Sustainable load of fibre delay line buffers

Reference:
Laevens K., van Houdt Benny, Blondia Christian, Bruneel H..- *Sustainable load of fibre delay line buffers*
Electronic letters - 40:2(2004)
Handle: http://hdl.handle.net/10067/460060151162165141

On the Sustainable Load of Fiber Delay Line Buffers

K. Laevens[*], B. Van Houdt[+], C. Blondia[+] and H. Bruneel[*]

Affiliations:

[*] SMACS Research Group, Gent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium.

[+] PATS Research Group, University of Antwerp, Middelheimlaan 1, B-2020 Antwerpen, Belgium.

Corresponding author:

K. Laevens

email: koenraad.laevens@ugent.be

Abstract

We discuss the sustainable load of fiber delay line buffers, defined as the load at which a system with infinite buffering capacity becomes unstable. Due to the particularities of fiber delay line buffers, namely their finite delay granularity, this sustainable load is generally less than 100%. We further show that the packet-size distribution has some impact too.

1. Introduction

In packet switching, inevitably output port contention needs to be resolved. One option is to buffer one (or more) of the packets in conflict, until the moment they can be safely transmitted. In the electronic domain, the use of electronic random access memory (RAM) allows packets to be stored for an arbitrary period of time, which in turn allows efficient scheduling on the output channel, without waste of capacity. Currently, this is not the case in the optical domain, due to the lack of optical RAM. In the foreseeable future, fiber delay lines (FDLs) seem to be the most viable alternative. In essence, they are pieces of fiber of fixed length, capable of delaying an optical signal during a fixed time (order of μsec). Clearly, not all delays will be realizable, typically a "buffer" will be implemented by providing delays of $0 \cdot D$, $1 \cdot D$, ... up to $N \cdot D$, see e.g. [1]. Here, D is the so-called granularity of the FDL buffer, and $N \cdot D$ can be considered as its capacity.

The drawback of this approach is that it does not admit perfect scheduling. Sometimes channel capacity will be wasted due to the fact that a packet was delayed longer than theoretically necessary to avoid output port contention. The best one can do is to delay the packet during $n \cdot D$ time units, where $n \cdot D$ is the smallest realizable delay that is greater than or equal to that theoretical time. This impacts, amongst others, the load the system can sustain. For finite buffers ($N < \infty$), this would be the maximum load the system can carry without too much loss, for inifinite buffers ($N = \infty$) it would be the maximum load the system can carry without becoming unstable. The latter is the subject of this letter.

2. Sustainable load

A quantity of interest in the study of FDL-based buffers, is the scheduling horizon [1],[2]. It is the earliest time, pending new arrivals, at which the channel will become available again. If observed just prior to arrivals, the evolution of the scheduling horizon at these embedded points, denoted H_k (for $k \geq 0$), can be described by

$$H_{k+1} = \left[B_k + D \cdot \left\lceil \frac{H_k}{D} \right\rceil - \tau_k \right]^+ ,$$

where B_k is the size of the k-th packet (expressed in time units), and τ_k the interarrival time between packet k and packet k+1, see Figure 1. One easily recognizes the operation

$[X]^+ = \max\{0,X\}$, frequently encountered in queueing problems. The other operation, $\lceil H_k/D \rceil$ (the ceiling of H_k divided by D) reflects the finite granularity of the FDL buffer. An arriving packet, say packet k, will have to be delayed by at least H_k to avoid contention on the outgoing channel. As mentioned before, typically this delay will not be realizable, the closest match being $D \cdot \lceil H_k/D \rceil$.

In [2], we analyzed the equilibrium distribution of H (i.e. H_k for $k \to \infty$) by means of a generating functions approach (see e.g. [3] for a broad introduction to the technique), under the assumptions that the system is time-slotted, that packet sizes are iid (independent and identically distributed) random variables, and that interarrival times are iid random variables with a geometric distribution. In keeping with the time-slotted approach, in the remainder, we will express all quantities of interest as integer multiples of the slot length.

