Broadband Networks

Prof. Karandikar

Electrical Engineering Department

Indian Institute of Technology, Bombay

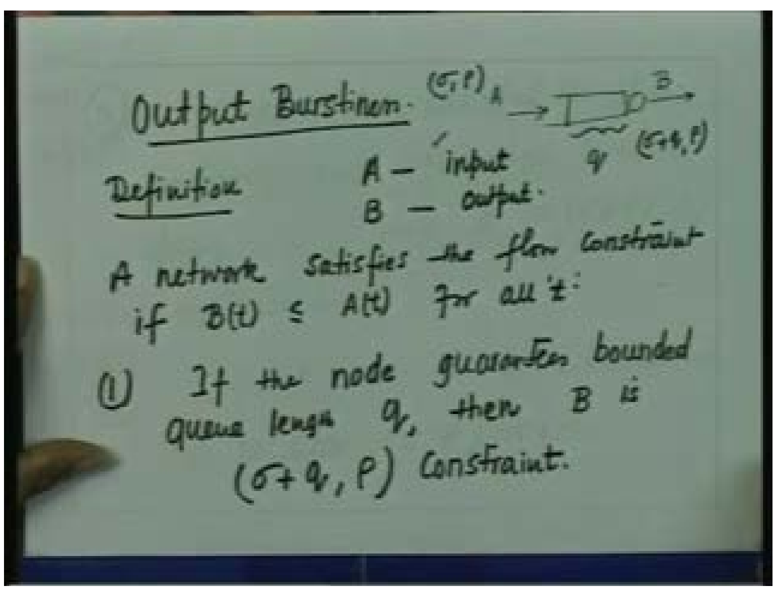
Lecture - 8

(6, p) Calculus For QoS - II

So, we continue our discussion with traffic characterization by a rho sigma parameters and what we had seen earlier is that when a FIFO scheduler or essentially a work on serving link with a transmitter which transmits at c bits or c packets per unit of time is fed by a rho sigma regulator, then what is the effect on the maximum queue length and what is the maximum bounded delay and how we can characterize the output departure process; that we had seen.

Now, in the previous lecture, we were also trying to see what is the effect on the output burstiness that how does the output burstiness get affected when the input to a work conserving link happens to be a row sigma regulated traffics and we had seen that result here.

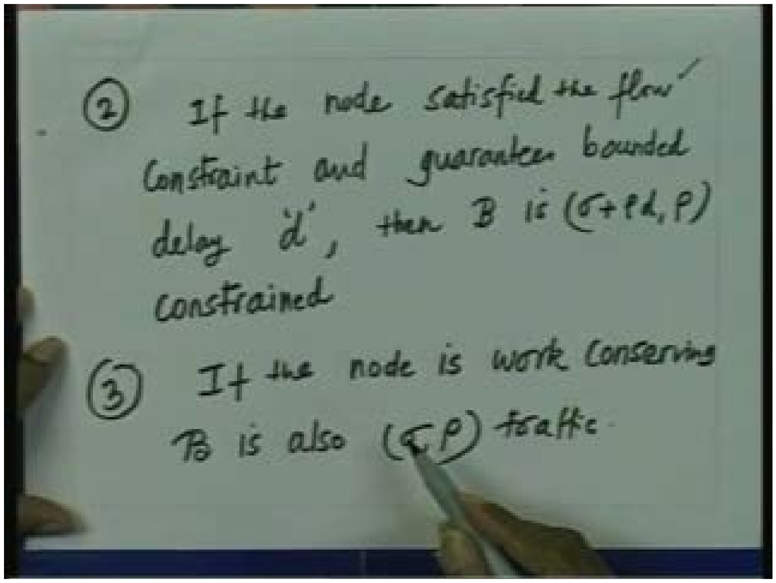
(Refer Slide Time: 1:41)



In the previous lectures I had explained to you the output burstiness and so we had first defined what is meant by satisfying the flow constraint. So, let us say that A is an input to a work conserving link. So, there happens to be a work conserving link. So, A is an input and B happens to be the output. So, we say that a network satisfies the flow constraint if the B (t) that is the output is less than or equal to A (t), for all t.

