

Stochastic Networks with Multipath Flow Control: Impact of Resource Pools on Flow-level Performance and Network Congestion

Vinay Joseph and Gustavo de Veciana
The University of Texas at Austin
Austin, U.S.A.

vinayjoseph@mail.utexas.edu, gustavo@ece.utexas.edu

ABSTRACT

Multipath flow control has been proposed as a key way to improve the Internet's performance, reliability, and flexibility in supporting changing loads. Yet, at this point, there are very few tools to quantify the performance benefits; particularly in the context of a stochastic network supporting best effort flows, e.g., file transfers and web browsing sessions, where the metric of interest is transfer delay. This paper's focus is on developing analysis tools to evaluate flow-level performance and to support network design when multipath bandwidth allocation is based on proportional fairness. To overcome the analytical intractability of such systems we study closely related multipath approximations based on insensitive allocations such as balanced fairness. We obtain flow-level performance bounds on the mean per bit delay, exhibiting the role of resource pooling in the network, and use these to explore scenarios where increased path diversity need not result in high gains. While insightful these results are difficult to use to drive network design and capacity allocation. To that end, we study the large deviations for congestion events, i.e., accumulation of flows, in networks supporting multipath flow control. We show that such asymptotics are determined by certain critical resource pools, and study the sensitivity of congestion asymptotics to the pool's capacity and traffic loads. This suggests a disciplined approach to a capacity allocation problem in multipath networks based on a linear optimization problem.

Categories and Subject Descriptors

C.2.1 [Computer Systems Organization]: Computer-Communication Networks—*Network Architecture and Design*

General Terms

Design, Performance, Theory

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGMETRICS'11, June 7–11, 2011, San Jose, California, USA.
Copyright 2011 ACM 978-1-4503-0262-3/11/06 ...\$10.00.

1. INTRODUCTION

There has recently been substantial interest in redesigning the Internet's data transport mechanisms to support multipath flow control so as to exploit path diversity; see e.g., [9, 7, 13, 12, 21]. Key benefits of doing so might include: improved reliability through path/provider diversity, as well as improved performance and flexibility in supporting highly variable network loads through *resource pooling*. Resource pooling refers to making, possibly distributed, resources appear to users as a single possibly shared one, typically resulting in a higher effective capacity and better statistical multiplexing. Similarly, in the context of wireless networks one can expect that concurrently exploiting multiple interfaces, e.g., cellular and Wi-Fi, would lead to performance benefits. Additional intriguing scenarios are being considered where a high bandwidth residential wireless mesh is used to enable substantially higher uplink wireline capacity to the Internet by sharing multiple limited capacity upstream residential broadband wireline links. Fig. 1 shows two scenarios where one might expect to reap benefits from supporting sessions with multipath flow control.

While the potential for exploiting path diversity through multipath flow control is intuitively clear, there exist very few works in the literature that shed light on flow-level performance and network design. Herein we refer to flow-level performance, as measures of transfer delays or throughput achievable in a stochastic network supporting best effort flows, i.e., file transfers and web browsing sessions. If this paradigm is to be adopted, there are many fundamental questions that need to be answered, for example:

- What flow-level throughput gains can we expect over traditional networks? How sensitive are they to system loads, and peak rate access network constraints, e.g., due to DSL or TCP?
- Are there abstractions that can be used to understand the interaction of multipath sessions with others (e.g., resource pooling or cutset bounds) and how do these affect performance, routing and network dimensioning?
- What measures of robustness are gained when traffic with multipath flow control is supported?
- What fraction of the traffic needs to have the extra flexibility of multipath flow control to achieve the practical benefits?
- In the residential wireless mesh setting described above, given cost of relaying traffic across neighboring home networks, how far should traffic be relayed to enable effective sharing of limited uplink capacity?

Our goal in this paper is to make some progress towards developing the tools needed to partially answer these questions. For simplicity, we will focus on wireline networks though our models can also be interpreted in terms of shared wireless resources.

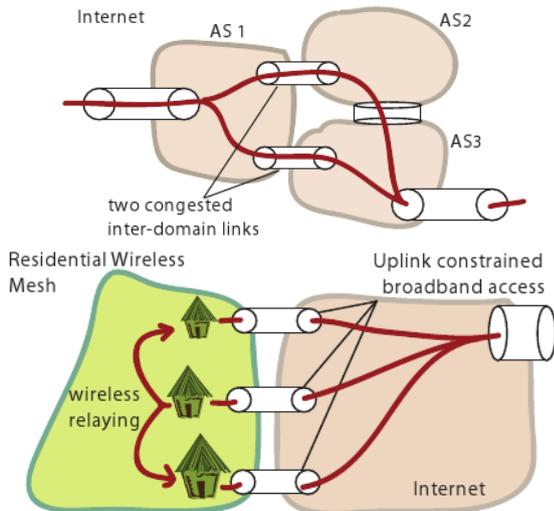


Figure 1: Top figure: Multipath routing could provide substantial benefits when there is limited shared inter-domain bandwidth among providers. Bottom figure: A residential wireless mesh network enables home users exploit multiple paths to overcome limited uplink links to the Internet.

Related Work.

There has been extensive recent work towards understanding how network protocols, e.g., TCP, allocate, or should allocate, bandwidth among competing flows. Much of this work has focused on associating a utility function with each user, and studying mechanisms to maximize the sum of the users' utilities. For a *fixed* set of users, the resulting allocation can be viewed as optimal in terms of optimizing network utility. The utility functions may be chosen to reflect users' valuation for allocated resources or can be used as a device to implement notions of fairness among competing flows, e.g., proportionally fair, max-min fair, α -fair, see e.g., [11, 1, 19]. Most notable among these are the allocations based on proportional fairness ([11]) due to its ease of implementation, and its performance, robustness and (approximate) insensitivity properties ([18], [10]). Note that an allocation is said to be insensitive if its steady state distribution is independent of all traffic characteristics except the average traffic loads.

Unfortunately, for networks supporting best effort flows, the relevant user perceived performance metrics will be averages taken over their sojourn in the network, such as file transfer delay or mean throughput, and the relationship between these flow averages and the utilities/fairness underlying the network design is not well understood. Early works in this area focused on the stability of such networks. See [22] for a survey of these results.

These early papers, however, do not provide concrete tools for assessing user perceived *performance* for network engi-

neering. More recent work in [5, 3] provides substantial insights into why, attempting to do so directly is difficult, and attribute this difficulty mainly to the dependence of the steady state distribution of utility based allocations on detailed traffic characteristics, e.g., flow size distribution. In [5], the same researchers propose *balanced fair* allocation. They introduce it as the most efficient insensitive allocation, in the sense that it is the unique insensitive allocation for which a link is saturated or a peak rate constraint is met in every state with at least one flow (see Section 2 for a detailed discussion of this allocation). Due to its insensitivity, flow level performance can either be explicitly computed, approximated or bounded, see e.g., [6, 4, 2].

The tractability of balanced fair allocation has been exploited to study the flow-level performance of intractable (sensitive) allocations like proportional fair allocation ([2]). Balanced fair allocation has been shown to be a good approximation of the proportional fair allocation especially at high loads ([3]). For a certain class of networks, balanced fair allocation coincides with proportional fair allocation ([5]). Further, [18] introduces modified proportional fair allocation that coincides with proportional fair allocation in an asymptotic sense, and shows that the modified proportional fair allocation and balanced fair allocation admit the same large deviation behavior.

Utility based allocations have been extended to accommodate multipath flow control to exploit the benefits of the increased path diversity. However, most of the progress made in this area concerns the stability and distributed implementation for these allocations ([9, 7, 20]). Only the following works study flow-level performance in networks using multipath flow control: [16], [15], [10] and [8]. In [16] and [15], the main performance metric considered is the blocking probability for flows in a system with admission control that blocks a flow if it cannot be allocated a certain minimum bandwidth. However, a weakness of the approach in [16] and [15] is that they mostly rely on complex optimization formulations and simulations, and fail to give insights into the elements in a network that are critical to the performance.

The performance metric considered in [10] is the mean delay experienced by flows in a network using multipath proportional fair allocation. Inspired by the relationship of this allocation to a related network of processor sharing queues, they propose an approximation for the mean delay experienced by a (possibly multipath) flow of class s

$$\sum_{j \in \mathcal{J}} \left(\frac{a_{js}}{\mu_s} \right) \frac{c_j}{c_j - \rho_j}$$

where \mathcal{J} is a set of virtual resources, a_{js} is the capacity required on virtual resource j per unit of bandwidth allocated to class s , $(\mu_s)^{-1}$ is the mean flow size, and c_j and ρ_j are the capacity of and load on virtual resource $j \in \mathcal{J}$ respectively. The above approximation suggests that the delay experienced by a flow is dictated by certain virtual resources referred to as resource pools, and is related to the delay experienced by a flow that traverses the resource pools in a store and forward manner. Their approach is to transform the multipath network into an equivalent network with single path routing comprised of virtual resources. This transformation is computationally demanding for large networks ([14]). Also, [10] fails to give useful insights into the transformation, the constituent pools or even the terms a_{js} and hence, the approximation above is difficult to use

for engineering purposes like capacity allocation, design of multipath routing/flow control schemes etc. A concrete understanding of the resource pools affecting flow's delays is needed to enable the engineering of such networks. In [8], multipath network under proportional fair allocation is considered in a heavy-traffic regime. However, [8] also relies on the transformation of the multipath network into an equivalent network with single path routing.

Our Contributions.

The main goal of this paper is to develop tools to evaluate mean delay experienced by flows and support network design when multipath flow control allocation is based on proportional fairness. However, a study of flow level dynamics of proportional fair allocation is in general hard due to its sensitivity ([5]). Thus, we adopt an approach that involves studying a more tractable network that uses multipath balanced fair allocation, a multipath generalization of the balanced fair allocation introduced in [5]. We obtain lower and upper bounds on the mean per bit delay, exhibiting its relationship to the capacity and load associated with certain resource pools. Our bounds take into account competing multipath traffic in the network, and thus our results go beyond max flow like arguments. Proving these bounds is a challenging extension of the bounds obtained in [2] for networks with single path routing, in part due to a key feature of the multipath balanced fair allocation which modifies the splitting of traffic along different routes in accordance with the state of the network. We also use the bounds to explore scenarios where peak rate constraints are likely to hurt the performance gains achievable by adding more routes.

While insightful, the performance bounds are difficult to drive network design. Hence, we use the large deviations characteristics of balanced fair allocation obtained in [18] to zero in on a collection of resources referred to as a critical pool that plays the dominant role in determining the likelihood of congestion events involving accumulations of a large number of flows. Further, these critical resources indeed behave as a single pooled resource. We establish this by studying the sensitivity of the large deviations exponent to the pool's capacity. We also use the large deviation behavior to present a disciplined approach to a capacity allocation problem for such networks that is based on a linear optimization problem.

The main contributions of this paper are summarized below:

- We develop tools to study delays experienced by flows and support network design for a network under multipath proportional fair allocation.
- We obtain a stability condition for multipath balanced fair allocation.
- For multipath balanced fair allocation, we obtain bounds on mean per bit delay that exhibit the impact of the capacity and load associated with resource pools on performance.
- We explore scenarios where peak rate constraints can reduce the performance gains from the addition of more routes.
- We obtain the most likely mix of the flows of different classes when there is an accumulation of aggregate number of flows.
- For the multipath setting, we studied the large deviation for congestion events to identify critical resource pools.
- We present a disciplined approach to a capacity allocation problem in the form of a linear optimization problem.

Organization of the paper.

We discuss the system model in Section 2. In Section 3, we discuss two insensitive multipath allocations Random routing and multipath balanced fair allocations and study their stability. In Section 4, we obtain bounds on the mean per bit delay under multipath balanced fair allocation. In Section 5, we use the large deviation characteristics of multipath balanced fair allocation to identify and study critical resource pools and study their sensitivity to the pool's capacity. We present some conclusions drawn from the main discussion of the paper in Section 6, which is followed by an Appendix which contains discussions of the proofs of some of the results in the main body of the paper.

2. SYSTEM MODEL

We begin by introducing some notation. Let \mathbb{Z}_+ denote the set of all non-negative integers. For any set \mathcal{B} , $|\mathcal{B}|$ denotes the number of elements in the set. For any vectors \mathbf{b}, \mathbf{n} whose elements are indexed by some set \mathcal{I} , i.e., $\mathbf{b} = (b_i)_{i \in \mathcal{I}}$ and $\mathbf{n} = (n_i)_{i \in \mathcal{I}}$, let

$$\mathbf{b}^{\mathbf{n}} = \prod_{i \in \mathcal{I}} b_i^{n_i} \text{ and } |\mathbf{a}| = \sum_{i \in \mathcal{I}} a_i.$$

We consider a network where possibly multipath flows arrive, utilize network resources and leave. The network is comprised of a set \mathcal{L} of links where each link $l \in \mathcal{L}$ has a capacity $c_l > 0$ bits per second. A flow ϑ arrives at some time t_ϑ^a , brings D_ϑ bits and leaves at some time t_ϑ^d given by $D_\vartheta = \int_{t_\vartheta^a}^{t_\vartheta^d} \sigma_\vartheta(t) dt$ where $\sigma_\vartheta(t)$ denotes the rate at which the flow ϑ is served at time t . We refer to D_ϑ as the size of flow ϑ .

