Computer Communications 31 (2008) 3789–3798

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Simulation studies on router buffer sizing for short-lived and pacing TCP flows

Go Hasegawa *, Takeshi Tomioka, Kentarou Tada, Masayuki Murata

Graduate School of Information Science and Technology, Osaka University, 1-32 Yamadaoka, Suita, Osaka 565-0871, Japan

ARTICLE INFO

Article history:
Available online 11 May 2008

Keywords:
Router
Buffer size
TCP
Bandwidth–delay product
Pacing TCP

ABSTRACT

Traditionally, the size of router buffers is determined by the bandwidth–delay product discipline (*normal discipline*), which is the product of the link bandwidth and average round-trip time (RTT) of flows passing through the router. However, recent research results have revealed that when the number of flows is sufficiently large, the buffer size can be decreased to the bandwidth–delay product divided by the square-root of the number of flows (*sqrtN discipline*), without introducing under-utilization of the link bandwidth. This assertion has been verified primarily for long-lived flows. In contrast, there has not been a thorough verification of short-lived flows, which make up the majority of Internet flows. Furthermore, the effects of network parameters, such as the link bandwidth and propagation delay, have not yet been investigated. In the present paper, we compare the performance of the above two disciplines by simulation experiments. We focus on the performance of both long-lived and short-lived TCP connections traversing the router under various network environments. We show that sqrtN discipline would degrade the TCP performance in terms of the packet loss ratio and file transmission delay, and it may be useful only when the size of the file being transferred is approximately 50–100 Kbytes or when the propagation delay between the sender and the receiver hosts is significantly small. In addition, we demonstrate that using pacing TCP cannot improve the network performance in many situations and that sqrtN discipline is not suitable for situations in which pacing and non-pacing TCP flows co-exist in the network.

1. Introduction

At present, many applications rely on Transmission Control Protocol (TCP) to avoid and resolve congestion in the Internet. Although other applications utilize User Datagram Protocol (UDP) to control the network congestion on their own, on the current Internet, the proportion of UDP traffic is very small compared with TCP traffic [1]. Furthermore, in the current Internet, the traffic volume of video streaming applications such as YouTube increases [2] and they utilize TCP, not UDP, for a transport-layer protocol. Therefore, evaluating the performance of TCP traffic on a network is very important. TCP performance is largely affected by the round-trip time (RTT) and packet loss ratio of the network path [3,4]. The output link buffer of almost all Internet routers deploys the First-In First-Out (FIFO) discipline, and the size of this buffer affects the RTT and packet loss ratio of TCP connections passing through the router. Packets can be accumulated at this buffer, which causes queuing delay and delay jitter. Furthermore, packet losses also occur when packets arrive at a fully-utilized buffer. Therefore, the packet loss ratio can be reduced by utilizing a larger-sized buffer, but this can cause a larger queuing delay because a larger number of packets are accumulated at the buffer.

The size of router buffers is traditionally determined based on a rule-of-thumb attributed to [5]. As stated in [5], the size of a router's buffer should be greater than $B_n = C \times \mathrm{RTT}$, that is, the product of the link bandwidth and the average RTT of flows that pass through the router. This is the bandwidth–delay product discipline (referred to herein as *normal discipline*), and many routers are equipped with buffers for which the size is determined by this discipline. This discipline is also described in a recent RFC [6].

However, according to [7], it is difficult to construct a router buffer based on this discipline due to the hardware limitation. Today's backbone networks generally carry more than 10,000 concurrent flows and have a link bandwidth of 2.5 Gbps or 10 Gbps [8]. If the average RTT equals 250 ms a 10 Gb/s router needs 250 ms $\times$ 10 Gbps = 2.5 Gbits for its buffer. The size of the largest commercial static RAM (SRAM) chip is currently 72 Mbits, which means that several dozen SRAM chips are needed to provide a 2.5-Gbps buffer. This results in large overhead in terms of board size, electrical power consumption, and monetary cost. On the other hand, the dynamic RAM (DRAM) chip is available up to 1 Gbps as well as significant advantages in monetary cost and board size. However, DRAM has a random access time of dozens of ns, which is from five times to ten times slower than that of SRAM. Therefore, the problem will become worse as line rates increase in the future. In addition, the electrical power consumption of DRAM is much larger than that of SRAM. In summary, it is

* Corresponding author. Tel.: +81 6 6850 6864; fax: +81 6 6850 6868.
E-mail address: hasegawa@ist.osaka-u.ac.jp (G. Hasegawa).

0140-3664/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2008.04.025

extremely difficult to build a router buffer for current and future high-speed networks based on normal discipline.

One possible solution for this problem is reported in [7]. It is shown that the router buffer size can be reduced to the bandwidth–delay product divided by the square-root of the number of flows, N, that is, $B_s = \frac{C \times \overline{RTT}}{\sqrt{N}}$, when there are many flows (500 or more) passing through the link. We call this guideline *sqrtN discipline*. The authors in [7] assert that this small buffer size is sufficient to maintain the link utilization as well as that in normal discipline. In [9], the authors state that we need only dozens of packets for the router buffer size when the input link bandwidth is significantly smaller than the output link bandwidth (for example, 10 Mbps input and 1 Gbps output), and/or when using pacing TCP [10] (paced TCP), in which the successive data packets are transmitted with some time intervals to prevent the packets from being sent in bursts. Pacing TCP packets would decrease the packet loss ratio at the bottleneck router, which may contribute to the decrease in buffer size while maintaining the link utilization.

