Traffic Engineering with AIMD in MPLS Networks *

Jianping Wang[1], Stephen Patek[2], Haiyong Wang[1], and Jörg Liebeherr[1]

[1] Department of Computer Science
[2] Department of Systems and Information Engineering,
University of Virginia, Charlottesville, VA 22904, U.S.A.
{jgwang@cs,hw6h@cs,jorg@cs,patek}@virginia.edu

Abstract. We consider the problem of allocating bandwidth to competing flows in an MPLS network, subject to constraints on fairness, efficiency, and administrative complexity. The aggregate traffic between a source and a destination, called a flow, is mapped to label switched paths (LSPs) across the network. Each flow is assigned a preferred ('primary') LSP, but traffic may be sent to other ('secondary') LSPs. Within this context, we define objectives for traffic engineering, such as fairness, efficiency, and preferred flow assignment to the primary LSP of a flow ('Primary Path First', PPF). We propose a distributed, feedback-based multipath routing algorithm that attempts to apply additive-increase and multiplicative-decrease (AIMD) to implement our traffic engineering objectives. The new algorithm is referred to as multipath-AIMD. We use *ns-2* simulations to illustrate the fairness criteria and PPF property of our multipath-AIMD scheme in an MPLS network.

1 Introduction

Multiprotocol Label Switching (MPLS) [20] has offered new opportunities for improving Internet services through traffic engineering. An important aspect of traffic engineering, defined as "that aspect of Internet network engineering dealing with the issue of performance evaluation and performance optimization of operational IP networks" [4], is the allocation of network resources to satisfy an aggregate measure of the demand for services and to obtain better network utilization. MPLS, in conjunction with path establishment protocols such as CR-LDP [2] or RSVP-TE [8], makes it possible for network engineers to set up dedicated label switched paths (LSPs) with reserved bandwidth for the purpose of optimally distributing traffic across a given network.

Figure 1 illustrates an MPLS network, where all traffic across the network is accounted for by a set of source/destination pairs, called flows, and multiple LSPs are in place for accommodating the demand for service. We consider a multipath routing scenario where sources can make use of multiple LSPs. For each source, one LSP is assigned as the primary path, and other LSPs can be used as secondary paths. We consider sets of sources where each source can use all primary paths of the other sources

* This work is supported in part by the National Science Foundation through grants ANI-9730103, ECS-9875688 (CAREER), ANI-9903001, DMS-9971493, and ANI-0085955.

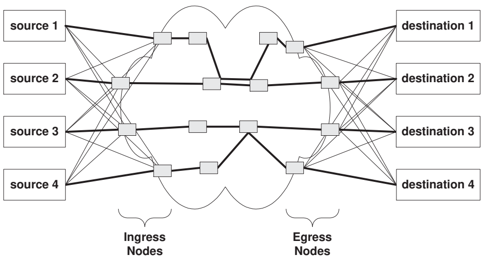

Fig. 1. An example of an MPLS network. The primary paths are indicated as thick lines.

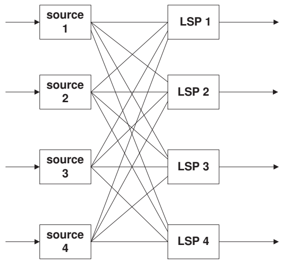

Fig. 2. The simplified network model. LSP i is the primary path for source i.

as its secondary paths. While Figure 1 presents a general view of an MPLS network, we concentrate in this paper on a simplified model, as illustrated in Figure 2. To begin, we assume that there are N sources and N LSPs, each serving as the primary path associated with exactly one source. In this context, the traffic engineering problem is the assignment of traffic of a flow to the primary path and the secondary paths, in such a way that a given set of traffic engineering objectives is satisfied. While a centralized solution to the given problem is quite straightforward, we strive to find distributed mechanisms for traffic engineering without central control. Specifically, we investigate if and to which degree binary feedback schemes and rate control schemes, such as, additive increase/multiplicative decrease (AIMD) [5, 6, 10–12, 19], can be used to achieve traffic engineering objectives.

Recently, considerable effort has been invested into scalable mechanisms for providing differentiated services in the Internet. For example, in [7], Elwalid et al. presented a multi-path adaptive traffic engineering mechanism, called MATE, designed for MPLS networks where several explicit LSPs have been established between ingress and egress nodes. MATE is intended to work for traffic that does not require bandwidth reservation and seeks to distribute traffic across the LSPs by dynamically adjusting the rate on each ingress node. The relationships between end-to-end congestion control and fairness are established in [9, 14, 15, 17, 18, 21]. The network model adopted in these papers supposes that all sources are greedy and each source sends traffic through a single path or a dedicated set of paths, and fairness is characterized by means of a social welfare-type optimization objective. These models generally give rise to differential equations that characterize the behavior of AIMD and AIPD[3] congestion control schemes.

Different from most of the related work on fairness and binary feedback, we consider sources as being either greedy or non-greedy. A source is greedy if it always has traffic backlogged. A non-greedy source has an upper bound on the desired sending rate. Moreover, we model LSPs as being either pooled or owned. In the owned case, each LSP gives high priority to the source for which the path is primary, and any remaining

[3] additive-increase / proportional-decrease

capacity is available to be shared equally among the remaining sources. In the pooled case, primary paths do not give priority to their respective sources.

In this paper, we propose congestion control mechanisms for dynamically adjusting the rates of all sources. In contrast to [9, 14, 15, 17, 18, 21], in an effort to minimize overhead and processing complexity, we only employ binary feedback mechanisms in adjusting flow rates. The feedback mechanisms we have developed seek to address the following traffic engineering characteristics.

• **Efficiency.** An allocation of network resources is said to be efficient if either all resources are completely consumed while the network is overloaded or all sources are completely satisfied while the network is underloaded.

• **Fairness.** The appropriate notion of fairness for MPLS traffic engineering varies with the nature of sources (greedy vs. non-greedy) and LSPs (pooled vs. owned). In Section 2, we formally define the corresponding fairness criteria for each case, based on the notion of fair-share resource allocation in [3]. With N greedy sources and N pooled resources, it is easy to prove the equivalence between our definition and the minimum potential delay fairness from [16, 18]. However, our fairness definitions are constructive, providing an easier way to characterize and achieve fair allocations than through the solution of nonlinear optimization models.

