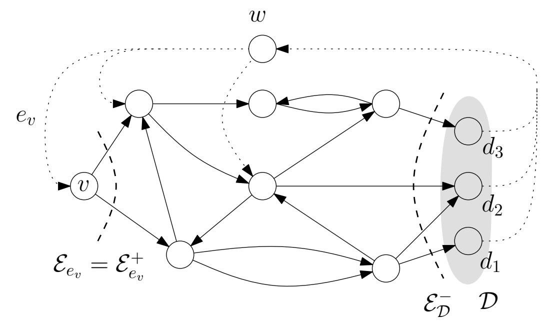

Fig. 2. An example of multi-destination network with cycles and parallel edges. The links added in the augmented graph $\mathcal{G}_\lambda$ are shown in dotted line.

where $\mathcal{V}$ and $\mathcal{E}$ stand for the node set and the link set, respectively, and are both finite. They are endowed with three vectors: $\sigma, \tau \in \mathcal{V}^{\mathcal{E}}$, and $C \in (0, +\infty]^{\mathcal{E}}$. For every $e \in \mathcal{E}$, σ_e and τ_e stand for the tail and head nodes respectively of link e and C_e for the positive (and possibly infinite) flow capacity of link e. We shall always assume that there are no self-loops, i.e., $\tau_e \neq \sigma_e, \forall e \in \mathcal{E}$. On the other hand, we allow for parallel links. For a node $v \in \mathcal{V}$, let $\mathcal{E}_v^+ := \{e : \sigma_e = v\}$ and $\mathcal{E}_v^- := \{e : \tau_e = v\}$. For a link $e \in \mathcal{E}$, let $\mathcal{E}_e^+ := \mathcal{E}_{\tau_e}^+$ be the set of links downstream to e and $\mathcal{E}_e^- := \mathcal{E}_{\sigma_e}^-$ be the set of links upstream to e. Put $\mathcal{E}_e := \{e\} \cup \mathcal{E}_e^+$. For a vector $x \in \mathbb{R}^{\mathcal{E}}$, we shall denote by $x^e := \{x_j : j \in \mathcal{E}_e\}$ its projection on $\mathcal{E}_e$. For a node subset $\mathcal{U} \subseteq \mathcal{V}$, define $\mathcal{E}_{\mathcal{U}}^+ := \cup_{u\in\mathcal{U}} \mathcal{E}_u^+$ and $\mathcal{E}_{\mathcal{U}}^- := \cup_{u\in\mathcal{U}} \mathcal{E}_u^-$. Let $\partial_{\mathcal{U}}^+ := \{e \in \mathcal{E} : \sigma_e \in \mathcal{U}, \tau_e \notin \mathcal{U}\}$ and $\partial_{\mathcal{U}}^- := \{e \in \mathcal{E} : \sigma_e \in \mathcal{V}\setminus\mathcal{U}, \tau_e \in \mathcal{U}\}$ be the set of links from $\mathcal{U}$ to $\mathcal{V} \setminus \mathcal{U}$ and from $\mathcal{V} \setminus \mathcal{U}$ to $\mathcal{U}$, respectively. See Figure 1 for an illustration of some of these notations.

II. Problem statement

A. *Static single-commodity network flows and the max-flow min-cut theorem*

We shall identify a network with a weighted directed multi-graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, C)$ and denote its set of destinations by $\mathcal{D} := \{v \in \mathcal{V} : \mathcal{E}_v^+ = \emptyset\}$ and the set of its feasible flows by $\mathcal{F}^* := \{x \in \prod_{e\in\mathcal{E}}[0, C_e] : \sum_{e\in\mathcal{E}_v^+} x_e - \sum_{e\in\mathcal{E}_v^-} x_e \geq 0,\ \forall v \in \mathcal{V} \setminus \mathcal{D}\}$. For $f^* \in \mathcal{F}^*$, the vector $\lambda(f^*) \in \mathbb{R}_+^{\mathcal{V}\setminus\mathcal{D}}$ with components $\lambda_v(f^*) := \sum_{e\in\mathcal{E}_v^+} f_e^* - \sum_{e\in\mathcal{E}_v^-} f_e^*$ will be referred to as the value of f^*.[1] For $\lambda \in \mathbb{R}_+^{\mathcal{V}\setminus\mathcal{D}}$, we introduce the *augmented* network $\mathcal{G}_\lambda = (\mathcal{V}_\lambda, \mathcal{E}_\lambda, C)$ (see Figure 2) with node and link sets $\mathcal{V}_\lambda = \mathcal{V} \cup \{w\}$ and $\mathcal{E}_\lambda := \mathcal{E} \cup \mathcal{O}_\lambda \cup \mathcal{E}_{\mathcal{D}}^+$, respectively, where $\mathcal{O}_\lambda := \{e_v := (w, v) : \lambda_v > 0\}$, $\mathcal{E}_{\mathcal{D}}^+ := \{e_d := (d, w) : d \in \mathcal{D}\}$, and $C_{e_v} = C_{e_d} = +\infty$ for all $v \in \mathcal{V} \setminus \mathcal{D}$ and $d \in \mathcal{D}$. The extra node w may be thought of as representing an external world, playing the double role of source of the flow for nodes with positive value of flow, and sink of the flow exiting from the destination nodes, respectively. We refer to links in $\mathcal{O}_\lambda$ as *origin links* and define $\mathcal{E}_{e_v} = \mathcal{E}_{e_v}^+ := \mathcal{E}_v^+$, for all $v \in \mathcal{V} \setminus \mathcal{D}$.

