

Buffer Requirements at ECN-Capable RED Gateways to Minimize Packet Losses

Haowei Bai*, David Lilja†

*Honeywell Labs
3660 Technology Drive
Minneapolis, MN 55418, USA
E-mail: haowei.bai@honeywell.com

†Department of Electrical and Computer Engineering
University of Minnesota
200 Union St. SE
Minneapolis, MN 55455, USA
E-mail: lilja@ece.umn.edu

Abstract—Explicit Congestion Notification (ECN) used with Random Early Detection (RED) is known to reduce delays for low-bandwidth delay-sensitive Transport Control Protocol (TCP) connections by avoiding unnecessary packet drops and retransmissions. However, choosing the optimum values for buffer size and RED parameters is still an ongoing research topic. In this paper, we present a model to determine the buffer size and RED parameters to minimize packet losses at a RED gateway.

I. INTRODUCTION

TCP sources implicitly interpret absence of acknowledgements as signal of network congestion. ECN (Explicit Congestion Notification) [1] was proposed to explicitly inform sources of network congestion, without the sources having to wait for either a retransmit timer timeout or three duplicate acknowledgements (ACKs) to infer a packet loss. ECN has been recommended to be used in conjunction with Random Early Detection (RED) [2] in the next generation Internet routers.

ECN-capable RED gateways use an exponential weighted moving average to calculate an average queue size from the instantaneous queue size, and two thresholds (*minimum* and *maximum*) to determine whether an arriving packet should be dropped. If the average queue size is greater than the maximum threshold, the packet is *dropped*. If the average queue size is between the minimum and the maximum thresholds, the packet is marked with a probability as a Congestion Experienced (CE) packet.

Packet losses due to the average queue size exceeding the maximum queue size at a RED gateway hurt TCP performance. The *objective* of this paper is to determine if packet losses at a router can be eliminated by optimally

dimensioning the buffer and selecting the two RED thresholds. Some preliminary work in this area has been done by authors in [3], [4]. The authors in [3], instead of using a linear drop function and two thresholds as in RED, used only one threshold to mark packets; a packet is marked as CE with a probability of one if the average queue level exceeds the threshold. The study therefore, does not apply to RED gateways. The authors in [4] did not show the effect of RED parameters on packet drops at a RED router.

In this paper, we develop a model to analyze the performance of ECN mechanism in RED gateways. Our main contribution is that we derive approximate expressions for the maximum buffer size requirement and the maximum threshold of a RED gateway to minimize packet loss. The *significance* of our study is that the buffer size, and consequently the queuing delay, could be much smaller than what has been proposed by previous researchers.

II. NOTATIONS

We consider a RED gateway fed by multiple TCP sources as shown in Fig. 1 for two sources. The link connecting routers R1 and R2 is the bottleneck link which causes congestion at R1. The sources, destinations and the RED gateways use ECN for end-to-end congestion control. The following notations will be used in our model:

- $Q(t), Q(t)_{max}$: *Instantaneous* and *maximum instantaneous* queue sizes respectively at the RED gateway at time t .
- \bar{Q}, \bar{Q}_{max} : *Average* and *maximum average* queue sizes respectively at the RED gateway.

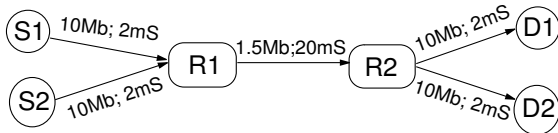


Fig. 1. Simulation topology.

- ω : Weighting factor for calculating \bar{Q} .
- $p(t)$: Marking probability at the RED gateway at time t .
- min_{th} , max_{th} : Minimum and maximum thresholds respectively of a RED gateway.
- m : total number of TCP flows.
- $W_i(t)$: Window size of the i^{th} TCP flow at time t , $t \geq 0$, $i = 1, \dots, m$.
- $SStresh_i$: Slow Start threshold for the i^{th} TCP flow, $i = 1, \dots, m$.
- r_i : Round Trip Time (RTT) for the i^{th} TCP flow, $i = 1, \dots, m$. r_i is replaced by r when all the RTTs are same.
- $\bar{\mu}_i$: Average share of bottleneck link bandwidth of the i^{th} TCP flow, $i = 1, \dots, m$.
- μ : Bandwidth of bottleneck link which is given by $\mu = \sum_{i=1}^m \bar{\mu}_i$.
- $T[1]$: Waiting time for the first marking event after the average queue size exceeds min_{th} [3].
- β_i : Number of window size increases during time $T[1]$ for the i^{th} TCP flow, $i = 1, \dots, m$.
- τ_i : Propagation delay from source i to the RED gateway, $i = 1, \dots, m$.
- t_0 : Time when the first packet is marked at the RED gateway [3].
- t_1 : Time when the last packet, which was sent just before the first window size reduction, arrives at the RED gateway [3].