Each flow is associated with some flow class $s \in \mathcal{S}$ where \mathcal{S} denotes the set of all flow classes. The flows associated with a class $s \in \mathcal{S}$ arrive as a Poisson process of rate ξ_s with a mean flow size $(1/\nu_s)$ bits. Let $\rho_s = \xi_s/\nu_s$ bits per second be the traffic intensity of class $s \in \mathcal{S}$. Let $\boldsymbol{\rho} = (\rho_s)_{s \in \mathcal{S}}$. The flows associated with class s are peak rate constrained (e.g., by an access link, or by the transport mechanism like in TCP where the finite receiver buffer effectively acts as a peak rate constraint.) to a rate $a_s \in [0, \infty]$ bits per second, i.e., $\sigma_\vartheta(t) \leq a_s$ for any flow ϑ of class $s \in \mathcal{S}$. If there are no peak rate constraints for class s , we set $a_s = \infty$. For any $s \in \mathcal{S}$, let R_s denote the set of possible routes for class s flows where a route is a subset of \mathcal{L} . We refer to the important special case in which $|R_s| = 1 \forall s \in \mathcal{S}$ as the single path routing case. Let \mathbf{A} be the route-link incidence matrix associated with the network, i.e., for $l \in \mathcal{L}$ and $r \in \cup_{s \in \mathcal{S}} R_s$, $A_{l,r} = 1$ if $l \in r$ (route r traverses link l) and 0 otherwise. The example in Fig. 2 should help the reader to get a feel for the notation (see the caption). The results presented in this paper are valid for a much more general model (see [5] for details) where the flows associated with each class are generated within sessions and the session arrivals correspond to independent Poisson processes. In this model, the flow size and number of flows per session of class s can have general distributions, and successive flow sizes can be correlated.

The network state is an integer valued $|\mathcal{S}|$ -tuple whose s^{th} component is the number of flows of class $s \in \mathcal{S}$. For each $s \in \mathcal{S}$ and $r_s \in R_s$, let $\phi_{sr_s}(\mathbf{x})$ denote the bandwidth allocated to the route r_s of class s in network state \mathbf{x} . For each $s \in \mathcal{S}$, let $\phi_s(\mathbf{x}) = \sum_{r_s \in R_s} \phi_{sr_s}(\mathbf{x})$ denote the total

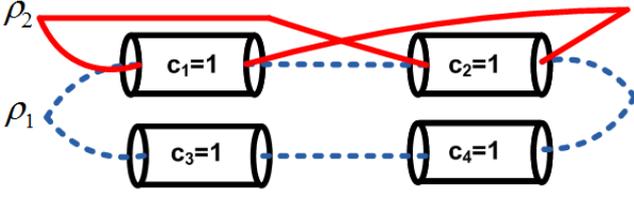


Figure 2: For the above network, $\mathcal{L} = \{1, 2, 3, 4\}$, $\mathcal{S} = \{1, 2\}$, $R_1 = \{r_{11}, r_{12}\}$, $R_2 = \{r_{21}, r_{22}\}$, $r_{11} = \{1, 2\}$, $r_{12} = \{3, 4\}$, $r_{21} = \{1\}$, $r_{22} = \{2\}$

bandwidth allocated to flows of class s in state \mathbf{x} , and assume it is shared equally among the x_s flows. Hence, there exists some $\mathbf{f}(\mathbf{x}) \in \mathcal{F}$ such that

$$\phi_{sr_s}(\mathbf{x}) = f_{sr_s}(\mathbf{x})\phi_s(\mathbf{x}), \forall s \in \mathcal{S}, r_s \in R_s,$$

where

$$\mathcal{F} = \prod_{s \in \mathcal{S}} \mathcal{F}_s,$$

$$\mathcal{F}_s = \{\mathbf{f}_s | f_{s,r_s} \geq 0, \forall r_s \in R_s; \sum_{r_s \in R_s} f_{s,r_s} = 1\} \text{ for each } s \in \mathcal{S}.$$

Thus, $(\phi(\mathbf{x}), \mathbf{f}(\mathbf{x}))$ where $\phi(\mathbf{x}) = (\phi_s(\mathbf{x}))_{s \in \mathcal{S}}$ fully characterizes the allocation in network state \mathbf{x} . We refer to $\mathbf{f} : \mathbb{Z}_+^{|\mathcal{S}|} \rightarrow \mathcal{F}$ as the splitting function. Here, the set \mathcal{F} captures all possible ways to split the bandwidth allocated to various classes across their respective routes.

The allocated bandwidths satisfy the following linear capacity constraints

$$\sum_{s \in \mathcal{S}, r_s \in R_s} A_{l,r_s} \phi_{sr_s}(\mathbf{x}) \leq c_l, \forall l \in \mathcal{L}, \forall \mathbf{x} \in \mathbb{Z}_+^{|\mathcal{S}|}, \quad (1)$$

and peak rate constraints given by

$$\phi_s(\mathbf{x}) \leq a_s x_s, \forall s \in \mathcal{S}, \forall \mathbf{x} \in \mathbb{Z}_+^{|\mathcal{S}|}. \quad (2)$$

Let

$$\mathcal{C}_P^M(\mathbf{x}) = \{(\lambda, \mathbf{f}) : \lambda = (\lambda_s)_{s \in \mathcal{S}}, \mathbf{f} \in \mathcal{F},$$

$$\left. \sum_{s \in \mathcal{S}, r_s \in R_s} A_{l,r_s} f_{sr_s} \lambda_s \leq c_l \forall l \in \mathcal{L}, \lambda_s \leq a_s x_s \forall s \in \mathcal{S} \right\}.$$

Then, (1) and (2) is equivalent to the condition

$$(\phi(\mathbf{x}), \mathbf{f}(\mathbf{x})) \in \mathcal{C}_P^M(\mathbf{x}).$$

In systems where there are no peak rate constraints, i.e., $a_s = \infty \forall s \in \mathcal{S}$, the above condition simplifies to

$$(\phi(\mathbf{x}), \mathbf{f}(\mathbf{x})) \in \mathcal{C}^M$$

where

$$\mathcal{C}^M = \{(\lambda, \mathbf{f}) : \lambda = (\lambda_s)_{s \in \mathcal{S}}, \mathbf{f} \in \mathcal{F} \text{ and } \left. \sum_{s \in \mathcal{S}, r_s \in R_s} A_{l,r_s} f_{sr_s} \lambda_s \leq c_l \forall l \in \mathcal{L} \right\}. \quad (3)$$

The multipath proportional fair allocation, for instance, satisfies (3) in which the allocation $(\lambda^{PF}(\mathbf{x}), \mathbf{f}^{PF}(\mathbf{x}))$ in a network state \mathbf{x} is a solution to the optimization problem

MULTIPATH-PF given below

$$\max_{(\lambda, \mathbf{f})} \left\{ \sum_{s \in \mathcal{S}} x_s \log(\lambda_s) \mid (\lambda, \mathbf{f}) \in \mathcal{C}^M \right\}.$$

The steady state distribution of multipath proportional fair (more generally, utility function based allocations) are sensitive to the detailed traffic characteristics which makes an analysis of their flow level dynamics hard ([5]). Balanced fair allocation introduced in [5] for the single path routing setting is much more tractable due to its insensitivity properties.

The balanced fair allocation belongs to a much larger class of insensitive allocations. In [5], it is shown that any insensitive allocation corresponds to a positive function $\Phi : \mathbb{Z}_+^{|\mathcal{S}|} \rightarrow [0, \infty)$ where the allocation for class $s \in \mathcal{S}$ in network state \mathbf{x} is given by

$$\phi_s(\mathbf{x}) = \frac{\Phi(\mathbf{x} - \mathbf{e}_s)}{\Phi(\mathbf{x})}$$

where the vector $\mathbf{e}_s \in \mathbb{Z}_+^{|\mathcal{S}|}$ has the u^{th} coordinate equal to one for $u = s$ and 0 otherwise. Further, it is shown that if

$$\sum_{\mathbf{x} \in \mathbb{Z}_+^{|\mathcal{S}|}} \Phi(\mathbf{x}) \rho^{\mathbf{x}} < \infty, \quad (4)$$

then the invariant distribution for a network state \mathbf{x} is given by

$$\pi(\mathbf{x}) = \pi(\mathbf{0}) \Phi(\mathbf{x}) \rho^{\mathbf{x}}$$

where $\pi(\mathbf{0})$ is the normalization constant of the distribution. In the rest of this section, we mainly focus on some of the important results for the single path routing setting. As each class is associated with a single route, in the rest of this section, we use s and the route associated with the class interchangeably. The balanced fair allocation for the single path routing setting is characterized by a balance function $\Phi : \mathbb{Z}_+^{|\mathcal{S}|} \rightarrow [0, \infty)$ given by ([5])

$$\Phi(\mathbf{x}) = \max \left\{ \max_{s \in \mathcal{S}: x_s > 0} \frac{\Phi(\mathbf{x} - \mathbf{e}_s)}{a_s x_s}, \max_{l \in \mathcal{L}} \sum_{s \in \mathcal{S}} \frac{A_{l,s}}{c_l} \Phi(\mathbf{x} - \mathbf{e}_s) \right\}$$

for $\mathbf{x} \in \mathbb{Z}_+^{|\mathcal{S}|} \setminus \{\mathbf{0}\}$ with $\Phi(\mathbf{0}) = 1$, $\Phi(\mathbf{x}) = 0$ outside the positive quadrant. The balanced fair allocation for class s in network state \mathbf{x} is then given by

$$\phi_s(\mathbf{x}) = \frac{\Phi(\mathbf{x} - \mathbf{e}_s)}{\Phi(\mathbf{x})}.$$

This allocation satisfies (1) and (2). In [5], it is shown that if ρ satisfies the following stability condition

$$\sum_{s \in \mathcal{S}, r_s \in R_s} A_{l,s} \rho_s < c_l \forall l \in \mathcal{L}, \quad (5)$$

then the invariant distribution for a network state \mathbf{x} is given by

$$\pi(\mathbf{x}) = \pi(\mathbf{0}) \Phi(\mathbf{x}) \rho^{\mathbf{x}} \quad (6)$$

where $\pi(\mathbf{0})$ is the normalization constant of the distribution. Note that the invariant distribution is *insensitive* to all the traffic characteristics except the traffic intensity. We refer to [5] for an extensive discussion on balanced fair allocation for the single path routing case. Except for certain networks ([6, 2, 4]), numerical evaluation of mean per-bit delay is

intractable for large $|\mathcal{S}|$ due to state space explosion. The simple explicit bounds for the mean per bit delay τ_s of a class $s \in \mathcal{S}$ obtained in [2] become valuable in such settings. The bounds are:

$$\begin{aligned} \tau_s &\geq \max \left(\frac{1}{a_s}, \max_{l \in \mathcal{L}} \frac{A_{l,s}}{c_l - \alpha_l} \right), \\ \tau_s &\leq \max \left(\frac{1}{a_s}, \max_{l \in \mathcal{L}} \frac{A_{l,s}}{c_l} \right) + \sum_{l \in \mathcal{L}} \frac{\alpha_l}{c_l} \frac{A_{l,s}}{c_l - \alpha_l} \end{aligned} \quad (7)$$

where $\alpha_l = \sum_s A_{l,s} \rho_s$ for $l \in \mathcal{L}$. Note the intuition here is as follows. The lower bound is at least that associated with the peak rate constraint of class s , or that of the bottleneck link along the route of class s . The upper bound includes the peak rate constraint or capacity constraint of links along the path, and an additive term that roughly corresponds to sending the bits, in a store and forward manner over the links along the route of class s . Hence, we have a weighted summation of the mean per bit delays on each link on the route.

3. RANDOM ROUTING VS MULTIPATH FLOW CONTROL

Our objective in this section and the next is to study and compare flow level delays in two systems. In the first system, for any class s , each flow of class s randomly chooses one route in R_s . We refer to this as *random routing*. In the second system, for any class s , each flow of class s can simultaneously use all the routes in the set R_s . The comparison allows us to gauge the benefits of balancing flow loads across paths vs balancing traffic (e.g., packet level) across multiple paths during the lifetime of a flow.