[1] The value of flow at a node is the same as the usual notion of external inflow at that node. In this paper, we use this terminology because of the necessity to interpret nodes with positive external inflow, i.e., origin nodes, as links.

Throughout this paper, we shall make the following assumptions on the network topology.

Assumption 1: The set of destinations $\mathcal{D}$ is nonempty, and the augmented network $\mathcal{G}_\lambda$ is strongly connected.

Assumption 1 is equivalent to the properties that, in $\mathcal{G}$, from every $v \in \mathcal{V} \setminus \mathcal{D}$ there exists at least one directed path to some destination node $d \in \mathcal{D}$, and there exists at least one directed path from some u with $\lambda_u > 0$ to every $v \in \mathcal{V}$.

A cut is a non-empty subset of non-destination nodes $\mathcal{U} \subseteq \mathcal{V} \setminus \mathcal{D}$. For a cut $\mathcal{U}$, we shall denote its capacity by $C_\mathcal{U} := \sum_{e \in \partial^+_\mathcal{U}} C_e$ and put $\lambda_\mathcal{U} := \sum_{v \in \mathcal{U}} \lambda_v$. The definition of $\mathcal{O}_\lambda$ implies that, under Assumption 1, there is no subset $\mathcal{A} \subseteq \mathcal{V}$ that is unreachable in $\mathcal{G}_\lambda$, i.e., it is not possible in $\mathcal{G}$ to have $\partial^-_\mathcal{A} = \emptyset$, and $\lambda_\mathcal{A} = 0$. Cut capacities determine potential bottlenecks for network flows. This is formalized in the celebrated max-flow min-cut theorem [14], [15], which states that, for $\mathcal{G} = (\mathcal{V}, \mathcal{E}, C)$ satisfying Assumption 1, it holds

$$\max_{f^*} \max_{\mathcal{U}} \{\lambda_\mathcal{U}(f^*) - C_\mathcal{U}\} = 0\,, \tag{1}$$

where the maximizations run over all feasible flows $f^* \in \mathcal{F}^*$, and cuts $\mathcal{U}$. For given $\mathcal{G} = (\mathcal{V}, \mathcal{E}, C)$ and λ, (1) gives a necessary and sufficient condition for the existence of a feasible flow with given value λ, namely, $\sum_{v \in \mathcal{U}} \lambda_v \leq C_\mathcal{U}$ for every cut $\mathcal{U}$.

B. Dynamical flow networks and monotone distributed routing

We now introduce dynamics over a network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, C)$. We associate, to each link $e \in \mathcal{E}$, a positive, and possibly infinite, buffer capacity $B_e \in (0, +\infty]$. Let $\mathcal{R} := \prod_{e \in \mathcal{E}} [0, B_e)$. For $e \in \mathcal{E} \cup \mathcal{O}_\lambda$, let $\overline{\rho}^e := \{B_j : j \in \mathcal{E}_e\}$, $\mathcal{R}_e := \prod_{j \in \mathcal{E}_e} [0, B_j)$, and $\mathcal{R}^\circ_e := \prod_{j \in \mathcal{E}_e} [0, B_j]$. Let $\mathcal{R}^\bullet_e$ be defined as $\mathcal{R}^\bullet_e = \mathcal{R}^\circ_e$ if $e \in \mathcal{E}^-_\mathcal{D}$, and $\mathcal{R}^\bullet_e = \mathcal{R}^\circ_e \setminus \{\overline{\rho}^e\}$ if $e \in (\mathcal{E} \cup \mathcal{O}_\lambda) \setminus \mathcal{E}^-_\mathcal{D}$. Finally, let the set of feasible flows on the outgoing links of e under capacity constraint be defined as $\mathcal{F}_e := [0, C_e]$ if $e \in \mathcal{E}^-_\mathcal{D}$ and $\mathcal{F}_e := \{x \in \mathbb{R}^{\mathcal{E}^+_e}_+ : \sum_{j \in \mathcal{E}^+_e} x_j \leq C_e\}$ if $e \in (\mathcal{E} \cup \mathcal{O}_\lambda) \setminus \mathcal{E}^-_\mathcal{D}$.

We shall consider a dynamical system with state vector $\rho(t) \in \mathcal{R}$ whose e-th component, $\rho_e(t) \in [0, B_e)$, represents the time-varying density on link $e \in \mathcal{E}$. Dynamics is driven by conservation of mass and by a distributed routing policy, which determines how the outflow from each link depends on the current density and how it gets split among its following links. We shall loosely use the phrase *a set of links getting congested* to refer to the fact that the densities on those links approach their respective buffer capacities.

