Computer Networks 55 (2011) 45–60

Contents lists available at ScienceDirect

Computer Networks

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

On the design of load factor based congestion control protocols for next-generation networks ☆,☆☆

Ihsan Ayyub Qazi *, Taieb Znati

Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15260, USA

ARTICLE INFO

Article history:
Received 3 February 2010
Received in revised form 25 June 2010
Accepted 19 July 2010
Available online 1 August 2010

Keywords:
Congestion control
TCP
AQM
ECN
Load factor

ABSTRACT

Load factor based congestion control schemes have shown to enhance network performance, in terms of utilization, packet loss and delay. In these schemes, using more accurate representation of network load levels is likely to lead to a more efficient way of communicating congestion information to hosts. Increasing the amount of congestion information, however, may end up adversely affecting the performance of the network. This paper focuses on this trade-off and addresses two important and challenging questions: (i) How many congestion levels should be represented by the feedback signal to provide near-optimal performance? and (ii) What window adjustment policies must be in place to ensure robustness in the face of congestion and achieve efficient and fair bandwidth allocations in high Bandwidth-Delay Product (BDP) networks, while keeping low queues and negligible packet drop rates?

Based on theoretical analysis and simulations, our results show that 3-bit feedback is sufficient for achieving near-optimal rate convergence to an efficient bandwidth allocation. While the performance gap between 2-bit and 3-bit schemes is large, gains follow the *law of diminishing returns* when more than 3 bits are used. Further, we show that using multiple back-off factors enables the protocol to adjust its fairness convergence rate, rate variations and responsiveness to congestion based on the degree of congestion at the bottleneck. Based on these insights, we design Multi-Level feedback Congestion control Protocol (MLCP). In addition to being efficient, MLCP converges to a fair bandwidth allocation in the presence of diverse RTT flows while maintaining near-zero packet drop rate and low persistent queue length. A fluid model for the protocol reinforces the stability properties that we observe in our simulations and provides a good theoretical grounding for MLCP.

1. Introduction

The congestion control algorithm in the Transmission Control Protocol (TCP) has been widely credited for the stability of the Internet. However, future trends in technology (*e.g.*, increases in link capacities [1] and incorporation of wireless WANs into the Internet), coupled with the need to support diverse QoS requirements, bring about challenges that are likely to become problematic for TCP. This is because (1) TCP reacts adversely to increases in bandwidth and delay and (2) TCP's throughput and delay variations makes it unsuitable for many real-time applications. These limitations may lead to the undesirable situation where most Internet traffic is not congestion-controlled; a condition that is likely to impact the stability of the Internet.

TCP was designed to suit an environment where the BDP was typically less than ten packets and any packet loss inside the network was assumed to be due to overflow of

☆ An earlier version of the paper was presented at IEEE INFOCOM 2008, April 2008, Phoenix, AZ.
☆☆ This work was supported by NSF under Grant 05-010684.
* Corresponding author. Tel.: +1 412 624 8442; fax: +1 412 624 8854.
E-mail addresses: ihsan@cs.pitt.edu (I.A. Qazi), tznati@nsf.gov (T. Znati).

1389-1286/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2010.07.010

router buffers at the bottleneck [2]. These assumptions are no longer true today. BDP of many Internet paths is orders of magnitude larger and in networks such as wireless LANs and WANs, congestion is no longer the only source of packet loss; instead bit errors, hand-offs, multi-path fading etc. account for a significant proportion of lost packets. A more fundamental problem with TCP (*e.g* Reno, NewReno, SACK) and its other variants (*e.g.*, Vegas, Fast) is the usage of packet loss and queueing delay as signals of congestion, respectively [3–6]. Packet loss is a binary signal and so provides little information about the level of congestion at the bottleneck, while (forward path) queuing delay is hard to measure reliably. Moreover, loss and delay are important performance metrics; using them as signals of congestion implies that action can only be taken after performance has degraded.

