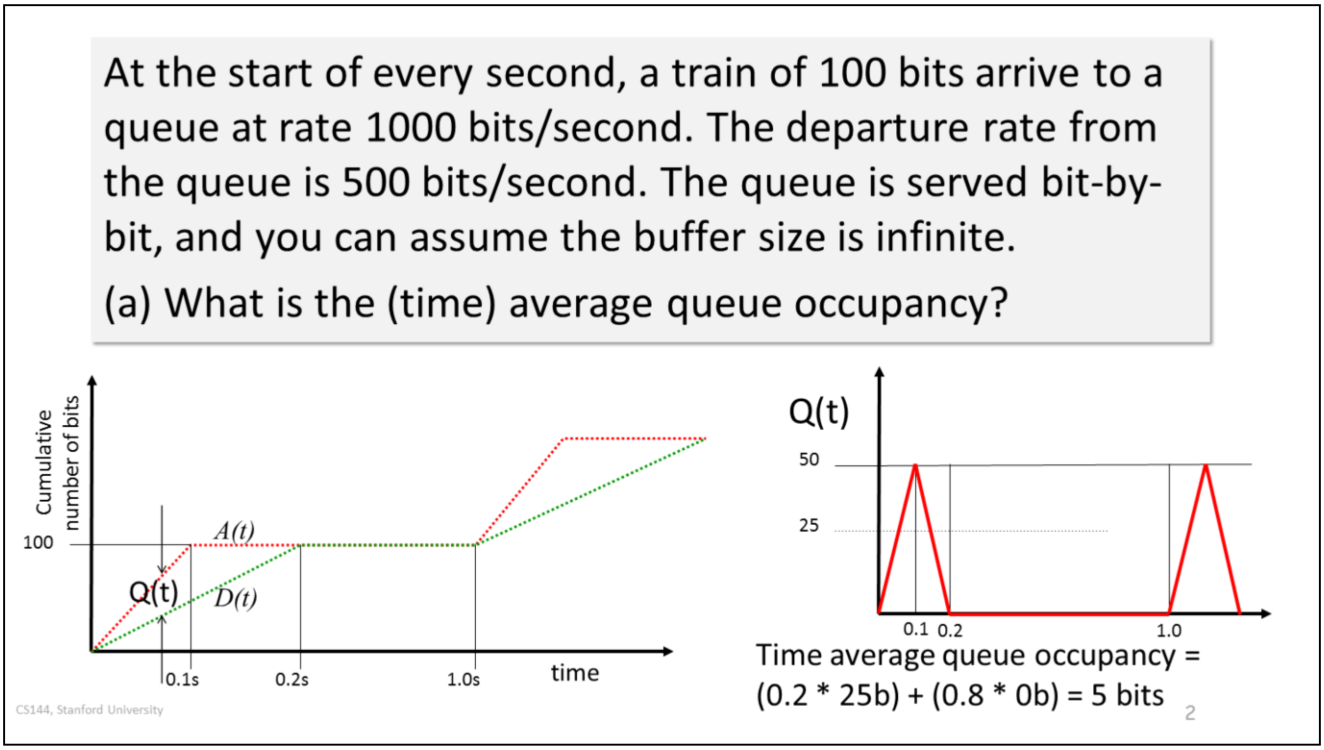
CS144
An Introduction to Computer Networks

Worked Example

Nick McKeown
Professor of Electrical Engineering
and Computer Science, Stanford University

The video included a quick worked example. In case you found the example hard to follow, in this short video I will explain the same example in more detail, and ask a few more questions about the same example.



"At the start of every second, a train of 100 bits arrive to a queue at rate 1000 bits/second. The departure rate from the queue is 500 bits/second. The queue is served bit-by-bit, and you can assume the buffer size is infinite.

(a)What is the average queue occupancy?"