Definition 1: Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, C)$ be a network satisfying Assumption 1. A distributed routing policy f with value $\lambda \in \mathbb{R}^{\mathcal{V} \setminus \mathcal{D}}_+$ and buffer capacities $\{B_e \in (0, +\infty] : e \in \mathcal{E}\}$ is a family of Lipschitz-continuous maps

$$f^e : \mathcal{R}^\bullet_e \to \mathcal{F}_e\,, \qquad e \in \mathcal{E} \cup \mathcal{O}_\lambda\,, \tag{2}$$

such that

$$f^e(\rho^e) = \begin{cases} \{f_{e \to j}(\rho^e)\}_{j \in \mathcal{E}^+_e} & \text{if} \quad e \notin \mathcal{E}^-_\mathcal{D} \\ f_{e \to e_d}(\rho_e) & \text{if} \quad e \in \mathcal{E}^-_d\,,\, d \in \mathcal{D} \end{cases}$$

$$f^{\text{out}}_e(\rho^e) := \sum_{j \in \mathcal{E}^+_e} f_{e \to j}(\rho^e)$$

satisfy

$$f^{\text{out}}_{e_v}(\rho^{e_v}) = \lambda_v\,, \qquad \rho^{e_v} \in \mathcal{R}^\bullet_{e_v}\,,\; v \in \mathcal{V} \setminus \mathcal{D}\,, \tag{3}$$

and, for all $e \in \mathcal{E}$ and $\rho^e \in \mathcal{R}^\bullet_e$,

$$\rho_e = 0 \qquad \Longrightarrow \qquad f^{\text{out}}_e(\rho^e) = 0\,, \tag{4}$$

$$\rho_e = B_e \qquad \Longrightarrow \qquad f^{\text{out}}_e(\rho^e) = C_e\,, \tag{5}$$

and, for all $e \in \mathcal{E} \cup \mathcal{O}_\lambda$, $k \in \mathcal{E}^+_e$, $\rho^e \in \mathcal{R}^\bullet_e$

$$\rho_k = B_k \qquad \Longrightarrow \qquad f_{e \to k}(\rho^e) = 0\,. \tag{6}$$

The functions $f_{e \to j}(\rho^e)$ specify both how the outflow f^{out}_e depends on the local density and how it gets split into the outgoing links of τ_e. Notice that the domain of f^e is $\mathcal{R}^\bullet_e$, thus if $e \notin \mathcal{E}^-_\mathcal{D}$ it is not defined at the point $\overline{\rho}^e = \{B_j : j \in \mathcal{E}_e\}$, where (6) and (5) cannot hold simultaneously. On the other hand, f^e is well defined when the density is strictly less than its buffer capacity at least on one link in $\mathcal{E}_e$. Also notice that, because of the structure imposed by (2), the functions $\{f_{e \to j}(\rho^e)\}$ depend on the local density only, and in particular for $e \in \mathcal{E} \setminus \mathcal{E}^-_\mathcal{D}$ the outflow $f^{\text{out}}_e(\rho^e)$ depends only the density on link e itself and the links downstream to it, and if $e \in \mathcal{E}^-_\mathcal{D}$ then $f^{\text{out}}_e(\rho^e)$ only depends on ρ_e. Similarly, the inflow $f^{\text{in}}_e(\rho) := f_{e_{\sigma_e} \to e}(\rho^{e_{\sigma_e}}) + \sum_{j \in \mathcal{E}^-_e} f_{j \to e}(\rho^j)$ of a link $e \in \mathcal{E}$ depends on the density on all the links in $\mathcal{E}$ incoming to or outgoing from σ_e (including link e itself). Also notice that the flow $f_{e_v \to j}$ from e_v to a link $j \in \mathcal{E}^+_v$ depends on the densities of the links in $\mathcal{E}^+_v$ only, and that by (3) it holds $f^{\text{out}}_{e_v}(\rho^{e_v}) \equiv \lambda_v$, i.e., the outflow from every link e_v is constantly equal to λ_v. Finally, (4) and (6) imply that $f^{\text{out}}_e(\rho^e) = 0$ if $\rho_e = 0$, i.e., there is no outflow from a link e which is empty, or if $\rho_j = B_j, \forall j \in \mathcal{E}^+_e$, i.e., if the densities on all the links outgoing from τ_e are at their buffer capacities.

