# Performance Analysis of TCP with RIO Routers

Naceur Malouch
INRIA Sophia-Antipolis
2004 route des Lucioles
BP 93, 06902 Sophia Antipolis
France
Naceur.Malouch@sophia.inria.fr

Zhen Liu
IBM T.J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
USA
zhenl@us.ibm.com

**Abstract**

We present an approach to analyzing the performance characteristics of TCP sessions in the presence of network routers which deploy the Random Early Detection (RED) mechanism with two *in* and *out* drop probability functions (RIO). We consider the case with a large number of TCP sessions which use token buckets for marking *in* and *out* packets at the entrance of the network. Under some simplifying assumptions we derive a set of equations that govern the evolution of these TCP sessions and the routers under consideration. We then solve these equations numerically using a fixed point method. Our analysis can capture characteristics of both RED and Tail Drop (TD) mechanisms in the RIO router. Our model is validated through simulations which show that less than $5\%$ error is achieved in most cases. Various performance analyses are then carried out using this approach in order to study the impact of the RIO parameters on the performance characteristics of TCP sessions. Our results show that the loss probability threshold of *out* packets has a significant effect on the TCP throughput and on the average queue length. Setting this parameter consists in trading off between the network utilization and the fairness among TCP connections. Our results also show that Tail Drop mechanism is particularly suitable for *in* packets to satisfy various QoS constraints.

**Keywords**

TCP, Random Early Detection, Tail Drop, Packet Marking, Fixed Point Method.

## I. INTRODUCTION

Differentiated Services (DiffServ) has been proposed for about half decade as a scalable mechanism of providing Quality of Service (QoS) in the Internet. A number of studies have been conducted to understand such an architecture. It is still not clear what services a Service Provider can offer using DiffServ mechanisms and also how they can provide them [1]. For example, exploiting the Assured Forwarding Per Hop Behavior (AF PHB) [9] service is quite intricate since the quality of service offered is statistically guaranteed. Random Early Detection with In and Out (RIO) plays a major role in the design and the implementation of the AF classes. RIO is a mechanism that includes both Active Queue Management for congestion control and preferential packet treatment for service differentiation. RIO is characterized by two non-decreasing drop probability functions. The most analyzed cases are piecewise linear functions defi ned by two thresholds and a maximum drop probability. When the queue size is below the lower threshold, no packet is dropped. When the queue size is in between the two thresholds, packets are dropped randomly according the drop probability function. Beyond the upper threshold, any incoming packets are dropped (Figure 1). Each function is assigned to one class, i.e. *in* or *out*. Hence, to give better service to *in* packets than *out* packets we need to set carefully the parameters of the drop probability functions. We are unaware of any previous work that explicitly examines the problem of how to set those parameters in order to satisfy a traffi c contract.

Many previous studies have been carried out to characterize the steady state and the transient behavior of the Random Early Detection (RED) in presence of TCP traffi c. Kuusela *et al.* [11] used differential equations to describe the dynamics of a RED queue in interaction with idealized TCP sources. Firoiu *et*

*al.* [7] presented a method to configure RED for congestion control based on TCP flows. In [16], Ziegler *et al.* developed a simple model enhanced by simulations to provide guidelines to set RED parameters in order to avoid severe oscillations of the queue size. Bu *et al.* [3] used a fixed point method to find the average queue length in RED routers. They focus on the early drop behavior of the RED queue. Besides, their model is applicable when congested routers are known and when the queue size oscillates between the *min* and the *max* thresholds of the RED algorithm.

The RIO mechanisms have also been analyzed in the literature. May *et al.* [15] studied analytically and by simulation the impact of RIO on the throughput of UDP-like traffic with two classes of packets. Kuusela *et al.* [12] used an ordinary differential equation approximation to describe the evolution of the expectations of the exponentially averaged queue lengths. They consider two Poisson streams from each class as input traffic. Fang [6] used extensive simulations to study the throughput of Internet-like traffic in presence of RIO routers.

