National University of Singapore
School of Computing
CS5229: Advanced Computer Networks
Semester I, 2021/2022

Lecture 5 Training
Network-assisted Congestion Control

In Lecture 5, we discussed how congestion control can be driven entirely by the network i.e. the network can inform senders of the rate at which they should send. In this lecture training, we will investigate how rate allocations and network utilization can change depending on the notion of fairness adopted by the network.

Fair Rate vs. Max-Min Fair

Consider a network with 4 links with different capacities shown in Figure 1.

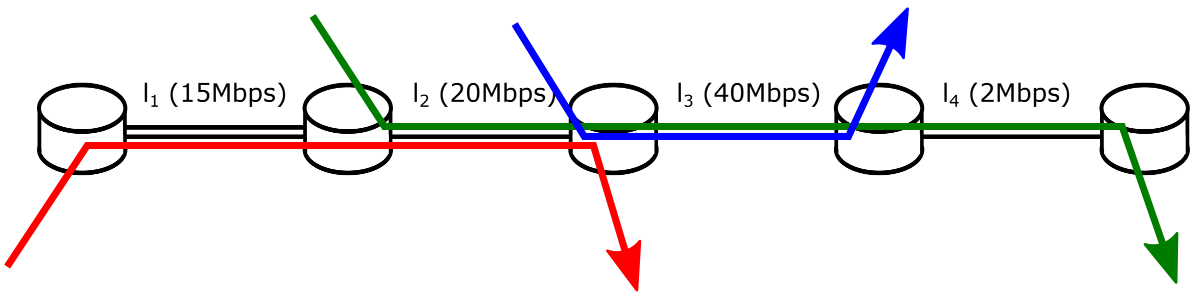

Figure 1: A simple network scenario

There are 3 flows (shown in red, blue and green) that share different bottleneck links in the network. Assume that the flows are all long running TCP flows and have reached a steady state in terms of the sending rate.

The switches in the network can choose to adopt one of the two types of rate allocations:

1. **Fair Rate:** With this rate allocation, each switch independently determines the number of flows sharing the bottleneck link and limits each flow to a rate no more than C/N, where C is the capacity of the bottleneck link and N is the number of flows sharing the link.
2. **Max-Min Fair:** With this rate allocation, the switches coordinate with each other and perform rate allocation so as to maximize the minimum rate achievable by each flow in the network.

Given this information, answer the following questions on Coursemology:

Q1 When the switches are doing fair rate allocation, what is the final (steady-state) throughput achieved by the Red, Blue and Green flows respectively?

Q2 Which links in the network are underutilized?

Next, suppose that the switches can coordinate with each other and do max-min fair allocation.

Q3 In this new scenario, what is the final (steady-state) throughput achieved by the Red, Blue and Green flows respectively?

Q4 In this new scenario, which links in the network are underutilized?

Fair Rate: Fixed vs. Instantaneous

In this section, we investigate the two ways of doing fair rate allocation and their implication on end-points with continuous and intermittent data flow. In order to do fair rate allocation for a link, the switch first needs to know the number of flows sharing the link. This estimation of the number of flows can be done in the following ways:

- **Fixed:** The switches count the number of flows based on the initial TCP handshake. So if a TCP flow establishes a connection through the 3-way handshake, the switch increases the flow count N by 1 and allocates a fair rate of C/N to all the flows, even if some of the TCP flows are not actively sending any data.
- **Instantaneous:** The switches count the number of flows based on which flows are actively sending data at the given moment. For example, if 3 flows have established a TCP connection, but at the current moment, only 2 flows are sending their data, then the allocated fair rate is $C/2$. After some time, if the third flow also starts sending data, then the rate allocation is changed to $C/3$ instantaneously.