For $\rho \in \mathcal{R}$, let $F(\rho) \in \mathbb{R}^{(\mathcal{E} \cup \mathcal{O}_\lambda) \times (\mathcal{E} \cup \mathcal{E}^+_\mathcal{D})}_+$ be defined as

$$F_{ej}(\rho) = \begin{cases} f_{e \to j}(\rho^e) & \text{if } j \in \mathcal{E}^+_e\,, \\ f_{e \to e_d}(\rho_e) & \text{if } e \in \mathcal{E}^-_d, d \in \mathcal{D}, j = e_d\,, \\ 0 & \text{otherwise.} \end{cases}$$

Imposing mass conservation $\dot{\rho}_e = f^{\text{in}}_e - f^{\text{out}}_e$ on every link $e \in \mathcal{E}$ leads one to consider the dynamical system

$$\dot{\rho} = (F(\rho)'\mathbb{1})_\mathcal{E} - (F(\rho)\mathbb{1})_\mathcal{E} = \Phi(\rho)\,. \tag{7}$$

We shall refer to it as the *dynamical flow network*. Observe that, thanks to the Lipschitzianity assumption on the routing policies, standard analytical results (Picard's Existence Theorem) imply, for every initial density $\rho(0) = \rho^\circ \in \mathcal{R}$, existence and uniqueness of a solution $\{\rho(t) : 0 < t < \kappa(\rho^\circ)\}$ of (7) up to $\kappa(\rho^\circ) := \sup\{t \geq 0 : \rho(t) \in \mathcal{R}, \rho(0) = \rho^\circ\}$, i.e., as long as $\rho(t)$ stays within $\mathcal{R}$. Moreover, (4) implies invariance of the nonnegative orthant, i.e., $\rho(t) \succeq 0$ for all $\rho^\circ \in \mathcal{R}$ and $t \leq \kappa(\rho^\circ)$. Hence, $\kappa(\rho^\circ)$ coincides with the instant the solution hits the buffer capacity on some link.

Remark 1: In this paper, we study the behavior of dynamical flow networks only for $t \in [0, \kappa(\rho^0))$. Some initial work on the complex behavior of dynamical flow networks, such as cascading failures, for $t > \kappa(\rho^0)$ is reported in our companion papers [16], [17].

Remark 2: In our previous work [12], [13], we considered routing policies under which the outflow from a link j is independent of the densities on the links downstream from link j. This, combined with the fact that all the links have infinite buffer capacities, implied that there is no backward propagation of congestion effects.

In this paper, we allow the routing policies to completely control (subject to capacity constraints) the flow transfer between links, i.e., the routing policies can also control the inflow arriving at nodes. This allows for backward propagation of congestion effects, and hence yields stronger results in comparison to [12], [13].

We shall be interested in a special class of distributed routing policies, as per the following.

Definition 2: A distributed routing policy f is *monotone* if, for all $e \in \mathcal{E} \cup \mathcal{O}_\lambda$, $\rho^e \in \mathcal{R}_e^\bullet$, the functions $\{f^e\}$ satisfy

$$\frac{\partial f_{e\to j}}{\partial \rho_k}(\rho^e) \geq 0, \qquad \forall\, j \in \mathcal{E}_e^+, k \in \mathcal{E}_e \setminus \{j\}\,, \tag{8}$$

$$\frac{\partial}{\partial \rho_k} f_e^{\text{out}}(\rho^e) \leq 0, \qquad \forall\, k \in \mathcal{E}_e^+\,, \tag{9}$$

for almost every $\rho^e \in \mathcal{R}_e$. A monotone distributed policy is *strongly monotone* if, for all $e \in \mathcal{E} \cup \mathcal{O}_\lambda$, and almost every $\rho^e \in \mathcal{R}_e$, the inequalities in (8) and (9) are strict.

Example 1: For every link $e \in \mathcal{E}$, let $\varphi_e : [0, B_e] \to [0, +\infty]$ be Lispchitz continuous, strictly increasing, and such that $\varphi_e(0) = 0$ and $\varphi_e(B_e) = +\infty$. Example of such a φ_e is $\varphi_e(\rho_e) = \beta_e \rho_e/(1 - \rho_e/B_e)$ if $B_e < +\infty$, and $\varphi_e(\rho_e) = \beta_e\rho_e$ if $B_e = +\infty$, for some $\beta_e > 0$, $e \in \mathcal{E}$. Define

$$f_{e\to j}(\rho^e) = \begin{cases} C_e\,(1-\gamma_e)\,\gamma_j/Z \text{ if } e \in \mathcal{E}\setminus\mathcal{E}_{\mathcal{D}}^-\,, \\ C_e\,(1-\gamma_e) \text{ if } e \in \mathcal{E}_d^-, d \in \mathcal{D}, j = e_d\,, \\ \lambda_v \gamma_j/Z \text{ if } e = e_v \in \mathcal{O}_\lambda\,, \end{cases}$$

with $\gamma_i := \exp(-\varphi_i(\rho_i))$ and $Z := \sum_{k\in\mathcal{E}_e}\gamma_k$. Then $\{f^e\}_{e\in\mathcal{E}\cup\mathcal{O}_\lambda}$ is a strongly monotone distributed routing policy.

Notice that, under monotone distributed routing policies, (7) defines a *cooperative* dynamical system in the sense of Hirsch [18], [19], i.e,

$$\frac{\partial \Phi_e}{\partial \rho_k}(\rho) \geq 0, \qquad \forall e, k \in \mathcal{E}, e \neq k\,. \tag{10}$$

Then, Kamke's theorem [19, Theorem 1.2], [20] implies that (7) is a monotone system [18], i.e.,

$$\rho(0) \preceq \tilde{\rho}(0) \quad \Rightarrow \quad \rho(t) \preceq \tilde{\rho}(t)\,, \quad \forall t \in [0, \kappa(\tilde{\rho}(0)))\,. \tag{11}$$

Also, observe that monotonicity implies that $\kappa(\rho^\circ) \leq \kappa(\mathbf{0})$ for all $\rho^\circ \in \mathcal{R}$.