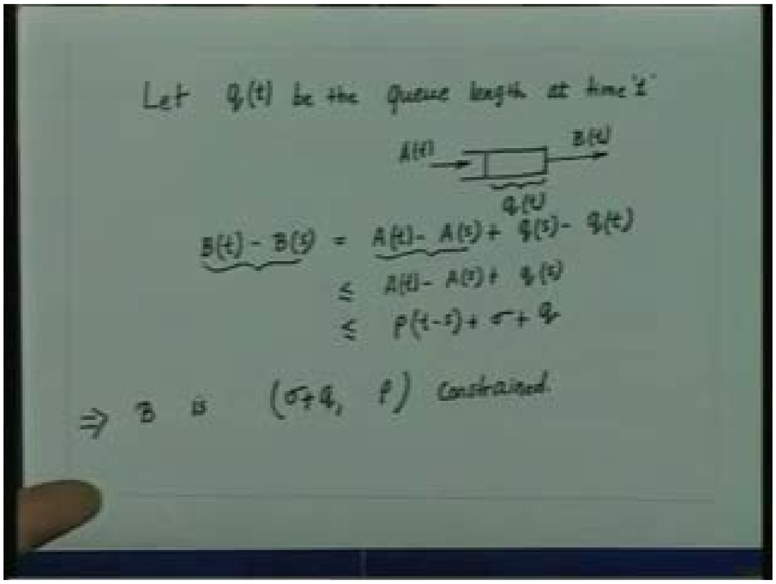
Then if this is the definition of the flow constraint, then we have 3 claims. One - if the node guarantees the bounded queue length q, that means the maximum queue length that is the maximum queue length is guaranteed to be q; then, the output B will be sigma plus q rho constraint. So, the input is sigma row and we say that the output will be sigma plus q rho. So, this is one result that we will shortly prove.

(Refer Slide Time: 2:55)



The other result that we had seen was that if the node satisfies the flow constraint and guarantees the bounded delay. So, in addition to satisfying the flow constraint, if it guarantees the bounded delay d, then the output is sigma plus rho d into rho constrained and lastly, if the node is a work conserving link, then B is also rho sigma traffic. So, these 3 results we would try to prove. So, first let us prove this result that if the node guarantees the bounded queue length q, then B is sigma plus queue or rho constrained. So, let me just prove this result and then we will prove the other two results.

(Refer Slide Time: 3:44)



So, just proof of this is very simple. So, we say that let q (t) be the queue length at time t at time t. So, what we are saying is that here is a queue and here is the input A (t) and here is the output B (t) and q (t) denotes this queue length. Then we know that B (t) minus B (s), note that B (t) is the cumulative number of departures by the time t. So B (t) minus B (s) is given by A (t) minus A (s) plus q (s) minus q (t).

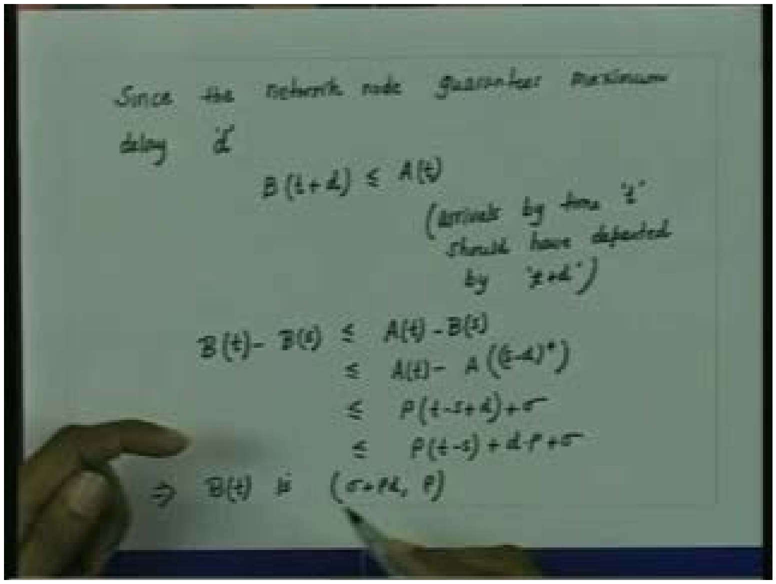
So, this is the number of departures that have occurred during the interval t to s. This is number of arrivals that have occurred during this interval. Now, A (s) A (t) minus A (s) is the number of arrivals that have occurred during this interval and q (s) minus q (t) is the total queue length change. So obviously, this should be equal to the number of departures.

Now, this is less than or equal to A (t) minus A (s) plus q (s) and note that this itself is less than or equal to rho (t minus s) plus sigma and this q is less than or equal to q, which is q is the maximum length then. So, that means that the departure process is sigma plus q into rho constrained.

So, we have proved this result that if a node guarantees the bounded queue length and let us say that maximum queue length is q and if the input is rho sigma traffic; then the output B is sigma plus q into rho, sigma plus q and the rho constrained, if the node is guarantying the maximum queue length of q.

Now, let us assume that this node satisfies the flow constraint that is the B (t) is less than or equal to A (t) and in addition the node guarantees the bounded maximum delay of d; then we will try to characterize what will be the output process B (t). So, we will prove now the second claim that we had stated. And, the second claim was that if the node satisfies the flow constraint and guarantees the bounded delay d, then B is sigma plus rho d into row constrained. So, we will to try to prove this result.

