On the ATM Adaptation Layer

Arne A. Nilsson

Żhi Cui

Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
North Carolina State University

TR-92/21
November 1992

On the ATM Adaptation Layer

Arne A. Nilsson and Zhi Cui

Center for Communications and Signal Processing
Dept. of Electrical and Computer Engineering
North Carolina State University
Raleigh, N.C. 27695-7914

Abstract

The impact of the Convergence Sublayer Protocol Data Unit (CS_PDU) size on the end-to-end network delay in a B-ISDN environment is investigated in this paper. Cell loss probabilities are assumed to be bursty and when a cell loss occurs a retransmission is assumed to happen at the CS_PDU level. The cell loss probability is assumed to be such that there are two different probabilities for cell loss. Typically the cell loss probability is very small, in a real ATM network 10^{-9} *or better. There are however instances when the cell loss probability will be significantly higher due to network congestion, such as buffer overflow, say a cell loss probability of* 10^{-3}. *An ATM network, for simplicity consisting of one single link, has been simulated and the simulation results indicate that it is possible to select a CS_PDU size that will minimize the end-to-end delay. The ATM network has also been analyzed by using mathematical modeling technique. Our first mathematical analysis approach assumed naively a Poisson arrival process. The resulting closed queueing network was solved approximately and an expression for the end-to-end delay as a function of the CS_PDU size was obtained. It was then shown that the delay is optimized for a particular CS_PDU size. For high speed networks, burstiness is a very important factor for system end-to-end delay and thus a Poisson arrival process is not a reasonable assumption, furthermore batch arrivals of CS_PDU should also be considered. It is found that queueing models are* $IBP^{[x]}/GEO/1/K$ *and* $IBP^{[x]}/D/1/K$ *systems. Numerical results for queue length distribut ions and system end-to-end delays for both queueing systems have been obtained and are used to obtain the optimal CS_PDU size.*

1 Introduction

The Asynchronous Transfer Mode (ATM) has been proposed as a fast packet switching and multiplexing technique for broadband ISDN. In an ATM network, all information ranging from narrowband voice and data traffic to broadband video traffic is transmitted using a fixed size "cell". The ATM Adaptation Layer (AAL) isolates the higher layers from the specific characteristics of the ATM layer by mapping the higher layer Protocol Data Units(PDUs) into the information field of the ATM cell and vice versa [1] [2]. In high speed networks, it is most likely that the traffic is highly bursty and packet sizes at the transport level will be fairly large (64KBytes to a few Megabytes). Since ATM cells only have a payload of 48 bytes or 44 bytes, this implies a segmentation and reassembly process involving potentially tens of thousands of cells per packet. For a transport layer protocol, without error correction capabilities which retransmit errored USER-PDUs, significant operational issues arise such as: how to deal with lost cells, how to reduce the system end-to-end delay, etc. [3].

In this paper, we propose an intermediate segmentation technique as a method to break up large transport layer packets into several sub-packets to reduce packet end-to-end delay. A critical design issue is to determine the optimal sub-packet size. In section 2, we describe the system and queueing model in detail and present simulation result for the optimal sub-packet size. For high speed networks, burstiness is a very important factor for system end-to-end delay and thus a Poisson arrival process is not a reasonable assumption, furthermore batch arrivals of CS_PDU should also be considered. It is found that the queueing models that we need to solve are $IBP^{[x]}/GEO/1/K$ and $IBP^{[x]}/D/1/K$ queues. In Section 3, numerical results for queue length distributions and system time for both queueing systems have been obtained. In section 4, the analytical results for system end-to-end delay and optimal CS_PDU size are obtained based upon $IBP^{[x]}/GEO/1/K$ and $IBP^{[x]}/D/1/K$ queues.

2 Model description

In this section, we describe in detail the CS_PDU and network architecture, cell-error probability and simulation result for the end-to-end delay as a function of CS_PDU size.

2.1 AAL and CS_PDU

The AAL consists of two sub-layers: Convergence Sub-layer(CS) and Segmentation And Re-assembly Sublayer(SAR) [4]. In CS, a big packet is broken up into several sub-packets so called CS-SDU:s. By adding headers and trailers to the CS-SDU, a CS_PDU is formed [7]. At the receiving side of the network, these CS_PDUs will be reassembled to packets. In the SAR-sublayer, we segment the CS_PDU into a size suitable for the information field of an ATM cell, and reassemble the contents of ATM cell information fields into CS information.

