Worst-case Ethernet Network Latency

Max Azarov, SMSC

26th September 2005

For 802.3 ResE study group

1 Assumptions

In order to derive mathematically a worst-case latency we will need to make certain assumptions about the network we are dealing with. Following assumptions are set forth:

- All switches on the network are straight store-and-forward and function as output queue switches
- All receiving and transmitting ports are functioning independently (HW router and full duplex)
- All traffic has the same priority.
- Processing time for packets (time between reception and start of transmission of the target port) inside switches is considered to be zero.
- All packets have the same size.
- PHYs on switches have identical speed and transmit/receive single packet in fixed amount of time of τ seconds.
- Packet source and sink are separated by N switches and each switch has n input ports with one source connected to each port.
- Network is never congested, we will define this as sum of incoming traffic to the switch targeted on the same output port over any period of time $n \cdot \tau$ does not exceed output port capacity. This should be true for each port of each participating switch. This effectively means that no more than n packets targeted for one output port come from all input ports during the period of time $n \cdot \tau$. This assumption put burden of traffic shaping on sources.

2 Theorem

With the assumptions set forth in section 1 worst-case propagation delay for the packet from the source to the sink will be

$$T = (n \cdot N + 1) \cdot \tau \tag{1}$$

3 Proof

Proof will consist of two parts:

1. construction of example network with delay expressed with formula 1,
2. proof from the opposite that worse propagation value is not possible

First let's build the network with needed propagation delay. For this we will imagine that our source of interest emitted a packet, which we refer to as marked packet. All interfering sources on the network, i.e. all sources but the one we measure propagation delay for, emit packets in such a fashion that they, for each switch, all arrive and get queued for transmission just before the marked packet. This means that when marked packet arrives there are $n - 1$ packets queued waiting to be transfered.

Since there are $n-1$ packets in the queue, it will effectively take $\tau \cdot (n-1) + \tau = n \cdot \tau$ seconds for packet to reach the next switch or sink, in case if this switch was last on the path.

Note that because of the way interfering sources emit their packets, by the time marked packet gets to the next switch on its path, packets that where in front of it in the previous switch get transmitted further down the network and same situation repeats where marked packet finds $n - 1$ packets waiting to be transmitted ahead of it.

Since by our assumption there are N switches between the source and the sink and every switch produces $n \cdot \tau$ seconds of delay, adding the time τ is takes for marked packet to reach a first switch, we get the value for the total propagation delay as

$$T = N \cdot n \cdot \tau + \tau = (N \cdot n + 1) \cdot \tau$$

, which is identical to the expression 1 we're trying to prove.

Now we shall show that given expression is the upper bound.

Let assume that this is not the case and there is a configuration which causes greater propagation delay $\tilde{T}$. This would mean that at least on one switch marked packet found more than $n-1$ packets enqueued when it arrived. At the minimum there was n packets queued, so adding marked packet we will get minimum $n + 1$ packets in the queue total.

Lets show that no congestion assumption made in section 1 is equivalent to demanding that at no time any output port on switch has more than n packets in queue.

Indeed, output port is a leaking bucket with a constant leakage rate, equal to the link capacity. Number of queued packets can be expressed as

$$m(t) = \begin{cases} \alpha = m(t - n \cdot \tau) - p_{tx} + p_{rx} & \text{—}\alpha > 0 \\ 0 & \text{—otherwise} \end{cases} \tag{2}$$

,where p_{rx} is a number of packets arrived from input ports and p_{tx} is the number of packets transmitted to the output port. By assumption of non-congested network $p_{rx} \leq n$.

Naturally $p_{tx} \leq n$ since n is the maximum number that can be transmitted at the maximum transmission rate and $p_{tx} \leq m(t - n \cdot \tau) + p_{rx}$ since switch cannot transmit more packets than available to be potentially transmitted. We can write these two relations in shorter form as

$$p_{tx} \leq \min(m(t - n \cdot \tau) + p_{rx}; n). \tag{3}$$

Let's consider a moment of time t_0 when $m(t_0) > n$ for the first time. This means that $\forall t < t_0, m(t) \leq n$. We can ensure that such t_0 exists if we assume than $m(t_s) = 0$, $\forall t_s < n \cdot \tau$, which means that initially router had no packets queued for the time period of $n \cdot \tau$.