• **Primary Path First Property.** While routing along multiple paths is an opportunity we seek to exploit, there are drawbacks associated with multi-path routing, such as overhead associated with label distribution, additional state information, classification, and potential out-of-sequence delivery. To address these issues, we formulate a traffic engineering objective that seeks to minimize the amount of traffic sent over secondary paths. We introduce a novel criteria for network performance, called the Primary Path First (PPF) property. Generally speaking, PPF refers to the desire to have each source exploit available capacity on secondary paths, but to refrain from using the secondary paths whenever possible. We make this notion precise in Section 3. The PPF property reflects that, given multiple feasible rate assignments to primary and secondary paths that satisfy fairness and efficiency criteria, the preferred assignment is the one that sends the most traffic of a source on the primary path of that source.

• **Distributed Implementation.** We seek to provide efficient, fair, and PPF allocations of network resources using simple distributed algorithms. Distributed mechanisms, which operate mainly on local state information, are preferred as they minimize network overhead and retain scalability. In Section 4, we describe a distributed scheme which allocates flow to resources in a fashion reminiscent of AIMD in [13]. In some cases, global information and coordination are required for specific goals such as PPF.

• **Stability and Convergence.** Traffic engineering mechanisms, such as those we seek to develop in this paper, often suffer from potential instability and oscillations within the network. We seek to prevent this type of behavior by requiring incremental adjustments to the flow allocations specified by our algorithms. However, in this paper we do not offer a formal proof of stability or convergence properties.

This paper makes two contributions. First, we introduce notions of fairness for MPLS traffic engineering, and we show how the AIMD algorithm of [13] can be extended, in a distributed fashion, to achieve fair allocations of network resources. We call the enhanced AIMD algorithm multipath-AIMD. Next, we introduce the PPF criterion which seeks to limit the administrative complexity associated with multipath-routing, by concentrating traffic on a designated LSP.

The reminder of this paper is organized as follows. Section 2 formally defines notions of fairness for both pooled and owned resources. Section 3 defines and analyzes the PPF criterion. Section 4 presents AIMD algorithms which experimentally converge to fair allocations of network capacity and suggests modifications to the AIMD schemes for achieving a PPF assignment of resources. Section 5 presents *ns-2* simulation results, and Section 6 concludes the paper.

2 Fairness Criteria

We consider a network of LSPs which correspond to the simplified model of Figure 2. In this model, there are N traffic sources and N LSPs.

At any time, each source $i = 1, \ldots, N$ has a load of λ_i ($\lambda_i \geq 0$) , which is the maximum desired sending rate of the source. If the traffic demand from source i is $\lambda_i = \infty$, we say that the source is *greedy*.[4] Each LSP $i = 1, \ldots, N$ has a maximum transmission capacity of B_i. LSP i is the primary path associated with source i, and all other LSPs are secondary paths with respect to source i. We use γ_i to denote the actual allocation of bandwidth to source i. The rate allocation consists of the allocation on the primary path and the secondary paths.

We distinguish two different allocation schemes for assigning bandwidth on the LSPs to sources:

- **Owned Resources:** Each source may consume the entire capacity of its primary path, i.e., B_i, and, in addition, it can obtain unused bandwidth on its secondary paths.
- **Pooled Resources:** The aggregate capacity on all LSPs, i.e., $\sum_{i=1}^{N} B_i$, is distributed across all sources, without regard to the capacity on primary paths.

The fairness criteria for networks with owned and pooled resources are specified in the following definition.

Definition 1. *Given a network as shown in Figure 2 with N sources and LSPs. Let B_i denote the capacity of LSP i ($1 \leq i \leq N$), and let $\lambda_i \geq 0$ and $\gamma_i \geq 0$ denote the load and the rate allocation of source i.*

1. *A rate allocation is a relation R = (λ_i, γ_i), $1 \leq i \leq N$ such that both $\gamma_i \leq \lambda_i$ and $0 \leq \sum_{i=1}^{N} \gamma_i \leq \sum_{i=1}^{N} B_i$.*

[4] Note that the values of λ_i vary with time; However, we do not carry the dependence on time in our notation, i.e., by writing '$\lambda_i(t)$'.

2. *A rate allocation is efficient if the following hold:*
   a) If $\sum_{i=1}^{N} \lambda_i < \sum_{i=1}^{N} B_i$*, then* $\sum_{i=1}^{N} \gamma_i = \sum_{i=1}^{N} \lambda_i$,
   b) If $\sum_{i=1}^{N} \lambda_i \geq \sum_{i=1}^{N} B_i$*, then* $\sum_{i=1}^{N} \gamma_i = \sum_{i=1}^{N} B_i$.
   If case b) holds, we say that the rate allocation is saturating.
3. *A rate allocation for pooled resources is fair if there exists a value* $\alpha^p > 0$ *such that for each source* i *it holds that* $\gamma_i = \min\{\lambda_i,\ \alpha^p\}$.
4. *A rate allocation for owned resources is fair if there exists a value* $\alpha^o > 0$ *such that for each source* i *it holds that* $\gamma_i = \min\{\lambda_i,\ B_i + \alpha^o\}$.

According to this definition, a rate allocation is fair if sources with low bandwidth requirements are fully satisfied while sources with high requirements obtain a fair share of the capacity according to the given fairness criteria. With pooled resources, the fair allocation to a given source depends only on the rate requirement of the source and the total capacity of all resources. With owned resources, the fair rate allocation takes into consideration the capacity B_i on the primary path of source i.

As we will discuss below, rate allocations for a network with pooled and owned resources that are fair and satisfy the respective fairness criteria are uniquely defined (with respect to the values of γ_i). Further, assuming knowledge of the load of all sources and the bandwidth of all LSPs, the rate allocations can be effectively constructed. Later in this paper, we attempt to achieve the desired rate allocations in a distributed fashion via a feedback loop, and without explicit knowledge of the traffic load of the sources.

2.1 Pooled Resources

The rate allocation distributes the aggregate capacity from all LSPs across all sources, regardless of the available capacity on the primary path of a source. Hence, the aggregate capacity on all LSPs can be thought of as a single pool of resources. We refer to α^p as the fair share of this rate allocation. The fair share α^p in a network with pooled resources is given by

$$\alpha^p = \begin{cases} \frac{\sum_{i=1}^{N} B_i - \sum_{j \in U} \lambda_j}{|O|} & \text{if } O \neq \emptyset \\ \infty & \text{otherwise}, \end{cases} \tag{1}$$

where

$$U = \{j \mid \lambda_j < \alpha^p\} \qquad \text{and} \qquad O = \{j \mid \lambda_j \geq \alpha^p\}. \tag{2}$$

One can think of U as the set of 'underloaded' sources that can satisfy their bandwidth demands, and of O as the set of 'overloaded' sources. Then, the fair rate allocation is obtained by subtracting the bandwidth demand from underloaded sources, and then dividing the remainder by the number of overloaded sources.