2.2 Queueing Network Architecture.

Based upon the AAL functions, the queueing network structure, as shown in Figure 1, can be derived. At the source, we use queues split1 and split2 to simulate the functions of the CS and SAR-sublayers, respectively. Cells are transmitted into the ATM-network which is assumed to have finite capacities and low cell loss probability. At the destination, the cells are reassembled to CS_PDU and packets. If an error happens either by cell loss or by bit error, the particular CS_PDU will be retransmitted.

2.3 Error Probability

In high speed networks, fiber optics is often the selected transmission medium. It is well known that in a fiber optic based network, the probability of bit error is very low, typically less than 10^{-9}. Therefore it is more likely that an ATM cell will be lost or destroyed due to buffer overflow or a lack of some network resource. We have adopted as a model for the probability of cell loss a simple two state model. In one of the states the cell loss probability is very small, typically negligible, and this happens when the amount of information in the network is smaller than the network system capacity. In the other state the cell loss probability is typically

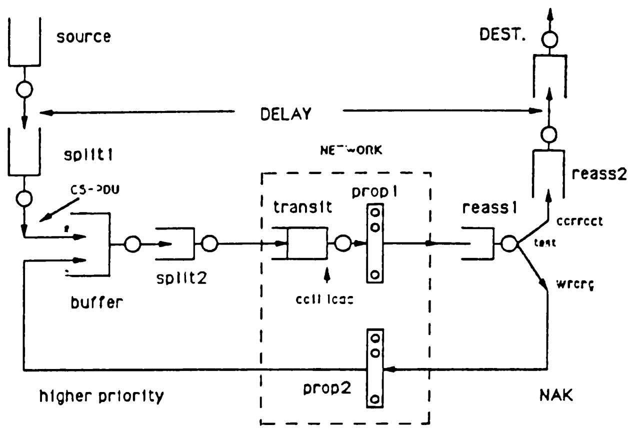

Figure 1: ATM Network Architecture.

high and this, of course, happens when the amount of information in the network is significant.

2.4 Simulation

Given a cell loss probability behavior as above, we simulate the queueing network [see Figure 1] to find the mean end-to-end delay and maximum end-to-end delay in terms of the CS_PDU size. The simulation result is presented in Figure 2, From the simulation result, we can conclude:

1) the mean end-to-end delay depends upon the error probability distribution and CS_PDU size;
2) an optimal CS_PDU size under a given cell loss distribution can be found;
3) CS_PDU size has much more influence on the maximum end-to-end delay than on the mean end-to-end delay.

3 $IBP^{[x]}/GEO/1/K$ and $IBP^{[x]}/D/1/K$ queues

3.1 Arrival process

Initially we analyzed the system by assuming the Poisson process as the packet arrival process to the CS layer. The analytical result for the mean end-to-end delay and optimal CS_PDU size has been obtained [7]. Since most of the traffic sources that an ATM network supports are bursty, a Poisson process may no