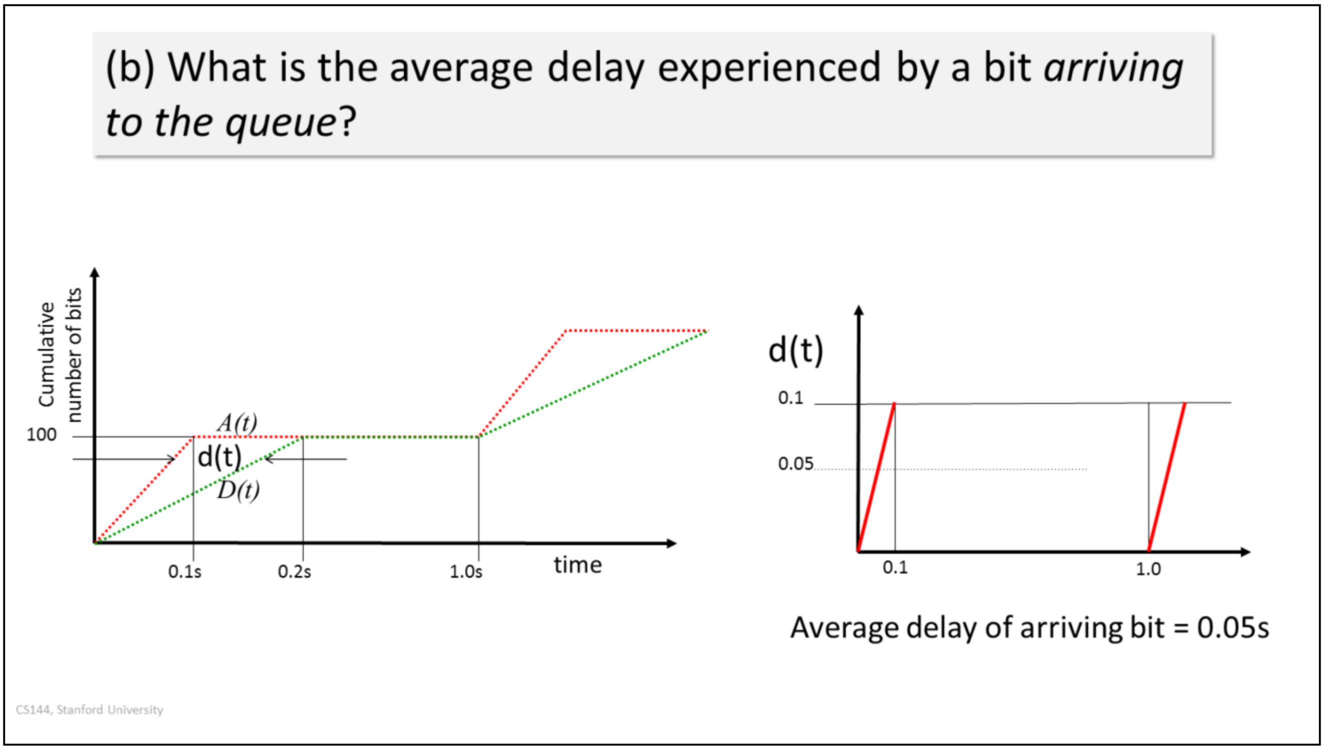
<click> The cumulative arrival and departure processes will look like this. First, look at the cumulative arrivals, A(t) shown in red. In the first 1/10th of a second, 100bits will arrive at rate 1000bits/second. There are no more arrivals until 1s, then a new 100bits arrive. Now let's look at the departure process D(t) shown in green. As soon as the first train of bits starts to arrive at rate 1000bits/second, bits will start departing, because the

question says they are served bit-by-bit. If we had to wait for a whole packet to arrive, then the departure process would not start increasing yet. The bits take 0.2 seconds to depart because they depart at 500bits/second which is half the rate at which new bits arrive. The queue therefore builds up, with a peak after 0.1seconds.

<click> To answer the question, let's look at the occupancy of the queue Q(t) as a function of time. <click> The queue occupancy is the vertical distance between A(t) and D(t) – it's the number of bits that have arrived up until this point, minus those that have departed.

<click> During the first 0.1s the queue builds up as new bits arrive, to a peak of 50bits. During the first 0.2s the average occupancy is 25bits. Then the queue drains from 0.1s to 0.2s until it is empty and all the bits have departed. The queue is empty for 0.8s then the same process repeats.

We can immediately calculate the time average occupancy of the queue. It spends 0.2s with an average occupancy of 25b, then 0.8s with an occupancy of zero. The time average is therefore 5bits.



The second question is: What is the average delay of a bit in the queue?

<click> Let's look at the evolution of A(t) and D(t) again. <click> The delay seen by a bit arriving at time t is d(t), the horizontal distance between the two lines.

<click> The first bit arriving at time zero experiences no delay. Whereas a bit arriving at 0.1s experiences a delay of 0.1s. Notice that no more bits arrive after 0.1s, so it makes no sense to consider the delay of a bit arriving between 0.1s and 1.0s. We are conditioning the probability on a bit arriving, which only happens in the first 0.1 of every second.

