

Buffer Bounds of a FIFO Multiplexer

TUD-FI03-15-November-2003

Jork Loeser

TU Dresden, Germany
contact@os.inf.tu-dresden.de

Abstract

The goal of this paper is to give upper bounds for the delay of a frame and upper bounds for the memory required in an Ethernet switch, depending on the traffic received by this switch.

We analyze input traffic characterized by two different arrival curves: 1) traffic described by a T-SPEC and 2) so-called on-off traffic.

We assume the switch to be output-buffered and to operate in FIFO mode without any priorities.

1 General Assumptions

Figure 1 shows a typical Ethernet switch. The switch has $N=4$ receive (rx) ports, control logic (switching fabric), buffer space and N queued transmit (tx) ports¹. The receive port can receive data with bandwidth C , this is 100MBit/s for Fast Ethernet and 1000MBit/s for Gigabit Ethernet. The transmit ports send data with bandwidth C .

When a frame arrives at the switch, the control logic determines the transmit port and tries to transmit the frame immediately. If the port is busy because another frame is already being sent, the frame is stored in the transmit ports queue, which is a first-in first-out (FIFO) queue. The memory to store pending frames is obtained from a shared memory pool. If no more memory is available, the received frame is dropped.

For the rest of this paper, we concentrate on one output port. We derive formulas for the buffer bounds of one output port. To get buffer bounds for the whole switch, one has to determine the bounds for each output port separately and accumulate them.

1.1 Some Definitions

For the rest of this paper, we use the following terms to refer to times related to frame transmission.

¹We do not consider using multiple priority queues in switches.

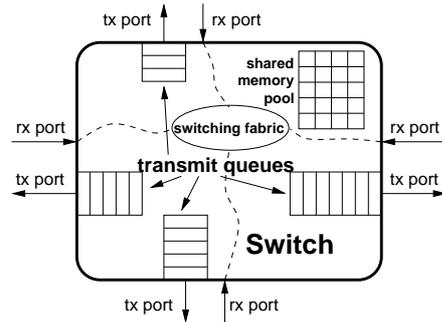


Figure 1: Buffering inside an output-queueing Switch. If queueing a frame is necessary, memory is allocated from a shared memory pool and assigned to the corresponding queue.

switch multiplexing delay (t_{mux}) is a switch-specific parameter describing the maximum delay (without queueing effects) after which the switch starts to transmit a frame once it is received. For almost all switches such a t_{mux} can be found. It is mainly determined by the switching fabric and is typically less than 0.1ms.

queueing delay (t_{queue}) is the time a queued frame sits in the queue of a switch plus the time needed to transmit it finally. With first-in-first-out queues (FIFOs), queueing delays solely depend on the queue length, and bounding this length results in bounded queueing delays.

switch delay (t_{switch}) is the time a frame is delayed at a switch. $t_{switch} = t_{mux} + t_{queue}$.

1.2 General bounds

$\alpha_k(t)$ denotes the *arrival curve* [1, 2] of the traffic received at port k – in any time interval of length t no more than $\alpha_k(t)$ bytes are received at port k ($k=1..N$).

$\beta(t)$ denotes the *service curve* [1] of the transmit port at the switch. We assume this service curve to be a *rate-latency* function $\beta(t) = C * (t - t_{mux})^+$, with $(t - t_{mux})^+$ defined to be 0 for $t < t_{mux}$. The rate-latency function basically means that the switch starts sending no later than t_{mux} time units

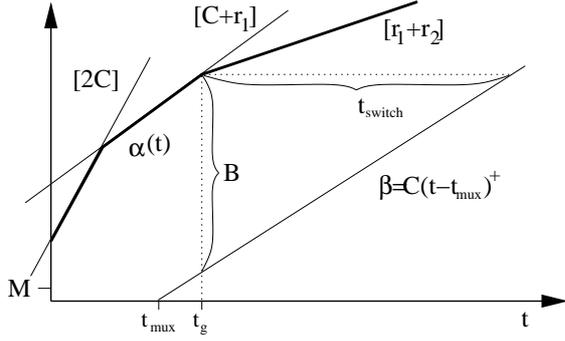


Figure 2: Illustration of the arrival curve $\alpha(t)$ (thick line) as the sum of two flows (M, C, r_1, b) and (M, C, r_2, b) . The slopes of the 3 parts of $\alpha(t)$ are $2C$, $C + r_1$ and $r_1 + r_2$.

after receiving the first byte, and that during a busy period its sends with a rate C .