If the total demand is less than the total available capacity, i.e., $\sum_{i=1}^{N} \lambda_i \leq \sum_{i=1}^{N} B_i$, then all sources are underloaded and $\alpha^p = \infty$. In the special case where all sources are greedy, i.e. $\lambda_i = \infty$ for all $i = 1, \ldots, N$, we have $U = \emptyset$ and $\gamma_i = \alpha^p = \sum_{i=1}^{N} B_i / N$ for all i.

Efficiency of this rate allocation can be verified by inspection. A proof of the efficiency and the uniqueness of this rate allocation is given in [3], which specifies a rate allocation for a shared bus metropolitan area network.

We can construct α^p as follows. Assume, without loss of generality, that the sources are ordered according to the generated load, that is, $\lambda_i \leq \lambda_j$ for $i < j$. Select $\hat{k}$ has the largest index k ($1 \leq k \leq N$) which satisfies

$$\lambda_k \leq \frac{\sum_{l=1}^{N} B_l - \sum_{l=1}^{k} \lambda_l}{N-k} \ . \tag{3}$$

Then, we can determine the fair share α^p by

$$\alpha^p = \begin{cases} \frac{\sum_{l=1}^{N} B_l - \sum_{l=1}^{\hat{k}} \lambda_l}{N-\hat{k}} & \text{if } \hat{k} < N \\ \infty & \text{otherwise} \ . \end{cases} \tag{4}$$

2.2 Owned Resources

Here, each source may consume all of the capacity on its primary path and, in addition, a fair share of the remaining unused capacity on all secondary paths. Hence, since each source can always consume all the resources on its primary path, the capacity of the primary path can be thought of as being 'owned' by the source. For the fairness definition, we distinguish between flows that use the entire bandwidth on the primary path and those that do not:

$$\tilde{U} = \{j \mid \lambda_j < B_j\} \qquad \text{and} \qquad \tilde{O} = \{j \mid \lambda_j \geq B_j\}. \tag{5}$$

Thus, the total surplus capacity, which can be distributed to sources in $\tilde{O}$, amounts to $C' = \sum_{i \in \tilde{U}} (B_i - \lambda_i)$. For a source where the demand is not satisfied by the primary path, i.e., $i \in \tilde{O}$, we define $\lambda_i' = \lambda_i - B_i$. Now the fair share of the surplus is given by

$$\alpha^o = \begin{cases} \frac{C' - \sum_{j \in U'} \lambda_j'}{|O'|} & if\ O' \neq \emptyset \\ \infty & otherwise, \end{cases} \tag{6}$$

where

$$U' = \{j \in \tilde{O} \mid \lambda_j' < \alpha^o\} \qquad \text{and} \qquad O' = \{j \in \tilde{O} \mid \lambda_j' \geq \alpha^o\}. \tag{7}$$

The rate allocation is obtained via

$$\gamma_i = \begin{cases} \lambda_i & i \in \tilde{U} \text{ or } i \in U' \\ B_i + \alpha^o & i \in O' \ . \end{cases} \tag{8}$$

Thus, a source either obtains enough bandwidth to satisfy its demand, or it obtains the resources on its primary path and a fair share of the surplus. If all sources are greedy, with owned resources, we have $\tilde{U} = \emptyset$, $C^{'} = 0$, $U^{'} = \emptyset$, $\alpha^o = 0$, and therefore $\gamma_i = B_i$ for $i = 1, \ldots, N$.

As with pooled resources, the proofs in [3] can be used to establish the efficiency of the rate allocation. A value for α^o can be constructed as follows. Assume, without loss of generality, that the sources in $\tilde{O}$ have index $1, 2, \ldots, |\tilde{O}|$, and are ordered according to the generated load, that is, $\lambda'_i \leq \lambda'_j$ for $i < j$ and $i, j \in \tilde{O}$. Select $\hat{k}$ as the largest index k ($1 \leq k \leq |\tilde{O}|$) which satisfies

$$\lambda'_k \leq \frac{C' - \sum_{l=1}^{k} \lambda'_l}{|\tilde{O}| - k} \,. \tag{9}$$

Then, we have

$$\alpha^o = \begin{cases} \frac{C' - \sum_{l=1}^{\hat{k}} \lambda'_l}{|\tilde{O}| - \hat{k}} & \text{if } \hat{k} < |\tilde{O}| \\ \infty & \text{otherwise} \,. \end{cases} \tag{10}$$

3 The Primary Path First (PPF) Property

For each source, the fairness and efficiency criteria presented in the previous section make statements about the total rate allocation to a source, but ignore how traffic is split between the primary path and the secondary paths. From a traffic engineering perspective, a rate allocation that transmits more traffic on the primary paths is more attractive, since routing traffic on secondary paths increases the fraction of out-of-sequence delivered packets, leading to higher administrative complexity.

To prevent the multipath routing scheme from spreading traffic across *all* available paths [21], we formulate an objective for our traffic engineering problem, which we call *Primary Path First (PPF)*. PPF refers broadly to the desire to limit the consumption of secondary paths in the MPLS network. In this paper, we focus on an instantiation of PPF where we seek to minimize the total volume of flow assigned to secondary paths.

To make the notion of PPF precise, we refer to a source and its primary path as a source-path pair. An $N \times N$ matrix M will be called a routing matrix if it describes the global assignment of path capacity to sources, i.e. $M(i, j)$ is the amount of traffic sent by source i to path j. Thus, the throughput for source i is $\gamma_i = \sum_{j=1}^{N} M(i, j)$, and the secondary traffic associated with source i is equal to $\sum_{j \neq i} M(i, j)$. The definition of an assignment of traffic that minimizes the total volume of flow assigned to secondary paths is as follows.

Definition 2. *Given a saturating rate allocation $\{(\lambda_i, \gamma_i) : i = 1, \ldots, N\}$, a routing matrix M is said to be PPF-optimal if it solves the following linear program.*

$$\min \quad \sum_{i=1}^{N} \sum_{j \neq i} M(i, j), \tag{11}$$

$$\begin{aligned} \textit{subject to} \quad \sum_{j=1}^{N} M(i, j) &= \gamma_i, && \forall\, i = 1, \ldots, N, \\ \sum_{i=1}^{N} M(i, j) &= B_j, && \forall\, j = 1, \ldots, N, \\ M(i, j) &\geq 0, && \forall\, i, j = 1, \ldots, N. \end{aligned}$$

From this definition, a routing matrix M is PPF-optimal if it achieves the given saturating allocation with minimum total volume of traffic sent along secondary paths. There is unnecessary use of secondary paths if either of the following cases hold.