3.1 Random routing

A random routing policy corresponds to a $\mathbf{p} \in \mathcal{F}$ where each flow of class $s \in \mathcal{S}$ is routed to route $r_s \in R_s$ independently with probability p_{sr_s} . We could choose \mathbf{p} to optimize a performance objective like load balancing. Since the route of each flow of any class s is chosen independently using \mathbf{p} , we can treat the traffic of class s on route $r_s \in R_s$ as that of an independent class indexed by (s, r_s) with traffic intensity $\rho_{(s, r_s)} = p_{sr_s} \rho_s$ and we let $\boldsymbol{\rho}^R = (\rho_{(s, r_s)})_{s \in \mathcal{S}, r_s \in R_s}$. The route level network state is a vector $\mathbf{x}^R \in \mathbb{Z}^{\sum_{s \in \mathcal{S}} |R_s|}$ where $x_{(s, r_s)}^R$ denotes the number of flows of class s routed along route $r_s \in R_s$. For a route level network state \mathbf{x}^R , let $\phi_{(s, r_s)}^R(\mathbf{x}^R)$ denote the total bandwidth allocated to flows of class s routed along $r_s \in R_s$, and assume this bandwidth is shared equitably by these flows. Further, the allocation must satisfy (1) and (2). For the route level balanced fair allocation, the allocation for the set of active flows of class s along route r_s is then given by

$$\phi_{(s, r_s)}^R(\mathbf{x}^R) = \frac{\Phi^R(\mathbf{x}^R - \mathbf{e}_{(s, r_s)}^R)}{\Phi^R(\mathbf{x}^R)}.$$

We refer to the (single path routing) balanced fair allocation for this setting as the route level balanced fair allocation, where the allocation is characterized by the balance function

$\Phi^R : \mathbb{Z}^{\sum_{s \in \mathcal{S}} |R_s|} \rightarrow [0, \infty)$ given by

$$\Phi^R(\mathbf{x}^R) = \max \left\{ \max_{s \in \mathcal{S}, r_s \in R_s: x_{(s, r_s)} > 0} \frac{\Phi^R(\mathbf{x}^R - \mathbf{e}_{(s, r_s)}^R)}{a_s x_{(s, r_s)}}, \max_{l \in \mathcal{L}} \sum_{s \in \mathcal{S}, r_s \in R_s} \frac{A_{l, r_s}}{c_l} \Phi^R(\mathbf{x}^R - \mathbf{e}_{(s, r_s)}^R) \right\} \quad (8)$$

for $\mathbf{x}^R \in \mathbb{Z}_+^{\sum_{s \in \mathcal{S}} |R_s|} \setminus \{\mathbf{0}\}$, $\Phi^R(\mathbf{0}) = 1$ and $\Phi^R(\mathbf{x}^R) = 0$ outside the positive quadrant. Here, $\mathbf{e}_{(s, r_s)}^R \in \mathbb{Z}^{\sum_{s \in \mathcal{S}} |R_s|}$ with 1 in the coordinate corresponding to the class (s, r_s) and zero otherwise.

The next result gives the stability condition and provides bounds on the mean per bit delay experienced by flows of a class $s \in \mathcal{S}$ under the route level balanced fair allocation. To that end, for a given \mathbf{p} , let

$$\Gamma(\mathbf{p}) = \left\{ \mu \in \mathbb{R}_+^{|\mathcal{S}|} : \sum_{s \in \mathcal{S}, r_s \in R_s} A_{l, r_s} p_{sr_s} \mu_s < c_l \quad \forall l \in \mathcal{L}, \right\}.$$

Then, we have the following result.

THEOREM 1. *If $\boldsymbol{\rho} \in \Gamma(\mathbf{p})$, the following hold:*

- (a) *The invariant distribution $\pi^R(\mathbf{x}^R)$ for the route level network state \mathbf{x}^R is given by*

$$\pi^R(\mathbf{x}^R) = \pi^R(\mathbf{0}) \Phi^R(\mathbf{x}^R) (\boldsymbol{\rho}^R)^{\mathbf{x}^R} \quad (9)$$

where $\pi^R(\mathbf{0})$ is the normalization constant for the distribution.

- (b) *The mean mean per bit delay τ_t for a class $t \in \mathcal{S}$ satisfies*

$$\begin{aligned} \tau_t &\geq \sum_{r \in R_t} p_{tr} \left(\max \left\{ \frac{1}{a_t}, \max_{l \in \mathcal{L}} \frac{A_{l, r}}{c_l - \alpha_l} \right\} \right) \text{ and} \\ \tau_t &\leq \sum_{r \in R_t} p_{tr} \left(\max \left\{ \frac{1}{a_t}, \max_{l \in \mathcal{L}} \frac{A_{l, r}}{c_l} \right\} + \sum_{l \in \mathcal{L}} \frac{\alpha_l}{c_l} \frac{A_{l, r}}{c_l - \alpha_l} \right) \end{aligned}$$

where for $l \in \mathcal{L}$,

$$\alpha_l = \sum_{s \in \mathcal{S}, r_s \in R_s} A_{l, r_s} p_{sr_s} \rho_s.$$

We skip the proof as it follows from (7). In particular, the bounds given above can be proved by noting that

$$\tau_t = \sum_{r_t \in R_t} p_{tr_t} \tau_{(t, r_t)}$$

and using the bounds given in (7) for the mean per bit delay.

3.2 Multipath flow control

Next, we consider the second system in which *each flow* of class s can simultaneously use all the routes in the set R_s by sending its data across on all the routes. Further, the splitting of the traffic corresponding to a class can change depending on the network state.

As pointed out in Section 2, corresponding to each insensitive allocation, there is a balance function. However, in the multipath setting, to fully characterize such an allocation, we need to further specify the splitting function $\mathbf{f}(\mathbf{x})$ for each network state \mathbf{x} . Thus, in the multipath setting, any insensitive allocation is fully characterized by a balance function

and a splitting function defined for all network states. Then, the allocation for class s in network state \mathbf{x} given by

$$\phi_s(\mathbf{x}) = \frac{\Phi(\mathbf{x} - \mathbf{e}_s)}{\Phi(\mathbf{x})} \quad (10)$$

and the bandwidth allocated on the route $r_s \in R_s$ is given by

$$\phi_{s,r_s}(\mathbf{x}) = f_{s,r_s}(\mathbf{x})\phi_s(\mathbf{x}).$$

We define a multipath balanced fair allocation for this system as the insensitive allocation obtained using the balance function $\Phi : \mathbb{Z}^{|\mathcal{S}|} \rightarrow [0, \infty)$ defined as

$$\Phi(\mathbf{x}) = \max \left\{ \max_{s \in \mathcal{S}: x_s > 0} \frac{\Phi(\mathbf{x} - \mathbf{e}_s)}{a_s x_s}, \min_{\mathbf{f} \in \mathcal{F}} \max_{l \in \mathcal{L}} \left(\sum_{s \in \mathcal{S}, r_s \in R_s} \frac{A_{l,r_s}}{c_l} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) \right) \right\} \quad (11)$$

for $\mathbf{x} \in \mathbb{Z}^{|\mathcal{S}|} \setminus \{\mathbf{0}\}$, $\Phi(\mathbf{0}) = 1$ and $\Phi(\mathbf{x}) = 0$ outside the positive quadrant. For any network state \mathbf{x} , the splitting function $\mathbf{f}(\mathbf{x})$ is equal to $\mathbf{f}^*(\mathbf{x})$ which is a minimizer of the optimization problem given below:

$$\min_{\mathbf{f} \in \mathcal{F}} \max_{l \in \mathcal{L}} \left(\sum_{s \in \mathcal{S}, r_s \in R_s} \frac{A_{l,r_s}}{c_l} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) \right). \quad (12)$$

Note that the splitting function need not be unique. Also, note that such allocations satisfy (1) and (2). Further, note that the splitting function $\mathbf{f}^*(\mathbf{x})$ of the allocation is state-dependent and recursively defined. Also,

$$\Phi(\mathbf{x}) = \max \left\{ \max_{s \in \mathcal{S}: x_s > 0} \frac{\Phi(\mathbf{x} - \mathbf{e}_s)}{a_s x_s}, \max_{l \in \mathcal{L}} \left(\sum_{s \in \mathcal{S}, r_s \in R_s} \frac{A_{l,r_s}}{c_l} f_{s,r_s}^*(\mathbf{x}) \Phi(\mathbf{x} - \mathbf{e}_s) \right) \right\}. \quad (13)$$

Also, Lemma 1 in Appendix A establishes that the balance function associated with multipath balanced fair allocation defined above is unique.

Next, we study the stability of the multipath balanced fair allocation. Let

$$\Gamma^M = \left\{ \mu \in \mathbb{R}_+^{|\mathcal{S}|} : \mu \in \Gamma(\mathbf{f}) \text{ for some } \mathbf{f} \in \mathcal{F} \right\}. \quad (14)$$

Then, the next result gives the stability region. See Appendix A for a proof.

THEOREM 2. *If the offered load $\boldsymbol{\rho} \in \Gamma^M$, then the invariant distribution for a network state \mathbf{x} is given by*

$$\pi(\mathbf{x}) = \pi(\mathbf{0})\Phi(\mathbf{x})\boldsymbol{\rho}^{\mathbf{x}} \quad (15)$$

where $\pi(\mathbf{0})$ is the normalization constant of the distribution.

From Lemma 1 (see Appendix A) and (15), we can conclude that the probability of having no flows in the system for the multipath balanced fair allocation is greater than or equal to that of any multipath insensitive allocation.

Also, note that $\Gamma(\boldsymbol{\rho}) \subseteq \Gamma^M$ for any $\boldsymbol{\rho} \in \mathcal{F}$. Thus, if a network using route level balanced fair allocation is stable, it will be stable when using multipath balanced fair allocation.

4. PERFORMANCE BOUNDS

In this section, we move beyond stability conditions and study the mean per bit delay for such a system. In this system too, the numerical evaluation of mean per bit delay for the balanced fair allocation is intractable for large $|\mathcal{S}|$. Thus, we resort to bounding the mean per bit delay. We consider bounds that are based on pooled capacity constraints. The following definitions capture how such pooling is considered.

DEFINITION 1. *A (possibly multipath) flow class s is said to be supported by $\mathcal{H} \subset \mathcal{L}$ if for each $r_s \in R_s$, $|r_s \cap \mathcal{H}| > 0$, i.e., each route of class s traverses at least one link in \mathcal{H} . Let $\mathcal{S}(\mathcal{H}) \subset \mathcal{S}$ denote the set of flow classes that are supported by \mathcal{H} .*

DEFINITION 2. *A (possibly multipath) flow class s is said to be partially supported by $\mathcal{H} \subset \mathcal{L}$ if for some $r_s \in R_s$, $|r_s \cap \mathcal{H}| > 0$, i.e., at least one route of class s traverses at least one link in \mathcal{H} . We let $\mathcal{P}(\mathcal{H}) \subset \mathcal{S}$ denote the set of flow classes that are partially supported by \mathcal{H} .*

Note that $\mathcal{S}(\mathcal{H}) \subset \mathcal{P}(\mathcal{H})$, i.e., if a flow is supported by \mathcal{H} , then it is also partially supported by \mathcal{H} . Next, for each class $s \in \mathcal{S}$, we define the following collections of resource pools:

$$\begin{aligned} \mathcal{B}_s &= \{ \mathcal{H} \subseteq \mathcal{L} : s \in \mathcal{S}(\mathcal{H}) \} \text{ and} \\ \mathcal{D}_s &= \left\{ \mathcal{H} \in \mathcal{B}_s : \text{there exists no } \mathcal{H}' \in \mathcal{B}_s, \right. \\ &\quad \left. \mathcal{H}' \neq \mathcal{H} \text{ such that } \mathcal{H}' \subset \mathcal{H} \right\}. \end{aligned}$$

Thus, \mathcal{B}_s is the collection of resource pools that support flow class s . The set \mathcal{D}_s corresponds to a minimal collection of pooled resources that support flow class s , i.e., pools which can not be reduced and still support class s . For the network in Fig. 2, $\mathcal{D}_1 = \{\{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\}$ and $\mathcal{D}_2 = \{\{1, 2\}\}$. In the next section, we will see that the bounds on mean per bit delay for a class $s \in \mathcal{S}$ will be dictated by the cumulative capacities and traffic associated the resource pools in \mathcal{B}_s and \mathcal{D}_s .

In general, a multipath flow may traverse a set of links \mathcal{H} multiple times, or may have constituent routes that traverse those resources different numbers of times. To account for these we define the following.

