Optimal Congestion Control of TCP Flows for Internet Routers *

K. Avrachenkov[a], U. Ayesta[b,c,d], J. Doncel[b,c,e], P. Jacko[e]

[a]INRIA Sophia-Antipolis, France
[b] LAAS-CNRS, Toulouse, France
[c] Université de Toulouse, LAAS, Toulouse, France
[d]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[e]BCAM- Basque Centre for Applied Mathematics, Bilbao, Spain

ABSTRACT

In this work we address the problem of fast and fair transmission of flows in a router, which is a fundamental issue in networks like the Internet. We model the interaction between a TCP source and a bottleneck queue with the objective of designing optimal packet admission controls in the bottleneck queue. We focus on the relaxed version of the problem obtained by relaxing the fixed buffer capacity constraint that must be satisfied at all time epoch. The relaxation allows us to reduce the multi-flow problem into a family of single-flow problems, for which we can analyze both theoretically and numerically the existence of optimal control policies of special structure. In particular, we show that for a variety of parameters, TCP flows can be optimally controlled in routers by so-called index policies. We have implemented index policies in Network Simulator-3 (NS-3) and compared its performance with DropTail and RED buffers. The simulation results show that the index policy has several desirable properties with respect to fairness and efficiency.

1. INTRODUCTION

In this paper we develop a rigorous mathematical framework to model the interaction between a TCP source and a bottleneck queue with the objective of designing optimal packet admission controls in the bottleneck queue. The TCP sources follow the general family of *Additive Increase Multiplicative Decrease* pattern that TCP versions like New Reno or SACK follow, but to keep the Markovian model simple we ignore the *slow-start* phase. A TCP source is thus characterized by the decrease factor γ, which determines the decrease factor of the congestion window in the event of a packet loss (in New Reno γ takes the value $\frac{1}{2}$). The objective is to design a packet admission control strategy that uses the resources efficiently and that provides satisfactory experience to users. Mathematically, we formulate the problem as a resource allocation problem within the Markov Decision Process (MDP) framework [5]. The main diffference of the proposed scheme with respect to the Active Queue Management schemes is that with our scheme we can achieve a very large spectrum of fairness criteria.

We have implemented our solution in NS-3 [1] and performed extensive simulations in a benchmark topology to explore and validate the properties of the algorithm, and to assess the improvement with respect to a DropTail and RED buffer.

Due to lack of space in this paper we are including some basic notions to understand the model and the first simulations in NS-3. For full details we refer to [3].

*Research partially supported by grant MTM2010-17405 (Ministerio de Ciencia e Innovación, Spain) and grant PI2010-2 (Department of Education and Research, Basque Government) and Inria Alcatel-Lucent Joint Lab.

2. MARKOV DECISION PROCESS MODEL

Let us consider the time slotted $t \in \mathcal{T} := \{0, 1, 2, \ldots\}$, where each time slot t corresponds to time periods of one round-trip time (RTT). Every flow can be allocated either the capacity required by its current congestion window (being *admitted*) and transmitted, or zero capacity (being *rejected*). We denote by $\mathcal{A} := \{0, 1\}$ the *action space*, where 0 corresponds to blocking and 1 corresponds to admitting. This action space is the same for every flow k.

Each flow k is defined independently of other flows as the tuple $(\mathcal{N}_k, (\boldsymbol{W}_k^a)_{a\in\mathcal{A}}, (\boldsymbol{R}_k^a)_{a\in\mathcal{A}}, (\boldsymbol{P}_k^a)_{a\in\mathcal{A}})$, where $\mathcal{N}_k := \{1, 2, \ldots, N_k\}$ is the *state space*, i.e., a set of possible congestion windows flow k can set; $\boldsymbol{W}_k^a := \left(W_{k,n}^a\right)_{n\in\mathcal{N}_k}$, where $W_{k,n}^a$ is the expected one-period capacity consumption (in number of packets), or *work* required by flow k at state n if action a is decided at the beginning of a period; $\boldsymbol{R}_k^a := \left(R_{k,n}^a\right)_{n\in\mathcal{N}_k}$, where $R_{k,n}^a$ is the expected one-period generalized α-fairness or *reward* earned by flow k at state n if action a is decided at the beginning of a period; in particular $R_{k,n}^0 := 0$ and

$$R_{k,n}^1 := \begin{cases} \dfrac{(1+n)^{1-\alpha} - 1}{1-\alpha}, & \text{if } \alpha \neq 1, \\ \log(1+n), & \text{if } \alpha = 1; \end{cases}$$

which is a convex function that depends on the parameter α of the generalized α-fairness criterion [2]; and $\boldsymbol{P}_k^a := \left(p_{k,n,m}^a\right)_{n,m\in\mathcal{N}_k}$ is the flow-k stationary one-period *state-transition probability matrix* if action a is decided at the beginning of a period, i.e., $p_{k,n,m}^a$ is the probability of moving to state m from state n under action a.