According to Le Boudec [1], the maximum backlog B , that is the maximum memory required for buffering, of a system that offers a service curve β to a flow that is constrained by an arrival curve α is given by the vertical deviation

$$B(\alpha, \beta) = \sup_{s \geq 0} \{\alpha(s) - \beta(s)\}. \quad (1)$$

As the data of all receive ports is multiplexed on the transmit port, the arrival curve α of the traffic for the switch output port is the sum of the arrival curves of all receive ports, thus $\alpha = \sum_{k=1}^N \alpha_k$.

2 Flows constrained by T-SPECs

We characterize the arrival curves using a *T-SPEC* (C, M, r_k, b_k) , a specific form of traffic description commonly used in the context of ATM: $\alpha_k(t) = \min(Ct + M, r_k t + b_k)$. A T-SPEC takes the maximum transmission rate C as well as the maximum packet size M into account. r_k specifies the long term average rate, and b_k allows for some burstiness. Typically, $M < b_k$ and $r_k < C$.

Obviously, the sum of the long term average input rates r_k of traffic for the switch output port must not exceed its output rate, thus

$$\sum_{k=1}^N r_k \leq C \quad (2)$$

must hold.

The arrival curve $\alpha = \sum_{k=1}^N \alpha_k$ is

$$\alpha(t) = \sum_{k=1}^N \min\{Ct + M, r_k t + b_k\}. \quad (3)$$

Figure 2 illustrates the arrival curve of a switch with two receive channels and its service curve. The maximum backlog is the maximum vertical distance between α and β . We

define g_k to be the time of the inflexion point of arrival curve α_k , thus

$$g_k = \frac{b_k - M}{C - r_k}. \quad (4)$$

We define g_{max} as the maximum of all g_k .

We use (1), (3) and (4). If $t_{mux} \leq g_{max}$, we get

$$B = \sum_{k=1}^N b_k + \sum_{k=1}^N r_k * g_{max} - C * (g_{max} - t_{mux}) \quad (5)$$

$$= \sum_{k=1}^N b_k - g_{max} * (C - \sum_{k=1}^N r_k) + C * t_{mux} \quad (6)$$

According to (2), the second addend in the last equation is negative or zero, and hence a safe bound (which is not tight) for the backlog for the case $t_{mux} \leq g_{max}$ can be given by

$$B_{est} = \sum_{k=1}^N b_k + C * t_{mux}. \quad (7)$$

If $t_{mux} > g_{max}$, then the backlog is given by $\alpha(t_{mux})$, which is

$$B' = \sum_{k=1}^N b_k + \sum_{k=1}^N r_k * t_{mux}. \quad (8)$$

Again we use equation (2) and conclude that a safe bound for the maximum backlog required at the switch can be given by

$$B'_{est} = \sum_{k=1}^N b_k + C * t_{mux} \quad (9)$$

which is the same as B_{est} . This means, the memory required in the switch can be estimated by the sum of the bursts in the T-SPECs plus a small fixed amount ($C * t_{mux}$).

According to Le Boudec [1], the maximum delay d of a system that offers a service curve β to a flow that is constrained by an arrival curve α and serviced in FIFO order, is given by the maximum horizontal deviation, as illustrated in Figure 2. This is for both cases $t_{mux} \leq g_{max}$ and $t_{mux} \geq g_{max}$ given by the distance between α and β at g_{max} , divided by the slope of β which is C .