In this paper we present an approach to analyzing the performance characteristics of TCP sessions in networks with RIO routers. We consider the case with a large number of TCP sessions which use token buckets for marking *in* and *out* packets at the entrance of the network. Under some simplifying assumptions we derive a set of equations that govern the evolution of these TCP sessions and the routers under consideration. We then solve these equations numerically using a fixed point method. We validate this model through simulations which show that less than $5\%$ error is achieved in the most cases. We then use this method to study the impact of the RIO parameters on the performance characteristics of TCP sessions. Our results show that the loss probability threshold of *out* packets has a significant effect on the TCP throughput and on the average queue length. Setting this parameter consists in trading off between the network utilization and the fairness among TCP connections. Our results also show that Tail Drop mechanism is particularly suitable for *in* packets to satisfy various QoS constraints.

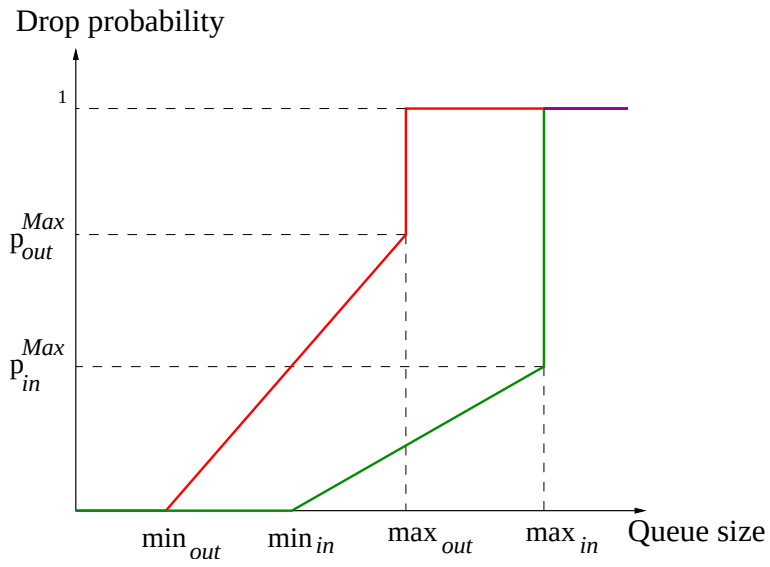


Fig. 1. Example of RIO loss probability functions

The paper is organized as follows. In section II below we describe the network model and the notation. In section III we derive the set of equations relating the performance metrics under investigation, then we present the fixed point method for the numerical resolution of these equations. In section IV we report validation results of our approach obtained through NS simulations. In section V we investigate the effect of the RIO parameters on the QoS and fairness of the TCP sessions. Conclusions are provided in section VI.

## II. NETWORK MODEL

We consider a RIO router in the network fed by $N$ long-lived TCP connections. In this paper, for simplicity of exposition, we shall assume that this router is the bottleneck router of these TCP sessions so that the round trip time (RTT) of these sessions are represented by the propagation delays and the queueing delay incurred in the router under consideration. However, as we shall see later on in the paper, the analysis techniques can be extended to the case of any arbitrary number of routers.

The router is modeled by a FIFO queue with RIO, see Figure 2. The RTTs are arbitrary and can be different for different TCP sessions. Every TCP session uses a token bucket (TB) to mark the packets *in* or *out*. The token bucket parameters are the rate generation of the tokens $r^i$ and the buffer size of the bucket $\sigma^i$. Roughly speaking, if the instantaneous rate of the connection is less than $r^i$, the packets are marked *in*, otherwise they are marked *out*. The buffer $\sigma^i$ allows for burst absorption. More detailed descriptions of token buckets mechanisms can be found in, e.g., [10] and [14].