For every packet arrival, the RED gateway estimates \bar{Q} using the following exponential weighted moving average algorithm

$$\bar{Q} \leftarrow (1 - \omega)\bar{Q} + Q(t)\omega, \quad (1)$$

and then calculates the packet marking/dropping probability $p(t)$ using

$$p(t) = \begin{cases} 0, & 0 \leq \bar{Q} < min_{th}; \\ \frac{\bar{Q} - min_{th}}{max_{th} - min_{th}} max_p, & min_{th} \leq \bar{Q} \leq max_{th}; \\ 1, & \bar{Q} > max_{th}. \end{cases} \quad (2)$$

III. ASSUMPTIONS

We make the following assumptions regarding RED gateways and TCP sources in our analytical model for minimizing packet losses in Secs. IV and V.

- For small ω (as suggested in [2]), \bar{Q} varies very slowly, so that consecutive packets are likely to experience the same marking probability [5].
- The random packet marking of packets in flow i is described by a Poisson process with time varying rate $\lambda_i(t) = p(t)W_i(t)/r_i(t)$ [6]. Accordingly, the waiting time ($T_i[n]$) for the n -th marking event of flow i , which is given by $T_i[n] = \sum_{k=1}^n X_i(k)$, is a Gamma distributed random variable. $X_i(k)$ is the time interval between $(k-1)$ and k -th marking events for flow i . Specifically, the expected value of the waiting time for the first marking event is $E[T_i[1]] = 1/\lambda_i(t)$.
- All TCP sources start sending at the same time, and all packet are of the same size (as used in [3]). The queue size is measured in packets.

IV. MAXIMUM BUFFER SIZE

Packet drops at an ECN-capable RED gateway are either due to buffer overflows ($Q(t)$ is equal to the buffer size) or $\bar{Q} > max_{th}$. In this section, we estimate the buffer size required for minimizing packet losses.

The congestion window size during the slow start phase increases very quickly. The average queue size (being the output of a low pass filter) of a RED gateway can not follow the quick change of $Q(t)$; as a result \bar{Q} stays less than min_{th} . Therefore, $Q(t)$ reaches the maximum value when the packet leaving the source at $t - \tau_i$ reaches the RED buffer. When this packet left the source, $W_i(t - \tau_i) = SStresh_i$ for $i = 1, 2, \dots, m$; the queue size is smaller when the sources are in congestion avoidance [3]. For m TCP flows, $Q(t)_{max}$ can be expressed as the output of a system with processing capacity of $\sum_{i=1}^m r_i \bar{\mu}_i$ and the maximum input rate when sources reach their slow start threshold.

$$Q(t)_{max} = \sum_{i=1}^m (W_i(t - \tau_i) - r_i \bar{\mu}_i) = \sum_{i=1}^m (SStresh_i - r_i \bar{\mu}_i). \quad (3)$$

$Q(t)_{max}$, as given by the above equation, is therefore, the **buffer size required to minimize packet loss at the RED gateway**.

V. max_{th} FOR RED GATEWAYS

Authors in [2] have recommended $max_{th} = 3 \times min_{th}$. In this section, we setup a model to estimate max_{th} for minimizing losses at the RED buffer. We start with the recommended RED parameter values, and end with values suggested by our model.

Fig. 2 shows our analytical model of a RED gateway. When the average queue size is in the steady-state condition (during which the sources are in the congestion avoidance phase), the instantaneous queue size at time

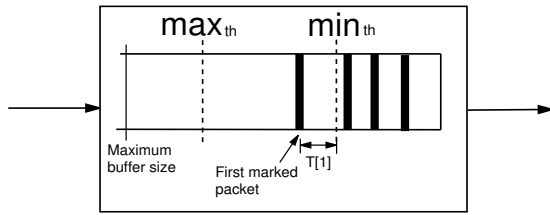


Fig. 2. Analytical model of a RED gateway.

t_0 is

$$Q(t_0) = \min_{th} + \sum_{i=1}^m \beta_i, \quad (4)$$

where β_i can be calculated as

$$\beta_i = \frac{E[T[1]]}{r_i} = \frac{1}{\lambda_i(t)r_i} = \frac{1}{p(t)W_i(t)}, i = 1, \dots, m. \quad (5)$$