The dynamics of flow k is thus captured by the *state process* $X_k(\cdot)$ and the *action process* $a_k(\cdot)$, which correspond to state $X_k(t) \in \mathcal{N}_k$ and action $a_k(t) \in \mathcal{A}$, respectively, at all time epochs $t \in \mathcal{T}$. The states $n \in \mathcal{N}_k$ denote possible levels

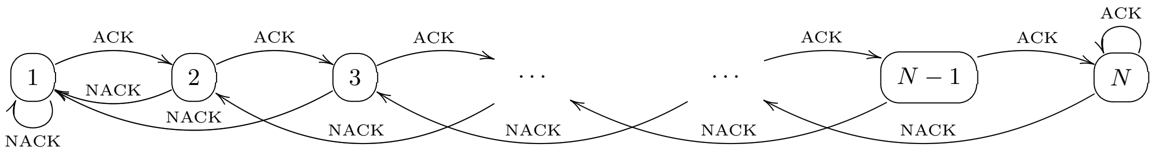

Figure 1: A model of an AIMD flow as a Markov chain. The arrows represent one-period transitions among the states $1, 2, \ldots, N$ after a congestion-free (ACK) and a congestion-experienced (NACK) transmission.

of the sending rate. In particular $W_n^{\text{sent}} := n$ can therefore be interpreted as the bandwidth capacity the flow requires for complete transmission at the current period.

The schematic behavior of the AIMD flow as a Markov chain is shown in Figure 1, where "ACK" represents a congestion-free delivery of the flow packets to the receiver (positive acknowledgments) and "NACK" represents a congestion-experienced transmission (negative acknowledgments).

3. OPTIMIZATION PROBLEM

Let Π be the set of all history-dependent randomized policies. Denote by the symbol $\mathbb{E}_{\boldsymbol{n}}^{\pi}$ the conditional expectation given that the initial conditions are $\boldsymbol{n} := (n_k)_{k\in\mathcal{K}}$, and the policy applied is $\pi \in \Pi$. Let $\overline{W}$ be the number of packets that can be served in one RTT, the router controller's problem to solve under the discounted criterion (if $\beta < 1$) is the following:

$$\max_{\pi\in\Pi} \mathbb{E}_{\boldsymbol{n}}^{\pi}\left[\sum_{t=0}^{\infty}\sum_{k\in\mathcal{K}} \beta^t R_{k,X_k(t)}^{a_k(t)}\right] \quad (1)$$

$$\text{subject to} \quad \mathbb{E}_{\boldsymbol{n}}^{\pi}\left[\sum_{t=0}^{\infty}\sum_{k\in\mathcal{K}} \beta^t W_{k,X_k(t)}^{a_k(t)}\right] \le \frac{\overline{W}}{1-\beta} \quad (2)$$

The standard solution of such a formulation is by solving for each ν the Lagrangian relaxation of (1)–(2), which is

$$\max_{\pi\in\Pi} \mathbb{E}_{\boldsymbol{n}}^{\pi}\left[\sum_{t=0}^{\infty}\sum_{k\in\mathcal{K}} \beta^t \left(R_{k,X_k(t)}^{a_k(t)} - \nu W_{k,X_k(t)}^{a_k(t)}\right)\right] + \nu\frac{\overline{W}}{1-\beta} \quad (3)$$

where ν is the Lagrangian parameter that can be interpreted as a per-packet *transmission cost*. The Lagrangian theory assures that there exists ν^*, for which the Lagrangian relaxation (3) achieves optimum of (1)–(2). Since for any fixed ν the flows are independent and the second term of (3) is constant, we can decompose (3) into K individual-flow

$$\max_{\pi_k\in\Pi_k} \mathbb{E}_{n_k}^{\pi_k}\left[\sum_{t=0}^{\infty} \beta^t \left(R_{k,X_k(t)}^{a_k(t)} - \nu W_{k,X_k(t)}^{a_k(t)}\right)\right] \quad (4)$$

If for a given parameter ν, each policy π_k^* for $k \in \mathcal{K}$ optimizes the individual-flow problem then π^* optimizes the multi-flow problem (3).