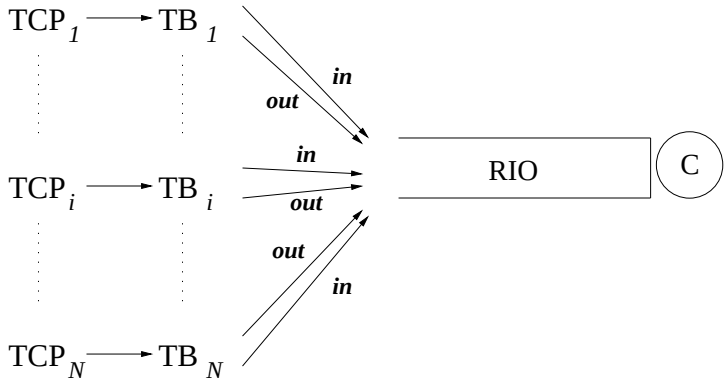


Fig. 2. The simplified network model

We now introduce the notation that will be used in the next section. The superscript $i$ refers to the connection number and the subscript $c$ refers to the class of packets (*in* or *out*).

- $N$: Total number of TCP connections
- $T^i$: Throughput of TCP connection $i$
- $T$: Total Throughput, i.e. $T = \sum_{i=1}^{N} T^i$
- $T_c^i$: Throughput of class $c$ packets of TCP connection $i$
- $T_c$: Total throughput of class $c$ packets
- $(r^i, \sigma^i)$: token bucket parameters of TCP connection $i$
- $C$: router capacity
- $D^i$: the round-trip propagation delay of connection $i$
- $RTT^i$: the round-trip time of TCP connection $i$
- $RTO^i$: the retransmission timeout of connection $i$
- $\bar{q}$: the average queue length in the router
- $\bar{p}_c$: average loss probability of class $c$ packets
- $p_c()$: loss probability function of class $c$ packets, where

$$p_c(q) = \begin{cases} 0 & if\ q \leq min_c \\ p_c^{max} \frac{q - min_c}{max_c - min_c} & if\ min_c < q < max_c \\ 1 & if\ q \geq max_c \end{cases}$$

Where $q$ is the instantaneous queue length. Indeed, we will not consider throughout this paper the effect of the exponential weighted moving averaging introduced with RED mechanism [8].

## III. Characteristic Equations and Computational Scheme

In this section we shall first derive a set of equations that relate the state variables of the TCP sessions and those of the router. We then use the fixed point method to numerically compute these performance metrics. Such an approach was first used to study a best-effort network in presence of TCP in [5].

### A. Equations of TCP Dynamics

Consider first the dynamics of the TCP control protocol. It follows from equations we developed in [14] that we can derive the expressions of the expected throughput of both *in* and *out* packets of each TCP connection as a function of the average loss probabilities, the token bucket parameters, the round-trip time and the retransmission timeout. For each connection $i$ we have:

$$T_{in}^i = \frac{(1-p_{TO}^i)S_{in}^i + p_{TO}S_{in}^{\sigma,i} + p_{TO}^i R^i}{(1-p_{TO}^i)Y^i + p_{TO}^i Y^{\sigma,i} + p_{TO}^i Z^i} \tag{1}$$

$$T_{out}^i = \frac{(1-p_{TO}^i)S_{out}^i + p_{TO}S_{out}^{\sigma,i}}{(1-p_{TO}^i)Y^i + p_{TO}^i Y^{\sigma,i} + p_{TO}^i Z^i} \tag{2}$$

where

- $S_c^i$ denotes the total number of class $c$ packets sent in a congestion avoidance period following a triple duplicate loss, $c \in \{in,\ out\}$; $Y^i$ is the duration of such a period;
- $S_c^{\sigma,i}$ is the total number of class $c$ packets sent in a congestion avoidance period following a timeout loss event, $c \in \{in,\ out\}$; $Y^{\sigma,i}$ is the duration of such a period;
- $p_{TO}^i$ the probability that a loss is detected by a Timeout;
- $Z^i$ the duration of the timeout retransmission period;
- $R^i$ the number of packets sent to retransmit a lost packet.

The formulae of these expected parameters can be found in [13]. A detailed analysis is presented in [14].

We have also $T = \sum_{i=1}^N T^i$, $T_{in} = \sum_{i=1}^N T_{in}^i$ and $T_{out} = \sum_{i=1}^N T_{out}^i$.