Geometric iid interarrival times amount to a Bernoulli arrival process, which can be characterized by p, the probability of having a new packet arrival during a given slot. A maximum arrival intensity p_{max} can then be defined, above which the system can not reach a steady state, i.e., above which the scheduling horizon grows without bound. This p_{max} was shown to be solution of

$$p_{max} = \left(E[B] + \frac{D-1}{2} + \sum_{k=1}^{D-1} \frac{1}{\varepsilon_k - 1} \frac{p_{max} B(\varepsilon_k)}{\varepsilon_k - \bar{p}_{max}} \right)^{-1} , \qquad (1)$$

where E[B] is the mean packet size, B(z) the generating function $E[z^B]$ of the packet-size distribution, and $\bar{p}_{max} = 1 - p_{max}$. The sum involves ε_k, the D different complex D-th order roots of unity (where, by definition, ε_0 equals 1). A first approximation, in which only E[B] appears, follows if one neglects the contribution of that sum:

$$p_{max} \approx \left(E[B] + \frac{D-1}{2} \right)^{-1} . \qquad (2)$$

This approximation also follows by an intuitive reasoning [1]: each packet contributes to the scheduling horizon by its own size, E[B], and by a so-called void, created due to its suboptimal scheduling in a finite-granularity buffer. Assuming this void is uniformly distributed on the set {0,...D-1} (recall we are studying the problem in a slotted-time setting), the average void is (D-1)/2. The term E[B]+(D-1)/2 can then be interpreted as an average equivalent packet size, in which void creation is accounted for. Approximation (2) then easily follows.

The exact relation (1) depends on the details of the packet-size distribution, here through

$$B(\varepsilon_k) = \sum_{n=0}^{\infty} \varepsilon_k^n \Pr[B = n] = \sum_{n=0}^{D-1} \varepsilon_k^n \left(\sum_{m=0}^{\infty} \Pr[B = m \cdot D + n] \right) = \sum_{n=0}^{D-1} \varepsilon_k^n \Pr[B \bmod D = n] .$$

Thus only the distribution of B mod D plays a role. Let us therefore introduce the quantities

$$\hat{b}(n) = \sum_{m=0}^{\infty} \Pr[B = mD + n] = \begin{cases} \Pr[B \bmod D = n] & n = 1, \ldots, D-1 \\ \\ \Pr[B \bmod D = 0] & n = D \end{cases} .$$

(Defining these $\hat{b}(n)$ for 1 ≤ n ≤ D, instead of for 0 ≤ n ≤ D-1, somewhat simplifies the results to be presented next.) One can then rewrite equation (1) in the following way:

$$E\left[\left\lceil \frac{B}{D} \right\rceil\right] = \frac{\sum_{n=1}^{D} \hat{b}(n) \overline{p}_{max}^{\,n-1}}{1 - \overline{p}_{max}^{\,D}} \quad . \qquad (3)$$

Once all parameters of interest are determined, numerical solution of this equation for p_{max} is straightforward.

3. Special cases

Some packet-size distributions allow for further simplification of equation (3). If, for instance, packet lengths are uniformly distributed on {1,..., K·D} for some K ≥ 1, one can show that approximation (2) becomes exact.

On the other hand, when packet sizes are deterministic of size B, we can always write B = a·D+c, with a ≥ 0 and 1 ≤ c ≤ D, and one easily finds

$$E\left[\left\lceil \frac{B}{D} \right\rceil\right] = a + 1 \quad ,$$

$$\hat{b}(n) = \delta_{c,n} \qquad 1 \le n \le D$$

and

$$a + 1 = \frac{\overline{p}_{max}^{c-1}}{1 - \overline{p}_{max}^{D}} \quad .$$

It is a simple exercise to establish that, for given a and D, the case c = 1 results in the highest p_{max} value. Reversing the argument, this shows that p_{max}, as a function of D, will exhibit "local maxima" whenever D ≈ (B-1)/a (a integer).