(Refer Slide Time: 7:08)



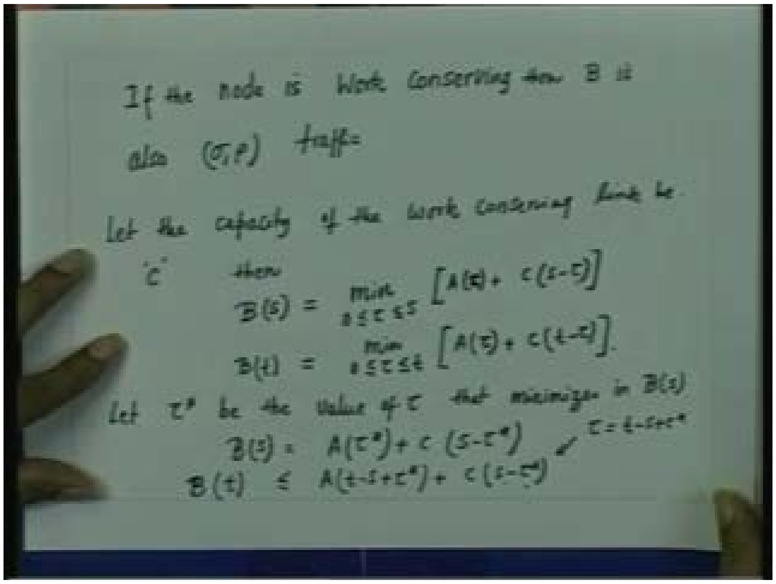
So, now since the network node satisfies or guarantees the maximum delay of d, then this should be true - B (t) plus d should be less than or equal to t. That means those arrivals, all those arrivals which occur by t, they must have left by t plus d. So, all the arrivals, so what does it say is that all the arrivals by time t should have departed should have departed by time t plus d. So, this is what this result is saying.

So, we now have the B(t) minus B(s) is less than or equal to A (t) minus B (s) because obviously B(t) is less than or equal to t, because we have proved we have assumed that the node is guarantying the flow constraint. Therefore, we have assumed that B (t) is less than or equal to A (t). So, with that we are saying that B (t) minus B (s) is less than or equal to A (t) minus B (s) which in addition is less than or equal to A (t) minus A of s minus d plus which indicates maximum of 0 or s minus d which in addition we know that since the process A is rho sigma constrained, this will be less than or equal to rho into t minus s plus d plus sigma which in addition we can show that rho t minus s plus rho into d plus sigma. So, that means that the process B (t) is sigma plus rho d into rho constrained.

So, we have proved this result that if a network node satisfies the flow constraint and in addition guarantees the maximum delay of d, then the output process B (t) is sigma plus rho d rho constrained. That means the maximum burst size will be sigma plus row d and of course, the average rate remains to be rho.

Now, we will prove the third result that if the node is work conserving then the output is also row sigma constrained. So, let us prove the third result.

(Refer Slide Time: 10:26)



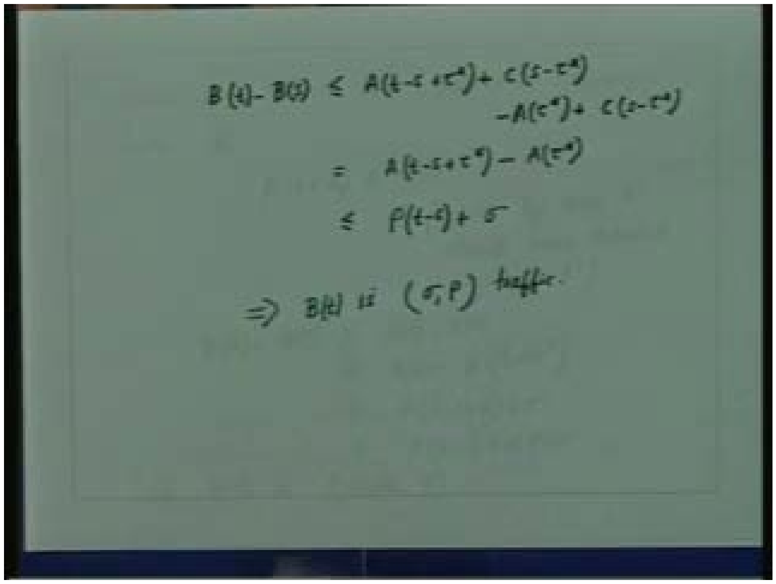
So, the third result is that if the node is work conserving, then B is also… If the node is work conserving, then B is also sigma rho constrained traffic. So, we will try to prove this result and so let us say that let the capacity of the work conserving link be C, then we had already proved that this the departure process B (s) will be equal to minimum of A (s) plus c (s minus tau).

Note that we had proved this result about how in a deterministic queuing system, the departure process behaves for a work conserving link and how the queue length behaves for a work conserving link. Both these results we had proved in our earlier lectures. So, that is what we are trying to say that this B (s) is actually equal to minimum over 0 to s A (s) plus c into s minus tau. And, by the similar … we have B (t) will be equal to minimum over 0 to t A sorry this should be A tau and so similarly, here it should be A tau plus c t minus tau.

