1. For the basic queueing model, produce a plot of the tail probability of the state distribution for \( B=25 \) and \( \rho=0.5, 0.7, .85, .99 \) and 1.1. More precisely, let \( T_j \) be the probability that the queue has at least \( j \) packets. Plot \( j \) on the x-axis and \( T_j \) on the y-axis for values of \( j \) from 0 to \( B \). Show separate curves for each value of \( \rho \) and use a log scale on the vertical axis. This is easy to do using a spreadsheet.

![Plot of tail probability](attachment:plot.png)

2. Suppose we let \( B\to\infty \) in the basic queueing model. In this case, what is the probability that the queue has \( j \) packets? What is the average number of packets in the queue?

*The probability of \( j \) packets is \( p_j=(1-\rho)^j \) and the average queue length is*

\[
\sum_{i>0} i(1-\rho)^i = \rho/(1-\rho)
\]

3. Consider a queue for which the input traffic varies as follows. For the first 5 seconds, the arrival rate is 70% of the link rate. For the next 10 seconds, the arrival rate is 110% of the link rate and for the next 5 seconds, the arrival rate is 90% of the link rate. Assuming the link can forward a million packets of average length each second, estimate the number of packets that will be lost in the 20 seconds of the queue’s operation, if the queue has room for 10 packets. Use the basic queueing model introduced in the notes. Note that the probability that an arriving packet is discarded is equal to the probability that the queue is full when the packet arrives.

*Since \( p_B = (1-\rho)^B \) the loss probability is 0.008645 during the first 5 seconds, 0.14 during the next 10 seconds and 0.0508 during the last 5 seconds, so the total number of lost packets is about 700,000×5×0.008645 + 1,100,000×10×0.14 + 900,000×5×0.0508 ≈ 6,052+154,000+45,720 = 1,798,860. This is out of a total of 19,000,000 that are sent in this period or 9.5%.*
4. TCP requires large buffers in routers to avoid underflow. The generally recommended minimum buffer length is the product of the link bandwidth times the round trip delay (in the absence of queueing). Within a local area network the round trip delay is generally less than 100 μs, while in a wide area network it can be 100 ms or more. What is the appropriate amount of buffering for a LAN with 1 Gb/s links? What is the appropriate amount for a WAN with 10 Gb/s links?

For the LAN, the minimum recommended buffer length is 100,000 bits or 12.5 KB. This is roughly 8 maximum length Ethernet packets. For the WAN, the recommended buffer length is 1 Gb or 125 MB. This is enough for about 80,000 maximum length packets.

Suppose, we use less than the recommended amount of buffering. What is the impact on the long-term utilization of a bottleneck link in a network, as TCP adjusts its sending rate. Note that for such a link, the aggregate traffic from all the TCP sources will oscillate between about 100% of the link rate and 50%. How do you think the cost of the lost link capacity compares to the cost of the memory required for buffering?

Since the TCP sources generate traffic that oscillates between 50% and 100% of the link rate, we would expect the average input traffic rate to be about 75%. So, we would expect about 25% of the link capacity to be left unused. If this is a 10 Gb/s WAN link, we need about 1 Gb of memory to store the packets. The cost of this memory today is at most $10. The cost of 25% of a 10 Gb/s WAN link is much larger than this, since it includes not only the electronics and optics that terminate the link, but the cost of optical fiber that may stretch over many hundred miles.

5. Why is it important for $B$ to be at least equal to the maximum packet length in the Deficit Round Robin packet scheduler?

$B$ is the amount of “credits” a queue is given on each pass through the scheduling list. If $B$ is smaller than the maximum length packet, a queue may have to wait for several passes before it acquires enough credits to send a packet (assuming for example, that its first packet is of maximum length). This means that the scheduler may spend time updating credits for a number of queues in a row without sending any packets. This can make it difficult for the scheduler to send packets fast enough to keep up with the outgoing link rate.

For the WDRR scheduler, suppose we have 90 queues that are assigned a rate of 1 Mb/s and one that is assigned a rate of 10 Mb/s, that are all sharing a 100 Mb/s link. What are appropriate values of $B_i$ to assign to each queue? Suppose the link is backlogged and all packets are of maximum length. What is pattern of transmission for the first 100 packets (that is, which queue sends a packet at each time)? How does this compare to the ideal transmission pattern?

For the 1 Mb/s queues, we could let $B_i=1,550$ bytes (just over the maximum packet length for Ethernet packets). For the 10 Mb/s queue, we would let $B_i=15,500$ bytes, in order to ensure that it gets ten times the link bandwidth of the others.

With these choices, the 90 slow queues will each send a max length packet followed by 10 packets from the fast queue. And this pattern will repeat. It would be more desirable to have the 10 packets from the fast queue spread out, instead of having them sent in a burst.