Justin Ciocoi

Oct 4, 2023

CSCI 379 Textbook Notes

Computer Networking

Textbook Notes

Section 3.7: TCP Congestion Control

Overview of TCP Congestion Control

- TCP Provides a reliable transport service between two processes which are running on different hosts
  - A congestion-control mechanism is also of paramount importance to the functioning of TCP
    - The approach that TCP takes in order to control congestion is to limit the rate of sent traffic as a function of perceived network congestion
    - This brings up important questions about TCP
      - *How* does TCP limit the rate of traffic that is being sent out
      - *How* does TCP *perceive* the level of congestion present on a network
      - *Which* algorithm should be used to reduce traffic
  - The TCP congestion-control mechanism operating at the *sender* keeps track of an additional variable, `cwnd`, or the congestion window, which imposes a constraint on the rate at which the TCP sender can send traffic into the network
    - This is how rate limiting is achieved under the TCP protocol
  - Next, we must consider how the TCP protocol is able to perceive the congestion on the bath between sender and recipient
    - Congestion can be perceived, roughly speaking, through the responses associated with dropped datagrams and *loss events*
  - TCP is said to be **self-clocking**

        - This means that when the protocol does not receive duplicate ACKs
        - The rate of arrival of these responses will indicate to TCP at which rate to increase its congestion window
            - The faster the rate at which ACKs are received, the higher the perceived bandwidth under TCP and thus the higher the rate limit goes

Guiding Principles of TCP Congestion Control

- A **lost segment** implies **congestion**
    - Thus, the TCP sender's rate should be **decreased** when a segment is lost
- An **acknowledged segment** indicates successful network delivery
    - When the sender receives an ACK for a previously unacknowledged segment, its rate will be **increased**
- The two previous principles imply TCP's mechanism of **bandwidth probing**
    - Roughly speaking, bandwidth probing can be explained as TCP's increasing of the data rate until a loss event occurs
    - This can be thought of as TCP continuously checking for the fastest achievable speed by slowly increasing the rate at which data is transmitted
- Slow Start
    - When a TCP connection is established, the value of `cwnd` is initialized as a small value of 1 MSS
        - If MSS = 500 bytes and RTT=200 msec
        - $InitialSendingRate = \frac{MSS}{RTT} \approx 20kbps$
- TCP's congestion control mechanism can be referred to as an **additive-increase, multiplicative-decrease (AIMD)** form of congestion control
    - The `cwnd` variable is increased linearly by TCP until the point at which a loss event occurs, where it will be decreased by a function of the rate at the previous loss event

TCP Fairness

- TCP can be thought of as a *fair* protocol

- In a situation with multiple connections to the same recipient, each connection is granted the same amount of bandwidth
- For this reason, many multimedia applications choose to run over **UDP** as losing the occasional packet during video or audio streaming is not as hurtful to the user experience as throttling speeds to achieve a *fair* connection environment

Explicit Congestion Notification (ECN): Network-Assisted Congestion Control

- Through more recent developments and implementations of TCP, the protocol can now make use of **ECNs** in order to allow the network layer to explicitly signal network congestion to both a sender and a receiver
- In theory, this provides a connection that can more optimally utilize the available bandwidth of the network