Now, let us say in these equations that let tau star be that value, be the value of tau that minimizes this, that minimizes in B (s). Then, we will have Bs which will be equal to A tau star plus c into s minus tau star and B (t) will be less than or equal to in that case A (t) minus s plus tau star plus c into s minus tau star. So, what we are saying? Let us say that tau star is the value that minimizes this; then we have B (s) is equal to A tau star plus c into s minus tau star that is what we have written.

Now, B (t) is this. So now, if we choose the tau to be equal to t minus s plus tau star in this equation, for the B (t) equation; then we will have B (t) is less than or equal to. Say instead of tau, we have put t minus s plus tau star and it has to be less than or equal to, because we are minimizing it over 0 to t. So, c (s minus tau star) minus, so us right, so we have put this.

(Refer Slide Time: 14:07)



So now, we have B (t) minus B (s). So, if you write B (t) minus B (s) that will then become less than or equal to A (t minus s plus tau star) plus c (s minus tau star) minus A (tau star) plus c into s minus tau star. This, we get simply by subtracting B (t) minus B (s). So, this A (t minus s tau star) minus A (t tau star) - from subtraction of this which actually is equal to A (t minus s plus tau star) minus A of tau stars which as you know is less than or equal to rho of t minus s plus sigma. So, that means the traffic B (t) is sigma rho traffic.

So, this is how we have proved the 3 results on the characterization of the output burstiness. One - if the node guarantees the maximum queue length q, then the output is sigma plus q rho constrained. If the node satisfies the flow constraint, we need an additional assumption of it satisfying the flow constraint and the bounded delay d; then the traffic is sigma plus row d into row constrained. If the node happens to be a work conserving link, then the output is also sigma row traffic. So, these three results actually completely characterize the burstiness of a node, where the input is rho sigma traffic.

So now, what we have seen till now in our discussion is that if the traffic is regulated by a rho sigma regulator; so the first we proved about the multiplexing. That is such traffics are multiplexed by an ideal multiplexer; then what will be the output process and what we saw? We saw that if such traffics are multiplexed by an ideal multiplexer, then the output is also a rho sigma traffic where this sigma is the sum of the sigma's of all the traffic and the rho is the sum of the rhos of the all the input traffics that is what our ideal multiplexer was. Then, we tried to characterize that what will be the characteristics if a node is fed by this rho sigma traffic, then what can we say about the queue length and then we proved that a maximum queue length will be bounded by this sigma. That is one result we proved.

The second result we proved that a delay will be bounded by sigma upon c minus rho where c is the capacity of this work conserving link and a rho that is the average rate happens to be strictly less than the capacity of the link.

In addition, if we invoke this assumption that the link happens to be a FIFO scheduler, then we can say that the maximum delay will be bounded by sigma by c, where now rho is strictly rho is less than or equal to c. So, that is how we characterize the queue length and the delay.

And thirdly, we have just now proved that what will be the effect on the output burstiness if the input happens to be a rho sigma regulator. The other results that we have proved is that this rho sigma traffic can be generated by a token bucket regulator whose maximum depth is sigma and whose token generation rate is rho.

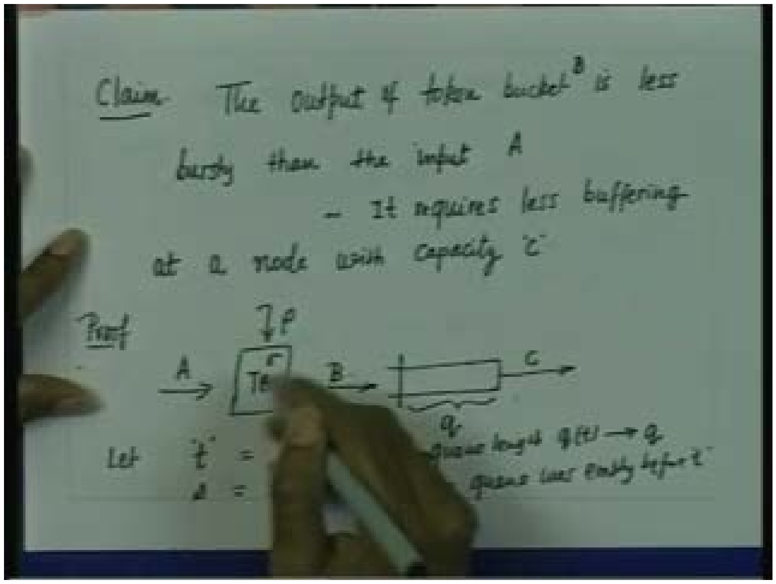
So, if we have a random bursty traffic and if we pass through a token bucket regulator whose bucket depth is sigma and its token generation rate is rho and every traffic you know takes away as many number of token as is the length of the packet; then the output of such a token bucket regulator will be a rho sigma traffic.

In addition, we have proved another result and that is very important result that if this random bursty traffic needs to be regulated by a token bucket regulator, if there are other FIFO controllers that also can generate the rho sigma regulated traffic; then among all these FIFO controllers, token bucket regulator is the best, token bucket is the best rho sigma regulator, in the sense that it will delay this traffic at the source the least among all the controllers that are trying to generate the rho sigma regulated traffic.