- **Case 1:** There is a sequence $(i_1, i_2, \cdots, i_k)$ with $k > 2$ such that $M(i_1, i_2) > 0$, $M(i_2, i_3) > 0, \ldots, M(i_{k-1}, i_k) > 0$. We call such a sequence a *chain*.
- **Case 2:** There is a *cycle* $(i_1, i_2, \cdots, i_k)$ with $i_k = i_1$ and $k > 2$, such that $M(i_1, i_2) > 0$, $M(i_2, i_3) > 0, \ldots, M(i_{k-1}, i_k) > 0$.

With these cases in mind, we can devise a procedure that reduces the total amount of traffic sent on secondary paths without altering the total rate allocation γ_i of any source i. Suppose the rate allocation is saturating and we identify a chain which satisfies the condition in Case 1. We can eliminate the chain, or at least cut the chain into two smaller chains, by setting

$$\begin{aligned} M(i_s, i_{s+1}) &\leftarrow M(i_s, i_{s+1}) - \min\{M(i_t, i_{t+1}) : t = 1, \ldots, k-1\}, \quad s = 1, \ldots, k-1, \\ M(i_s, i_s) &\leftarrow M(i_s, i_s) + \min\{M(i_t, i_{t+1}) : t = 1, \ldots, k-1\}, \quad s = 2, \ldots, k-1, \\ M(i_1, i_k) &\leftarrow M(i_1, i_k) + \min\{M(i_t, i_{t+1}) : t = 1, \ldots, k-1\}. \end{aligned}$$

Suppose we identify a cycle which satisfies the condition in Case 2. Then, we can eliminate the cycle by setting

$$\begin{aligned} M(i_s, i_{s+1}) &\leftarrow M(i_s, i_{s+1}) - \min\{M(i_t, i_{t+1}) : t = 1, \ldots, k-1\}, \\ M(i_s, i_s) &\leftarrow M(i_s, i_s) + \min\{M(i_t, i_{t+1}) : t = 1, \ldots, k-1\}, \end{aligned}$$

for all $s = 1, \ldots, k-1$. By adjusting the routing matrix M in this fashion, the cycle disappears, and the total volume of secondary-path traffic is reduced. However, some new shorter chain might be created.
By repeating the above steps of eliminating cycles and chains, we reduce the total volume of secondary traffic, and, eventually, obtain a routing matrix where no chains or cycles exist that satisfy the conditions of Case 1 or 2.

The following proposition shows that a necessary and sufficient condition for a routing matrix to be PPF-optimal is the absence of chains or cycles that satisfy the conditions of Cases 1 or 2.

Proposition 1. *Given a saturating rate allocation $\{(\lambda_i, \gamma_i) : i = 1, \ldots, n\}$. A routing matrix M^* that achieves this allocation is PPF-optimal if and only if there does not exist a chain or cycle as defined above.*

Proof: (Sufficiency) Consider any routing matrix M that achieves the saturating rate allocation $\{(\lambda_i, \gamma_i) : i = 1, \ldots, N\}$. Since $M(i, i) \leq B_i$ and $\sum_{j=1}^{N} M(i, j) = \gamma_i$, we have that $\sum_{j \neq i} M(i, j) \geq \max\{\gamma_i - B_i, 0\}$. Summing this lower bound across all sources, we have that

$$\sum_{i=1}^{N} \sum_{j \neq i} M(i, j) \geq \sum_{i=1}^{N} \max\{\gamma_i - B_i, 0\}. \tag{12}$$

Now consider the routing matrix M^* described in the statement of the proposition. Since there exist no chains or cycles of source-path pairs that satisfy Case 1 or 2, we can follow that, if a source i sends secondary flow, then path i does not receive secondary flow from any other source. Thus, $\sum_{k \neq i} M^*(k,i) = 0$. Since the rate allocation is saturating, we have that $B_i = \sum_{k=1}^{N} M^*(k,i)$, which implies $M^*(i,i) = B_i$, therefore $\sum_{j \neq i} M^*(i,j) = \gamma_i - B_i = \max\{\gamma_i - B_i, 0\}$. Conversely, if path i receives secondary flow from some other source, then source i itself does not send secondary flow. Then, $\sum_{j \neq i} M^*(i,j) = 0$ and $\gamma_i = M^*(i,i) \leq B_i$. Thus, $\sum_{j \neq i} M^*(i,j) = 0 = \max\{\gamma_i - B_i, 0\}$. Again, since the rate allocation is saturating, at least one of the two cases above holds for each $i = 1, \ldots, N$. Thus,

$$\sum_{i=1}^{N} \max\{\gamma_i - B_i, 0\} = \sum_{i=1}^{N} \sum_{j \neq i} M^*(i,j), \tag{13}$$

which, combined with Equation (12), implies that

$$\sum_{i=1}^{N} \sum_{j \neq i} M(i,j) \geq \sum_{i=1}^{N} \sum_{j \neq i} M^*(i,j). \tag{14}$$

Consequently, M^* is PPF-optimal with respect to the saturating rate allocation $\{(\lambda_i, \gamma_i) : i = 1, \ldots, N\}$.

(Necessity) Suppose M is PPF-optimal. If there exists a chain or cycle of source-path pairs which satisfies the conditions of either Case 1 or Case 2, then we can reduce the total volume of secondary traffic by reducing the length of the chain or by eliminating the cycle. However, this would contradict the fact that M is PPF-optimal. Thus, PPF-optimality implies no chains or cycles of source-path pairs satisfying the conditions of Case 1 or 2. ∎

From the proof, if there is a routing matrix M that achieves the saturating rate allocation $\{(\lambda_i, \gamma_i) : i = 1, \ldots, N\}$, then a lower-bound of the total secondary-path traffic is $\sum_{i=1}^{N} \max\{\gamma_i - B_i, 0\}$. From the definition of the PPF criterion, we obtain the following corollary:

Corollary 1. *A routing matrix M is PPF-optimal if it satisfies*

$$\sum_{i=1}^{N} \sum_{j \neq i} M(i,j) = \sum_{i=1}^{N} \max\{\gamma_i - B_i, 0\}\ . \tag{15}$$

4 Multipath-AIMD

Additive-Increase Multiplicative-Decrease (AIMD) feedback algorithms are used extensively for flow and congestion control in computer networks [5, 6, 11, 10, 12, 19] and are widely held to be both efficient and fair in allocating traffic to network paths. These algorithms adjust the transmission rate of a sender based on feedback from the network following an *additive increase/multiplicative decrease* rule. If the network is free of congestion, the transmission rate of the sender is increased by a constant amount. If the network is congested, the transmission rate is reduced by an amount that is proportional to the current transmission rate. Note that in earlier instantiations of the AIMD rule the sending rate for a given source is adjusted as though only one path exists for end-to-end communication.