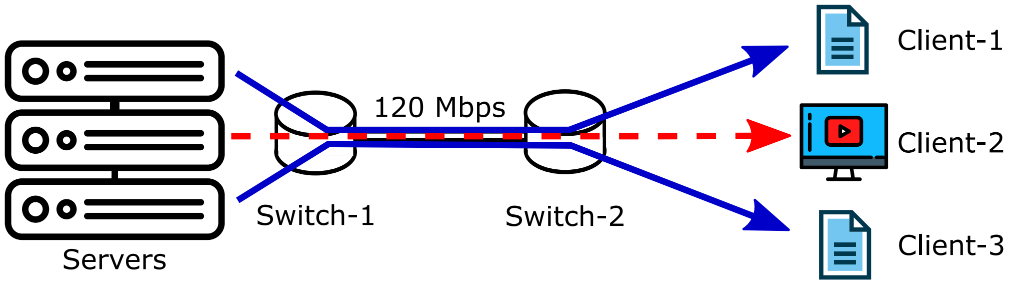

Figure 2: Network scenario with a DASH client

Now consider the scenario in Figure 2. Three clients are downloading data from servers while sharing a bottleneck of 120 Mbps. Switch-1 performs fair rate allocation – either fixed or instantaneous. Client-1 and client-3 are long running file download clients, each of them downloading a very huge file. Therefore, client-1 and client-3 have a continuous data flow. Client-2 is a browser-based video player which is streaming a long running live video stream using DASH (Dynamic Adaptive Streaming over HTTP). Client-2 has an intermittent data flow as explained below.

DASH Streaming: In DASH streaming, the video is divided into small segment files which the client downloads over HTTP. In this scenario, each segment is 5 secs worth of video content and the size of each segment is constant and fixed at 5 MB. After the TCP handshake is done and the HTTP connection is established over TCP, the video player (aka DASH client) downloads the first video segment (Seg-0). During this first download, the DASH client records the time[1] (say δ) it took to download the segment. After this download is complete, the DASH client will start playing the first segment. We denote this point of starting the video playback as $t = 0$ and the DASH client (client-2) would be

[1] Note that irrespective of fixed or instantaneous fair rate allocation, the time δ to download a video segment would remain the same for each segment.

in a steady state[2] here onward. Since each segment is 5 seconds long, at $t = 5$, the DASH client would need to play the next segment. Since it takes δ time to download a segment, the DASH client will start downloading the next segment at $t = 5 - \delta$. This behavior of the DASH client will create an intermittent data flow on the network as shown in Figure 3.

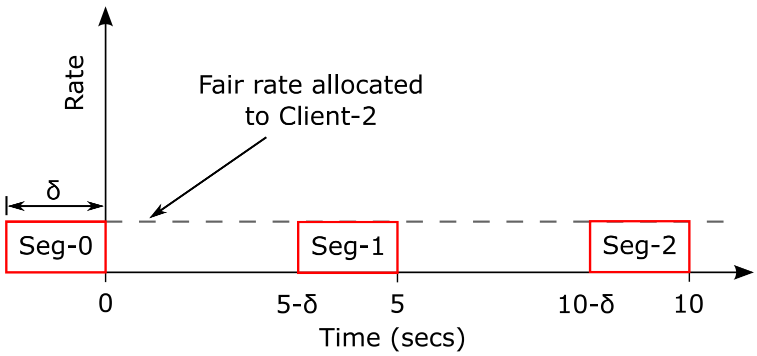

Figure 3: DASH client's data flow

Given this information, answer the following questions on Coursemology:

Q5 Suppose switch-1 is doing fixed fair rate allocation. In this case, what is the average utilization of the link between switch-1 and switch-2?

Q6 What is the average throughput observed by each of the three clients?

Now suppose that the switch-1 is doing instantaneous fair rate allocation.

Q7 For this new scenario, what is the average utilization of the link between switch-1 and switch-2?

Q8 What is the average throughput observed by each of the three clients?

[2] steady state in terms of throughput/rate when averaged over a sufficient amount of time