Now, we are going to prove the last result of this rho sigma traffic characterizations and that result is that the output, the output of the row sigma regulator will always be less burstier than the input traffic - that is what we will try to prove. I mean, the input traffic happens to be some kind of a random traffic and the regulated traffic, input traffic happens to be an unregulated traffic. That means, by putting this row sigma regulator, we will always need a less buffering at the downstream network nodes. That means token bucket regulator is essentially acting as some kind of a smoother. So, that is what we will try to prove this last result.

So, let me just prove, state this result and then well, we will prove this result.

(Refer Slide Time: 19:46)

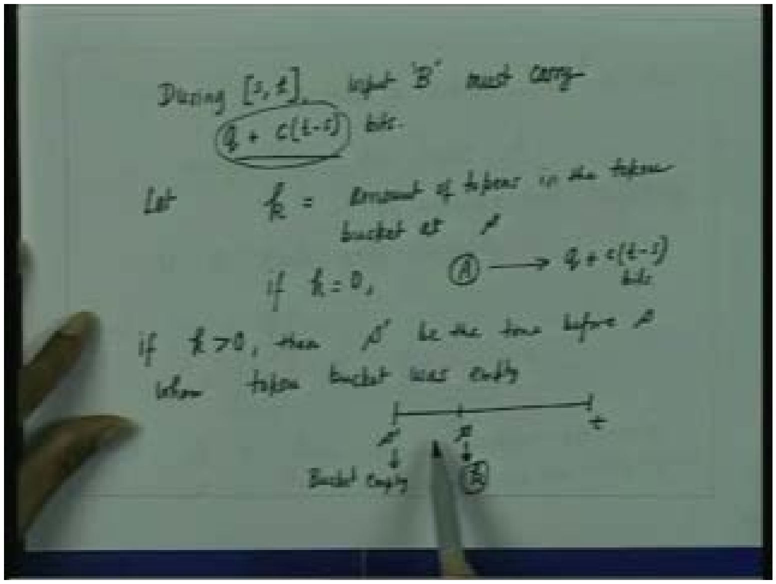


The claim is the output of a token bucket regulator is… so let say, let us call this output to be some B is less bursty than the input A. So, what you mean by less bursty? By saying that less bursty means it requires less buffering at a node with the capacity c. This is what we mean.

So, what we are essential trying to say that if this A happens to be an on the regulated traffic and the output of this token bucket regulator is now a regulated traffic B, what we are try to prove is that B is less burstier than the A. If it is less burstier, obviously it will require less buffers at the downstream network nodes. That is what we will try to prove. So, we will prove this.

So essentially, this is your downstream, possible downstream network nodes which is serving at a rate of c and this has a queue length, let us say a q. The input to this is B. So, let us say that t happens to be the time when the queue length queue length q (t) is b is q and let us say s is the last time when queue was empty before t. So, at time t, the queue length has reached queue and s was the last time the queue was empty. So during the period from s to t, the B must have carried how many numbers of bits? The B must have carried q plus c into t minus s.

(Refer Slide Time: 22:49)



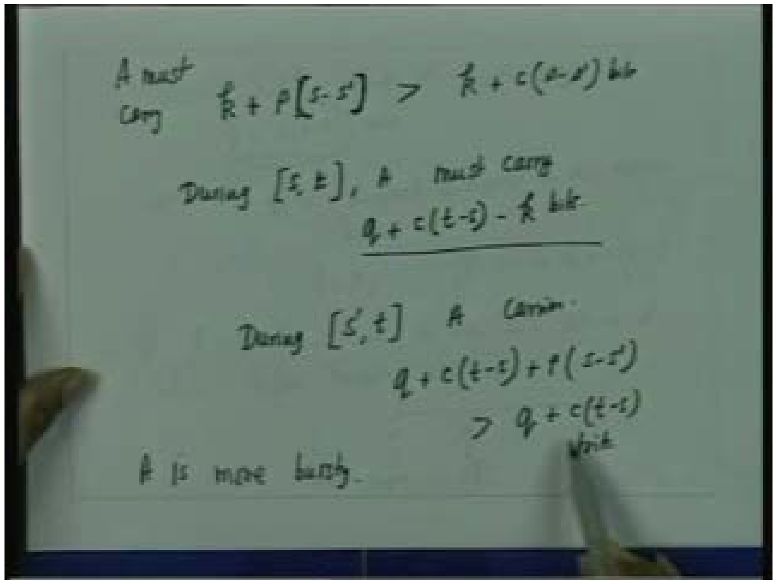
So, we will just write by saying that therefore during this interval s to t; this input B, input B to the node which is the output of this token bucket regulator, the input B must carry q plus c into t minus s bits. Now, let us say k is the amount of tokens in the token bucket at time s. Now, note that token bucket regulator so here is the token bucket and the input is A and this is like, the token rate is rho and the bucket depth here is happens to be sigma.