In this section, we generalize the AIMD rule to account for multiple paths between the sender and the receiver. The resulting algorithm, called *multipath-AIMD*, is intended to provide an efficient and fair mechanism for allocating bandwidth in an MPLS network. We assume that each LSP in the network periodically sends binary congestion-state information to all sources, similar to the DECbit scheme [12]. In the following, we develop two versions of the multipath-AIMD algorithm: basic multipath-AIMD (cf. Section 4.1) and multipath-AIMD with PPF correction (cf. Section 4.2). In the basic multipath-AIMD algorithm, each source uses the original AIMD rule to periodically adjust the rate at which it sends traffic along each LSP, providing a simple, distributed scheme for allocating network paths. Our simulation experiments indicate that the basic scheme is very robust, generally converging to an efficient and fair rate allocation within a reasonable interval of time. One undesirable feature associated with basic multipath-AIMD, especially in the case of greedy sources, is that it tends to allocate flow from all sources to all paths, completely ignoring the PPF criterion. Multipath-AIMD with PPF correction seeks to address this issue by modifying the AIMD adjustment on each LSP according to additional binary feedback information which informs sources of opportunities for reducing secondary path traffic.

4.1 Basic Multipath-AIMD

The basic multipath-AIMD algorithm consists of two parts: a feedback mechanism provided by the network and a rate adjustment mechanism implemented by the sources. The feedback mechanism is similar to the DECbit scheme [12]. Each LSP $j = 1, \ldots, N$ periodically sends messages to all sources containing a binary signal $f_j = \{0, 1\}$ indicating its congestion state. Congestion is defined in terms of the capacity B_j of LSP j: if the utilization of path j meets or exceeds B_j, then the source will receive a signal $f_j = 1$; otherwise the source receives a signal $f_j = 0$. We assume that the source receives signals on the congestion state from each path at regular intervals of length $\Delta_{LSP} > 0$ (a parameter), asynchronously with respect to all other paths.

The rate adjustment mechanism is based upon the original AIMD algorithm [13] with a slight modification to account for non-greedy sources. Each source updates its sending rates to all paths at regular intervals of length $\Delta_{src} > 0$ (also a parameter), asynchronously with respect to all other sources. In this mechanism, the most recent feedback signals received and stored by a source are used in the rate adjustments. We usually set $\Delta_{src} = \Delta_{LSP}$ so that a feedback signal is used by a source only one time. Let x_{ij} denote the rate at which source i sends traffic to path j. The formula for the adjustment depends upon whether resources are pooled or owned, as follows.

Pooled Resources. Each source i adjusts its sending rate to LSP j based on the received feedback signals according to

$$x_{ij} \leftarrow \begin{cases} x_{ij} + k_a, & \text{if } \sum_{l=1}^{N} x_{il} < \lambda_i \text{ and } f_j = 0, \\ x_{ij} & \text{if } \sum_{l=1}^{N} x_{il} \geq \lambda_i \text{ and } f_j = 0, \\ x_{ij} \cdot (1 - k_r) & \text{if } f_j = 1. \end{cases} \tag{16}$$

where $k_a > 0$ and $k_r \in (0, 1]$ are the additive increase and multiplicative decrease parameters, respectively, and where f_j is the latest congestion signal received for LSP j.

Owned Resources. There are two cases to consider. First, if the desired sending rate of source i does not exceed the capacity of its primary path, i.e. $\lambda_i \leq B_i$, then it adjusts its rate to LSP i according to

$$x_{ii} \leftarrow \begin{cases} \min\{x_{ii} + k_a, \lambda_i\}, & \text{if } x_{ii} \leq \lambda_i, \\ x_{ii} \cdot (1 - k_r), & \text{if } x_{ii} > \lambda_i. \end{cases} \tag{17}$$

Note that no flow is ever assigned from source i to any other LSP $j \neq i$. For sources i that demand more than the capacity of their primary paths, i.e. $\lambda_i > B_i$, then, after receiving feedback signals from all paths, source i adjusts its sending rate according to

$$x_{ii} \leftarrow \min\{x_{ii} + k_a, B_i\}, \tag{18}$$

and for $j \neq i$

$$x_{ij} \leftarrow \begin{cases} x_{ij} & \text{if } (x_{ii} < B_i) \text{ or } (x_{ii} = B_i \text{ , } \sum_{l=1}^{N} x_{il} \geq \lambda_i \text{ and } f_j = 0), \\ x_{ij} + k_a, & \text{if } x_{ii} = B_i \text{ , } \sum_{l=1}^{N} x_{il} < \lambda_i \text{ and } f_j = 0, \\ x_{ij} \cdot (1 - k_r) & \text{if } x_{ii} = B_i \text{ , and } f_j = 1. \end{cases} \tag{19}$$

Thus, in the owned case, a source never sends traffic to secondary paths before it makes full use of its primary path. Moreover, the traffic sent from a source to its primary path is independent of traffic sent from other sources.

4.2 Multipath-AIMD with PPF Correction

In the case of owned paths, the basic multipath-AIMD algorithm requires all sources to consume first the capacity of their respective primary paths, and secondary paths are utilized only when the primary paths are insufficient to meet the desired sending rate. As a result, the basic multipath-AIMD algorithm automatically produces PPF-optimal routing matrices. However, this is not the case for pooled resource. Therefore, to enforce the PPF criterion for pooled resources, we develop an alternative algorithm, called multipath-AIMD with PPF correction.

As with the basic scheme, multipath-AIMD with PPF correction consists of a feedback mechanism and a rate adjustment mechanism. As before, multipath-AIMD with PPF correction takes binary feedback $f_j = \{0, 1\}$ from all LSPs. However, in this case, extra feedback information is required to allow the sources to coordinate in attempting to reduce the total volume of secondary traffic. Each source $i = 1, \ldots, N$

periodically sends messages to all other sources containing a binary (routing) vector $m_i = (m_{i1}, \ldots, m_{iN})$, where $m_{ij} = 1$ if source i is currently sending traffic to LSP j, and $m_{ij} = 0$ otherwise. Each source i retains the routing vector m_j associated with all other sources and uses this information to modify the basic AIMD update of Equation (16). Each source transmits its routing vector at regular intervals of length $\Delta_{PPF} > 0$ (a parameter), asynchronously with respect to all other sources.