To address these issues, researchers have proposed transport protocols that can be placed into three broad categories, (a) end-to-end (e2e) schemes with implicit feedback, (b) e2e schemes with explicit feedback and (c) network-based solutions. e2e schemes with implicit feedback treat the network as a black box and infer congestion via implicit signals such as loss and delay. Research studies have shown that using only packet loss and/or delay as a signal of congestion poses fundamental limitations in achieving high utilization and fairness while maintaining low bottleneck queue and near-zero packet drop rate on high BDP paths [4,7]. The benefit of using such schemes is in the ease of their deployment because they require modifications only at the end-hosts. e2e schemes with explicit feedback (such as TCP + AQM/ECN proposals [8–12] and VCP [13]) use one or few bits of explicit feedback from the network, however, the bulk of their functionality still resides at the end-hosts. They typically require changes at the end-hosts with incremental support from the network. Such schemes have been shown to perform better than their counterparts with implicit feedback. However, it is unclear how the amount of congestion feedback information affects performance; a question we make an attempt to answer in this work. In network-based schemes (*e.g.*, XCP [14], RCP [15]), fairness and congestion control are enforced inside the network, therefore, these schemes are likely to induce more overhead on routers. Moreover, such schemes require significant changes in the routers and end-hosts which makes their deployment difficult.

VCP is an e2e scheme that uses two bits of explicit feedback from the network. It is a generalization of the one-bit ECN that uses load factor (ratio of demand to capacity) as a signal of congestion [13]. However, VCP's rate of convergence to an efficient bandwidth allocation is far from optimal, which considerably increases the AFCT of short flows (see Section 2). VCP's usage of a single, fixed Multiplicative Decrease (MD) parameter reduces responsiveness to congestion in high load and causes slow convergence to fairness. Further, in the presence of diverse RTT flows, VCP becomes considerably unfair as shown by simulation results in Section 4. A closer look at the VCP analysis reveals that (1) more refined spectrum of congestion levels is necessary to avoid inefficiencies on high BDP paths, (2) The window adjustment policies in high load regions should adapt to the degree of congestion, to provide smooth rate variations and to ensure robustness in the face of congestion and (3) mechanisms should be in place to achieve good fairness while maintaining low queues in the presence of RTT heterogeneity. This, however, raises few fundamental questions about load factor based congestion control schemes: (i) *What representation of the network load provides the best trade-off between performance gains and the adverse effects due to the larger amount of feedback?* (ii) *What window increase/decrease policies must be in place to ensure efficient and fair bandwidth allocations in high BDP networks while keeping low queues and near-zero packet drop rate?* This paper addresses these issues and uses the insights gained by the analysis to design **M**ulti-**L**evel Feedback **C**ongestion Control **P**rotocol.

The theoretical analysis and simulations carried out as part of this work show that using 3-bit representation of the network load levels is sufficient for achieving near-optimal rate of convergence to an efficient bandwidth allocation. While the performance improvement of 3-bit over 2-bit schemes is large, the improvement follows the "*law of diminishing returns*" when more than three bits are used. Our results also show that using multiple levels of MD enables the protocol to adjust its rate of convergence to fairness, rate variations and responsiveness to congestion according to the degree of congestion at the bottleneck. Guided by these fundamental insights, we design MLCP, in which each router classifies the level of congestion in the network using 4-bits while employing *load factor* as a signal of congestion [16]. In addition, each router also computes the *mean RTT* of flows passing through it, to dynamically adjust its load factor measurement interval. These two pieces of information are tagged onto each outgoing packet using *only* seven bits. The receiver then echoes this information back to the sources via acknowledgment packets. Based on this feedback, each source applies one of the following window adjustment policies: Multiplicative Increase (MI), Additive Increase (AI), Inversely-proportional Increase (II) and Multiplicative Decrease (MD). MLCP like XCP decouples efficiency control and fairness control by applying MI to converge exponentially to an efficient bandwidth allocation and then employing AI-II-MD control law for providing fairness among competing flows [14]. MLCP adjusts its aggressiveness according to the spare bandwidth and the feedback delay which prevents oscillations, provides stability in the face of high bandwidth or large delay, and ensures efficient utilization of network resources. Dynamic adaptation of the load factor measurement interval allows MLCP to achieve high fairness in the presence diverse RTT flows. In addition, MLCP decouples loss recovery from congestion control which facilitates distinguishing error losses from congestion-related losses; an important consideration in wireless environments. MLCP has low algorithmic complexity, similar to that of TCP and routers maintain no per-flow state.

Using extensive packet-level ns2 [17] simulations, we show that MLCP achieves high utilization, low persistent queue length, negligible packet drop rate and good fairness. We use an approximate fluid model to show that the proposed protocol is globally stable for any link capacity, feedback delay or number of sources for the case of a single bottleneck link shared by identical RTT flows. The