So, what we are saying is that let us say that k is the amount of tokens in the bucket at s. Now, k is 0, then the A should have carried at least q plus c into t minus s bits. So, we have trivially proved this result. So, what we are trying to say is that there is downstream node which is being served the capacity of c and this node, let us say at time t has reached a capacity of q of a queue length. So obviously, the input must have transmitted this q bits plus during the interval t to s, those many bits; how many bits? c into t minus s bits must have been outputted because at time t, the queue was empty.

Now, this input traffic B itself is coming from a token bucket regulator itself is coming from a token bucket regulator. So, since it is coming from a token bucket regulator, we say let us say that at time t, when the last time this downstream buffer was empty, the k is the amount of tokens in the token bucket. So, if k is the amount of tokens in the token bucket and that k was 0, then this input A must have transmitted at least q plus c into t minus s of bits.. So, I trivially proved.

However, if k is greater than 0, then let us say that s prime be the time before s when the token bucket was empty. So, it is something like this that here is the s, here is t, here is s prime. So here, the token bucket has a token equal to q and here the bucket was empty. Now, to accumulate k amounts of tokens, k units during this, the m that the input A, it carries at least how many bits?

(Refer Slide Time: 26:47)



To accumulate these many amount of fluids, the input A must carry at least k plus rho into s minus s prime. So, what we are saying is that to accumulate this k units, the input A must carry this. So, A must carry these many number of bits which is greater than k plus c into s minus s prime bits. This is rho into s minus s.

So, during therefore you know s to t interval, A therefore must carry q plus c into t minus s minus k bits and hence during s prime to t, A carries q plus c into t minus s plus rho into s minus s prime which is greater than q plus c into t minus s bits. So, that means A is more bursty. Let me just re-sketch the proof again.

So, what we are saying is that during s to t, input B must carry these many number of bits because the queue length is q, the maximum queue length is q and the link capacity is c, the buffer was empty at time t and the buffer has a queue length of q at time s. So, during that interval of s to t, you have accumulated queue bits and you have transmitted c into t minus s bits. So, input B must carry this.

Now, let us say that k is amount of tokens in the bucket at time s. That is this k is 0, then obviously A also carries the same number of bits. So, that means A is at least as bursty as B or other A is as smooth as B. But now let us say that this k, that is the amount of tokens in the bucket was not 0 at time s and this was greater than 0. Then let us say that s prime be the time before s when the token bucket was empty. Now, this is so.

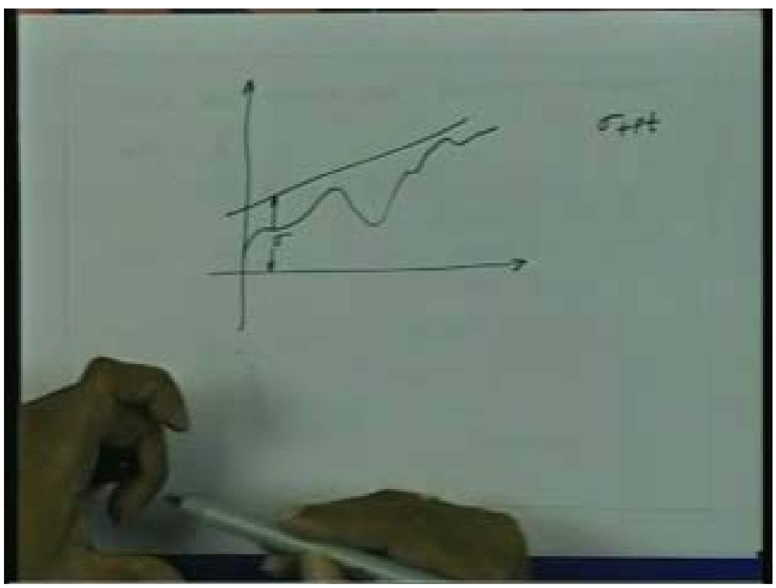
Now here, the token has accumulated k amount of fluid or k amount of token and here the bucket was empty. So, that means during this interval, A must carry k plus rho into s minus s. Remember, the token generation rate is rho. So, during this interval, these many tokens would have accumulated.

So, A therefore must have transmitted these many bits and they are greater than k plus c into s minus s bits. So, that means during s to t interval, A must carry these many bits and therefore during s prime to t interval, A must carry q plus c into t minus s plus rho into s minus s bits and that means they are greater than q plus s into t minus s bits. So, we have proved that A is actually more burstier than the output B. Now, this completes our discussion of rho sigma regulated traffic. As I have already pointed out is that rho sigma is a deterministic way of characterizing the traffic.

In practice, as we know that it is not possible to give a characterized a traffic statistically, in the sense either by its distribution functions or density functions or even by its effective bandwidth, it is difficult to determine the effective bandwidth of an otherwise a statistical traffic source. So typically, in practice, what will be done is that traffic source will be regulated by this deterministic envelope and the commonly accepted deterministic envelope is a linearly bounded arrival process which is parameterized by sigma and rho and a linearly bounded arrival process can be very easily generated by a token bucket regulator which has a bucket depth of sigma and the token generation rate of rho.