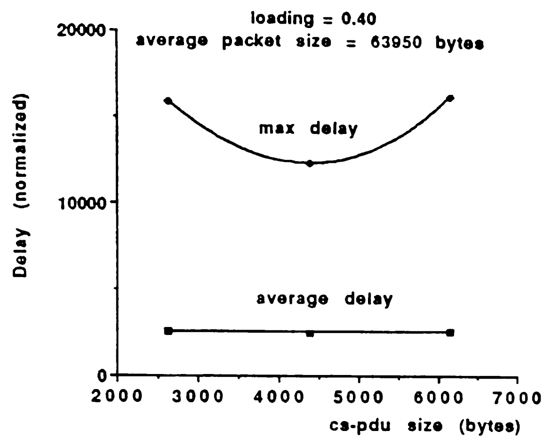

Figure 2: Maximum end-to-end delay and mean end-to-end delay

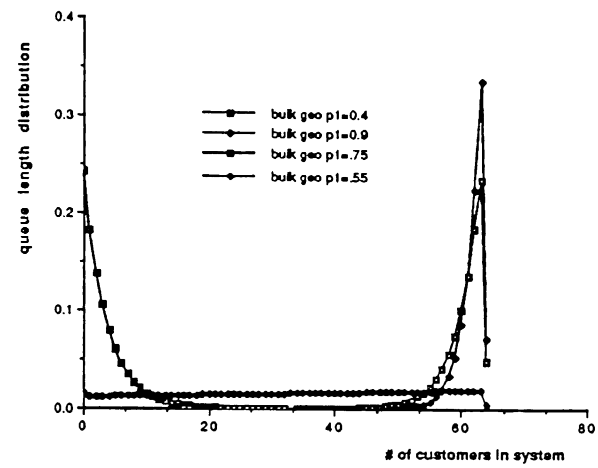

Figure 3: $IBP^{[x]}/GEO/1/K$ queue length distribution

longer be suitable for describing the network traffic. An Interrupted Bernoulli Process (IBP) captures the burstiness of the arrival process and may therefore be a good candidate for modeling the arrival process. IBP is a generalization of the Bernoulli process. It is governed by a discrete time Markov chain with two states, an active state and an idle state. If the process is in the idle state (active state), then in the next time slot it remains in the idle state (active state) with probability q (p), or it will change to the active state (idle state) with probability 1-q (1-p). In the idle state, no arrival occurs. In the active state, arrivals occur in a Bernoulli fashion with rate α. Furthermore, because the big packet is broken up into several CS_PDUs at CS layer, the arrival process is characterized by IBP with bulk arrival denoted by $IBP^{[x]}$. The bulk size is assumed to be geometrically distributed with failure probability p1.

3.2 Queue length distributions for the $IBP^{[x]}/GEO/1/K$ and $IBP^{[x]}/D/1/K$ queues

In order to get steady state queue length distribution of $IBP^{[x]}/GEO/1/K$ queue, we observe the system at the slot points and construct the Markov chain. In this Markov chain, there are 2(K+1) states denoted by (m,A) and (m,I) (m=0,1,...,K), where (m,A) represents an active state with m customers in the system; (m,I) represents an idle state with m customers in the system. The state changes can only be caused by:

1) a state change between active and idle; this can only happen at the beginning of the current slot, or,
2) customer arrival; this can only happen after the potential active-idle state change point, or,
3) customer departure, this can only happen at the slot point.

Let us denote by σ the service rate in the $IBP^{[x]}/GEO/1/K$ queue. By solving the Global Balance Equations of 2(K+1) state Markov chain, we obtain the analytical result for queue length distribution of $IBP^{[x]}/GEO/1/K$ queue. Figure 3 shows the queue length distribution of the $IBP^{[x]}/GEO/1/K$ queue for the given parameters p, q, α .

When the system service time is constant, i.e., σ =1, $IBP^{[x]}/GEO/1/K$ queue can be reduced to $IBP^{[x]}/D/1/K$. The result for the queue length distribution of $IBP^{[x]}/D/1/K$ queue be obtained similarly as for the $IBP^{[x]}/GEO/1/K$ queue.

4 Analytical results

Having developed a method to solve highly bursty traffic system, we can use the technique obtained above to evaluate the mean end-to-end delay of the ATM network analytically. We first simplify the original network [Fig. 1] to the queueing model shown in Figure 4.