### B. Equations of the RIO Router

For simplicity of analysis, we shall assume that the traffic entering the router is Poisson with rate equal to the sum of the throughputs of all TCP connections. We shall also assume that the service times in the router are exponentially distributed with parameter equal to the capacity of the router. The Poisson assumption seems to be justified when the TCP connection rate increases [4]. These assumptions allow us to derive analytically the expressions relating the average queue length and the loss probabilities. Under such assumptions we can also determine the departure process out of the queue so that we can resolve a system of multiple routers.

It is easy to see that the queue length is a birth-death process so that it has the stationary distribution expressed as

$$\pi(i) = \pi(0)\left(\frac{T}{C}\right)^i \prod_{j=0}^{i-1}\left[1-p(j)\right], \quad i = 1 \cdots max_{in} \tag{3}$$

where

$$p(i) = \frac{T_{in}p_{in}(i) + T_{out}p_{out}(i)}{T_{in}+T_{out}}.$$

We assume, without loss of generality, that the minimal condition to give preferential service to *in* packets is $max_{out} \leq max_{in}$. Hence, $\pi(0)$ is given by the normalization equation

$$\pi(0) = \left[1 + \sum_{i=1}^{max_{in}}\left(\frac{T}{C}\right)^i \prod_{j=0}^{i-1}\left[1-p(j)\right]\right]^{-1}.$$

We shall also approximate the average loss probability of *in* packets and *out* packets observed for each connection $\bar{p}_{in}^i$ and $\bar{p}_{out}^i$ by the loss probabilities observed in the RIO queue $\bar{p}_{in} = \sum_{i=0}^{max_{in}} p_{in}(i)\pi(i)$ and $\bar{p}_{out} = \sum_{i=0}^{max_{in}} p_{out}(i)\pi(i)$. However, the average loss probability observed by each connection including both *in* and *out* is different from the loss probability observed at the RIO queue.

The next two equations determine the round-trip time and the retransmission timeout.

$$RTT^i = D^i + \frac{\bar{q}+1}{C} \quad (4)$$

$$RTO^i = RTT^i + \alpha^i \quad (5)$$

This last equation is an approximation to estimate the $RTO$. Actually $RTO$ is estimated with $RTO = SRTT + 4RTTVAR$ where SRTT is a smoothed estimate of RTT and RTTVAR is a smoothed estimate of the variation of RTT. We observed in our simulations that this variation is negligible compared to the Round Trip Time. We observed also that the average RTT plus the preset lower bound is a good approximation of the average RTO. Since in many TCP implementations, e.g. BSD, the RTO is lower bounded by 1 second [2] we choose $\alpha^i$ equal to 1 second.

### C. Fixed Point Method

The above equations are now solved using a fixed point method. Each iteration of the algorithm contains only two steps. In the first step we use the values of the throughput to determine a new load of the system. This load determines a new stationary distribution of the queue length and new values of the loss probabilities. In the second step we use the formulae of TCP throughput to update the throughput of one connection at a time which leads to a small update of the total throughput. All connections contribute on the total throughput after N iterations. This method is necessary in order to avoid a significant increase/decrease of the throughput at each iteration which could result in oscillations.

## IV. Model Validation

In this section we validate our model using the NS-2 simulator. We simulate $100$ TCP connections with $1Mb/s$ access link for each connection. The bottleneck link capacity is set to $48Mb/s$. The Round-Trip propagation delays are chosen between 100 ms and 300 ms such that the Round-Trip propagation delay $= 100 + 2i$ $i \in 0 \cdots N-1$. $r^i$ and $\sigma^i$ are fixed to $40$ *packets/sec* and $20$ *packets*. The packet size is equal to $1000$ bytes. We run simulations for 3 scenarios of the RIO configuration corresponding to whether the loss probability functions fully overlap, partially overlap or do not overlap. Due to space constraints, we only present results of the fully overlapped case. A detailed presentation is available in [13].