Finally, if packet sizes follow a geometric distribution, $Pr[B=n]=\alpha(1-\alpha)^{n-1}$ for $n \ge 1$ (such that $E[B]=1/\alpha$), equation (3) reduces to

$$1 = \frac{\alpha}{1 - \overline{p}_{max}^{D}} \cdot \frac{1 - (\overline{\alpha}\overline{p}_{max})^{D}}{1 - \overline{\alpha}\overline{p}_{max}}$$

where $\overline{\alpha} = 1 - \alpha$. For typical parameter values, i.e., $1 \ll D/2 < 1/\alpha$, approximation (2) becomes sufficiently accurate once more.

4. Numerical example

In Figure 2, we have plotted ρ_{max} = p_{max} E[B] as a function of D, for deterministic packet sizes of size 100, and also the approximation given in (2). The "local maxima" D ≈ (B-1)/a for the deterministic case clearly show (at D=99, 49, 33, ...). The curve (not shown in Figure 2) for geometrically distributed packet sizes, with mean E[B]=100, is indistinguishable from the approximation (see above).

Clearly, in terms of sustainable load, the granularity D should be chosen as small as possible, i.e. D=1. However, in reality, i.e. for finite FDL buffers, there is a tradeoff to be made [1],[2]. A small D implies many FDLs have to be used to arrive at a given capacity N·D, or, vice versa, when N is fixed, a small D will lead to a small capacity, resulting in large loss figures. On the other hand, a large D requires less FDLs or realize a larger capacity, resulting in lower loss, but seriously impairs performance due to the typically large size of the voids created due to suboptimal scheduling.

5. Conclusions

We have shown that, in general, infinite FDL buffers cannot sustain loads up to 100%, due to their finite granularity. This granularity creates voids in the channel scheduling, wasting transmission capacity. An implicit formula for the sustainable load was derived, in which the details of the packet-size distribution appear. An approximation, based on intuitive reasoning, was shown to be good in most cases, exact in some special ones.

The arrival process considered here - iid geometric interarrival times and iid packet sizes - served only as a good first model. Currently, we are investigating the impact of correlation between (consecutive) interarrival times or (consecutive) packet sizes, and also that of cross-correlation between packet sizes and interarrival times. We hope to report on these results in the near future.

References

[1] F. Callegati, "Optical Buffers for Variable Length Packets", IEEE Communications Letters, vol. 4, pp. 292-294, September 2000.

[2] K. Laevens and H. Bruneel, "Analysis of a single-wavelength optical buffer", Proceedings Infocom 2003 (San Francisco, April 2003).

[3] H. Bruneel and B.G. Kim, *Discrete-Time Models for Communication Systems Including ATM*, Kluwer Academic Publishers (Boston), 1993.

Figure captions:

Figure 1: Evolution of the scheduling horizon

Figure 2: Sustainable load as function of the granularity D (E[B]=100)

Figure 1: Evolution of the scheduling horizon

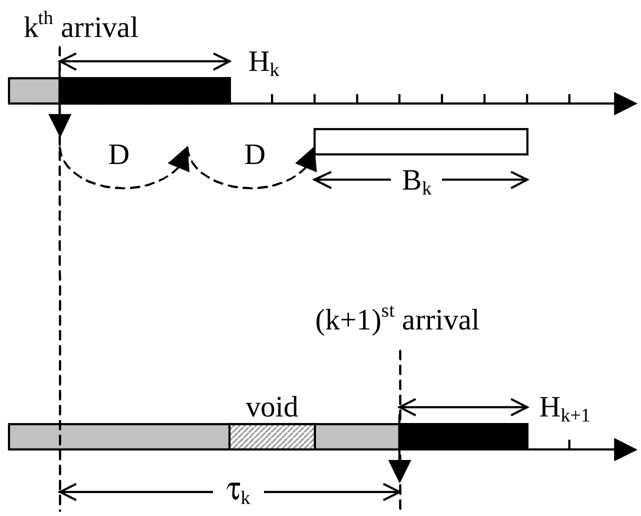

Figure 2: Sustainable load as function of the granularity D (E[B]=100)