So, as a result what we have seen is that we have a arrival process whose envelope is bounded.

(Refer Slide Time: 32:00)



So, this is like an arrival process whose envelope is bounded. So, this is a sigma and the rho is the token generation rate. So, this is like a straight line equation of sigma plus rho t. So, these are with respect to the time and these are number of bits which are generated. An arrival process may have something like this. So, what we are giving is an envelope to this arrival process and by having this deterministically bounded arrival process, we have seen that if this arrival process, now that is the rho sigma regulated traffic is given to queuing system which is or a work conserving link which is transmitting these packets; then we can ensure the queue length, we can ensure the bounded delay and we can predict the output burstiness of the traffic also.

So, as a result what we have essentially seen is that if this traffic regulator sorry if this traffic is a rho sigma regulated, the network node will be able to give certain quality of service guarantees to the traffic source in terms of the maximum delay or the maximum queue length.

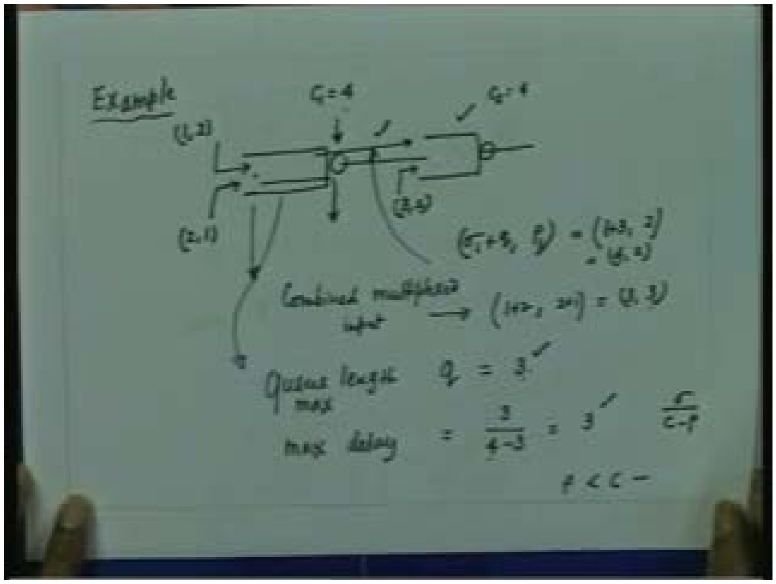
So, as a result the network is able to give quality of service guarantees. Now, we have just considered a very simplified case where we have seen that the network node essentially happens to be just a simple common buffer. So, it multiple such rho sigma traffic is transmitting to this simple common buffer, then we can apply these results. But we have not considered the case when there are multiple queues and there is a scheduler which is scheduling these queues. Those cases we will take up in our subsequent discussions, we have not considered that case.

Similarly, we have not considered that case that what happens when these nodes are connected together in a multi node network which is typically the case in a practical network where we have a network of such nodes. So, we have not considered that but these results are applicable, these results can be extended and generalized to a multi node network case and we can show that the quality of service guarantees can be given the end to end delay bounds can be given and an end to end burstiness can be maintained if the input traffic happens to be a rho sigma regulated traffics.

So, hence the importance of this rho sigma regulated traffic which have been the subject of extensive research both in the theory as well as in the practical implementations and there have been; therefore lot of debates as to what should be the appropriate values of the rho sigma parameters that a traffic source should choose so that it not only gets a quality of service guarantees which the network is offering to it but at the same time, it is able to characterize the traffic source accurately enough. That is that means these parameters are representative of the traffic source parameters.

So, with this we conclude our discussion of the rho sigma traffic characterizations.

(Refer Slide Time: 35:19)



We will see in this rho sigma regulated traffic characterization with an example. So, let us take an example. So, let us say there are 2 nodes. This c_1 has a capacity of 4 and this c_2 also has a capacity of 4. There is an input here, there is an input here of a rho sigma which let us say that this has a (1, 2) characteristics and there is another input which has (2, 1) and let us say that this input goes out here and however, this input continues here.

Now, there is an input (1, 2), there is an input 2 1, we apply the principles of multiplexing; what will be the combined input? The combined input will be combined multiplexed input here at this node. What will be that? The 2 sigmas will be added, that is what we had said. In an ideal multiplexer if the input is rho 1, sigma 1 and if another input is rho 2, sigma 2; then the multiplexers output will be sigma 1 plus sigma 2 and rho 1 plus rho 2.

So, in this particular case, what do we get? We get 2 plus 1, 3 and 2 plus 1, 3. So, in that case, 1 plus 2 that is sigma 1 plus sigma 2 and rho 1 plus rho 2 that is 2 plus 1. So, we actually get 3 into 3 is the output. Now, what will be queue length that will get bounded here? The queue length, the maximum queue length, the maximum queue length is bounded by sigma. So, the queue length q is, maximum queue length is 3 and what is the maximum delay? The maximum delay was sigma. So, that is in this case; 3 upon c that is 4, minus rho that is 3, which is again 3. So, the queue length is 3 and the delay is 3.