III. Main results

In this section, we present the main contributions of the paper. The first result is Theorem 1, which states a dichotomy. If the inflow is less than the capacity of every cut, then there exists a globally asymptotically stable equilibrium density $\rho^* \in \mathcal{R}$. Otherwise, the network is divided in two parts by a cut $\mathcal{S}$, such that the densities on the links in $\mathcal{E}_{\mathcal{S}}^+$ approach their buffer capacities simultaneously.

Theorem 1: Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, C)$ be a network satisfying Assumption 1, and f be a monotone distributed routing policy with value λ. For $\rho^\circ \in \mathcal{R}$, let $\{\rho(t) : 0 \leq t < \kappa(\rho^\circ)\}$ be the solution of the dynamical flow network (7) with initial condition $\rho(0) = \rho^\circ$. Then,

(i) if $\max_{\mathcal{U}}(\lambda_{\mathcal{U}} - C_{\mathcal{U}}) < 0$, then $\kappa(\rho^\circ) = +\infty$ for every initial density $\rho^\circ \in \mathcal{R}$; moreover, if the distributed routing policy is strongly monotone, then there exists an equilibrium density $\rho^* \in \mathcal{R}$ such that $\lim_{t\to\infty}\rho(t) = \rho^*$ for every initial density vector $\rho^\circ \in \mathcal{R}$.

(ii) if $\max_{\mathcal{U}}\{\lambda_{\mathcal{U}} - C_{\mathcal{U}}\} > 0$, or if $\max_{\mathcal{U}}\{\lambda_{\mathcal{U}} - C_{\mathcal{U}}\} = 0$ and the routing policy is strongly monotone, then, for every initial density $\rho^\circ \in \mathcal{R}$, there exists a cut $\mathcal{S}$ such that

$$\lim_{t\to\kappa(\rho^\circ)} \rho_e(t) = B_e, \quad \forall e \in \mathcal{E}_{\mathcal{S}}^+\,. \tag{12}$$

Remark 3: Part i) of Theorem 1 strengthens results on stability of dynamical flow networks from our previous work [12], [13] as follows. First, in [12], [13], we considered acyclic network topologies and infinite buffer capacities on links, whereas Theorem 1 is valid for cyclic network topologies, and infinite as well as finite buffer capacities. And, second, the routing policies in [12], [13] do not give strong guarantees for existence or stability of equilibria, whereas (strongly) monotone routing policies guarantee existence and (global) asymptotic stability of equilibria when $\max_{\mathcal{U}}(\lambda_{\mathcal{U}} - C_{\mathcal{U}}) < 0$. At an equilibrium ρ^*, the throughput of the dynamical network is $\sum_{d\in\mathcal{D}} f_{e_d}^{\text{out}}(\rho^*) = \sum_{v\in\mathcal{V}\setminus\mathcal{D}}\lambda_v$. Therefore, in conjunction with the max-flow min-cut theorem, part i) of Theorem 1 implies that monotone distributed routing policies are throughput optimal. It is important to emphasize that this throughput optimality is achieved under a 'distributed' routing policy. The stronger results in this paper are possible due to backward propagation of congestion effects facilitated by a routing policy architecture under which the outflow from a link j depends on the density of links downstream from j.

A. Overload behavior with finite buffer capacities

The following proposition gives a more detailed characterization of what happens when the capacity constraints are violated in the case of finite buffer capacities.

Proposition 1: Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, C)$ be a network satisfying Assumption 1, and f be a monotone distributed routing policy with value λ and finite buffer capacities $B_e \in (0, +\infty)$, $e \in \mathcal{E}$. Assume that $\max_{\mathcal{U}}(\lambda_{\mathcal{U}} - C_{\mathcal{U}}) > 0$. Then, for every

$\rho^\circ \in \mathcal{R}$,

$$\kappa(\rho^\circ) \leq \min_{\mathcal{U}:\, \lambda_{\mathcal{U}} > C_{\mathcal{U}}} \frac{\sum_{e \in \mathcal{E}^+_{\mathcal{U}}} (B_e - \rho^\circ_e)}{\lambda_{\mathcal{U}} - C_{\mathcal{U}}}\,, \tag{13}$$

and there exists a cut $\mathcal{S}$, possibly depending on ρ°, such that $\lambda_{\mathcal{S}} > C_{\mathcal{S}}$ and

$$\begin{aligned} &\rho_e(t) < B_e\,,\, \forall e \in \mathcal{E}\,,\; 0 \leq t < \kappa(\rho^\circ)\,, \\ &\lim_{t \to \kappa(\rho^\circ)} \rho_e(t) = B_e, \quad \forall e \in \mathcal{E}^+_{\mathcal{S}}\,, \end{aligned} \tag{14}$$

where $\{\rho(t) : 0 \leq t < \kappa(\rho^\circ)\}$ is the solution of the dynamical flow network (7) with initial condition $\rho(0) = \rho^\circ$.

