

# Sweet TCP: Battling the Bottleneck Bufferbloat

Name Surname and Name Surname, *Senior Member, IEEE*

**Abstract**—The transfer control protocol is designed to fill the *pipe* between the source and destination for an efficient use of network resources. As a side effect, TCP also fills in the bottleneck buffer, which is the buffer that precedes the slowest link in the path. Permanently filling the buffer does not offer any performance advantage. On the contrary, permanently full buffers impair the operation of TCP causing excessive delays and timeouts. This problem is known in the literature as bufferbloat. In this paper, we make use of some of the insights developed by the community around the bufferbloat problem to propose a TCP congestion avoidance protocol that fills in the pipe but not the buffer. By continuously monitoring improvements of losses in terms of throughput and delay, Sweet TCP finds and stabilizes around the optimum operation point that simultaneously maximizes the throughput and minimizes the delay. A simple algorithm that adjusts the contention window is proposed. In contrast with traditional congestion avoidance behaviour, the proposed algorithm reduces the congestion window as soon as symptoms of bufferbloat are detected, and increases the congestion window as soon as bufferbloat disappears.

**Index Terms**—TCP, congestion avoidance, bufferbloat, delay

## I. INTRODUCTION

is a very nice introduction introducing the topics that will be covered in the paper. And this [1] is a reference to the best paper ever written..

## II. OPERATING IN TCP'S SWEET POINT

Fig. ?? is a copy from [2] and explain the behaviour of TCP as a function of the number of packets in-flight. If the number of packets is too low, the link is not efficiently used. If it is too high, it fills up the buffer preceding the bottleneck link and increases the end-to-end delay.

TCP behaviour is to increase the number of packets in-flight as long as no packet loss occurs. Roughly speaking, the number of packets in flight increases by one every RTT. This means that, unless Random Early Detection is used, TCP has a tendency to fill the buffer that precedes the bottleneck link and therefore unnecessarily increasing the delay of all the flows that traverse such link.

The problem is that correctly configuring RED is not trivial. If it is too aggressive, tcp slows down to a point in which the link is not efficiently used. If RED is too quiescent, the queue builds up and results excessive delays.

We propose changing the mechanism that adjusts the congestion window mechanism in the congestion avoidance phase. Instead of blindly growing the congestion window until a packet loss occurs, we suggest to increase the congestion as long as it results in performance benefits. When a performance loss is detected, probably in the form of same throughput and

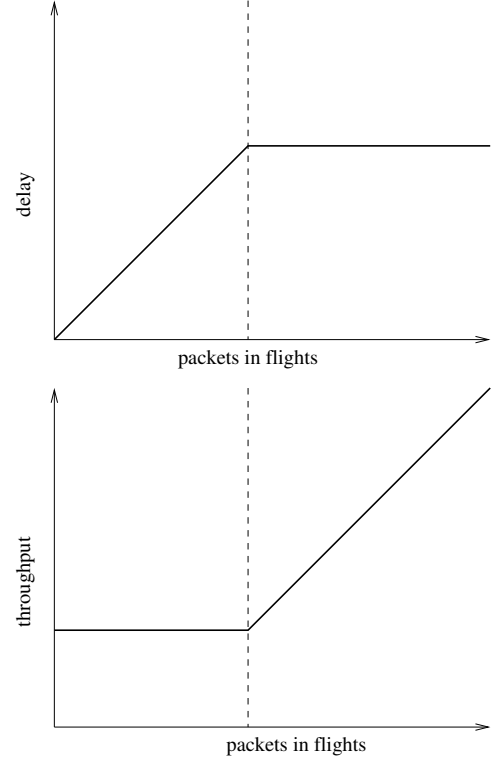


Fig. 1. Typical delay and throughput curves as a function of the size of the congestion window.

higher delay, TCP will start decreasing the congestion window. And it will keep decreasing it as long as this behaviour results in a performance gain. When decreasing the window results in a performance loss, TCP will revert to increasing. This mechanism will force TCP to operate in the sweet point where the throughput is maximum and the delay minimum.

The idea is to take measures of throughput and delay every RTT. We name these measures  $T_i$  and  $D_i$  where the  $T$  stands for throughput and the  $D$  for delay. The integer  $i$  is an index. Then we compute the relative variation of throughput and delay as  $T_{rv} = \frac{T_i - T_{i-1}}{T_{i-1}}$  and  $D_{rv} = \frac{D_i - D_{i-1}}{D_{i-1}}$ , respectively. Finally, we measure the performance improvement as  $PIMP = T_{rv} - D_{rv}$ .

Note that if the current operation point is at the left of the dashed line in Fig. 1 and we increase the congestion window (we move to the left) we measure a positive performance improvement. Similarly, if the current operation point is at the right of the dashed line and we reduce the congestion window (we move to the right), we also measure a positive performance improvement. The performance improvement is positive whenever we move closer to the optimal operation point. Conversely, the performance improvement is negative whenever we move further from the operation point.

We can use these properties to control whether TCP should increase or decrease the congestion window. The behaviour of the current TCP every RTT is  $CWND_i = CWND_{i-1} + MSS$  whenever there is no packet loss. It always increases as long as there is no packet loss.

We suggest  $CWND_i = CWND_{i-1} + STAT * MSS$  where  $STAT$  stands for state and can be 1 or -1. The  $CWND$  increases when  $STAT$  is positive and decreases when  $STAT$  is negative.  $STAT$  is simply the sign of the previously defined performance improvement  $STAT = SIGN(PIMP)$ .

### III. CONCLUSION

#### ACKNOWLEDGMENT

Thanks to innpacto for giving us the opportunity to talk about this.

#### REFERENCES

- [1] J. Barcelo, B. Bellalta, C. Cano, and M. Oliver, "Learning-BEB: Avoiding Collisions in WLAN," in *Eunice*, 2008.
- [2] J. Gettys and K. Nichols, "Bufferbloat: Dark buffers in the internet," *Queue*, vol. 9, no. 11, p. 40, 2011.