We set $min_{in} = min_{out} = 20$ and $max_{in} = max_{out} = 100$. We choose 0.01 and 0.06 for $p_{in}^{max}$ and $p_{out}^{max}$, respectively. Table I compares the results obtained by a 30-minutes simulation with the numerical results obtained by the model. We observe that our model predicts the average parameters very accurately. We should notice also that we run many other simulations where we vary the loss probability thresholds and we observed that the *relative* error percentage is less than 5% for the throughput.

TABLE I
THE FULLY OVERLAPPED CASE

| | Analytical | Simulation |
|---|---|---|
| $T$ $(pkts/sec)$ | 5927.97 | 5810.57 |
| $T_{in}$ $(pkts/sec)$ | 3764.09 | 3707.57 |
| $\bar{p}_{in}$ | 0.004134 | 0.003260 |
| $\bar{p}_{out}$ | 0.015095 | 0.015937 |
| $\bar{q}$ $(pkts)$ | 33.79 | 31.18 |

## V. Impact of RIO Parameters on QoS and Fairness

In this section we use our analytical model to study the ways to set the RIO parameters and the effect of these parameters on the QoS and fairness of TCP sessions. The performance metrics under consideration are essentially the achieved throughput, the loss probability and the network delays. These quantities could be the key elements in a Service Level Agreement (SLA). Following the model in section II, there are eight parameters that could have effect on these performance metrics. In this section we will focus on the RIO parameters. In the following numerical examples we consider mainly the same network confi guration used for simulation in the previous section.

### *A. Throughput*

In this DiffServ framework, each TCP connection attempts to achieve its reservation rate. In other words the constraint is $throughput^i \geq r^i_{SLA}$. In the experiment presented in section IV, all the connections have throughputs above the reservation rate. The throughput of TCP connection is inversely proportional to the RTT. A TCP connection with a large RTT may not be able to achieve its target rate. To illustrate that we add 5 connections to the set of 100 connections. We assign the following RTTs 0.4s, 0.5s, 0.6s, 0.7s and 0.8s. Figure 3 shows that neither of the connections could achieve the reservation rate. This is due to the fact that connections with small RTTs are very aggressive and they send many *out* packets beyond the reservation.

In order to reduce this unfairness, one solution is to increase the drop probability of *out* packets. Thereby, causing a decrease in the throughput of large RTT TCP connections and hence an increase in the throughput of small RTT connections. Figure 3 shows that by setting $p^{max}_{out}$ to 1, 4 connections achieve the reservation. Generally, if we set RIO parameters such that $\bar{p}_{out} = 1$ and if there are still connections which cannot achieve the reservation, then either the capacity of the link or the buffer thresholds should be re-provisioned.

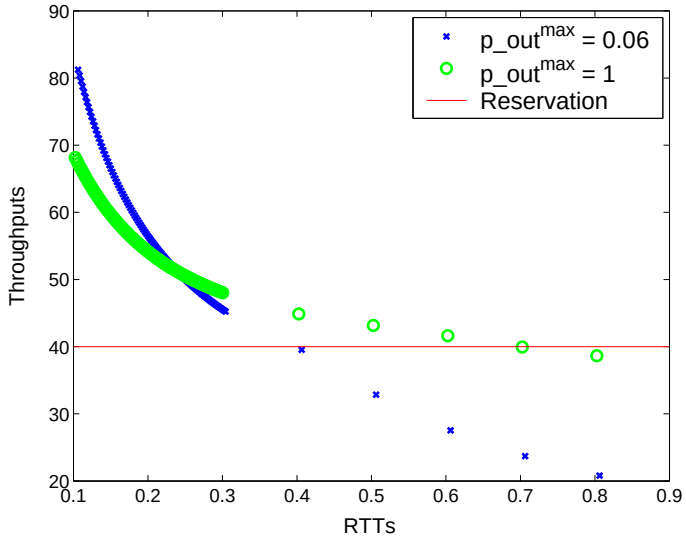


Fig. 3. Throughput of TCP connections with variable RTTs

### *B. Delay*

We can transform the constraints on the delay and the average delay to the queue length and the average queue length. To satisfy the strict constraint $delay \leq delay_{SLA}$, one simple way is to set $max_{in} = delay_{SLA} * C$. Then we can focus on the other parameters to satisfy other constraints.