Proposition 1 states that, if the buffer capacities are finite and some cut constraints are violated, then, for every initial density ρ°, all the links in $\mathcal{E}^+_{\mathcal{S}}$, where $\mathcal{S}$ is a cut such that $\lambda_{\mathcal{S}} > C_{\mathcal{S}}$, will reach their buffer capacities simultaneously at time $\kappa(\rho^\circ)$. It is important to stress that, when there are multiple cuts violating the capacity constraint, then the cut $\mathcal{S}$ in the proposition may depend on the initial condition ρ°. Observe that dependence on the initial density ρ° is also evident in (13). While it may be tempting to identify the cut $\mathcal{U}$ minimizing the right hand side of (13) with the cut $\mathcal{S}$ of (14), it is worth stressing that (13) is merely an upper bound on $\kappa(\rho^\circ)$. In fact, in contrast to the right-hand side of (13), the cut $\mathcal{S}$ of (14) may depend on finer details of the routing policy, rather than just its value and buffer capacities.

Remark 4: It is interesting to compare Proposition 1 with the framework of our previous work [16], where we consider links with finite buffer capacities, but the routing policies are such that the outflow from a link j is independent of densities on links downstream from j. As a consequence, in the framework of [16], even if $C_{\mathcal{U}} > \lambda_{\mathcal{U}}$ for every cut $\mathcal{U}$, there might be a link e on which the density hits the buffer capacity which, in turn, could trigger a backward cascade. Part i) of Theorem 1 implies that this cannot happen under the routing policies presented in this paper if $C_{\mathcal{U}} > \lambda_{\mathcal{U}}$ for every cut $\mathcal{U}$. Moreover, as part ii) of Theorem 1 and Proposition 1 imply, the whole cut $\mathcal{S}$ fills simultaneously when $\max_{\mathcal{U}} \{\lambda_{\mathcal{U}} - C_{\mathcal{U}}\} > 0$, and hence the network collapse is abrupt, and does not involve any *cascading* phenomena.

B. Overload behavior with infinite buffer capacities

Similarly to Proposition 1, the following result characterizes overload in case of infinite buffer capacities.

Proposition 2: Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, C)$ be a network satisfying Assumption 1, and f be a strongly monotone distributed routing policy with value λ and buffer capacities $B_e = +\infty$, for $e \in \mathcal{E}$. Assume that $\max_{\mathcal{U}} (\lambda_{\mathcal{U}} - C_{\mathcal{U}}) \geq 0$. Let $\mathcal{U}^* := \bigcup_{\mathcal{U} \in \mathcal{M}} \mathcal{U}$, where $\mathcal{M} := \mathrm{argmax}_{\mathcal{U}} (\lambda_{\mathcal{U}} - C_{\mathcal{U}})$. Then, for every $\rho^\circ \in \mathcal{R}$, the solution $\rho(t)$ of the dynamical flow network (7) with initial condition $\rho(0) = \rho^\circ \in \mathcal{R}$ is such that $\kappa(\rho^\circ) = +\infty$ and

$$\begin{aligned} &\lim_{t \to +\infty} \rho_e(t) = +\infty\,, \qquad \forall e \in \mathcal{E}^+_{\mathcal{U}^*}\,, \\ &\lim_{t \to +\infty} \frac{1}{t} \sum_{e \in \mathcal{E}^+_{\mathcal{U}^*}} \rho_e(t) = \lambda_{\mathcal{U}^*} - C_{\mathcal{U}^*}\,. \end{aligned} \tag{15}$$

Moreover, there exist $\rho^*_e \in [0, +\infty)$, $e \in \mathcal{E} \setminus (\mathcal{E}^+_{\mathcal{U}^*} \cup \partial^-_{\mathcal{U}^*})$, such that

$$\lim_{t \to +\infty} \rho_e(t) = \rho^*_e\,, \qquad \forall e \in \mathcal{E} \setminus (\mathcal{E}^+_{\mathcal{U}^*} \cup \partial^-_{\mathcal{U}^*})\,, \tag{16}$$

for every initial density $\rho^\circ \in \mathcal{R}$.

Proposition 2 implies that, when the buffer capacities on all the links are infinite, then there exists a cut $\mathcal{U}^*$, independent of initial condition ρ° such that, asymptotically, all the links in $\mathcal{E}^+_{\mathcal{U}^*}$ get congested. This is to be contrasted with the finite buffer capacity case, which has a similar result, however, the cut there depends on the initial condition ρ°. Proposition 2 also implies that the total density in $\mathcal{E}^+_{\mathcal{U}^*}$ grows linearly in time, and that the densities on the links which do not get congested approach a unique limit point.