The rate adjustment mechanism for multipath-AIMD with PPF correction includes all of the rate adjustments from the basic scheme (i.e. updates of the type in Equation (16)) plus extra adjustments based on routing vector information received from the other sources. In particular, after each basic multipath-AIMD rate adjustment, each source i will engage in a PPF correction step as follows.

$$x_{ij} \leftarrow \begin{cases} \max\{x_{ij} - K, 0\} & \text{if } \sum_{l \neq i} m_{li} > 0, \\ x_{ij} & \text{otherwise,} \end{cases} \tag{20}$$

$$x_{ii} \leftarrow x_{ii} + \sum_{j \neq i} \min\{K, x_{ij}\}, \tag{21}$$

where $K > 0$ is the additive PPF correction parameter. Thus, if source i is making use of the secondary LSP $j \neq i$ and if LSP i is receiving secondary flow from some other source, then source i will reduce traffic on the secondary LSP j by K and, at the same time, increase its traffic to the primary LSP i by K.

Equations (20) and (21) have the effect of reducing flow along chains or cycles of source-path pairs with unnecessary secondary path flow. In fact, the PPF correction is inspired from the secondary path reduction scheme discussed in Section 3, which involves breaking chains or cycles of source-path pairs that satisfy either Case 1 or 2 from Section 3. By reducing the total flow along each chain or cycle with unnecessary secondary path utilization, the PPF correction creates the opportunity for subsequent basic multipath-AIMD rate adjustments to modify the solution toward efficiency, fairness, and PPF-optimality. Unfortunately, while the PPF correction creates this opportunity, it does not represent a complete solution. The basic problem is that the PPF correction tends to push flow onto primary paths, interfering with the natural tendency of AIMD to arrive at a fair distribution of the load. In fact, the basic multipath-AIMD rate adjustment (see Equation (16)) and the PPF correction (see Equations (20)-(21)) tend to compete with one another as the system evolves to a final rate allocation. In practice, one must be careful in choosing values for k_a, k_r, and K. If K is large compared to k_a and k_r, then the PPF correction will dominate, and the resulting rate allocation will show low utilization on the secondary paths, but may also be quite far from the fair-share rate allocation. The simulation results in Section 5 illustrate this tradeoff.

5 Simulation Results

Here we present *ns-2* [1] simulation results to evaluate multipath-AIMD as applied to an MPLS network with five sources and five LSPs. In Section 5.1 we present results for the basic multipath-AIMD algorithm, and in Section 5.2 we present results for the revised algorithm, multipath-AIMD with PPF correction. Our simulations indicate that (1) the basic algorithm achieves an efficient and fair allocation of capacities to sources, (2) the basic algorithm yields a PPF-optimal solution in the case of owned resources, and (3) the revised algorithm, multipath-AIMD with PPF correction, can reduce secondary path utilization in the pooled resources case at the expense of reduced fairness.

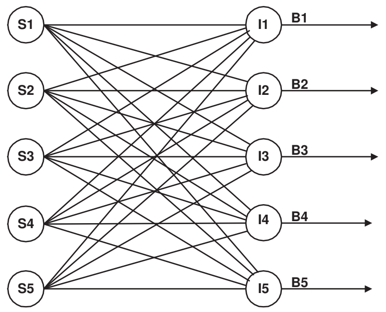

Fig. 3. Simulated network topology.

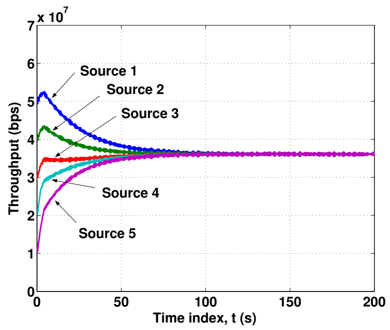

Fig. 4. Experiment 1 - Basic multipath-AIMD applied with greedy sources and pooled resources.

	Initial Scenario, $t \in [0, 80)$ sec			Final Scenario, $t \in [80, 200)$ sec			
Source i	Load λ_i	Throughput, γ_i (Pooled)	Throughput, γ_i (Owned)	Load λ_i	Throughput, γ_i (Pooled)	Throughput, γ_i (Owned)	Capacity of LSP i
1	10	10	10	10	10	10	50
2	30	30	30	50	46.7	50	40
3	50	50	50	50	46.7	45	30
4	60	60	60	60	46.7	45	30
5	30	30	30	30	30	30	30

Table 1. Experiment 2 - Predicted efficient and fair rate allocations for the pooled and owned cases. The predictions are based on the results from Sections 2 and 3. All values are expressed in Mbps. Note that the desired sending rate for Source 2 changes from 30 to 50 at time $t = 80$ sec.

Experimental Setup. Figure 3 illustrates the topology of the MPLS network simulated in *ns-2*. The nodes S1, S2, S3, S4, and S5 are source nodes and I1, I2, I3, I4 and I5 are ingress nodes of an MPLS network. The LSPs associated with the ingress nodes have bandwidths ($B_i \mid i = 1, \ldots, 5$) = (50, 40, 30, 30, 30) Mbps. We modified the ns-2 code in two ways. First, ingress nodes periodically send feedback messages to the sources indicating the congestion state of the corresponding LSPs, as described in Section 4. Second, source nodes generate CBR traffic with a rate specified by our multipath-AIMD scheme in response to the received feedback messages. The bandwidth and propagation delay for each link between the source nodes and the ingress nodes are set to 100 Mbps and 5 ms, respectively. Resources send congestion feedback every $\Delta_{LSP} = 5$ ms. Sources update their sending-rates every $\Delta_{src} = 5$ ms. For the experiments involving multipath-AIMD with PPF correction, the topology of Figure 3 is augmented to include full-duplex links between all source pairs, with bandwidth and propagation delay of 100 Mbps and 1 ms, respectively. Sources exchange binary routing vectors every $\Delta_{PPF} = 5$ ms. All packets in the simulation are 50 bytes in length and are treated as UDP packets (i.e. no flow control). Finally, for all experiments in this section, we set $k_a = .1$ Mbps and $k_r = .01$ as values for the additive increase and multiplicative decrease parameters, respectively.

5.1 Basic Multipath-AIMD

Experiment 1: Basic Multipath-AIMD with Greedy Sources and Pooled Resources. Figure 4 shows the outcome from basic multipath-AIMD applied to the case of greedy sources and pooled resources. The plot shows the evolution of the throughput γ_i achieved by each source i. Note that all sources converge within 90 seconds to a throughput of 36 Mpbs, the fair-share allocation for this case. We point out that the final routing matrix achieved by basic multipath-AIMD is not PPF-optimal. The total volume of secondary traffic in the final resource allocation (which is not shown in a graph) is 143.5 Mbps. This is much larger than the fair PPF-optimal allocation which requires only 18 Mbps of traffic on secondary paths.