DEFINITION 3. *For any flow class s , route $r_s \in R_s$ and set $\mathcal{H} \subseteq \mathcal{L}$, let $n_{s,r_s}(\mathcal{H}) = |r_s \cap \mathcal{H}|$, i.e., $n_{s,r_s}(\mathcal{H})$ denotes the number of links (possibly zero) in \mathcal{H} that r_s traverses. Let $n_s(\mathcal{H}) = \min_{r \in R_s} n_{s,r}(\mathcal{H})$ be the minimum multiplicity with which the class s is supported by \mathcal{H} . Similarly, let $\bar{n}_s(\mathcal{H}) = \max_{r \in R_s} n_{s,r}(\mathcal{H})$ be the maximum multiplicity with which the class s is supported by \mathcal{H} .*

For the network in Fig. 2, if $\mathcal{H} = \{1, 2\}$, then $n_1(\mathcal{H}) = 0$, $\bar{n}_1(\mathcal{H}) = 2$ and $n_2(\mathcal{H}) = \bar{n}_2(\mathcal{H}) = 1$. Finally, let

$$c(\mathcal{H}) = \sum_{l \in \mathcal{H}} c_l$$

denote the pooled capacity of an $\mathcal{H} \subset \mathcal{L}$,

$$\rho(\mathcal{H}) = \sum_{s \in \mathcal{S}(\mathcal{H})} n_s(\mathcal{H})\rho_s$$

denote the aggregate load of classes supported by \mathcal{H} accounting for multiplicities, and

$$\bar{\rho}(\mathcal{H}) = \sum_{s \in \mathcal{P}(\mathcal{H})} \bar{n}_s(\mathcal{H})\rho_s$$

be an upper bound on the aggregate load partially supported by \mathcal{H} also accounting for multiplicities.

In this section, we obtain bounds on the mean per bit delay of a flow of class $t \in \mathcal{S}$ under balanced fair allocation with multipath flow control assuming that the stability condition holds, i.e., the offered load $\boldsymbol{\rho} \in \Gamma^M$. From Little's theorem and (15), the mean per bit delay τ_t for any class $t \in \mathcal{S}$ satisfies

$$\tau_t = \frac{E[X_t]}{\rho_t} = \frac{1}{\rho_t} \sum_{\mathbf{x}} x_t \pi(\mathbf{x}) = \frac{1}{\rho_t} \frac{\sum_{\mathbf{x}} x_t \Phi(\mathbf{x}) \boldsymbol{\rho}^{\mathbf{x}}}{\sum_{\mathbf{x}} \Phi(\mathbf{x}) \boldsymbol{\rho}^{\mathbf{x}}}$$

where $\sum_{\mathbf{x}}$ stands for $\sum_{\mathbf{x} \in \mathbb{Z}_+^{|\mathcal{S}|}}$. The numerical evaluation of mean per bit delay using the above expression becomes intractable as the number of classes of flows becomes large. Also, note that evaluation of $\Phi(\mathbf{x})$ using (11) involves linear programming for each network state \mathbf{x} . Hence, the simple explicit bounds on mean per bit delay obtained in the rest of this section are invaluable towards understanding for the flow level performance in a general network with multipath routes and flow control. These bounds are generalizations of those in [2] which only consider the single path routing setting. The proofs of the bounds presented in this section use ideas from [2], but require substantial additional development to show, due to the state dependent nature of the splitting function $\mathbf{f}^*(\mathbf{x})$.

4.1 Lower bound

The following theorem gives a lower bound on the mean per bit delay τ_t of a class $t \in \mathcal{S}$.

THEOREM 3. *For any class $t \in \mathcal{S}$,*

$$\tau_t \geq \max \left\{ \frac{1}{a_t}, \max_{\mathcal{H} \in \mathcal{B}_t} \left(\frac{n_t(\mathcal{H})}{c(\mathcal{H}) - \rho(\mathcal{H})} \right) \right\}.$$

PROOF. For an $\mathcal{H} \subset \mathcal{L}$, consider a link $l \in \mathcal{H}$. From (13),

$$\Phi(\mathbf{x}) \geq \sum_{s \in \mathcal{S}, r \in R_s} \frac{A_{l,r}}{c_l} f_{s,r}^*(\mathbf{x}) \Phi(\mathbf{x} - \mathbf{e}_s).$$

Using (15), for any link $l \in \mathcal{H}$, we have

$$\pi(\mathbf{x}) \geq \sum_{s \in \mathcal{S}, r \in R_s} \frac{A_{l,r}}{c_l} f_{s,r}^*(\mathbf{x}) \rho_s \pi(\mathbf{x} - \mathbf{e}_s).$$

Further, we can write

$$\begin{aligned} \pi(\mathbf{x}) &= \sum_{l \in \mathcal{H}} \frac{c_l}{c(\mathcal{H})} \pi(\mathbf{x}) \\ &\geq \sum_{l \in \mathcal{H}} \frac{c_l}{c(\mathcal{H})} \sum_{s \in \mathcal{S}, r \in R_s} \frac{A_{l,r}}{c_l} f_{s,r}^*(\mathbf{x}) \rho_s \pi(\mathbf{x} - \mathbf{e}_s) \\ &\geq \frac{1}{c(\mathcal{H})} \sum_{s \in \mathcal{S}, r \in R_s} n_{s,r}(\mathcal{H}) f_{s,r}^*(\mathbf{x}) \rho_s \pi(\mathbf{x} - \mathbf{e}_s) \\ &\geq \frac{1}{c(\mathcal{H})} \sum_{s \in \mathcal{S}(\mathcal{H})} n_s(\mathcal{H}) \rho_s \pi(\mathbf{x} - \mathbf{e}_s). \end{aligned}$$

Thus, we have (using $\sum_{\mathbf{x}}$ for $\sum_{\mathbf{x} \in \mathbb{Z}_+^{|\mathcal{S}|}}$)

$$\begin{aligned} E[X_t] &= \sum_{\mathbf{x}} x_t \pi(\mathbf{x}) \geq \sum_{\mathbf{x}} \frac{x_t}{c(\mathcal{H})} \sum_{s \in \mathcal{S}(\mathcal{H})} n_s(\mathcal{H}) \rho_s \pi(\mathbf{x} - \mathbf{e}_s) \\ &= \sum_{s \in \mathcal{S}(\mathcal{H})} \frac{n_s(\mathcal{H}) \rho_s}{c(\mathcal{H})} \sum_{\mathbf{x}} x_t \pi(\mathbf{x} - \mathbf{e}_s) \\ &= \sum_{s \in \mathcal{S}(\mathcal{H}) \setminus \{t\}} \frac{n_s(\mathcal{H}) \rho_s}{c(\mathcal{H})} \sum_{\mathbf{x}} x_t \pi(\mathbf{x} - \mathbf{e}_s) \\ &\quad + \frac{n_t(\mathcal{H}) \rho_t}{c(\mathcal{H})} \left(\sum_{\mathbf{x}} (x_t - 1) \pi(\mathbf{x} - \mathbf{e}_t) + \sum_{\mathbf{x}} \pi(\mathbf{x} - \mathbf{e}_t) \right) \\ &= \sum_{s \in \mathcal{S}(\mathcal{H}) \setminus \{t\}} \frac{n_s(\mathcal{H}) \rho_s}{c(\mathcal{H})} E[X_t] + \frac{n_t(\mathcal{H}) \rho_t}{c(\mathcal{H})} (E[X_t] + 1) \\ &= \frac{\rho(\mathcal{H})}{c(\mathcal{H})} E[X_t] + \frac{n_t(\mathcal{H}) \rho_t}{c(\mathcal{H})}. \end{aligned}$$

From the above inequality, we conclude

$$\tau_t = \frac{E[X_t]}{\rho_t} \geq \frac{n_t(\mathcal{H})}{c(\mathcal{H}) - \rho(\mathcal{H})}.$$

Since the above inequality holds for any subset \mathcal{H} of \mathcal{L} , $n_t(\mathcal{H}) = 0$ for $\mathcal{H} \notin \mathcal{B}_t$, and $\tau_t \geq 1/a_t$, the theorem follows. \square

Intuitively, the lower bound for mean per bit delay is at least that associated with the peak rate constraint of class t , or that of the bottleneck resource pool associated with the routes of class t . We consider the network in Fig. 2 to stress that a lower bound based on pools in \mathcal{D}_t alone may not be tight. Let $\rho_1 = 1$ and $\rho_2 = 2 - \epsilon$. Suppose we are interested in bounding the performance of flow class 1. Any lower bound for flow class 1 using a pool in \mathcal{D}_1 , involves two resources with sum capacity 2, say link 1 and link 3, i.e., $\mathcal{H} = \{1, 3\}$, whence $c(\mathcal{H}) = 2$ and $\rho(\mathcal{H}) = 1$ since only flow class 1 is supported by \mathcal{H} . The lower bound is thus

$$\tau_1 \geq \frac{1}{c(\mathcal{H}) - \rho(\mathcal{H})} = 1.$$

Now consider the set $\mathcal{H}' = \{1, 2, 3\}$. Both flow classes are supported by \mathcal{H}' and thus we get a much tighter lower bound

$$\tau_1 \geq \frac{n_1(\mathcal{H}')}{c(\mathcal{H}') - \rho(\mathcal{H}')} = \frac{1}{3 - 1 - (2 - \epsilon)} = \frac{1}{\epsilon}.$$

4.2 Upper bound

One can also show an upper bound on the mean per bit delay of a class $t \in \mathcal{S}$. Let

$$b_t = \max \left\{ \frac{1}{a_t}, \left(\max_{\mathcal{H} \in \mathcal{D}_t} \frac{\bar{n}_t(\mathcal{H})}{c(\mathcal{H})} \right) \right\}$$

where b_t roughly is the bottleneck resource/peak rate constraint. Our upper bound is given in the following theorem:

THEOREM 4. *For any $t \in \mathcal{S}$, if $\bar{\rho}(\mathcal{H}) < c(\mathcal{H}) \forall \mathcal{H} \in \mathcal{D}_t$, then*

$$\tau_t \leq b_t + \sum_{\mathcal{H} \in \mathcal{D}_t} \frac{\bar{\rho}(\mathcal{H})}{c(\mathcal{H})} \left(\frac{\bar{n}_t(\mathcal{H})}{c(\mathcal{H}) - \bar{\rho}(\mathcal{H})} \right).$$

The proof of the upper bound is quite long and due to space constraints, we are not able to include the complete argument in this paper. However, the key intermediate results and a sketch of the proof are given in Appendix B. We can see that the upper bound includes the peak rate constraint or capacity constraint of pooled resources along the path, and an additive term that roughly corresponds to sending the bits, in a store and forward manner over the pooled resources along the route of class s . The expression for the upper bound provides some rough insights into how a flow is delayed in such a system.

Though the bound can be loose (especially when $|\mathcal{D}_t|$ is large or when $\bar{\rho}(\mathcal{H})$ is close to $c(\mathcal{H})$ for some $\mathcal{H} \in \mathcal{D}_t$), it is the only non-trivial upper bound available for mean per bit delay in a system using multipath routing with multipath flow control. Further, we can use the upper bound to provide mean delay guarantees to a class in such a system.

Another approach to obtain bounds on performance for networks using multipath routing with multipath flow control is to transform the original network to an equivalent network (for e.g., see [10], [8]) with single path routing, and apply bounds in [2] to it. We feel that our lower bound is going to be close to the lower bound obtained using this approach. However, the upper bound obtained for the equivalent network can be much better than our upper bound. But, as pointed out in Section 1, this approach is computationally demanding for large networks.

4.3 Comparisons and examples

In Section 3, we compared the throughput of a system using random routing with balanced fair allocation against a system using multipath routing and multipath flow control with balanced fair allocation. We saw that the latter has a larger stability region. In this section, we compare mean per bit delay for the two systems using the bounds obtained in the previous sections.

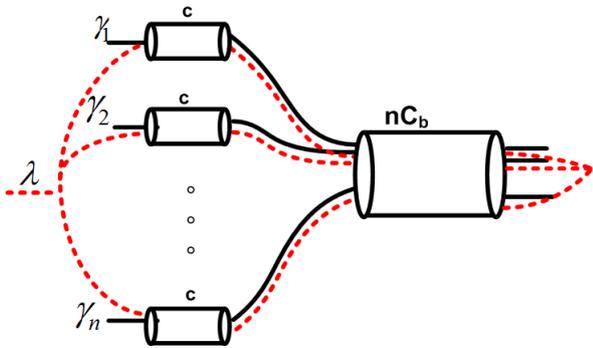


Figure 3: Example: One multipath class competing with n unipath classes

Consider the network shown in Fig. 3 being shared by one multipath class with peak rate constraint a and n unipath classes without any peak rate constraints. Each route in the network comprises a link of capacity c and a link of capacity c_b where $c \ll c_b$. The routes used by the classes should be clear from the figure: the multipath class is using n routes and each unipath class is using a different route. Note that the above network captures a scenario that can arise in the

residential wireless mesh network in Fig. 1 in which there is only one home user using multipath routing.