<click> Therefore, the average delay seen by a bit arriving to the queue is simply 0.05s.

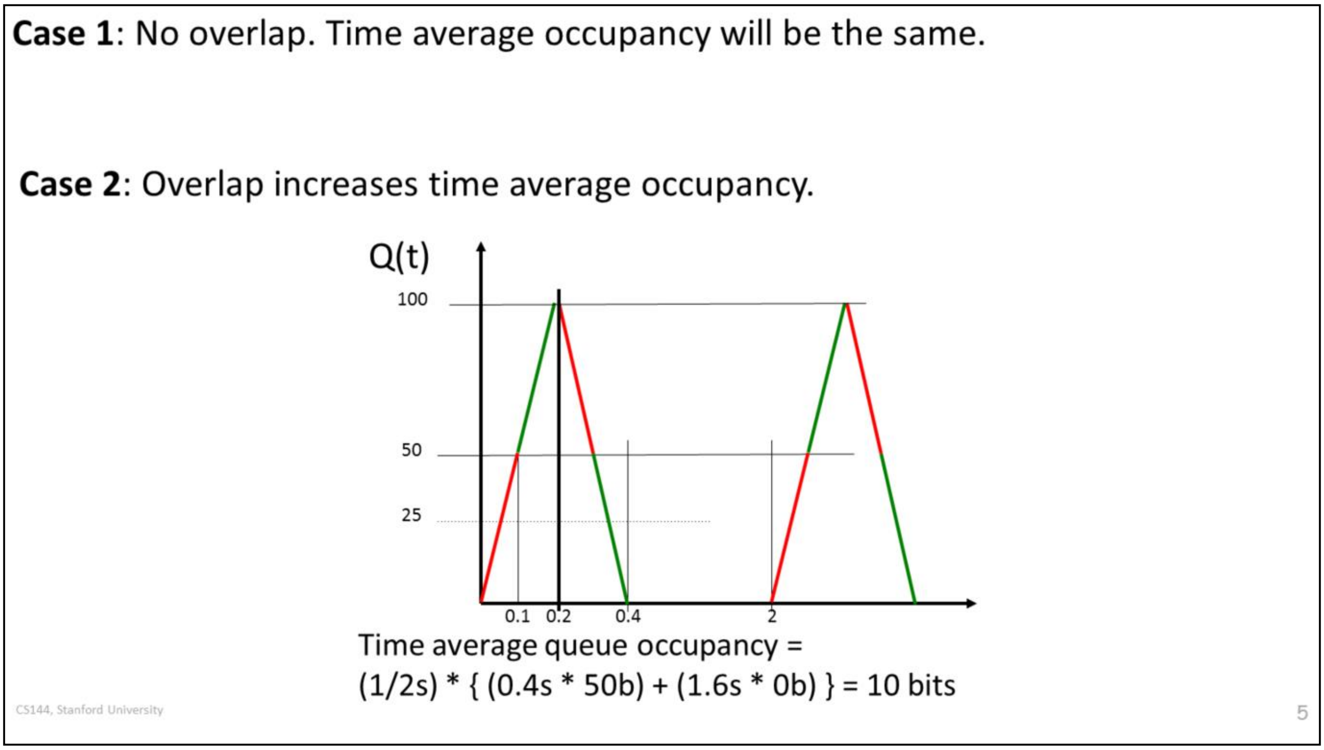
(c) If the trains of 100 bits arrived at random intervals, one train per second on average, would the average queue occupancy be the same, lower or higher than in part (a)?

Higher. When two trains arrived deterministically in part (a), they never overlapped and so there were never bits from two trains in the queue at the same time. But if they arrive randomly, we might have bits from two or more trains in the queue at the same time, pushing up the occupancy, and hence the average occupancy.

Staying with the same queue, the third question is: If the trains of 100 bits arrived at **random** intervals, one train per second on **average**, would the average queue occupancy be the same, lower or higher than in part (a)?

<click> The time average occupancy will be Higher. When two trains arrived deterministically in part (a), they never overlapped and so there were never bits from two trains in the queue at the same time. But if they arrive randomly, we might have bits from two or more trains in the queue at the same time, pushing up the occupancy, and hence the average occupancy.

This might be surprising. Let's look at an example to see why this is true.



Consider two cases.