However, these studies consider only the utilization of the bottleneck link bandwidth as a performance metric in the simulations and implementation experiments, and the performance of TCP flows passing through the router is almost ignored. In addition, the network environments in these experiments are quite limited and the effects of various network parameters, such as link bandwidth and propagation delay, have not been investigated. Furthermore, we believe that the conditions stated in [9] cannot be satisfied in future networks: the link bandwidth of the access network is increasing rapidly in recent years.

Therefore, in the present paper, we evaluate the effect of the buffer size on the following, in addition to link utilization: the packet loss ratio and queuing delay at the router, and the performance of TCP flows passing through the router. In particular, we focus on the performance of short-lived TCP connections when a small-sized buffer is used at the bottleneck link, since the performance of a short-lived TCP data transfer is affected not only by the bottleneck link utilization, but also by factors including the RTT, packet loss ratio and available bandwidth. We investigate the effect of other network parameters such as the propagation delay and physical capacity of the bottleneck link, and derive the parameter ranges in which sqrtN discipline is effective or ineffective. In addition, we explore the effectiveness of pacing TCP for decreasing the router buffer, in situations in which only pacing TCP flows exists in the network and in which pacing and non-pacing TCP flows co-exist in the network.

To our knowledge, the effect of the router buffer size on the performance of short-lived TCP connections has only discussed in one paper [11], which revealed that the packet loss ratio becomes larger when we use the smaller-sized buffer recommended in [7], and it sometimes hinders the performance of TCP data transfer. However, the abovementioned study [11] was performed with a fixed network environment, and the authors only considered congested networks with approximately 100% link bandwidth utilization. On the other hand, in the present study, we investigate the effects of the network parameters and consider under-utilized networks where the link utilization is far below 100%. We also consider the realistic distribution of the file sizes that TCP connections transmit, unlike the fixed value for transferred file sizes used in [11].

We believe that for the complete comparison of the two discipline in buffer sizing, we should evaluate them from various points of view. It includes the effect of network parameters: the effect of RTT (propagation delay), access/bottleneck link bandwidth, with homogeneous/heterogeneous situation. It also includes the effect of various type of TCP flows: short/long-lived and paced/non-paced, and their mixture situation. Among them, we select the following cases in this paper: the effect of long/short-lived flow, the effect of paced TCP and its mixture situation, and the effect of network parameters with homogeneous situation. This is because we would like to reveal the fundamental characteristics of the two discipline.

The remainder of the paper is organized as follows. Section 2 reviews the two disciplines for determining router buffer size: normal discipline and sqrtN discipline. Section 3 describes the network model, parameter setting, and evaluation metric for the simulations. In Section 4, we show extensive simulation results and discuss router buffer sizing. Section 5 discuss the effect of pacing TCP in router buffer sizing. Section 6 concludes the present paper and gives some future areas of study.

2. Guidelines for router buffer sizing

2.1. Normal (bandwidth–delay product) discipline

The traditional guideline for setting the buffer size based on the bandwidth–delay product is described in [5]. We call this guideline normal discipline. In what follows, we introduce the fundamental reasons for normal discipline. For a detailed explanation, please refer to [5].

The changes in the congestion window size of a TCP connection in the congestion avoidance phase can be modeled as additive-increase and multiplicative-decrease (AIMD) in versions of TCP such as Reno [12] and NewReno [13]. Fig. 1 presents the typical behavior of a single TCP-Reno flow passing through a single-bottlenecked-router network. The top graph shows the time evolution of the queue length at the bottleneck router buffer, and the bottom graph shows the changes in the congestion window size of the TCP connection, where B_{max} is the buffer capacity. We assume the bottleneck link bandwidth to be C. From time t_1, the sender starts filling the buffer until a packet is dropped because of the full buffer (at time t_2). Approximately one RTT later, the sender receives duplicate ACKs. The sender then retransmits the lost packet, and halves its window size from W_{max} to $W_{max}/2$ (at time t_3). Before time t_3, the sender is allowed to have W_{max} outstanding packets. However, after time t_3, the sender is only allowed to have $W_{max}/2$ outstanding packets. Therefore, the sender must stop sending packets until it receives $W_{max}/2$ ACK packets. This means that the number of packets in the buffer decreases while the sender stops sending packets (from time t_3 to time t_4). After time t_4, the sender increases its window size, so the number of packets in the buffer again increases after time t_5.

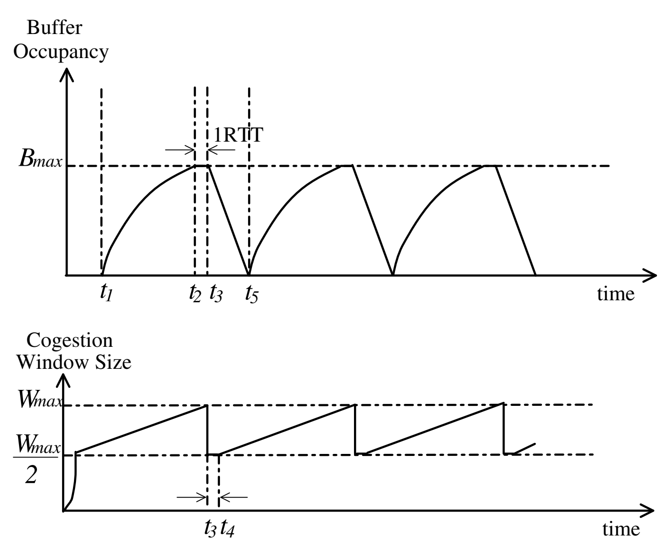

Fig. 1. The time evolution of the congestion window and the queue length.