If the network uses random routing where the multipath class chooses a route for a flow independently and with equal probability, the following upper bound τ_r^{UB} on mean per bit delay τ_r of the multipath class can be obtained using Theorem 1

$$\tau_r \leq \tau_r^{UB} = \max\left(\frac{1}{a}, \frac{1}{c}\right) + \left(\frac{1}{c}\right) \frac{\gamma + \lambda}{c - (\gamma + \lambda)} + \left(\frac{1}{nc_b}\right) \frac{\gamma + \lambda}{c_b - (\gamma + \lambda)}. \quad (16)$$

If the network uses multipath flow control, the following lower bound on mean per bit delay τ_m of the multipath class can be obtained using Theorem 3

$$\tau_m \geq \max\left(\frac{1}{a}, \frac{1}{n(c - (\gamma + \lambda))}\right).$$

The improvement in delay by using multipath flow control is captured by $G = \frac{\tau_r}{\tau_m}$ which can be upper bounded using the above bounds as given below

$$G \leq \min(a, n(c - (\gamma + \lambda))) \tau_r^{UB}.$$

Intuitively, we would expect G to increase with the number of routes n of the multipath class due to increased statistical multiplexing. However, this need not be the case always as shown below. Since $c \ll c_b$, the first two terms in (16) dominate τ_r^{UB} and thus, τ_r^{UB} does not change much with n . From the above expression, we can conclude that we can expect a linear improvement in mean per bit delay for multipath flow control only if $a > n(c - (\gamma + \lambda))$. Thus, as far as improvement in delay performance with multipath flow control is concerned, a key factor is whether the peak rate constraint exceeds the average spare capacity $n(c - (\gamma + \lambda))$. Thus, adding more routes to a system may not always help.

This can also be inferred from Fig. 4 where we compare the mean per bit delay of the multipath class for $n = 2$ and $n = 4$. The data points are obtained from a discrete event simulation (an event comprises of an arrival or a departure of a flow) carried out for different values of traffic intensity λ of the multipath class. The active flows are served at a rate determined by the balanced fair allocation. Each data point corresponds to a simulation run involving roughly 10^5 flows of each class. We set $c = 1$, $c_b = 2$, $a = 0.5$ and $\gamma = 0.25$ and simulated the networks with multipath balanced fair allocation. We can see that the improvement in delay performance by adding two more routes is minimal for low loads where the peak rate constraints and the average spare capacity are comparable whereas the improvement for larger loads where the peak rate constraint dominates the spare capacity. We have also plotted the lower and upper bounds for the mean per bit delay of the multipath class obtained using Theorem 3 and 4 which are quite good for this example.

5. BALANCED FAIR: LARGE DEVIATIONS

In this section, we use the large deviation characteristics of networks under single path and multipath balanced fair allocations to obtain insights that are useful in network design. As we have seen in the previous sections, directly studying the mean delay is often very difficult and hence, we consider meaningful alternatives. We focus on the events in a network

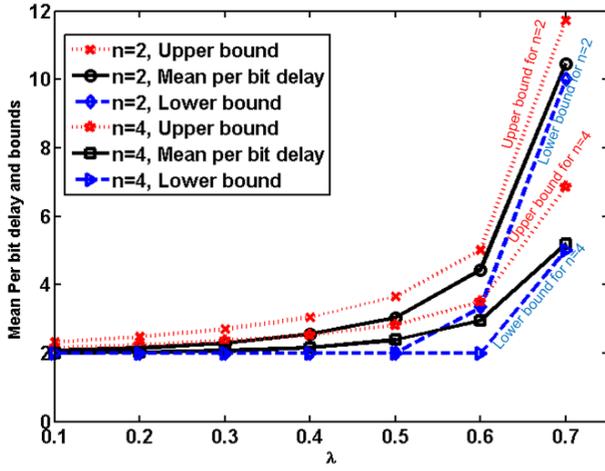


Figure 4: A comparison of mean per bit delay of the multipath class for different n

where there are accumulations in the aggregate number of flows. Accumulation of flows in the network would loosely translate to an increase in the delay seen by the flows. However, large deviations in delays could be due to several other reasons too (see [17]). In Sections 5.1 and 5.2, we study large deviations in the total number of flows. Though single path routing is a special case of multipath routing, for clarity, we start by considering it on its own. The notation used for the two settings are similar and hence, we use a superscript or (sometimes) subscript U for the terms associated with single path routing, and a superscript or (sometimes) subscript M is used for the multipath setting. Also, in this section, we assume that there are no peak rate constraints, i.e., $a_s = \infty$, $\forall s \in \mathcal{S}$ and hence, we ignore them.

5.1 Single route networks

We consider the single path routing setting where each class is associated with a single route. In the following, we use s and the route associated with the class interchangeably. Let

$$\Lambda^U = \left\{ \lambda \in \mathbb{R}^{|\mathcal{S}|} : \sum_{s \in \mathcal{S}} A_{l,s} \lambda_s \leq c_l \quad \forall l \in \mathcal{L} \right\}.$$

Define set $\mathcal{K}^U \subset \mathbb{R}_+^{|\mathcal{S}|}$ as follows:

$$\gamma = (\gamma_s)_{s \in \mathcal{S}} \in \mathcal{K}^U \iff \{e^{\gamma_s}\}_{s \in \mathcal{S}} \in \Lambda^U.$$

Let the function $\delta_{\mathcal{K}^U}$ be given by

$$\delta_{\mathcal{K}^U}(\gamma) = \begin{cases} 0, & \gamma \in \mathcal{K}^U, \\ \infty, & \gamma \notin \mathcal{K}^U. \end{cases}$$

Let $\delta_{\mathcal{K}^U}^*$ denote the convex conjugate function of $\delta_{\mathcal{K}^U}$, i.e.,

$$\delta_{\mathcal{K}^U}^*(\mathbf{x}) = \sup_{\gamma \in \mathbb{R}^{|\mathcal{S}|}} (\langle \gamma, \mathbf{x} \rangle - \delta_{\mathcal{K}^U}(\gamma)).$$

Note that $\delta_{\mathcal{K}^U}^*(\mathbf{x})$ is simply an alternative way to write the maximum value of sum log utility subject to the capacity constraints Λ^U . A proportional fair allocation (note that proportional fair allocation need not be unique for a state \mathbf{x} with $x_s = 0$ for some $s \in \mathcal{S}$ as the choice of allocation for

class s will not affect the objective function as long as we do not violate any capacity constraints) $(\lambda_s^{PF}(\mathbf{x}))_{s \in \mathcal{S}}$ in state \mathbf{x} is an optimizer to the above problem and hence,

$$\delta_{\mathcal{K}^U}^*(\mathbf{x}) = \sum_{s \in \mathcal{S}} x_s \log(\lambda_s^{PF}(\mathbf{x})).$$

In [18], the following large deviation result was proved for the single path routing setting:

$$\lim_{n \rightarrow \infty} \frac{\log(\pi_{BF}^U(n\mathbf{x}))}{n} = -L^U(\mathbf{x}). \quad (17)$$

where $\pi_{BF}^U(\cdot)$ denotes the steady state distribution under balanced fair allocation and

$$L^U(\mathbf{x}) = \delta_{\mathcal{K}^U}^*(\mathbf{x}) - \sum_{s \in \mathcal{S}} x_s \log(\rho_s).$$

The above result presents an interesting relationship between the large deviation characteristics of balanced fair allocation and the proportional fair allocation. Hence,

$$\pi_{BF}^U(n\mathbf{x}) = \zeta^U(n) e^{-nL^U(\mathbf{x})} \text{ for large } n \quad (18)$$

and some sub-exponential function $\zeta^U(n)$. Thus, the likelihood of accumulating a large number of flows, and the associated mix is determined by

$$\mathbf{x}^* = \operatorname{argmin}_{\{\mathbf{x}: x_s \geq 0 \quad \forall s \in \mathcal{S} \text{ and } \sum_{s \in \mathcal{S}} x_s = 1\}} L^U(\mathbf{x}) \quad (19)$$

This mix is a useful vector which helps us to find the classes and resources that require attention from the capacity allocation point of view. For instance, if $|\mathcal{S}| = 3$ and $\mathbf{x}^* = [0.2 \ 0.8 \ 0]$, then a typical congestion event will involve a large number of flows of the second class, and thus we can take steps to remedy this. Also, from (18), $L^U(\mathbf{x}^*)$ is roughly the negative of the rate of exponential decay along the direction \mathbf{x}^* . We will refer to $L^U(\mathbf{x}^*)$ as the LD exponent. The next result gives an expression for \mathbf{x}^* and $L^U(\mathbf{x}^*)$.

THEOREM 5. For the single path routing setting,

$$x_s^* = b^U n_s^U \rho_s \quad \forall s \in \mathcal{S} \text{ and } L^U(\mathbf{x}^*) = \log(d_U^*),$$

$$\text{where } b^U = \frac{1}{\sum_{u \in \mathcal{S}} n_u^U \rho_u}, \quad n_s^U = \sum_{l \in \mathcal{L}_{crit}^U} A_{l,s}$$

$$\mathcal{L}_{crit}^U = \left\{ l : \sum_{s \in \mathcal{S}} A_{l,s} \rho_s d_U^* = c_l \right\}, \quad d_U^* = \min_{l \in \mathcal{L}} \left(\frac{c_l}{\sum_{s \in \mathcal{S}} A_{l,s} \rho_s} \right).$$

The above result can be proved using following three facts:

(i) $\delta_{\mathcal{K}^U}^*(\cdot)$ is a convex function and hence, optimization problem in (19) is convex.

(ii) From [18], the subgradient set of $\delta_{\mathcal{K}^U}^*(\cdot)$ satisfies

$$\left\{ \gamma : [e^{\gamma_s}]_{s \in \mathcal{S}} \in \Lambda^{PF}(\mathbf{x}^*) \right\} \subset \partial \delta_{\mathcal{K}^U}^*(\mathbf{x}^*)$$

where

$$\Lambda_{PF}^U(\mathbf{x}^*) = \left\{ \boldsymbol{\lambda} : \boldsymbol{\lambda} \text{ is a proportional fair allocation in the state } \mathbf{x}^* \right\}.$$

(iii) A proportional fair allocation $\boldsymbol{\lambda}$ for state \mathbf{x}^* satisfies ([11])

$$\lambda_s = \frac{x_s^*}{\sum_{l \in \mathcal{L}} A_{l,s} p_l} \quad \forall s \in \{u : x_u^* > 0\}$$

where $\forall l \in \mathcal{L}$, $p_l \geq 0$ and $(\mathbf{A}\boldsymbol{\lambda})_l \leq c_l$, and $\sum_{l \in \mathcal{L}} p_l (c_l - (\mathbf{A}\boldsymbol{\lambda})_l) = 0$.

The result suggests that for a class $s \in \mathcal{S}$, the traffic intensity ρ_s and the number of links in \mathcal{L}_{crit}^U traversed by the class are the two critical factors that decide its contribution to accumulation in aggregate number of flows.

This large deviation characterization motivates an approach to capacity allocation. Here, we consider the problem of assigning link capacities c_l to maximize the LD exponent so that $\sum_{l \in \mathcal{L}} c_l \leq c_{tot}$ for some $c_{tot} > 0$. Thus, we are roughly minimizing the probability that a large number of flows accumulate. Since $L^U(\mathbf{x}^*) = \log(d_{UV}^*)$, we can use the definition of d_{UV}^* to get an equivalent linear optimization problem:

$$\min_{k, (c_l)_{l \in \mathcal{L}}} \left\{ k \mid \sum_{l \in \mathcal{L}} c_l = c_{tot}, k \sum_{s \in \mathcal{S}} A_{ls} \rho_s \geq c_l \text{ for all } l \in \mathcal{L} \right\}.$$

It can be shown (using the KKT optimality conditions) that the optimal allocation is given by $c_l^* = k^* \sum_{s \in \mathcal{S}} A_{ls} \rho_s \forall l \in \mathcal{L}$ where $k^* > 0$ is chosen so that $\sum_{l \in \mathcal{L}} c_l^* = c_{tot}$. The result suggests a simple rule of thumb: allocating bandwidth in proportion to the load being carried by the links minimizes the likelihood of network congestion.