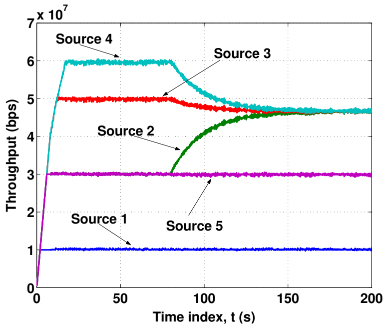

Fig. 5. Experiment 2 - Basic Multipath-AIMD applied to the case of pooled resources.

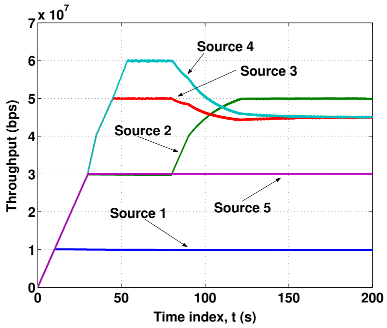

Fig. 6. Experiment 2 - Basic Multipath-AIMD applied to the case of owned resources.

Experiment 2: Basic Multipath-AIMD with Non-Greedy Sources. Here we consider the case of non-greedy sources, and we apply the basic algorithm, where the capacities of the paths are either pooled or owned. In this experiment, the desired sending rates $(\lambda_i \mid i = 1, \ldots, 5)$ for the sources all start out at values (10, 30, 50, 60, 30) Mbps. At time $t = 80$ sec, source 2 switches its desired sending rate from $\lambda_2 = 30$ Mbps to 50 Mbps. The theoretical fair-shares for all sources (before and after the switch) are shown in Table 1, with results for both the pooled and user-owned cases. Figures 5 and 6 illustrate the evolution of the algorithm in terms of throughput achieved by each source. The results in Figure 5 apply to the case of pooled resources, whereas Figure 6 describes the outcome for owned resources. In both figures the throughput values for each source converge to the appropriate theoretic fair-share value listed in Table 1, both before and after the switch in λ_2 at $t = 80$ sec. As before, the figures do not indicate the total volume of secondary traffic associated with the final routing matrices at $t = 200$ sec. It turns out that the final solution for the pooled resources is not PPF-optimal, with a total volume of secondary path traffic equals to 146.4 Mbps which is larger than the PPD-optimal value of 40 Mbps for the corresponding fair rate allocation for pooled resources. On the other hand, the final solution for the owned case is PPF-optimal, achieving the optimal secondary path utilization value of 40 Mbps for the fair allocation to owned resources.

5.2 Multipath-AIMD with PPF Correction

In this subsection, we present results from two experiments where we apply the revised algorithm, multipath-AIMD with PPF correction, to an MPLS network with non-greedy sources and pooled resources. In both experiments, the desired sending rates ($\lambda_i \mid i = 1, \ldots, 5$) are (10, 50, 50, 60, 30) Mbps, and the corresponding fair-share rate allocation appears under the "Final Scenario" heading in Table 1. In Experiment 3, we set the PPF correction parameter K to a very small numerical value, $K = .00001$ Mbps, which, like basic multipath-AIMD, results in a fair but PPF-suboptimal routing matrix. In Experiment 4, we set the PPF correction parameter more aggressively, $K = .01$ Mbps, resulting in a final routing matrix which is PPF-optimal. Here, however, the corresponding rate allocation deviates from the fair-share rate allocation predicted in Table 1.

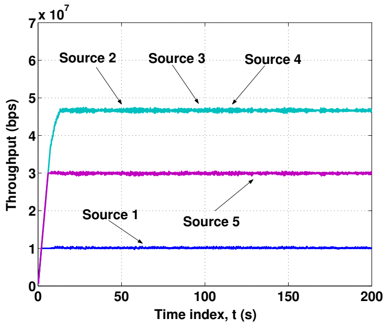

Fig. 7. Experiment 3 - Throughput achieved by multipath-AIMD with PPF correction ($K = .00001$ Mbps).

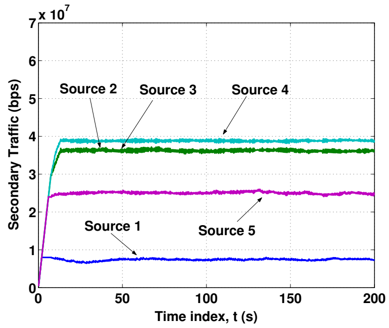

Fig. 8. Experiment 3 - Secondary traffic achieved by multipath-AIMD with PPF correction ($K = .00001$ Mbps).

Experiment 3: Multipath-AIMD with PPF correction with $K = .00001$ Mbps. The results of this experiment are shown in Figures 7 and 8. Figure 7 shows the evolution of throughput for each of the five sources. The network settles within 25 seconds to a rate allocation consistent with the predicted values in Table 1. Figure 8 shows the total allocation of each source to secondary paths. Note that the final routing matrix results in 145.7 Mbps total secondary traffic, which is larger than the PPF-optimal value of

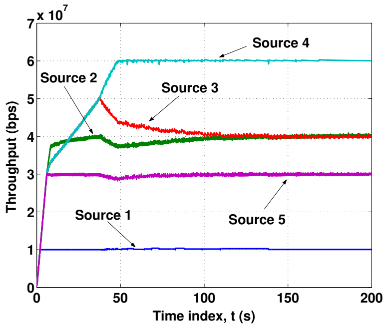

Fig. 9. Experiment 4 - Throughput achieved by multipath-AIMD with PPF correction ($K = .01$ Mbps).

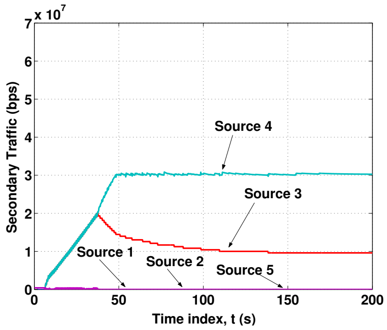

Fig. 10. Experiment 4 - Secondary traffic achieved by multipath-AIMD with PPF correction ($K = .01$ Mbps).

40 Mbps. Evidently, because K is so small, the PPF correction in this experiment does not have much impact in guiding the system to a PPF-optimal routing matrix.

Experiment 4: Multipath-AIMD with PPF correction with $K = .01$ Mbps. Here we apply the revised algorithm, multipath-AIMD with PPF correction, to the same problem as in Experiment 3, with a more aggressive value for the PPF correction parameter, $K = .01$ Mbps. The results are shown in Figures 9 and 10. The evolution of throughput γ_i for each of the five sources appears in Figure 9. We observe that the final rate allocation deviates from the predicted allocation from Table 1, especially with regard to sources 2, 3, and 4. Thus, the PPF correction causes the system to evolve to an unfair distribution of path resources. On the other hand, the final routing matrix is PPF-optimal with respect to the final rate allocation. This can be seen in Figure 10 which shows the evolution of secondary traffic allocated by each of the five sources. Note that the final routing matrix involves a 40 Mbps total secondary traffic, which is PPF-optimal for the rate allocation given in Figure 9.