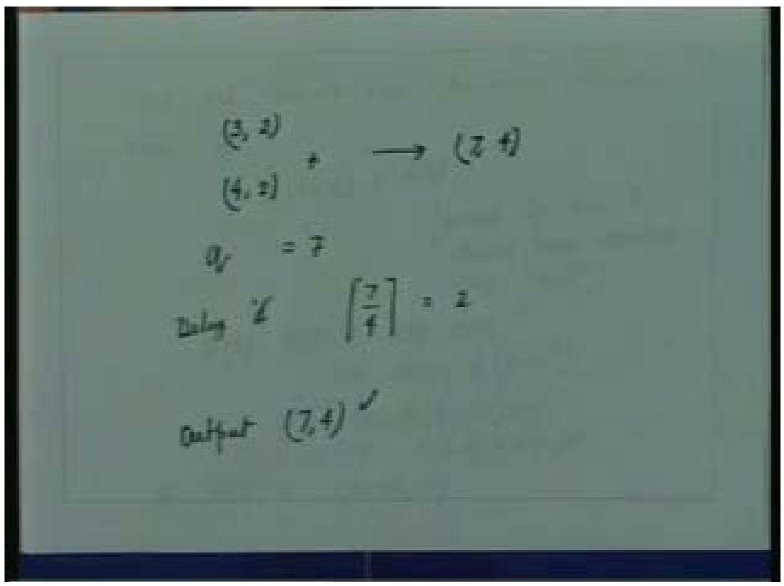
So, this is we have used the formula as sigma upon c minus rho. So, sigma is 3, this 3 and c is 4 which is this 4 and rho is 3 of the combined input and therefore this 3. Now, so this A_1 output goes here, so what will be this output characterization we would like to know? What will be this output characterization?

Now, note that a maximum queue length is 3 and therefore if you assume that all these packets, the worst case belongs to only A; then this output process which will be characterized which is going as an input to this will be what? Sigma 1 plus q into rho 1; that is what we had proved that is the queue length is bounded by q, the output will be sigma 1 plus q into rho.

So now, what is sigma 1? Sigma 1 is the input A traffic that is 1. What is q? The q is maximum length 3. So, that is 3 here and what is rho 1? Rho 1 is here, so we have 3 plus 1 that is the (4, 2) traffic. So, this is the input traffic.

Now, let us say that this input traffic is something like you know, (3, 2). So, this is (3, 2), this is (4, 2), so what will be the combined traffic?

(Refer Slide Time: 39:23)



The combined traffic will be, the multiplexed traffic will be, so this one is (3, 2), another is (4,2), when they are multiplexed; we get is (7, 4). So, input to this case is given by the (7, 4) traffic. So now at the node 2, what will be the queue length, the maximum queue length? The queue length will be sigma. So, the maximum queue length is equal to sigma. What will be the delay? What will be the maximum delay?

Now, if you see here, maximum delay, note that rho is strictly less than c, here in this case. That is because rho is 3 and c is 4. But what happens at the second link? Second link, the c is 4 and the rho is how much? The rho is also 4. So therefore, c is actually equal to rho. So then, we need to make an additional assumption here as we had seen previously, because in this case we cannot apply this result of the delay because otherwise in this case, the delay will be bounded by infinity and we do not get any proper delay bound.

So, then we have to invoke an additional assumption that let us say that at the second link, the scheduling policy is FIFO. So, the scheduling policy is FIFO; then the maximum delay will be bounded by sigma by c. So therefore, the delay d will be bounded by sigma which is 7 here by 4.

We will take an integer part of this, so which is which is 2. So, therefore this output is given by a 4.

What is the output characterization? We can well, we can say the output characterization; since the maximum queue length is 7 here, the output will be 14 plus 4. But we can get a better bound. Since this is the work conserving link, the output will also be a rho sigma regulated traffic and therefore the output of the second link is also (7, 4) which gives us a better bound. So, in this manner what I was trying to say is that in this manner we can characterize, I mean I have just given you an example. We can characterize that if we have this various network nodes which are connected together, then we can characterize the output burstiness and we can characterize the departure process, we can characterize the delay at each of the nodes and as a result, we can characterize the end to end delays of a multi node network.

We have of course considered a single queue here as I was just trying to mention that when we considered a non first in first out scheduling scheme that means that there are several queues and there is one scheduler which is trying to schedule out of these queues; we will see that it is possible to guarantee or characterize the delay bound if these input you know, if the input to all these queues happen to be the rho sigma regulated traffic.

So, if they happen to be rho sigma regulated traffic, it should be possible to characterize the delays even in the case of a non FIFO, non first in first out scheduler and with a multiple queuing classes. So, that we will see in our subsequent lectures but this concludes our complete discussion on the rho sigma regulated traffic.