Question 1:

The figure below shows how RTT and throughput varies with the amount of data in flight (data sent but not yet acknowledged).

![Graph showing RTT and Throughput vs Data in Flight]

a) What does each region separated by the dashed vertical lines in the plot signify?

b) At which point along the x-axis does loss-based congestion control operate?

c) What is the optimal point along the x-axis, that BBR aims to operate at, for minimizing RTT and maximizing throughput?

Solution:

a) From left to right, the regions are app-limited (when the app does not push enough data to fill the pipe), bandwidth-limited (the pipe is full; additional in flight data add to the queue at the bottleneck) and buffer-limited (packets drops occur due to exceeding the buffer capacity).
b) Loss-based congestion control operates at the right edge of the bandwidth-limited region. It delivers full bottleneck bandwidth but incurs higher delays and frequent packet loss.

c) BBR aims to operate at the left edge of the bandwidth-limited region. This is known as Kleinrock’s optimal operating point which maximizes bandwidth and minimizes delay and loss.

Question 2:

The first of the 2 plots below shows how the capacity of inter-DC networks are wasted. The networks are provisioned for peak usage while the average utilization is less than 50%. If we further look into the traffic characteristics in the second plot, we see that a large part of this traffic is background traffic which does not need strict QoS from the networks. How can this observation be used to optimize inter-DC network provisioning?

Solution:

If background traffic is adapted to only use unused capacity (after non-background traffic is assigned required capacity), the traffic can fit in much lower capacity (more than 90% vs. less than 50% of the peak utilization).
Question 3:

You are operating a small WAN between the 4 sites, with 5 links, as shown in the figure. You want to move from the path assignment for flows on the left to the one on the right. All links have 1.5 units of capacity, and each flow carries 1 unit traffic. Draw the minimal sequence of flow reassignments (flow splitting across multiple paths is allowed) to achieve the configuration on the right such that there is no congestion in the network at any point in the sequence. Does such a sequence also exist if the flows were carrying 1.5 units of traffic each?

Figure adapted from: Achieving high utilization with software-driven WAN, Hong et al. ACN SIGCOMM 2013

Solution:

Sequence does not exist if flows carry 1.5 units of traffic each.

Question 4:

You are operating the WAN in the below figure. Your routing decisions use the following policy: whenever a flow arrives, you assign it to the shortest path with any available capacity (without any flow splitting, so each flow takes one path.) All links are 1 unit capacity, and all flows will need 0.75 unit capacity. For three flows $R_1 \rightarrow R_6$, $R_3 \rightarrow R_6$, $R_4 \rightarrow R_6$ that could arrive in an arbitrary order, illustrate:

a) the worst flow-arrival order and assignment of flows to paths (in the left diagram)

b) the best flow-arrival order and assignment of flows to paths (in the right diagram)
Solution:
There can be multiple solutions. Below are 2 such orderings and assignments for flows $R_1 \rightarrow R_6(F_A)$, $R_3 \rightarrow R_6(F_B)$, $R_4 \rightarrow R_6(F_C)$:
Worst (left diagram below): $F_A$, $F_B$, $F_C$
Best (right diagram below): $F_B$, $F_A$, $F_C$