Since the difference between t_0 and t_1 is one RTT, and the window size of a source is increased by one per RTT during the congestion avoidance phase, the instantaneous queue size at time t_1 can be expressed as

$$Q(t_1) = \min_{th} + \sum_{i=1}^m (\beta_i + 1). \quad (6)$$

The average queue size is estimated using an exponential weighted moving average as shown in Eqn. (1). If time is discretized into time slots with each slot being equal to one RTT, the RED's average queue size estimation algorithm at the k -th slot can be expressed as

$$\bar{Q}[k+1] = (1-\omega)\bar{Q}[k] + Q[k]\omega. \quad (7)$$

In practice, ω is very small, and the congestion window size increases by one every RTT during the congestion avoidance phase. Therefore, before the first marking event happens (i.e., no congestion control) it is reasonable to consider both the instantaneous queue size and the average queue size to be constant within a very short time period (see the first assumption in Section III). Thus, by plugging $Q(t_1)$ (slot k is equal to t_1 in time) into Eqn. (7) and assuming that the average queue sizes during the two previous consecutive time slots are the same, the average queue size estimated at time t_1 can be solved iteratively, which is

$$\bar{Q}_{max} = \bar{Q} = \min_{th} + \sum_{i=1}^m (\beta_i + \omega). \quad (8)$$

The first marking event is followed by many random ECN marking events, which make TCP sources adjust their congestion window sizes. The average queue size stays at a certain level smaller than the average queue size at time t_1 , as will be shown by our simulation results

TABLE I
COMPARISON BETWEEN SIMULATION AND ANALYTICAL RESULTS
FOR MAXIMUM BUFFER SIZE (PACKETS).

Sim cases	SSt_{h1}	SSt_{h2}	$r = r_1 = r_2$ (ms)	μ (Mbps)	$Q(t)_{max}$	
					Anlyt	Sim
Case 1	15	15	59	1.5	19	18
Case 2	15	20	59	1.5	24	23
Case 3	15	15	99	1.5	12	13

TABLE II
COMPARISON BETWEEN SIMULATION RESULTS AND ANALYSIS
RESULTS FOR MAXIMUM THRESHOLD (I.E., MAXIMUM AVERAGE
QUEUE SIZE).

Sim cases	ω	\min_{th} (Pkts)	\max_{th} (Pkts)	\max_p	\bar{Q}_{max} (Pkts)	
					Anlyt	Sim
Case 1	0.002	5	15	0.1	7.3	7.6
Case 2	0.002	5	15	0.2	6.7	7.1
Case 3	0.002	7	21	0.1	9.6	9.9

in Fig. 3 later. Therefore, **Eqn. (8) gives the maximum average queue size for minimizing packet losses, i.e. this is our suggested value of \max_{th} .**

VI. SIMULATION VERIFICATION

We have simulated the topology of Fig. 1 using *ns-2*. Packet size of 1000 bytes have been used in our simulation. Other parameters of the simulation vary depending on the three cases we have studied as described below.

A. Verification of maximum buffer size

To verify the maximum buffer size suggested by our model in Sec. IV, we have run simulations for three different cases with different values of r and SSt_{thresh} . We have measured $Q(t)_{max}$ and the results are shown in Table I along with $Q(t)_{max}$ predicted by simulation. It is seen that values from simulation and analytical model are close, thereby *validating the maximum buffer size suggested by our analytical model.*

B. Verification of \max_{th}

The RED parameters shown in Table II are used to verify the correctness of the value of \max_{th} suggested by our model in Sec V. Case 1 uses recommended RED parameters. To make the different cases comparable, we choose RTT of all TCP connections to be the same (59 mS). It is seen that \bar{Q}_{max} (which we have suggested in Eqn. (8) as the value to be used for \max_{th}) obtained from our analytical model agrees with that obtained from simulation.

The congestion window size, instantaneous queue size, and average queue size for Case 1 in Table II are shown in Fig. 3 and Fig. 4. We can see that the

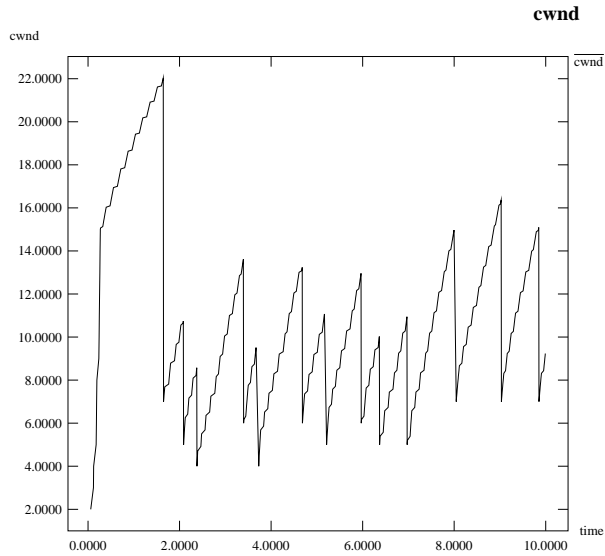


Fig. 3. Congestion window evolution for source 1 in case 1.