5.2 Multipath networks

Here, we consider a more general setting where the classes can send their traffic through more than one route. Let

$$\Lambda^M = \left\{ \lambda \in \mathbb{R}_+^{|\mathcal{S}|} : \text{for some } \mathbf{f} \in \mathcal{F}, (\lambda, \mathbf{f}) \in \mathcal{C}^M \right\}.$$

As done above, we define the set $\mathcal{K}^M \subset \mathbb{R}^{|\mathcal{S}|}$ as follows:

$$\gamma = \{\gamma_s\}_{s \in \mathcal{S}} \in \mathcal{K}^M \Leftrightarrow \lambda = \{e^{\gamma_s}\}_{s \in \mathcal{S}} \in \Lambda^M.$$

Let the function $\delta_{\mathcal{K}^M}$ be given by

$$\delta_{\mathcal{K}^M}(\gamma) = \begin{cases} 0, & \gamma \in \mathcal{K}^M, \\ \infty, & \gamma \notin \mathcal{K}^M. \end{cases}$$

Let $\delta_{\mathcal{K}^M}^*$ denote the convex conjugate function of $\delta_{\mathcal{K}^M}$, i.e.,

$$\delta_{\mathcal{K}^M}^*(\mathbf{x}) = \sup_{\gamma \in \mathbb{R}^{|\mathcal{S}|}} (\langle \gamma, \mathbf{x} \rangle - \delta_{\mathcal{K}^M}(\gamma)).$$

An optimizer to the above problem corresponds to a solution $(\lambda_s^{PF}(\mathbf{x}))_{s \in \mathcal{S}}$ of the optimization problem MULTIPATH-PF given in Section 2.

Although (17) is proved in [18] for the case where a flow class uses only a single route, it can be shown that a similar result holds for the case where each flow class routes its traffic through multiple routes, i.e.,

$$\lim_{n \rightarrow \infty} \frac{\log(\pi_{BF}^M(n\mathbf{x}))}{n} = -L^M(\mathbf{x})$$

$$\text{where } L^M(\mathbf{x}) = \delta_{\mathcal{K}^M}^*(\mathbf{x}) - \sum_{s \in \mathcal{S}} x_s \log(\rho_s),$$

and π_{BF}^M denotes the steady state distribution of the multipath balanced fair allocation obtained in Section 3.2. Thus, the most likely mix \mathbf{x}^* of the flows of different classes when there is an accumulation in the aggregate number of flows is given by

$$\mathbf{x}^* = \operatorname{argmin}_{\{\mathbf{x}: x_s \geq 0 \forall s \in \mathcal{S} \text{ and } \sum_{s \in \mathcal{S}} x_s = 1\}} L^M(\mathbf{x}).$$

Next we consider the following linear optimization problem referred to as OPT-MAXMIN in the sequel:

$$\max_{d_M, \mathbf{f}} d_M \text{ such that} \\ \sum_{s \in \mathcal{S}} \sum_{r_s \in R_s} A_{lr_s} f_{sr_s} \rho_s \leq c_l \quad \forall l \in \mathcal{L}; \quad (20)$$

$$\sum_{r_s \in R_s} f_{sr_s} = d_M \quad \forall s \in \mathcal{S}; \quad (21)$$

$$f_{sr_s} \geq 0 \quad \forall r_s \in R_s, \forall s \in \mathcal{S}. \quad (22)$$

Let (d_M^*, \mathbf{f}') be a solution to the above problem, and let $(p_l')_{l \in \mathcal{L}}$, $(\alpha_s)_{s \in \mathcal{S}}$ and $(\beta_{sr_s})_{s \in \mathcal{S}, r_s \in R_s}$ be corresponding optimal Lagrange multipliers associated with (20), (21) and (22) respectively. Then, the next result gives the most likely direction for overflow. We skip a discussion of the proof as it is similar in flavor to that for the single path routing setting.

THEOREM 6. *For the multipath routing setting,*

$$x_s^* = b^M n_s^M \rho_s \quad \forall s \in \mathcal{S} \text{ and } L^M(\mathbf{x}^*) = \log(d_M^*)$$

$$\text{where } b^M = \frac{1}{\sum_{u \in \mathcal{S}} n_u \rho_u}, n_s^M = \sum_{l \in \mathcal{L}} p_l' \sum_{r_s \in R_s} A_{lr_s} f_{sr_s}' \text{ and}$$

d_M^* is the maximum value of the objective function in OPT-MAXMIN.

Let $\mathcal{L}_{crit}^M = \{l \in \mathcal{L} : p_l' > 0\}$ be the set of links that are critical to the exponent. Similar to the single path setting, for a class $s \in \mathcal{S}$, the traffic intensity ρ_s and the number of critical links traversed by the class are the two factors that decide its contribution to an accumulation in the aggregate number of flows. As the classes can split their traffic along multiple routes, there is more interdependence between the flow classes. Hence, we can expect more critical links and classes to contribute to the most likely mix than in the single path setting.

Like in the single path routing case, we use the large deviation behavior of balanced fair allocation to design networks in which the probability of overflows in the aggregate number of flows is less. We consider the problem of assigning link capacities c_l to maximize the LD exponent so that $\sum_{l \in \mathcal{L}} c_l \leq c_{tot}$ for some $c_{tot} > 0$. Since $L^M(\mathbf{x}^*) = \log(d_M^*)$ and d_M^* is obtained by optimizing OPT-MAXMIN, we can maximize $L^M(\mathbf{x}^*)$ by solving the following linear program:

$$\max_{d_M, (c_l)_{l \in \mathcal{L}}, \mathbf{f}} d_M \text{ such that} \\ \sum_{s \in \mathcal{S}} \sum_{r_s \in R_s} A_{lr_s} f_{sr_s} \rho_s \leq c_l \quad \forall l \in \mathcal{L}; \quad \sum_{r_s \in R_s} f_{sr_s} = d_M \quad \forall s \in \mathcal{S}; \\ f_{sr_s} \geq 0 \quad \forall r_s \in R_s, \forall s \in \mathcal{S}; \quad \sum_{l \in \mathcal{L}} c_l = c_{tot}.$$

5.3 Sensitivity of LD exponent

In this section, we study the sensitivity of the LD exponent to link capacities and the capacity of collections (pools) of links. This study of sensitivity will help us to identify the resources/pools of resources that are critical to the LD exponent and thus, to the accumulation of flows in the network. It is intuitive to expect that addition of capacity to certain links or resource pools should reduce the likelihood

of congestion. However, we show that the rate of increase in the exponent decreases with the capacity of associated link/resource pool. One can also show that the LD exponent is insensitive to the capacities of links $l \notin \mathcal{L}_{crit}^U$ for the single path routing setting and to the capacities of links $l \notin \mathcal{L}_{crit}^M$ in the multipath routing setting. We are now left to study the relationship between LD exponent and the remaining links.

First, we consider a simple but insightful example shown in Fig. 5. For all positive values of c and ρ_s for $s \in \mathcal{S}$, $L(\mathbf{x}^*) = \log(d_U^*) = \log\left(\frac{c}{\sum_{u \in \mathcal{S}} \rho_u}\right)$ is a differentiable function of c and ρ_s . Hence,

$$\frac{\partial L^U(\mathbf{x}^*)}{\partial c} = \frac{1}{c} \text{ and } \frac{\partial L^U(\mathbf{x}^*)}{\partial \rho_s} = -\frac{1}{\sum_{u \in \mathcal{S}} \rho_u}.$$

Thus the rate of decrease of $L^U(\mathbf{x}^*)$ with c decreases as we increase c . Thus if the link already has a large capacity, we require a large addition of capacity to obtain a significant increase in the LD exponent. From the above result, we can also infer that if the cumulative traffic intensity is high, we need a large decrease in the offered load to achieve a significant reduction in the LD exponent.

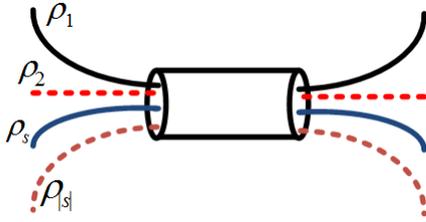


Figure 5: Example: A single link shared by $|\mathcal{S}|$ classes

A similar result holds for the single path routing setting when $|\mathcal{L}_{crit}^U| = 1$, i.e., there is only a single critical link. Let $\mathcal{L}_{crit}^U = \{l_B\}$. Then, using Theorem 5, we can show that $d_U^* = \log\left(\frac{c_{l_B}}{\sum_{s \in \mathcal{S}} A_{l_B} \rho_s}\right)$, and that we can partially differentiate d_U^* with respect to c_{l_B} at c_{l_B} to obtain

$$\frac{\partial L(\mathbf{x}^*)}{\partial c_{l_B}} = \frac{1}{c_{l_B}} \text{ and } \frac{\partial L(\mathbf{x}^*)}{\partial c_l} = 0 \forall l \in \mathcal{L} \setminus \{l_B\},$$

and that we can partially differentiate d_M^* with respect to ρ_s for $s \in \mathcal{S}$ at ρ_s to obtain

$$\frac{\partial L(\mathbf{x}^*)}{\partial \rho_s} = -\frac{A_{l_B}}{\sum_{u \in \mathcal{S}} A_{l_B} \rho_u}.$$

We close the discussion for the single path routing setting by pointing out that for the case in which $|\mathcal{L}_{crit}^U| > 1$, any small decrease in the capacity of a link $l \in \mathcal{L}_{crit}^U$ will result in a network with only l as the critical link.

5.3.1 Resource pools in the multipath routing setting

Next, let us consider the multipath flow control setting and obtain a collection of links $\mathcal{P}(\rho) \subset \mathcal{L}$ which behaves almost like a single pooled resource as far as LD exponents are concerned. This is a key feature associated with the notion of resource pooling discussed in [21]. As we will see, the pools can vary with changes in the offered load and hence,

the dependence on ρ . However, for notation simplicity, we use \mathcal{P} for $\mathcal{P}(\rho)$. In the sequel, we obtain $\mathcal{P} \subset \mathcal{L}$ such that

$$\frac{\partial L(\mathbf{x}^*)}{\partial c_l} = \frac{1}{\sum_{l' \in \mathcal{P}} c_{l'}} \forall l \in \mathcal{P}.$$

On comparing above expression with (23), we see that it is as if there is a single shared resource \mathcal{P} with capacity $c_{\mathcal{P}} = \sum_{l' \in \mathcal{P}} c_{l'}$ such that

$$\frac{\partial L(\mathbf{x}^*)}{\partial c_{\mathcal{P}}} = \frac{1}{c_{\mathcal{P}}}.$$

In the following, we refer to \mathcal{P} as the critical pool in the network.

Let $\mathcal{S}_{crit} = \{s \in \mathcal{S} : x_s^* > 0\}$. From Theorem 6, we have

$$\mathcal{S}_{crit} = \left\{ s \in \mathcal{S} : \rho_s > 0, \sum_{l \in \mathcal{L}} p_l \sum_{r_s \in R_s} A_{l r_s} f'_{s r_s} > 0 \right\}. \quad (23)$$

For a class $s \in \mathcal{S}_{crit}$, we define the set of links critical with respect to class s as

$$\mathcal{L}_{crit}(s) = \left\{ l \in \mathcal{L} : p_l \sum_{r_s \in R_s} A_{l r_s} f'_{s r_s} > 0 \right\}.$$

From (23), we see that $\mathcal{L}_{crit}(s)$ is non-empty for any $s \in \mathcal{S}_{crit}$. Further, we can show that $\cup_{s \in \mathcal{S}_{crit}} \mathcal{L}_{crit}(s) = \mathcal{L}_{crit}$. We define the Class Coupling Graph (CCG) as (\mathcal{S}_{crit}, E) where for $s_1, s_2 \in \mathcal{S}_{crit}$, $E_{s_1, s_2} = 1$ if $|\mathcal{L}_{crit}(s_1) \cap \mathcal{L}_{crit}(s_2)| > 0$, i.e., the classes s_1 and s_2 share a critical link.

To obtain the single resource pool, we make the following assumptions:

A1: The CCG is strongly connected.

A2: For any $s \in \mathcal{S}_{crit}$, $A_{l r_s} \in \{0, 1\} \forall l \in \mathcal{L}$ and $r_s \in R_s$.

A3: For any $s \in \mathcal{S}_{crit}$, $f'_{s, r_s} > 0 \forall r_s \in R_s$.

A4: For any $s \in \mathcal{S}_{crit}$ and any $r_s \in R_s$, $|r_s \cap \mathcal{L}_{crit}(s)| = 1$.

For $s \in \mathcal{S}_{crit}$ and $r_s \in R_s$, let $l^*(r_s)$ be the element in $r_s \cap \mathcal{L}_{crit}(s)$. The next result gives the critical pool. The proof is long and due to space constraints, we are not able to include it in the paper.