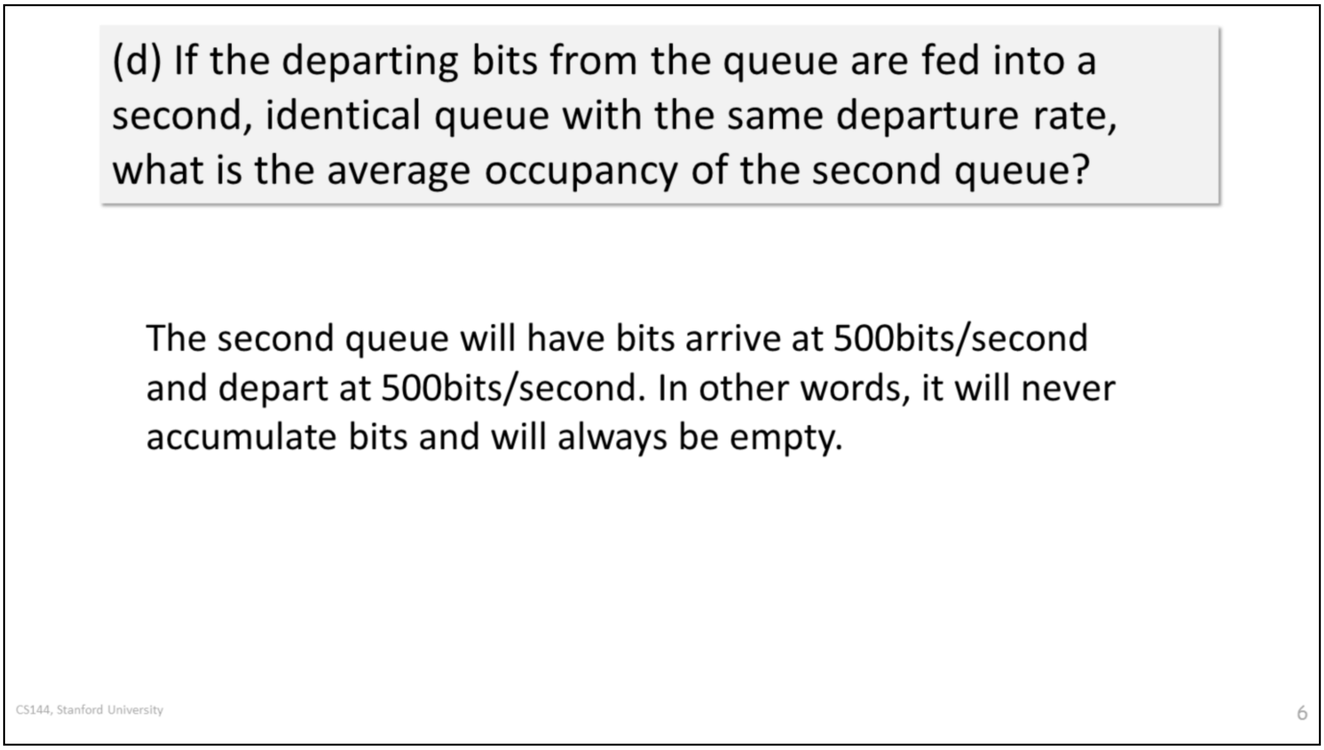
<click> Case 1: Assume the trains arrive randomly, but no two trains ever overlap. The average queue occupancy would be the same as before, which was 5bits.

<click> As soon as two trains overlap, even once, the average will increase. To see why, consider the example here. The red line shows the queue occupancy of tone train which we assume arrives at time 0. Let's say the second train arrives at time 0.1s just when the queue has 50bits in it still from the 1st train. The queue will keep growing because bits are arriving twice as fast as they are leaving. The queue won't drain until time 0.4s. If this happened every two seconds, the arrival rate would be the same as before, but the time average

queue occupancy would now be as follows.

<click> For 0.4s the time average occupancy is 50b, then for 1.6s it is empty. TO get the time average, we divide by two seconds. The time average occupancy is 10bits, which is double what it was before.

Why is that? It’s because the queue only drains at rate 500bits/second and so not only does it fill to twice as much as before, it also takes twice as long to drain. The triangle showing when the queue is non-empty has four times the area as before.



The fourth part of the question is: If the departing bits from the queue are fed into a second, identical queue with the same departure rate, what is the average occupancy of the second queue?

<click> The second queue will have bits arrive at 500bits/second and depart at 500bits/second. In other words, it will never accumulate bits and will always be empty.