A New Congestion Control Algorithm for Bandwidth Guaranteed Networks

Lin Han, Lijun Dong, Yingzhen Qu, Richard Li
Huawei USA – Futurewei Technologies, Inc.
Santa Clara, California, U.S.A
email: {lin.han, lijun.dong, yingzhen.qu, renwei.li}@huawei.com

***Abstract*— Future Internet can require bandwidth guaranteed services, thus the network resources need to be reserved before a Transmission Control Protocol (TCP) session starts transmitting data. The paper proposes a new TCP congestion control algorithm to assure the information rate for a flow. It is an extension to, yet different from the current TCP standards. The congestion window size changes at the sender side due to events, such as OAM (Operations, Administration and Management) congestion alarm, duplicate ACKs, and timeout.**

Keywords- QoS; TCP; IP; in-band signaling; bandwidth guaranteed networks; congestion control; congestion detection component.

I. INTRODUCTION

The current Internet, more generally a Transmission Control Protocol/Internet Protocol (TCP/IP) [1][2] network, was designed as a best-effort network, i.e., without any assurances as to Quality of Service (QoS), bandwidth, latency, processing time, etc. In other words, the IP service makes its "best effort" to deliver segments between two hosts, but it has no guarantees. On top of IP, TCP offers several additional services to applications. First and foremost, it provides reliable data transfer with flow control, sequence numbers, acknowledgments, and timers. TCP also provides congestion control [3], which prevents any one TCP connection from overwhelming the links and routers with a superabundant amount of traffic. TCP makes effort to provide each connection traversing an overloaded link with a fairly same proportion of the link bandwidth. The major components of congestion control in widely used TCP Reno include: slow start, congestion avoidance, and fast recovery. Figure 1. shows the congestion window at the sender side which changes over the time, as well as the state transfer due to the events.

With the unprecedented mobile applications emerging, such as Augmented Reality (AR) and Virtual Reality (VR), remote diagnosis and surgery, autonomous driving and road safety, guaranteeing the QoS in terms of bandwidth, latency, jitter etc. presents the unavoidable challenge to the current Internet's best-effort principle. In this paper, we use the term of "bandwidth guaranteed networks" to describe networks in which the bandwidth can be reserved for a particular flow. This can be achieved by the existing QoS mechanisms and frameworks: Integrated Services (IntServ) [4] with prior out of band signaling by RSVP [5], Differentiated Services Architecture (DiffServ) [6] with resource provisioning with the help of Service Layer Agreements (SLAs), Multiprotocol Label Switching (MPLS) [7] with Label Distribution Protocol (LDP) [8] and Resource Reservation Protocol-Traffic Engineering (RSVP-TE) and the in-band signaling protocol proposed in [9]. The common objective of all these solutions is to have network resources/bandwidth reserved before data is transmitted. In bandwidth guaranteed networks, the data transmission for a flow can be guaranteed at the committed information rate (CIR), but not above. When the data rate is between CIR and PIR (peak information rate), the shared resources are used. No traffic above PIR rate will be allowed to enter the network.

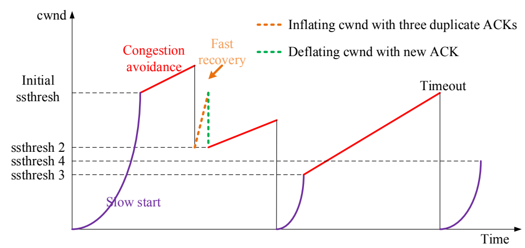

Figure 1. Congestion window in Reno

The paper proposes a new congestion control algorithm for the future Internet that builds upon TCP Reno, but considers the characteristics of bandwidth guaranteed networks as stated above. Section 2 explains the details of the new algorithm, and section 3 concludes this short paper.

II. NEW CONGESTION CONTROL ALGORITHM

The proposed congestion control algorithm has four components, which is introduced below. The congestion window size at the sender side is presented in Figure 2.

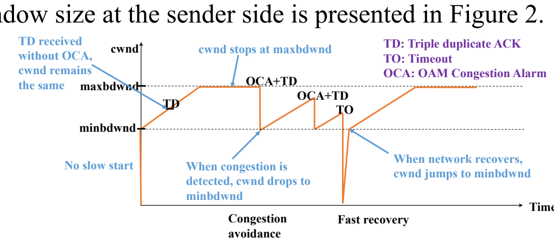

Figure 2. Congestion window in new congestion control algorithm for bandwidth guaranteed networks

A. Immediate Start

The proposed congestion control requires that OAM is used to constantly report on the network condition parameters, such as number of hops, Round Trip Time (RTT). This might be done through setting up a measuring TCP connection. The measuring TCP connection does not have user data, and it is only used to measure the key network parameters. As the network status is constantly changing, after a TCP session is

 ISBN: 978-1-61208-662-0

established, these parameters need to be updated. This requires a sender to periodically or consistently embed TCP data packet with OAM option. Consequently, in bandwidth guaranteed networks, the slow start component is not needed and removed from the proposed congestion control mechanism.