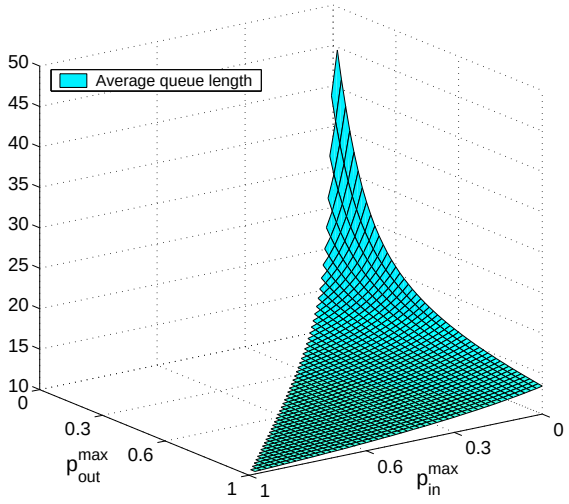


Fig. 4. Effect of the loss probability thresholds on the average queue length

Figure 4 illustrates the average length vs. $p_{in}^{max}$ and $p_{out}^{max}$. We can notice that when $p_{in}^{max} = 0$, i.e. when TD mechanism is deployed, the average is higher. Also that the $p_{out}^{max}$ controls more the queue utilization. If $p_{out}^{max}$ decreases, we increase signifi cantly the utilization. This is due to the fact that we allow *out* packets to fi ll the buffer of the router. This is opposite to the fairness goal mentioned in the previous paragraph since for that goal we need to increase the $p_{out}^{max}$. To study more the impact of $\bar{q}$, we keep $p_{in}^{max}$ and $p_{out}^{max}$ fi xed and we vary $min_{in}$ and $min_{out}$ to cover all the cases where the drop functions partially overlapped. We do that for two values of $p_{out}^{max}$. We see from Figure 5 that the utilization does not rely much on $min_{in}$, though the performance is a little better when $min_{in} = max_{in}$.

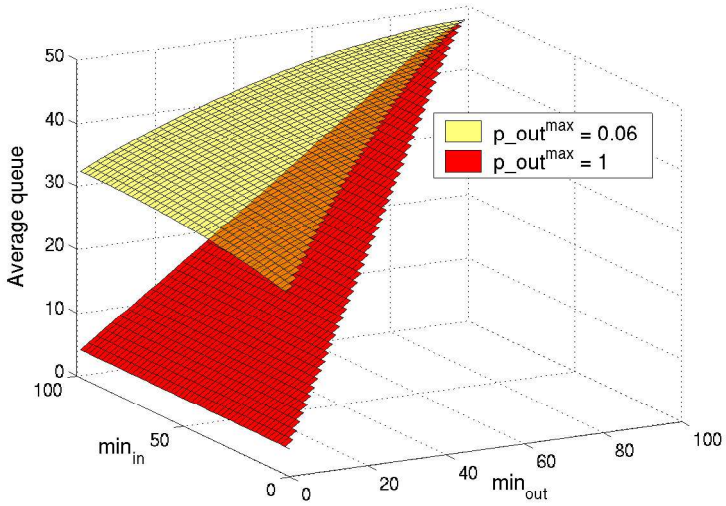


Fig. 5. Effect of the minimum thresholds on the average queue length

### C. Drop probability

We fi rst consider again the fully-overlapped case of section IV. Suppose that in the SLA we want to ensure that the loss probability of *in* packets $\bar{p}_{in}$ is less than $p_{SLA}$. Figure 6 shows the possible values of the two loss probability thresholds of RIO ($p_{in}^{max}$, $p_{out}^{max}$) that achieve the $p_{SLA} = 0.006$. Note that the

condition $p_{in}^{max} = 0$ is not sufficient to ensure the desired $p_{SLA}$. In this scenario, we need to set $p_{out}^{max}$ to at least $2.49\%$.