4. INDEXABILITY AND THRESHOLD POLICIES

Definition 1 (Optimality of Threshold Policies). *We say that problem* (4) *is* optimally solvable by threshold policies, *if for every real-valued ν there exists threshold state $n \in \mathcal{N}_k \cup \{0\}$ such that threshold policy admitting the flow in states $\mathcal{S}^{(n)} := \{m \in \mathcal{N}_k : m \le n\}$ and rejecting otherwise is optimal for problem* (4).

Of our interest will be the index proposed by Whittle [5]. We adopt the definition of indexability from [4].

Definition 2 (Indexability). *We say that ν-parameter problem* (4) *is* indexable, *if there exist unique values $-\infty \le \nu_{k,n} \le \infty$ for all $n \in \mathcal{N}_k$ such that the following holds for every state $n \in \mathcal{N}_k$:*

(i) if $\nu_{k,n} \ge \nu$, then it is optimal to admit flow k in state n, and

(ii) if $\nu_{k,n} \le \nu$, then it is optimal to reject flow k in state n.

The function $n \mapsto \nu_{k,n}$ is called the (Whittle) index, *and $\nu_{k,n}$'s are called the* (Whittle) index values.

An immediate consequence of the two definitions is formulated in the following previously known result [4].

Proposition 1. *If problem* (4) *is indexable and the index is nonincreasing, i.e., $\nu_{k,1} \ge \nu_{k,2} \ge \cdots \ge \nu_{k,N_k}$, then problem* (4) *is optimally solvable by threshold policies. Moreover, for a given ν the optimal threshold policy is $\mathcal{S}^{(n^*)}$ with $n^* \in \mathcal{N}_k \cup \{0\}$ such that $\nu_{k,n^*} \ge \nu \ge \nu_{k,n^*+1}$ (defining $\nu_{k,0} := -\infty, \nu_{k,N_k+1} := \infty$).*

4.1 Analytical and Numerical Results

From the analytical point of view, it was proven in [4] that one- and two-state flows are always indexable and solvable under threshold policies. In three-states flow, if the decrease factor γ is less than $\frac{2}{3}$, then it was also shown in [4] that the scheme is always indexable and solvable by threshold policies.

We show that three-state flow with $\gamma \ge \frac{2}{3}$ is always indexable, see [3] for details. We give a closed-form expression of the indices in this instance and we conclude that the value of the indices changes depending on α parameter as follows. If $\alpha < 1$, the threshold policies are optimal and the values of the indices are

- $\nu_{k,1} = \frac{R_{k,1}}{W_{k,1}}$,
- $\nu_{k,2} = \frac{R_{k,2}-\beta R_{k,1}}{W_{k,2}-\beta W_{k,1}}$,
- $\nu_{k,3} = \frac{R_{k,3}+\beta(R_{k,3}-R_{k,2})}{W_{k,3}+\beta(W_{k,3}-W_{k,2})}$.

If $\alpha \ge 1$, threshold policies are not optimal in general ($\nu_{k,1} > \nu_{k,3} > \nu_{k,2}$) and the values of the indices are