6 Conclusions and Future Work

We have studied the problem of allocating LSP resources in an MPLS network. Our network model assumes that each traffic source has a primary path and may utilize the capacity of other, secondary, paths. We accommodate both greedy and non-greedy traffic sources and allow the capacity of each LSP to be considered as either a shared ("pooled") resource or as a resource "owned" by the corresponding traffic source.

With regard to the allocation of resources, we have defined fairness criteria based on the notion of fair-share allocation in [3], with special consideration as to whether the resources are pooled or owned. In addition to the fairness criteria, we have also introduced a secondary objective, the PPF criterion, to be achieved in the final allocation of resources. The PPF criterion is defined with respect to a given throughput allocation as an optimization model where the objective is to minimize the total volume of traffic sent along secondary paths. We provide a characterization of the PPF solution in terms of the existence of secondary path chains and cycles, and in principle this provides an algorithm that can minimize the total volume of secondary path traffic without affecting the throughput of each source.

To achieve a fair and PPF-optimal rate allocation in a distributed fashion, we propose multipath-AIMD as an extension to the earlier work of [13]. Multipath-AIMD comes in two flavors: (1) basic multipath-AIMD, which seeks to provide a fair allocation of throughput to each source, without special consideration of the PPF criterion, and (2) multipath-AIMD with PPF correction, which augments the basic algorithm to reduce the volume of secondary path traffic. Both algorithms rely upon binary feedback information regarding the congestion state of each of the LSPs and, for the second version of the algorithm, a binary routing vector associated with each source. Simulation experiments with multipath-AIMD show that the basic algorithm converges to an efficient and fair allocation of resources and also yields a PPF-optimal solution in the case of owned resources. The revised algorithm, multipath-AIMD with PPF correction, can reduce secondary path utilization (for the pooled resources case) at the expense of fairness. From the perspective of Internet traffic engineering, multipath-AIMD seems to provide a practical mechanism for improving the utilization of LSP resources, while maintaining fairness and minimizing the complexity associated with multipath routing.

This paper presents a step towards a definition of traffic engineering criteria for MPLS networks. While these initial results are promising, there are limitations to our model which must be addressed in subsequent research. First, we assume that all sources have access to all LSPs, which is clearly unrealistic in many networking contexts. This is more than an assumption of convenience, since the appropriate notion of fairness for the case where each source has access to a subset of the resources is somewhat unclear. A second limitation of our network model is that it assumes that the flows along LSPs do not interact. While it is possible to rationalize the simplified model of Figure 2, future work in this area should address the full set of interactions possible in Figure 1.

References

1. *ns-2* network simulator. http://www.isi.edu/nsnam/ns/.
2. O. Aboul-Magd, L. Andersson, and P. Ashwood-Smith. Constraint-based LSP setup using LDP. http://www.ietf.org/internet-drafts/draft-ietf-mpls-cr-ldp- 05.txt, February 2001.
3. I. F. Akyildiz, J. Liebeherr, and A. Tantawi. DQDB+/-: A fair and waste-free media access protocol for dual bus metropolitan networks. *IEEE Transactions on Communications*, 41(12):1805–1815, December 1993.
4. D. O. Awduche, A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao. Overview and principles of Internet traffic engineering. http://www.ietf.org/internet-drafts/draft-ietf-tewg- principles-02.txt, November 2001.
5. F. Bonomi and W. Fendick. The Rate-Based Flow Control Framework for the Available Bit Rate ATM Service. *IEEE Network*, 9(2):25–39, March/April 1995.
6. D. Chiu and R. Jain. Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks. *Computer Networks and ISDN Systems*, 17:1–14, 1989.
7. A. Elwalid, C. Jin, S. Low, and I. Widjaja. MATE: MPLS adaptive traffic engineering. In *Proceedings of IEEE INFOCOM 2001*, volume 3, pages 1300–1309, 2001.

8. D. O. Awduche et. al. RSVP-TE: Extensions to RSVP for LSP tunnels. http://www.ietf.org/internet-drafts/draft-ietf-mpls-rsvp-lsp- tunnel-09.txt, August 2001.
9. P. Hurley, J.-Y. Le Boudec, and P. Thiran. A note on the fairness of additive increase and multiplicative decrease. In *Proceedings of ITC-16*, Edinburgh, UK, June 1999.
10. V. Jacobson. Congestion avoidance and control. In *Proceedings of ACM Sigcomm '88, August, 1988*, pages 314–329, 1988.
11. R. Jain. Congestion control and traffic management in ATM networks: Recent advances and a survey. *Computer Networks and ISDN Systems*, 28(13):1723–1738, October 1996.
12. R. Jain and K. K. Ramakrishnan. Congestion avoidance in computer networks with a connectionless network layer: Concepts, goals and methodology. *Proceedings of the Computer Networking Symposium; IEEE; Washington, DC*, pages 134–143, 1988.
13. R. Jain, K. K. Ramakrishnan, and D.-M. Chiu. Congestion avoidance in computer networks with a connectionless network layer. December 1988. Digital Equipment Corporation, Technical Report DEC-TR-506.
14. F. P. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8:33–37, 1997.
15. F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.
16. S. Kunniyur and R. Srikant. End-to-end congestion control: Utility functions, random losses and ECN marks. In *Proceedings of IEEE INFOCOM 2000*, pages 1323–1332, March 2000.
17. K.-W. Lee, T.-E. Kim, and V. Bharghavan. A comparison of end-to-end congestion control algorithms: the case of AIMD and AIPD. In *Proceedings of IEEE Globecom 2001*, San Antonio, Texas, November 2001.
18. L. Massoulie and J. Roberts. Bandwidth sharing: Objectives and algorithms. In *Proceedings IEEE INFOCOM 1999*, New York, March 1999.
19. K. K. Ramakrishnan and R. Jain. A Binary Feedback Scheme for Congestion Avoidance in Computer Networks. *ACM Transactions on Computer Systems*, 8(2):158–181, 1990.
20. E. Rosen, A. Viswanathan, and R. Callon. Multiprotocol label switching architecture. draft-ietf-mpls-arch-07.txt, ftp://ftp.isi.edu/in-notes/rfc3031.txt, January 2001.
21. M. Vojnovic, J.-Y. Le Boudec, and C. Boutremans. Global fairness of additive-increase and multiplicative-decrease with heterogeneous round-trip times. In *Proceedings of IEEE INFOCOM 2000*, volume 3, pages 1303–1312, 2000.