More generally, we can determine the upper and lower bounds of $\bar{p}_{in}()$. Figure 7 plots the average loss probability of *in* packets as function of the loss probability thresholds $p_{in}^{max}$ and $p_{out}^{max}$. The lower and upper bounds are mentioned in the Figure. If $p_{SLA}$ is greater than the upper bound, we can satisfy the constraint $\bar{p}_{in} \leq p_{SLA}$ and thus achieve the SLA for all $(p_{in}^{max}, p_{out}^{max})$ settings. In this case, we can set $(p_{in}^{max}, p_{out}^{max})$ to control the level of differentiation between *in* packets and *out* packets. If $p_{SLA}$ is less than the lower bound we can not achieve the SLA. However, if $p_{SLA}$ is between the two bounds, we can achieve only if we set correctly the parameters $(p_{in}^{max}, p_{out}^{max})$. We notice that the operating point should be close to the line which delimits the feasible region in order to increase the utilization of the network, i.e., we should choose $p_{out}^{max}$ as low as possible and $p_{in}^{max} = 0$ (which means that the TD is in place).

An important fact is that decreasing $p_{in}^{max}$ does not contradict the rate goals, i.e., for all the QoS constraints, a very small value of $p_{in}^{max}$ yields good performance. Also we notice that in the partially overlapped case, we obtain better performance in term of loss probability and utilization when $min_{in} = max_{in}$ which corresponds to TD too.

In contrast, $p_{out}^{max}$ has significant and *different* effects on the overall QoS. For example, we can configure differently the AF classes. If we know that in one AF class we will have TCP connections of almost the same RTTs we choose a low value of $p_{out}^{max}$. In an other AF class we can aggregate heterogeneous TCP connections that have tighter constraints on the throughput.

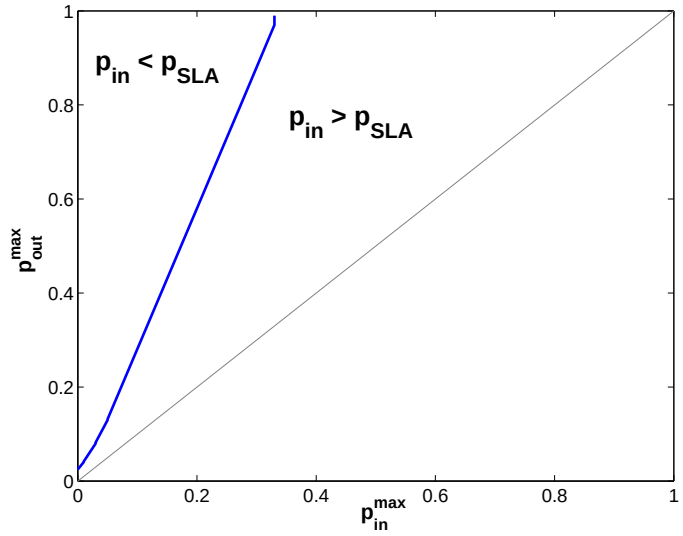


Fig. 6. Feasible region for the loss probability constraint

## VI. Conclusion

In this paper we developed and validated a method to analyze the steady state behavior of long lived TCP connections in interaction with RIO router. Using the model we examined the impact of RIO thresholds on the QoS and fairness of TCP connections. We have shown that the loss probability threshold of *out* packets has a significant effect on the TCP throughput and on the average queue length. Setting this parameter consists in trading off between the network utilization and the fairness among TCP connections. We have also shown that Tail Drop mechanism is particularly suitable for *in* packets to satisfy various QoS constraints.

Our method can easily be extended to handle the case with arbitrarily connected routers. We shall also study the case with UDP connections competing with TCP. Another question interesting to inves-

tigate is the short-lived TCP (or HTTP-like) sessions. In particular, it is interesting to see whether such TCP connections could still achieve their reservation rates in the over-booking case, and so, under what conditions.