Hence, the delay bound is

$$t_{switch} = \sum_{k=1}^N \frac{b_k}{C} - g_{max} * (1 - \sum_{k=1}^N \frac{r_k}{C}) + t_{mux} \quad (10)$$

By using equation (2) we can give a save estimation bound by

$$t_{est} = \sum_{k=1}^N \frac{b_k}{C} + t_{mux}. \quad (11)$$

This means, an estimation for the maximum delay of the switch is given by the time needed to transmit the bursts of the T-SPECs with the channel bandwidth plus the delay imposed by the electronics of the switch.

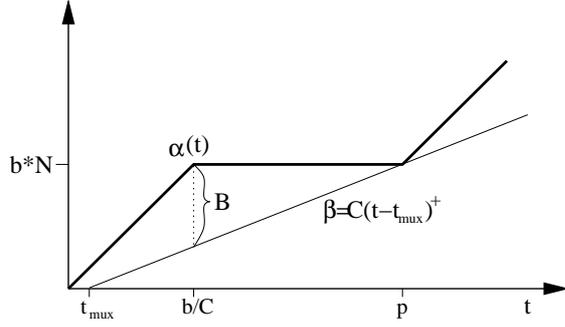


Figure 3: Arrival curve $\alpha(t)$ (thick line) as the sum of two on-off flows with burst length b and minimal burst period p . The maximum backlog is at time $t = b/C$.

Practical Considerations

In practice, g_{max} will hardly be less than t_{mux} . For example for Fast Ethernet $C=100\text{MBit/s}$ or 12207KByte/s , $M=1518\text{Byte}$. A flow typically contains bursts of multiple frames, we assume b_k to be at least $2 * M$. A lower bound g_{min} for g_k is then given by

$$\begin{aligned} g_{min} &\geq \frac{3 * M - M}{C} \\ &\geq 2 * M / C \\ &\geq 121\mu\text{s} \end{aligned}$$

Typical switch multiplexing delays t_{mux} are in the order of $45\mu\text{s}$ for Fast Ethernet, and we measured $25\mu\text{s}$ for a Gigabit Ethernet switch. We conclude that in general $t_{mux} \leq g_{max}$, and thus equation (6) can safely be used. The estimator given in equation (7) is valid in all situations.

3 On-Off Flows

On-off flows consist of bursts sent with pauses of a minimal length in between. We derive a buffer bound for N on-off flows with identical burst lengths b and identical minimal burst periods p . Obviously $b \leq p * C$ must hold.

Writing the time t as multiples of this period plus an offset within the period $t = n * p + o$, the arrival curve can be given as

$$\alpha_k(n * p + o) = \min\{C * (n * p + o), (n + 1) * b\} \quad (12)$$

Figure 3 shows the arrival curves of two on-off flows and the service curve of the switch. We assume that $t_{mux} \leq b/C$. Then it is easy to see that the maximum vertical distance between the arrival curves and the service curve is at b/C , this is $n=0$ and $o=b/c$. The value for the maximum backlog B is $\alpha(b/C) - \beta(b/C)$ or

$$B = \sum_{k=1}^N b - (b/C - t_{mux}) * C \quad (13)$$

$$= (N - 1) * b + t_{mux} * C \quad (14)$$

Practical Considerations

For a Fast Ethernet switch with $t_{mux}=45\mu\text{s}$, the second addend of equation (14) becomes 562Bytes.

For large values of b and such small values of t_{mux} an estimation of the buffer requirement of N on-off flows can be given as

$$B_{est} = (N - 1) * b \quad (15)$$

References

- [1] J.-Y. Le Boudec and P. Thiran. *Network Calculus*. Springer Verlag Lecture Notes in Computer Science volume 2050, July 2001.
- [2] Rene L. Cruz. A calculus for network delay, part i: Network elements in isolation. *IEEE Transactions on Information Theory*, 37(1):114–131, January 1991.