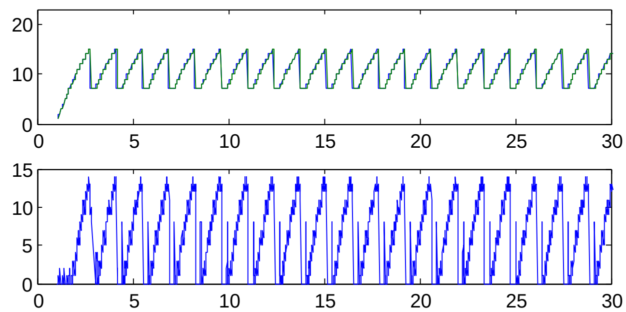

Figure 2: Congestion Window and size of the queue in droptail

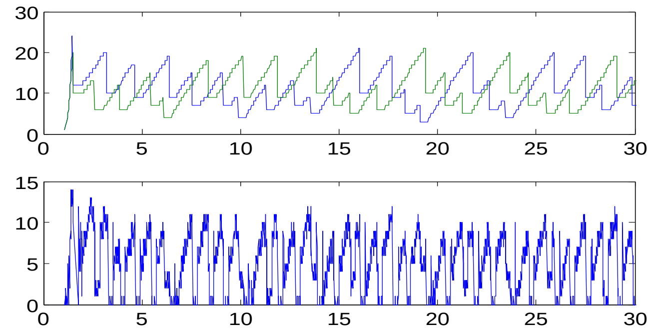

Figure 3: Congestion Window and size of the queue in RED

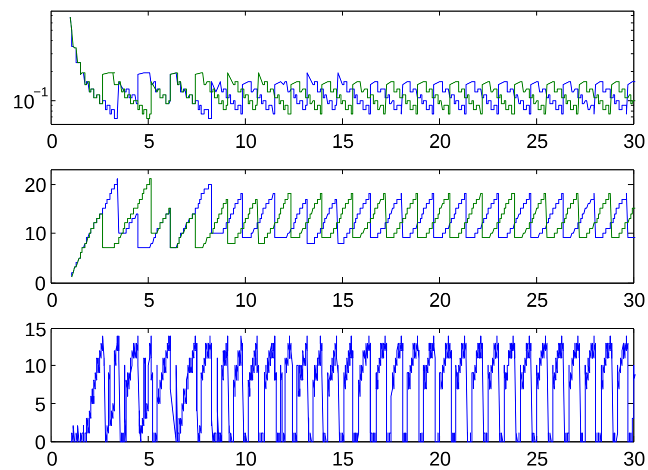

Figure 4: Indices, Congestion Window and size of the queue in index policy with $\alpha = 1$ model

- $\nu_{k,1} = \frac{R_{k,1}}{W_{k,1}}$,
- $\nu_{k,2} = \frac{R_{k,2}+\beta(R_{k,3}-R_{k,1})+\beta^2(R_{k,3}-R_{k,2})}{W_{k,2}+\beta(W_{k,3}-W_{k,1})+\beta^2(W_{k,3}-W_{k,2})}$,
- $\nu_{k,3} = \frac{R_{k,3}-\beta^2 R_{k,1}}{W_{k,3}-\beta^2 W_{k,1}}$.

From the numerical point of view, we have tested the indexability of the problem over a large number of flows with different parameters. The numerical analysis established that in all tested cases the problem was indexable. We conjecture that the scheme as defined in this paper is always indexable.

5. SIMULATION RESULTS

In this section we present simulation results from implementing index policy in NS-3 [1]. We define the following heuristic index policy to be implemented in the Internet routers:

Heuristic index policy at packet level: Upon a packet arrival, if the buffer is not full, then accept the packet. Otherwise, drop the packet (either the new one or from the queue) with smallest index value. In case of ties, drop the packet that has been the longest in the queue.

As the measure of fairness we employ the Jain's fairness index.

5.1 Two Symmetric Users

As a baseline scenario, we consider two symmetric (equal) users that are sending data to a server through a bottleneck router. The delay of the access link of each of the users is 10ms. The delay of the bottleneck link in this scenario is 10ms and the bandwidth capacity of the bottleneck link is 1500kb/s. The packet size is 536 Bytes. The buffer size is set to 14, that corresponds to the Bandwidth-Delay product of a single user.

Each user $k = 1, 2$ is halving its congestion window, i.e., $\gamma_k = 1/2$. We set $\beta = 0.9999$ which approximates the time-average criterion.

In the figures, user 1 is depicted with a blue and thick solid line, and user 2 with a green and thin line.

We depict the evolution of the congestion window and the size of the queue of the router in time for DropTail and RED in Figure 2 and in Figure 3, respectively. We show the evolution of the indices, the congestion window and the size of the queue of the router in time for the index policy in Figure 4. We observe users with the index policy become perfectly unsynchronized. The throughput increases comparing to RED and Droptail, caused by the more efficient buffer management, while fairness remains essentially the same as under the DropTail policy.

References

Network simulator, ver.3, (ns-3). http://www.nsnam.org/.

E. Altman, K. Avrachenkov, and A. Garnaev. Generalized α-fair resource allocation in wireless networks. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 2414–2419, Cancun, Mexico, December 2008.

K. Avrachenkov, U. Ayesta, J. Doncel, and P. Jacko. Congestion control of flows in internet routers by means of index policies. *arXiv:1209.3638v1*, 2012.

P. Jacko and B. Sansò. Optimal anticipative congestion control of flows with time-varying input stream. *Performance Evaluation*, 69(2):86–101, 2012.

P. Whittle. Restless bandits: Activity allocation in a changing world. *A Celebration of Applied Probability, J. Gani (Ed.), Journal of Applied Probability*, 25A:287–298, 1988.