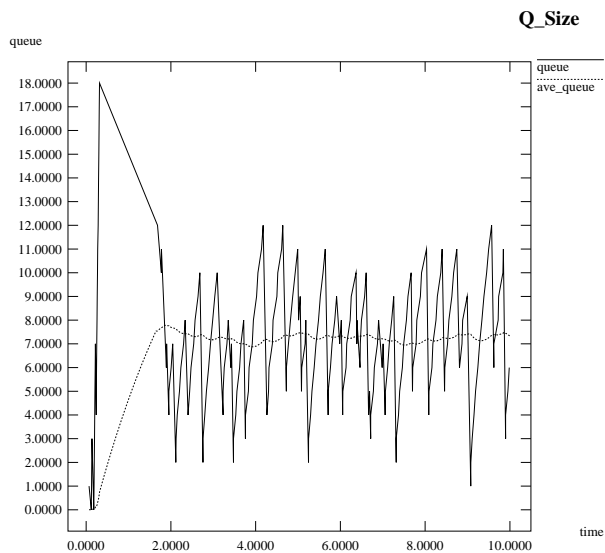


Fig. 4. Instantaneous queue size and average queue size for case 1.

instantaneous queue size is maximum when the congestion window size reaches the slow start threshold. The average queue size is maximum just before the first marking event at time $t = 1.9S$. This proves the validity of our statement in Sec. IV about the instantaneous queue size reaching the maximum. The same observation can be made for the other cases; due to space limitations we do not present the results for the other cases.

C. Verification of the effectiveness

The third simulation shown in Table III is used to verify the effectiveness our estimated max_{th} . Case 1 in Table III uses recommended RED parameters. It works perfectly to control the TCP congestion without

TABLE III
SIMULATION RESULTS OF PACKET DROPS WITH DIFFERENT RED
PARAMETERS.

Simulation cases	ω	min_{th} (Packets)	max_{th} (Packets)	max_p	Pkt drops
Case 1	0.002	5	15	0.1	No
Case 2	0.002	5	8	0.1	No
Case 3	0.002	5	7	0.1	Yes

unnecessary packet drops. However, case 2 in Table III uses our proposed max_{th} , which is much smaller than the recommended value. The simulation result shows it achieves the same congestion control effects as the recommended value does. In addition, one of the benefits of using our estimated value of max_{th} is that the queuing delay and buffer size can be significantly reduced (though it is not shown by simulation here). Case 3 is used to verify the validity of our proposed model. The value of max_{th} is reduced from 8 estimated using our proposed model to 7. The result of this reduction is that the objective of no packet drops can never be achieved.

VII. CONCLUSION

We have setup an analytical model to estimate the buffer size requirement as well as the maximum threshold of a RED gateway. The analytical model has been proved to be accurate by simulations. The significance of our study is that the buffer size, and consequently the queuing delay, could be much smaller than what has been proposed. The using of more complex network topology for analysis as well as simulation verification could be one of the extensions of this work.

REFERENCES

- [1] S. Floyd, "TCP and explicit congestion notification," *ACM Computer Communication Review*, vol. 24, no. 5, pp. 10–23, October 1994.
- [2] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transaction on Networking*, vol. 1, pp. 397–413, August 1993.
- [3] C. Liu and R. Jain, "Improving explicit congestion notification with the mark-front strategy," *Computer Networks*, vol. 35, no. 2-3, pp. 185–201, February 2001.
- [4] S. Kunniyur and R. Srikant, "End-to-end congestion control schemes: Utility functions, random losses and ECN marks," *INFOCOM*, Tel Aviv, Israel, pp. 1323–1332, March 2000.
- [5] T. Bonald, M. May, and J.C. Bolot, "Analytic evaluation of RED performance," *INFOCOM*, Tel-Aviv, Israel, pp. 1415–1424, March 2000.
- [6] V. Misra, W.B. Gong, and D. Towsley, "Fluid-based analysis of a network of AQM routers supporting TCP flows with an application to RED," *ACM SIGCOMM*, Stockholm, Sweden, pp. 151–160, 2000.