A comparison is due with [6], which studies an acyclic queuing network with set of queues $\mathcal{Q}$ employing max-weight algorithm. It is shown that if $q(t) \in \mathbb{R}^{\mathcal{Q}}_+$ is the vector of queue lengths, then $q(t)/t \to \hat{q}$ where $\hat{q} \in \mathbb{R}^{\mathcal{Q}}_+$ is the solution to an optimization problem related to the parameters of the max-weight algorithm.

IV. Numerical Simulations

In this section we provide numerical simulations confirming the predicted overload behavior of the network.

Consider the directed cyclic network with one destination node d depicted in Figure 3. Let $C_e = 1, \forall e \in \mathcal{E} \setminus \{3, 4\}$ and $C_3 = C_4 = 0.5$. The min cut capacity is attained at $\mathcal{S}$ such that $\partial^+_{\mathcal{S}} = \{3, 4\}$ and $\partial^-_{\mathcal{S}} = \{5\}$, and its value is $\mathcal{C}_{\mathcal{N}} := C_3 + C_4 = 1$. Following Example 1, the distributed routing policy used for the simulations is given by:

$$f_{e \to j}(\rho^e) = C_e \left(1 - e^{-3\varphi_e(\rho_e)}\right) \frac{e^{-5\varphi_j(\rho_j)}}{\sum_{k \in \{e\} \cup \mathcal{E}^+_e} e^{-5\varphi_k(\rho_k)}}\,,$$

where $\varphi_e(\rho_e) = \frac{\rho_e}{1 - \rho_e / B_e}, \forall e \in \mathcal{E}$. The value of the routing policy is given by $\lambda_o = 2.5 > \mathcal{C}_{\mathcal{N}}$, and $\lambda_v = 0$ for all $v \neq o$. Notice that $\max_{\mathcal{U}} \{\lambda_{\mathcal{U}} - C_{\mathcal{U}}\} = \lambda_0 - \mathcal{C}_{\mathcal{N}} > 0$. We run two sets of simulations: (i) $B_e = +\infty, \forall e \in \mathcal{E}$, and (ii) $B_e = 2, \forall e \in \mathcal{E}$. In case (ii), we consider two sets of initial conditions: $\rho^1_e(0) = 0, \forall e \in \mathcal{E}$, and $\rho^2_1(0) = \rho^2_2(0) = 1.9, \rho^2_e(0) = 0, e = 3, \ldots, 7$.

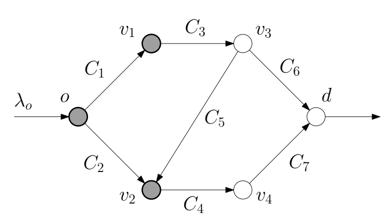

Fig. 3. Topology of the network used in the simulations. Grey marked nodes form the cut attaining the min-cut capacity $C_3 + C_4$.

The results of the simulations are shown in Figure 4 and Figure 5. Since $\max_{\mathcal{U}} \{\lambda_{\mathcal{U}} - C_{\mathcal{U}}\} > 0$, in case (ii), Proposition 2 implies that $\frac{1}{t} \sum_{e \in \mathcal{E}^+_{\mathcal{S}}} \rho_e(t) \to \lambda_o - \mathcal{C}_{\mathcal{N}} = 1.5$, as shown in Figure 4. In case (ii), Proposition 1 implies that there exists a cut such that all the links in the cut hit their buffer capacities simultaneously, and that the time at which this happens as well as this cut in general depend on the initial condition ρ°. This dependence on ρ° is illustrated in Figure 5. The two cuts whose links hit the buffer capacities are given by: $\mathcal{S}(\rho^1(0)) = \{o, v_1, v_2\}$ is such that $\mathcal{E}^+_{\mathcal{S}(\rho^1(0))} = \{1, 2, 3, 4\}$, while $\mathcal{S}(\rho^2(0)) = \{o\}$ is such that $\mathcal{E}^+_{\mathcal{S}(\rho^2(0))} = \{1, 2\}$. The times at which the links in the cuts attain the buffer capacities simultaneously are $\kappa(\rho^1(0)) = 4.7849$ and $\kappa(\rho^2(0)) = 0.347$, respectively. As expected, since $\rho^1(0) \prec \rho^2(0)$, we have that $\kappa(\rho^2(0)) < \kappa(\rho^1(0))$.

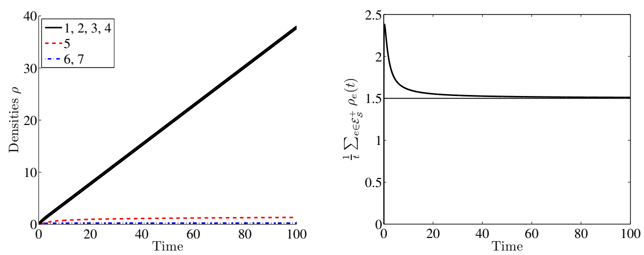

Fig. 4. Evolution of densities when links have infinite buffer capacities. Left: link-wise densities. Right: $\frac{1}{t}\sum_{e\in\mathcal{E}^+_{\mathcal{S}}} \rho_e(t)$ approaches $\lambda_o - \mathcal{C}_{\mathcal{N}} = 1.5$ as $t \to +\infty$.