This allows us to write particularly that $m(t_0 - n \cdot \tau) \leq n$. This in turn means that

$$p_{tx} \geq m(t_0 - n \cdot \tau) \tag{4}$$

, since at the minimum all packets queued at the time $t_0 - n \cdot \tau$ would have been transmitted by the time t_0 because there was less then n of them.

Now if we put together 3 and 4 we get

$$m(t_0 - n \cdot \tau) \leq p_{tx} \leq \min(m(t_0 - n \cdot \tau) + p_{rx}; n) \tag{5}$$

Now using 2 we will write an expression for maximum value of $m(t_0)$

$$\max m(t_0) = \max(\max(m(t_0 - n \cdot \tau) - p_{tx} + p_{rx}); 0) \tag{6}$$

Now we will re-write 5, by subtracting $m(t_0 - n \cdot \tau) + p_{rx}$ as

$$-p_{rx} \leq p_{tx} - (m(t_0 - n \cdot \tau) + p_{rx}) \leq \min(m(t_0 - n \cdot \tau) + p_{rx}; n) - (m(t_0 - n \cdot \tau) + p_{rx}) \Rightarrow$$

$$m(t_0 - n \cdot \tau) + p_{rx} - \min(m(t_0 - n \cdot \tau) + p_{rx}; n) \leq m(t_0 - n \cdot \tau) + p_{rx} - p_{tx} \leq p_{rx} \Rightarrow$$

$$m(t_0 - n \cdot \tau) + p_{rx} - p_{tx} \leq p_{rx} \Rightarrow$$

$$\max(m(t_0 - n \cdot \tau) - p_{tx} + p_{rx}) \leq p_{rx} \leq n$$

, here we used $p_{rx} \leq n$ (assumption of non-congested network).

Substituting to 6 we get

$$\max m(t_0) \leq \max(n; 0) \leq n.$$

Thus we get contradiction with means that such t_0 does not exist and our initial proposition that at least one of switches would have at least $n+1$ packets queued contradicts with our assumptions of the network not getting congested.

Theorem is proved.

4 Formula generalization

We've assumed in theorem above that traffic is shaped to not exceed network capacity on intervals of time $n \cdot \tau$. It is easy to show by rewriting expressions 2-6 that if traffic is shaped on intervals of time $k \cdot \tau$, $k > n$, $k \in \aleph$, it would be equivalent to requiring that at no time any switch/port would have more than than k packets queued in the output queue.

This generalization suggests that under new assumption formula 1 will look like

$$\hat{T} = (k \cdot N + 1) \cdot \tau. \tag{7}$$

Proof that this expression is an *exact* upper bound is not provided here, but it's easy to show that this will be an upper bound.

Indeed, if there's no switch on marked packet's path that have more than k packets queued at any time, than packet will spend at most time $k \cdot \tau$ traversing each switch. Counting in initial time τ to reach the first switch we come to formula 7.

In order to prove that this is an exact upper bound one merely would need to construct an example.

In essence formula 7 suggests that if we shape traffic at sources more coarsely than propagation delay upper bound will increase.

5 Conclusions

We have produced an exact upped bound (worst-case) for propagation delay under assumptions outlined in section 1. Formula 1 suggests that worst-case propagation delay does not change with changing link utilization. It does change with number of hops between source and the sink, number of ports on each switch and also with the granularity of traffic pacing.

Assuming standard five-port switches ($n = 5$) and one packet transmission time of $\tau = 100\mu s$ we will get following propagation figures (using formula 1):

number of hops N	1	2	3	4	5
propagation μs	600	1100	1600	2100	2600

This is under original assumption about proper traffic shaping on the interval of $n \cdot \tau = 500\mu s$.