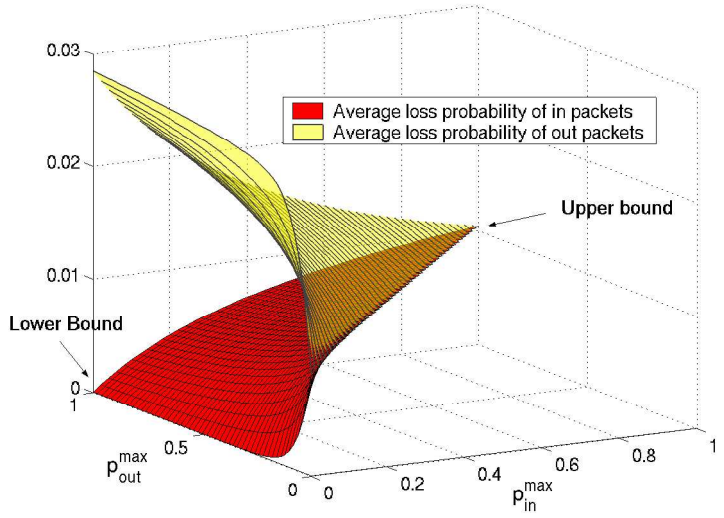


Fig. 7. Effect of loss probability thresholds on the average loss probabilities

## REFERENCES


[1] Diffserv discussion archive: The relation between pdb's and the slss. *http://www1.ietf.org/mail-archive/working-groups/diffserv/*, Apr 2001.

[2] Mark Allman and Vern Paxson. On estimating end-to-end network path properties. *ACM SIGCOMM'99*, 1999.

[3] Tian Bu and Don Towsley. Fixed point approximations for tcp behavior in aqm network. *ACM Sigmetrics 2001, Cambridge, MA USA*, June 2001.

[4] Jin Cao, William S. Cleveland, Dong Lin, and Don X. Sun. On the nonstationarity of internet traffi c. *in Proceedings of Sigmetrics*, 2001.

[5] Claudio Casetti and Michela Meo. A new approach to model the stationary behavior of tcp connections. *Proceedings of the 2000 IEEE INFOCOM*, Mar 2000.

[6] Wenjia Fang. Differentiated services: Architecture, mechanisms and an evaluation. *PhD thesis, The university of Princeton, Departement of Computer Science*, Nov, 2000.

[7] Victor Firoiu and Marty Borden. A study of active queue management for congestion control. *INFOCOM 2000*, Mar 2000.

[8] Sally Floyd and Van Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, August 1993.

[9] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski. Assured forwarding phb group. *Request For Comment: 2597*, June 1999.

[10] Juha Heinanen and Roch Guerin. A two rate three color marker. *Request For Comment: 2698*, September 1999.

[11] P. Kuusela, P. Lassila, J. Virtamo, and P. Key. Modeling red idealized tcp sources. $9^{th}$ *IFIP Conference on Performance Modeling and Evaluation of ATM & IP Networks. Budapest*, June 2001.

[12] P. Kuusela and J. T. Virtamo. Modeling red with two traffi c classes. *in Proceedings of Fifteenth Nordic Teletraffic Seminar, Lund, Sweden. pp. 271-282*, August 2000.

[13] Naceur Malouch and Zhen Liu. Performance analysis of tcp with rio routers. *Research Report, INRIA Sophia-Antipolis, May 2002. Available at http://www-sop.inria.fr/rapports/sophia/RR-4469.html*.

[14] Naceur Malouch and Zhen Liu. On steady state analysis of tcp in networks with differentiated services. *in Proceedings of Seventeenth International Teletraffic Congress, ITC'17.*, December 2001.

[15] Martin May, Jean-Chrysostome Bolot, Alain Jean-Marie, and Christophe Diot. Simple performance models of differentiated services schemes for the internet. *IEEE Infocom'99, New York, NY*, March 99.

[16] Thomas Ziegler, Serge Fdida, and Ulrich Hofmann. Stability criteria for red with bulk-data tcp traffi c. *IFIP ATM & IP Working Conference, Budapest*, June 2001.