There are two important window sizes proposed for the new congestion control mechanism: $minbdwnd$ and $maxbdwnd$, which are calculated as below:

$$minbdwnd = CIR * RTT/MSS \quad (1)$$

$$maxbdwnd = PIR * RTT/MSS \quad (2)$$

The RTT is the time taken to send a data packet to the destination and receive a response packet, and MSS is Maximum Segment Size. After a TCP session is established, the sender can immediately start transmitting data at an initial window size of *minbdwnd* as shown in Figure 2, if the receiver window (*rwnd*) is not a limiting factor.

Since the network status is constantly changing, RTT is updated using the following formula, with a is a number between 0 and 1:

$$RTT = a * old\,RTT + (1 - a) * new\,RTT \quad (3)$$

The initial RTT can be obtained by using a measuring TCP connection, or configured based on the historical data.

B. Congestion Avoidance

In bandwidth guaranteed networks, there is no slow-start, so congestion avoidance state is entered right after the initial start. During congestion avoidance, for every newly received ACK, *cwnd* is increased by one RTT/MSS until it reaches *maxbdwnd*. The value of *cwnd* stays constant at *maxbdwnd* afterwards, until packet loss is detected. This means a TCP sender is never allowed to send data at a rate larger than PIR.

C. Fast Retransmit and Fast Recovery

In the new congestion control algorithm, upon receiving duplicate ACKs the fast retransmit and fast recovery follow the following rules: (1) when a sender receives the first and second duplicate ACKs, the value of *cwnd* is not changed, and the sender continues to send traffic; (2) when a sender receives the third duplicated ACK, if the retransmission timer has not expired and a previous OAM congestion alarm has been received, it is likely a segment is lost due to congestion. The sender will perform a retransmission of the lost segment, and the value of *cwnd* is set to be *minbdwnd*; (3) when a sender receives the third duplicated ACK, but no previous OAM congestion alarm has been received, then it is considered that a segment is lost due to random failure instead of congestion. In this case, the value of *cwnd* is not changed.

D. Timeout Handling

If a retransmission timer in a TCP sender expires, in bandwidth guaranteed networks, this most likely indicates a physical failure no matter whether a duplicate ACK is received or not. In this case, the value of *cwnd* is set to be one, and the TCP sender will retransmit the lost segment. This retransmitted packet also serves the function of probing the network status. If there is really a network failure, no ACK will be received for this packet and the retransmission timer will expire again. Upon receiving an expected ACK after the retransmission(s), it indicates that the network has recovered from the physical failure, and the value of *cwnd* will be set to be *minbdwnd*.

III. SUMMARY AND OUTLOOK

A bandwidth guaranteed network is defined to be able to provide guaranteed bandwidth service with at least two bandwidth parameters: a Minimum Bandwidth or Committed information rate (CIR), and a Maximum Bandwidth or Peak information rate (PIR). The proposed congestion control algorithm to be used in bandwidth guaranteed networks comprises immediate start, congestion avoidance, fast retransmit and fast recovery, and timeout handling components. The detection of OAM signaling, duplicate ACKs, and timeout are used to infer the packet loss caused by random or permanent physical failure, or by congestion.

The proposed algorithm can coexist with current TCP congestion control mechanisms. Time sensitive TCP flows should achieve resource reservation before start sending data, and this guarantees bandwidth and latency especially when network is congested. Regular TCP sessions will share the remaining network resources.

In future works, we plan to implement the proposed congestion control in Huawei routers to prove the concept and verify that the guaranteed data rate of a flow can be achieved and not affected by other TCP flows. Moreover, we will extend the concepts and algorithms to realize guaranteed maximum latency for individual flow, which is extremely important and useful for latency sensitive applications.

REFERENCES

[1] S. Deering and R. Hinden, “RFC 8200: Internet Protocol, Version 6 (IPv6) Specification,” IETF, Jul. 2017.

[2] M. Allman, V.Paxson, and E. Blanton, “RFC 5681: TCP Congestion Control,” IETF, Sep. 2009.

[3] J. F. Kurose and K. W. Ross, “Computer Networking, A Topdown Approach”, Sixth Edition, ISBN: 9780132856201, 0132856204, published by Addison-Wesley Publishing Company.

[4] R. Braden, D. Clark, S. Shenker, “RFC 1663: Integrated Services in the Internet Architecutre: an Overview,” IETF, Jun. 1994.

[5] R. Braden, L. Zhang., S. Berson, S. Herzog, and S. Jamin, "Resource ReSerVation Protocol (RSVP)-Version 1 Functional Specification", RFC 2205, DOI 10.17487/RFC2205, September 1997.

[6] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, “RFC 2475: An Architecture for Differentiated Services,” IETF, Dec. 1998.

[7] E. Rosen, A. Viswanathan, and R. Callon, “RFC 3031: Multiprotocol Label Switching Architecture,” IETF, 2001.

[8] L. Andersson, I. Minei, and B. Thomas, “RFC 5036: LDP Specification,” IETF, Oct. 2007.

[9] L. Han, Y. Qu, L. Dong, and R. Li, Flow-Level QoS Assurance via IPv6 In-Band Signalling, WOCC 2018.