THEOREM 7. (Single critical pool: Sensitivity) Under Assumptions A1-A4,

$$x_s^* = \begin{cases} \frac{\rho_s}{\sum_{u \in \mathcal{S}_{crit}} \rho_u} & , \forall s \in \mathcal{S}_{crit}, \\ 0 & , \forall s \notin \mathcal{S}_{crit}, \end{cases} \quad (24)$$

$$\text{and } d_M^* = \frac{\sum_{l \in \mathcal{L}_{crit}} c_l}{\sum_{s \in \mathcal{S}_{crit}} \rho_s}. \quad (25)$$

Further, there is a collection of links

$\mathcal{P} = \{l^*(r_s) : r_s \in R_s \text{ for some } s \in \mathcal{S}_{crit}\}$ that satisfies

$$\frac{\partial L(\mathbf{x}^*)}{\partial c_l} = \frac{1}{\sum_{l' \in \mathcal{P}} c_{l'}} \forall l \in \mathcal{P} \text{ and } \frac{\partial L(\mathbf{x}^*)}{\partial c_l} = 0 \forall l \notin \mathcal{P}.$$

Assumptions A1 and A3 together ensure that for a small enough change in the capacity of a link in \mathcal{P} , traffic going through other links in \mathcal{P} can be shifted so that the effect of the change in capacity gets spread over all the links in \mathcal{P} . Consider the example in Fig. 6 where three multipath classes of traffic intensities 2, 1 and 1 bits per second are sharing the network. The capacities (measured in bits per second) of the links are indicated near the link in the figure.

Using Theorem 7, we can show that the shaded links form the critical resource pool. This example illustrates the usefulness of Theorem 7: intuition would only suggest that we need to allocate more capacities to links being used by class 1. However, the critical pool contains the starred link too which can be attributed to the coupling between flow classes 1 and 2.

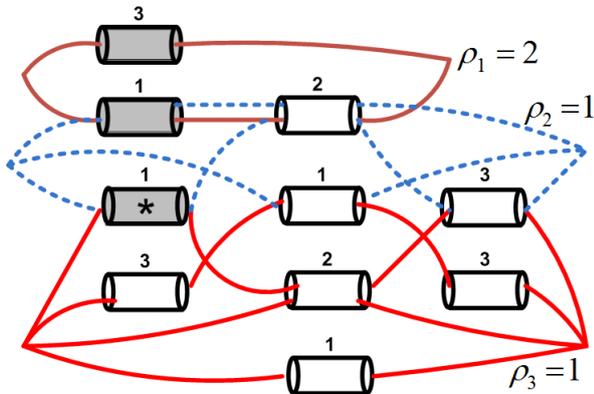


Figure 6: Example: $\mathbf{x}^* = [0.6667, 0.3333, 0]$ and the critical links are shaded.

6. CONCLUSIONS

We developed the first flow level performance bounds for networks supporting multipath flow control. Further, we studied large deviation for congestion events, and presented a possible approach to capacity allocation for such networks. Some practical implications of our work are listed below:

- (a) Theorem 3 and Section 4.3 suggest that when the spare capacity in the most congested resource pool of a multipath class is close to the access rate constraint of that class, gains from additional routes will be negligible.
- (b) Theorems 3 and 4 suggest that the resource pools in \mathcal{B}_t play a critical role in determining the performance of class t , and the spare capacity ($\approx c(\mathcal{H}) - \rho(\mathcal{H}) \approx c(\mathcal{H}) - \bar{\rho}(\mathcal{H})$) loosely captures the role of a pool $\mathcal{H} \in \mathcal{B}_t$.
- (c) Using Theorem 7, we can obtain the resource pool that is most critical in terms of resulting in an accumulation of the aggregate number flows in the network.
- (d) Section 5.2 provides a simple, intuitive capacity allocation scheme that minimizes the chances of an accumulation of the aggregate number of flows in the network. The scheme essentially allocates capacity in proportion to the load carried by links under an optimized multipath flow allocation.

This work provides initial steps towards tackling the many open questions (see Section 1) associated with possible adoption of multipath flow control in future networks. Addressing these challenging questions will be part of our future work.

7. ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation Award CNS-0917067. Prof. de Veciana would like to thank IMDEA Networks for hosting him when this work took birth, as well as Drs Banchs and Serrano for valuable discussions. We would also like to thank Dr. B. Godfrey for his comments which have served to improve this work.

8. REFERENCES

- [1] D. Bertsekas and R. Gallager. *Data networks*. Prentice-Hall, Inc., 1992.
- [2] T. Bonald. Throughput performance in networks with linear capacity constraints. In *Proc. of CISS 2006*, pages 1–6, 2006.
- [3] T. Bonald, L. Massoulié, A. Proutière, and J. Virtamo. A queueing analysis of max-min fairness, proportional fairness and balanced fairness. *Queueing Systems*, 53:65–84, 2006.
- [4] T. Bonald, A. Penttinen, and J. Virtamo. On light and heavy traffic approximations of balanced fairness. In *Proceedings of SIGMETRICS/Performance 2006*, pages 109–120, June 2006.
- [5] T. Bonald and A. Proutière. Insensitive bandwidth sharing in data networks. *Queueing Systems*, 44:69–100, 2003.
- [6] T. Bonald and J. Virtamo. Calculating the flow level performance of balanced fairness in tree networks. *Performance Evaluation*, 58:1–14, 2004.
- [7] H. Han, S. Shakkottai, C. V. Hollot, R. Srikant, and D. Towsley. Multi-path TCP: a joint congestion control and routing scheme to exploit path diversity in the internet. *IEEE/ACM Trans. Netw.*, 14:1260–1271, December 2006.
- [8] W. N. Kang, F. P. Kelly, N. H. Lee, and R. J. Williams. State space collapse and diffusion approximation for a network operating under a fair bandwidth sharing policy. *Annals of Applied Probability*, 19:1719–1780, 2009.
- [9] F. Kelly and T. Voice. Stability of end-to-end algorithms for joint routing and rate control. *Queueing Systems*, 35:5–12, 2005.
- [10] F. P. Kelly, L. Massoulié, and N. S. Walton. Resource pooling in congested networks: proportional fairness and product form. *Queueing Syst. Theory Appl.*, 63:165–194, December 2009.
- [11] F. P. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *Journal of Operation Research*, 49(3):237–252, 1998.
- [12] P. Key, L. Massoulié, and D. Towsley. Multipath routing, congestion control and dynamic load balancing. In *Proc. ICASSP*, pages 109–120, April 2007.
- [13] P. Key, L. Massoulié, and D. Towsley. Path selection and multipath congestion control. In *Proc. IEEE INFOCOM*, pages 109–120, May 2007.
- [14] C. N. Laws. Resource pooling in queueing networks with dynamic routing. *Advances in Applied Probability*, 24:699–726, 1992.
- [15] J. Leino and J. Virtamo. Insensitive traffic splitting in data networks. In *Proceedings of ITC*, pages 1355–1364, Beijing, China, 2005.
- [16] J. Leino and J. Virtamo. Insensitive load balancing in data networks. *Comput. Netw.*, 50:1059–1068, June 2006.
- [17] M. Mandjes and B. Zwart. Large deviations of sojourn times in processor sharing queues. *Queueing Syst. Theory Appl.*, 52(4):237–250, 2006.
- [18] L. Massoulié. Structural properties of proportional fairness: Stability and insensitivity. *Annals of Applied Probability*, 2007.
- [19] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8:556–567, October 2000.
- [20] W. H. Wang, M. Palaniswami, and S. H. Low. Optimal flow control and routing in multi-path networks. *Perform. Eval.*, 52:119–132, April 2003.
- [21] D. Wischik, M. Handley, and M. B. Braun. The resource pooling principle. *SIGCOMM Comput. Commun. Rev.*, 38:47–52, September 2008.
- [22] Y. Yi and M. Chiang. Stochastic network utility maximization. *European Transactions on Telecommunications*, 19:421–442, June 2008.

APPENDIX

A. STABILITY REGION FOR MULTIPATH BALANCED FAIR ALLOCATION

We prove the result using an approach similar to that in [3]. We start by identifying a special property satisfied by the balance function associated with multipath balanced fair allocations which also establishes its uniqueness.

LEMMA 1. *For any balance function $\tilde{\Phi} : \mathbb{Z}^{|\mathcal{S}|} \rightarrow [0, \infty)$ with $\tilde{\Phi}(\mathbf{0}) = 1$, and splitting function $\mathbf{f} : \mathbb{Z}^{|\mathcal{S}|} \rightarrow \mathcal{F}$, if the corresponding insensitive allocation satisfies (1) and (2), then*

$$\Phi(\mathbf{x}) \leq \tilde{\Phi}(\mathbf{x}), \forall \mathbf{x} \in \mathbb{Z}^{|\mathcal{S}|}.$$

PROOF. We prove the result using induction on the total number of flows $|\mathbf{x}|$ in the network state \mathbf{x} . Clearly the result holds for $\mathbf{x} = \mathbf{0}$. Suppose that the result is true for network states \mathbf{x} such that $|\mathbf{x}| = n - 1$. For network states \mathbf{x} such that $|\mathbf{x}| = n$, from (11), we have

$$\Phi(\mathbf{x}) \leq \max \left\{ \max_{s \in \mathcal{S}: x_s > 0} \frac{\Phi(\mathbf{x} - \mathbf{e}_s)}{a_s x_s}, \max_{l \in \mathcal{L}} \left(\sum_{s \in \mathcal{S}, r_s \in R_s} \frac{A_{l,r_s}}{c_l} f_{s,r_s}(\mathbf{x}) \Phi(\mathbf{x} - \mathbf{e}_s) \right) \right\}.$$

Using the induction assumption,

$$\Phi(\mathbf{x}) \leq \max \left\{ \max_{s \in \mathcal{S}: x_s > 0} \frac{\tilde{\Phi}(\mathbf{x} - \mathbf{e}_s)}{a_s x_s}, \max_{l \in \mathcal{L}} \left(\sum_{s \in \mathcal{S}, r_s \in R_s} \frac{A_{l,r_s}}{c_l} f_{s,r_s}(\mathbf{x}) \tilde{\Phi}(\mathbf{x} - \mathbf{e}_s) \right) \right\}.$$

Then,

$$\Phi(\mathbf{x}) \leq \tilde{\Phi}(\mathbf{x}) \max \left\{ \max_{s \in \mathcal{S}: x_s > 0} \frac{\tilde{\phi}_s(\mathbf{x})}{a_s x_s}, \max_{l \in \mathcal{L}} \left(\sum_{s \in \mathcal{S}, r_s \in R_s} \frac{A_{l,r_s}}{c_l} f_{s,r_s}(\mathbf{x}) \tilde{\phi}_s(\mathbf{x}) \right) \right\}.$$

Since, the insensitive allocation corresponding to $\tilde{\Phi}(\mathbf{x})$ satisfies (1) and (2), we conclude that

$$\Phi(\mathbf{x}) \leq \tilde{\Phi}(\mathbf{x}).$$

□

Now, we complete the proof of Theorem 2 using steps similar to that in [3]. Suppose $\boldsymbol{\rho} \in \Gamma^M$. Then, $\boldsymbol{\rho} \in \Gamma(\mathbf{f}')$ for some $\mathbf{f}' \in \mathcal{F}$, and $(1 + \epsilon)\boldsymbol{\rho} \in \Gamma(\mathbf{f}')$ for some $\epsilon > 0$. Consider the insensitive allocation corresponding to the splitting function \mathbf{f}' and the balance function

$$\tilde{\Phi}(\mathbf{x}) = \prod_{s \in \mathcal{S}} \frac{1}{((1 + \epsilon)\rho_s)^{x_s}} \quad \forall \mathbf{x} \in \mathbb{Z}^{|\mathcal{S}|}.$$

Since $\tilde{\Phi}_s(\mathbf{x}) = (1 + \epsilon)\rho_s \quad \forall \mathbf{x} \in \mathbb{Z}^{|\mathcal{S}|}$ and $(1 + \epsilon)\boldsymbol{\rho} \in \Gamma(\mathbf{f}')$, the allocation is feasible. Hence, the stability condition (4) is satisfied since (using Lemma 1)

$$\sum_{\mathbf{x} \in \mathbb{Z}_+^{|\mathcal{S}|}} \Phi(\mathbf{x}) \boldsymbol{\rho}^{\mathbf{x}} \leq \sum_{\mathbf{x} \in \mathbb{Z}_+^{|\mathcal{S}|}} \tilde{\Phi}(\mathbf{x}) \boldsymbol{\rho}^{\mathbf{x}} = \sum_{\mathbf{x} \in \mathbb{Z}_+^{|\mathcal{S}|}} \prod_{s \in \mathcal{S}} \frac{1}{((1 + \epsilon)\rho_s)^{x_s}} < \infty.$$

B. A SKETCH OF THE PROOF OF THE UPPER BOUND

A sketch of the proof of Theorem 4 is given below. We use the following intermediate results to prove the Theorem.