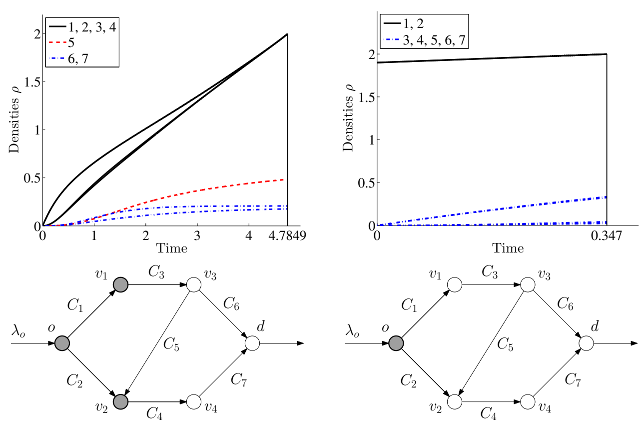

Fig. 5. Evolution of densities (top) and corresponding cut (bottom) when links have infinite buffer capacities, for initial conditions: (left) $\rho^1(0) = 0, \forall e \in \mathcal{E}$, and (right) $\rho^2_1(0) = \rho^2_2(0) = 1.9$, $\rho^2_e(0) = 0, e = 3, 4, 5, 6, 7$.

V. Conclusion

This paper studies dynamical flow networks with distributed monotone routing policies. It is shown that whenever the capacity constraints in the network are satisfied, there exists a unique equilibrium flow which is throughput optimal. We also characterize the overload behavior of the network when the external inflow at the origin nodes violates capacity constraint of some cut in the network.

There are several directions of research that we plan to pursue in the future. We plan to derive appropriate conditions under which monotone routing policies can optimize secondary objectives, such as steady-state delay, without comprising throughput optimality. We also plan to formally interpret monotone routing policies as combinations of physical properties and control policies in various application domains, and utilize the characteristic properties of monotone routing policies for synthesis of appropriate control policies. Finally, we plan to extend our formalism to the multi-commodity case.

Acknowledgements

The authors would like to thank Prof.s Bo Bernhardsson, Munther A. Dahleh, Emilio Frazzoli, Sanjoy Mitter, and Anders Rantzer for several valuable comments on this work.

References

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

[2] G. Como, E. Lovisari, and K. Savla, "Throughput optimality and overload behavior of dynamical flow networks under monotone distributed routing," 2013, available online at `http://arxiv.org/abs/1308.1993`.

[3] A. V. Goldberg and R. E. Tarjan, "A new approach to the maximum-flow problem," *Journal of the ACM (JACM)*, vol. 35, no. 4, pp. 921–940, 1988.

[4] 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, December 1992.

[5] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, October 2000.

[6] D. Shah and D. Wischik, "Fluid models of congestion collapse in overloaded switched networks," *Queuing Systems*, vol. 69, 2011.

[7] P. Giaccone, E. Leonardi, and D. Shah, "Throughput region of finite-buffered networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 2, pp. 251 – 263, 2007.

[8] L. B. Le, E. Modiano, and N. B. Shroff, "Optimal control of wireless networks with finite buffers," *IEEE/ACM Transactions on Networking*, vol. 20, no. 4, pp. 1316–1329, 2012.

[9] C. F. Daganzo, "The cell transmission model, part II: network traffic," *Transportation Research B*, vol. 29B, no. 2, pp. 79–93, 1995.

[10] M. Gugat, M. Herty, A. Klar, and G. Leugering, "Optimal control for traffic flow networks," *Journal of optimization theory and applications*, vol. 126, no. 3, pp. 589–616, 2005.

[11] M. Garavello and B. Piccoli, *Traffic Flow on Networks*. American Institute of Mathematical Sciences, 2006.

[12] 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, 2013.

[13] ——, "Robust distributed routing in dynamical networks - part ii: Strong resilience, equilibrium selection and cascaded failures," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, 2013.

[14] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, pp. 399–404, 1956.

[15] 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.

[16] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, "Robust distributed routing in dynamical networks with cascading failures," in *Proceedings of 51st Conference on Decision and Control, CDC'12*, December 2012.

[17] K. Savla, G. Como, M. A. Dahleh, and E. Frazzoli, "Distributed resilient control of flow networks under deterministic cascade dynamics," in *Proceedings of 52st Conference on Decision and Control, CDC'13*, December 2013, to appear.

[18] M. Hirsch and H. Smith, "Monotone dynamical systems," ser. Handbook of Differential Equations: Ordinary Differential Equations, Chapter 4, P. D. A. Canada and A. Fonda, Eds. North-Holland, 2006, vol. 2, pp. 239 – 357.

[19] ——, "Competitive and cooperative systems: A mini-review," *Positive Systems. Lecture Notes in Control and Information Sciences*, vol. 294, 2003.

[20] E. Kamke, "Zur theorie der systeme gewoknlicher differentialgliechungen, ii," *Acta Mathematica*, vol. 58, pp. 57–85, 1932.