The first result roughly states that if we ignore the peak rate constraints and assume $\bar{n}_t(\mathcal{H}) = 1$, then we can treat class t as if it is traversing the pooled resources in \mathcal{D}_t instead of the links in the set $\{l : A_{l,r_t} > 0 \text{ for some } r_t\}$.

LEMMA 2. *For any class $t \in \mathcal{S}$,*

$$\min_{\mathbf{f} \in \mathcal{F}} \max_{l \in \mathcal{L}} \left(\sum_{s \in \mathcal{S}} \sum_{r_s \in R_s} \frac{A_{l,r_s}}{c_l} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) \right) \leq \min_{\mathbf{f} \in \mathcal{F}} \max \left(\max_{l \in \mathcal{L}} \sum_{s \in \mathcal{S}, s \neq t} \sum_{r_s \in R_s} \frac{A_{l,r_s}}{c_l} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s), \max_{\mathcal{H} \in \mathcal{D}_t} \left(\frac{\sum_{l \in \mathcal{H}, s \neq t, r_s \in R_s} A_{l,r_s} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) + \bar{n}_t(\mathcal{H}) \Phi(\mathbf{x} - \mathbf{e}_t)}{c(\mathcal{H})} \right) \right).$$

PROOF. Let ψ^* be the optimal value of the optimization problem in (12). Thus, we can show that ψ^* is also the optimal value of an equivalent optimization problem:

min ψ such that

$$\begin{aligned} \sum_{s, r_s \in R_s} A_{l,r_s} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) &\leq \psi c_l \quad \forall l \in \mathcal{L} \setminus \bigcup_{H \in \mathcal{D}_t} H; \\ \sum_{s, r_s \in R_s} A_{l,r_s} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) &\leq \psi c_l \quad \forall l \in \bigcup_{H \in \mathcal{D}_t} H; \\ \sum_{l \in H, s \neq t, r_s \in R_s} A_{l,r_s} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) + \\ \sum_{r_t \in R_t} \sum_{l \in H} A_{l,r_t} f_{t,r_t} \Phi(\mathbf{x} - \mathbf{e}_t) &\leq \psi C(H) \quad \forall H \in \mathcal{D}_t; \\ \mathbf{f}_s &\in \mathcal{F}_s \quad \forall s \in \mathcal{S}. \end{aligned}$$

We tighten some of the constraints to get an optimization problem with optimal ψ_1^* : **OPT-1**

min ψ such that

$$\begin{aligned} \sum_{s, r_s \in R_s} A_{l,r_s} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) &\leq \psi c_l \quad \forall l \in \mathcal{L} \setminus \bigcup_{H \in \mathcal{D}_t} H; \\ \sum_{s, r_s \in R_s} A_{l,r_s} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) &\leq \psi c_l \quad \forall l \in \bigcup_{H \in \mathcal{D}_t} H; \\ \sum_{l \in H, s \neq t, r_s \in R_s} A_{l,r_s} f_{s,r_s} \Phi(\mathbf{x} - \mathbf{e}_s) + \\ \bar{n}_t(H) \Phi(\mathbf{x} - \mathbf{e}_t) &\leq \psi C(H) \quad \forall H \in \mathcal{D}_t; \\ \mathbf{f}_s &\in \mathcal{F}_s \quad \forall s \in \mathcal{S}. \end{aligned}$$

Clearly, $\psi^* \leq \psi_1^*$. Now, we loosen some constraints from above optimization problem to obtain the following problem

with optimal ψ_2^* : **OPT-2**

min ψ such that

$$\begin{aligned} \sum_{s, r_s \in R_s} A_{l, r_s} f_{s, r_s} \Phi(\mathbf{x} - \mathbf{e}_s) &\leq \psi c_l \quad \forall l \in \mathcal{L} \setminus \bigcup_{H \in \mathcal{D}_t} H; \\ \sum_{s \neq t, r_s \in R_s} A_{l, r_s} f_{s, r_s} \Phi(\mathbf{x} - \mathbf{e}_s) &\leq \psi c_l \quad \forall l \in \bigcup_{H \in \mathcal{D}_t} H; \\ \sum_{l \in H, s \neq t, r_s \in R_s} A_{l, r_s} f_{s, r_s} \Phi(\mathbf{x} - \mathbf{e}_s) &+ \\ &+ \bar{n}_t(H) \Phi(\mathbf{x} - \mathbf{e}_t) \leq \psi C(H) \quad \forall H \in \mathcal{D}_t; \\ \mathbf{f}_s \in \mathcal{F}_s \quad \forall s \in \mathcal{S} \setminus \{t\}. \end{aligned}$$

Clearly, $\psi_2^* \leq \psi_1^*$. Next, we show that $\psi_2^* = \psi_1^*$. Let $\mathbf{f}^* \in \prod_{s \in \mathcal{S} \setminus \{t\}} \mathcal{F}_s$ be the optimal fraction vector for OPT-2. Next, we obtain a $\mathbf{f}' \in \mathcal{F}$ using \mathbf{f}^* and ψ_2^* which satisfies all constraints in OPT-1. Let

$$d_l = \left(\psi_2^* c_l - \sum_{s \neq t, r_s \in R_s} A_{l, r_s} f_{s, r_s}^* \Phi(\mathbf{x} - \mathbf{e}_s) \right).$$

Consider the following optimization problem: **OPT-2.1**

$$\begin{aligned} \max \sum_{r_t \in R_t} p_{r_t} \quad \text{such that} \\ \sum_{r_t \in R_t} A_{l, r_t} p_{r_t} \leq \frac{d_l}{\Phi(\mathbf{x} - \mathbf{e}_t)} \quad \forall l \in \bigcup_{H \in \mathcal{D}_t} H; \quad (26) \\ p_{r_t} \geq 0 \quad \forall r_t \in R_t. \end{aligned}$$

Let \mathbf{p}^* be the optimal solution to **OPT-2.1**. In the solution to the above optimization problem, for each $r_t \in R_t$, there is a link $l_{r_t} \in r_t \cap \bigcup_{H \in \mathcal{D}_t} H$ such that the constraint corresponding to the link l_{r_t} in (26) is satisfied with equality (otherwise, the variable p_{r_t} can be increased to strictly improve on the optimal solution). Now, let $\tilde{\mathcal{H}}$ be the set of all such links. Then, from the above construction, $\tilde{\mathcal{H}} \in \mathcal{B}_t$. Let $\mathcal{H}^* \in \mathcal{D}_t$ such that $\mathcal{H}^* \subseteq \tilde{\mathcal{H}}$. Further, since the links in \mathcal{H}^* also satisfy (26) with equality,

$$\begin{aligned} \sum_{l \in \mathcal{H}^*} \sum_{r_t \in R_t} A_{l, r_t} p_{r_t}^* \Phi(\mathbf{x} - \mathbf{e}_t) &= \sum_{l \in \mathcal{H}^*} d_l \\ &= \sum_{l \in \mathcal{H}^*} \left(\psi_2^* c_l - \sum_{s \neq t, r_s \in R_s} A_{l, r_s} f_{s, r_s}^* \Phi(\mathbf{x} - \mathbf{e}_s) \right) \\ &\geq \bar{n}_t(\mathcal{H}^*) \Phi(\mathbf{x} - \mathbf{e}_t) \end{aligned}$$

Also,

$$\sum_{l \in \mathcal{H}^*} \sum_{r_t \in R_t} A_{l, r_t} p_{r_t}^* \Phi(\mathbf{x} - \mathbf{e}_t) \leq \bar{n}_t(\mathcal{H}^*) \Phi(\mathbf{x} - \mathbf{e}_t) \sum_{r_t \in R_t} p_{r_t}^*.$$

Hence, $\sum_{r_t \in R_t} p_{r_t}^* \geq 1$. Now, let $\mathbf{f}'_s = \mathbf{f}_s^* \quad \forall s \neq t$, and let $f'_{t, r_t} = \frac{p_{r_t}^*}{\sum_{r_t \in R_t} p_{r_t}^*} \quad \forall r_t \in R_t$.

Since \mathbf{p}^* is the optimal solution to **OPT-2.1**, $f'_{t, r_t} = \frac{p_{r_t}^*}{\sum_{r_t \in R_t} p_{r_t}^*}$ and $\sum_{r_t \in R_t} p_{r_t}^* \geq 1$, we have that

$$\sum_{r_t \in R_t} A_{l, r_t} f'_{t, r_t} \leq \sum_{r_t \in R_t} A_{l, r_t} p_{r_t}^* \leq \frac{d_l}{\Phi(\mathbf{x} - \mathbf{e}_t)} \quad \forall l \in \bigcup_{H \in \mathcal{D}_t} H.$$

Also, from (26) and $\psi_2^* \leq \psi_1^*$,

$$\begin{aligned} \sum_{r_t \in R_t} A_{l, r_t} f'_{t, r_t} \Phi(\mathbf{x} - \mathbf{e}_t) \\ \leq \psi_1^* c_l - \sum_{s \neq t, r_s \in R_s} A_{l, r_s} f'_{s, r_s} \Phi(\mathbf{x} - \mathbf{e}_s) \quad \forall l \in \bigcup_{H \in \mathcal{D}_t} H. \end{aligned}$$

Thus, all constraints removed when obtaining **OPT-2** from **OPT-1** still hold. Hence, $\psi_2^* \geq \psi_1^*$. Thus, $\psi^* \leq \psi_2^*$, and from the formulation **OPT-2**, the result follows. \square

The next result is mainly useful for a class $u \in \mathcal{S}$ with $|\mathcal{D}_u| = 1$ in which case it lower bounds the allocated bandwidth going through the resource pool in \mathcal{D}_u . It can be proved using induction on $|\mathbf{x}|$ and Lemma 2.

LEMMA 3. For any class $u \in \mathcal{S}$ with no peak rate constraints, i.e., $a_u = \infty$, and state \mathbf{x} where $x_u > 0$, we have

$$\Phi(\mathbf{x}) \leq \sum_{H \in \mathcal{D}_u} \frac{1}{c(H)} \sum_{s \in \mathcal{P}(H)} \bar{n}_s(H) \Phi(\mathbf{x} - \mathbf{e}_s).$$

The next result can be proved using Lemma 3 and some steps similar to that in the mean per bit delay lower bound.

LEMMA 4. Consider a class $u \in \mathcal{S}$ with $a_u = \infty$ and $|\mathcal{D}_u| = 1$. If $\bar{\rho}(H) < c(H)$, then

$$\tau_u \leq \frac{\bar{n}_u(H)}{c(H) - \bar{\rho}(H)}.$$

We follow the approach in [2] and modify the original network by introducing new classes s_0 and $s_H \quad \forall H \in \mathcal{D}_t$ referred to as virtual classes. Virtual class s_0 has the same routes as class t and has the same peak rate constraints, i.e., $R_0 = R_t$ and $a_0 = a_t$. A virtual class s_H uses routes R_H where $R_H = \{\{l\} : l \in H\}$ and has no peak rate constraint, i.e., $a_H = \infty$. Let $\tilde{\rho}$ be the vector denoting the traffic intensity associated with the virtual classes, i.e., $\tilde{\rho}_H$ is the traffic intensity of s_H and $\tilde{\rho}_0$ is the traffic intensity of s_0 . Further, let $\tilde{\Phi}$ denote the balance function associated with the new network. Then, we have the following result which can be proved using Lemma 3 and induction on $|\mathbf{x}|$.

LEMMA 5. For any class $t \in \mathcal{S}$ and state \mathbf{x} ,

$$\begin{aligned} \tilde{\Phi}(\mathbf{x}, \mathbf{e}_0) &\leq b_t \tilde{\Phi}(\mathbf{x}, \mathbf{0}) + \\ &\sum_{H \in \mathcal{D}_t} \frac{\bar{n}_t(H)}{c(H)} \sum_{s \in \mathcal{P}(H)} \bar{n}_s(H) \tilde{\Phi}(\mathbf{x} - \mathbf{e}_s, \mathbf{e}_H). \end{aligned}$$

The upper bound given in Theorem 4 mainly follows from Lemma 5 by applying Lemma 4 (on each virtual class s_H , $H \in \mathcal{D}_t$), letting $\tilde{\rho} \rightarrow 0$ noting that

$$\tau_t = \lim_{\tilde{\rho} \rightarrow 0} \tilde{\tau}_0.$$

and following some steps similar to that in the lower bound. Since we let $\tilde{\rho} \rightarrow 0$, the performance of the flows in the modified network is same as that in the original network.

In proving Lemmas 3 and 5, additional work is required to deal with the dependency of the splitting function on the network state.