The packet arrival rate and the average big packet

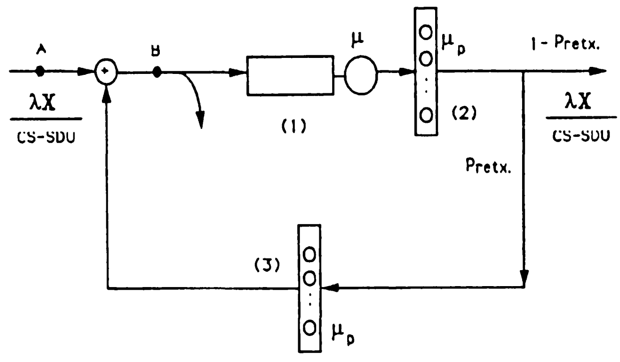

Figure 4: Analytical queueing network model

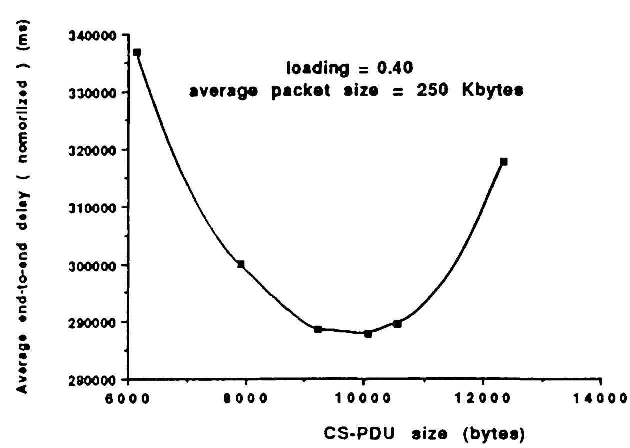

Figure 5: Analytical result for average end-to-end delay

size are λ and X, respectively. The arrival process to point A is assumed to be $IBP_1^{[x]}$ with parameters p,q,α; The bulk size is geometrically distributed with mean X/CS-SDU, where CS-SDU is used to denote the length of the service data unit. The arrival process, $IBP_2^{[x]}$ at point B, is the superposition of the feedback arrival process and the $IBP_1^{[x]}$. We use an $IBP_2^{[x]}/D/1/K$ to model queue 1 approximately , where the $IBP_2^{[x]}$ has the same bulk size distribution and the same parameters p,q as $IBP_1^{[x]}$ has, and with $\alpha_1 = \alpha/(1\text{-}P_{reTx})$.
Furthermore we have

$$\begin{aligned} P_{reTx} &= Prob(retransmitting\ a\ CS\_PDU) \\ &= P_{block} + P_{bit-error} \\ P_{block} &= Prob(CS\_PDUs\ are\ blocked\ at\ point\ B) \\ &= \sum_{j=0}^{K} B_j \left(1 - \frac{CS\_PDU}{X}\right)^{K-j} \end{aligned}$$

where CS_PDU in this equation indicates the length of the protocol data unit, and B_j is the probability that an arrival finds j CS_PDU:s in the system;

$$\begin{aligned} B_j &= \frac{(\pi_{j,1}p + \pi_{j,2}(1-q))(2-p-q)}{1-q}; \\ K &= the\ buffer\ size\ in\ CS\_PDU; \\ \pi_{j,1} &= Prob(j\ CS\_PDUs\ in\ system,\ at\ A\ state); \\ \pi_{j,2} &= Prob(j\ CS\_PDUs\ in\ system,\ at\ I\ state); \\ \frac{1}{\mu} &= transmission\ time\ for\ one\ CS\_PDU; \\ \frac{1}{\mu_p} &= propagation\ delay. \end{aligned}$$

It is apparent that P_{reTx} is a function of P_{block}, and vice versa. Given an initial value of P_{reTx}, P_{reTx} and P_{block} can be obtained iteratively, and will converge to certain values. Let D be the mean end-to-end delay for a message. For an open queueing network [Fig. 4], according to Little's result [6]

$$\begin{aligned} D &= \frac{average\ number\ of\ CS\_PDUs\ in\ network}{\lambda} \\ &= \frac{N1 + N2 + N3}{\lambda} \end{aligned}$$

where N_i is the average number of CS_PDUs in queue i. N_1 can be obtained numerically from Sec.3, by solving the queue length distribution of $IBP^{[x]}/D/1/K$ with parameters p,q,α_1. For the queues 2 and 3 the model used is the M/M/∞

$$\begin{aligned} N_2 &= \rho_2 \\ &= \frac{\lambda X}{(1 - P_{reTx}) * CS\_PDU * \mu_p}. \end{aligned}$$

$$\begin{aligned} N_3 &= \rho_3 \\ &= \frac{\lambda\, X P_{reTx}}{(1 - P_{reTx}) * CS\_PDU * \mu_p}. \end{aligned}$$

Fig. 5. shows the numerical result for the mean end-to-end delay as a function of CS_PDU size, from which the optimal CS_PDU size is obtained.

5 Conclusion

In this paper, we have presented results obtained from a simulation program that permits us to demonstrate the effect of how big packets should be broken up into smaller packets (sub-packets) before being fragmented into ATM cells. For high speed networks, the queueing models $IBP^{[x]}/GEO/1/K$ and $IBP^{[x]}/D/1/K$ were investigated, and numerical results for queue length distributions and system end-to-end delays for both queueing systems have been obtained. This technique is used to obtain the optimal CS_PDU size.

References

[1] S.E. Minzer, "Broadband ISDN and Asynchronous Transfer Mode (ATM)", *IEEE Communications Magazine*, vol.27, no.9, 1989.

[2] M. J. Rider, "Protocols for ATM Access Networks", *GLOBECOM*, vol.1, 1988.

[3] "VISTAnet: A Very High Bandwidth Prototype Network for Interactive 3D Medical Imaging ", *Research Plan by Bellsouth Services, GTE Corp., MCNC, UNC-Chapel Hill*, 1990.

[4] CCITT Recommendations Drafted by Working Party XVIII/8, no.6, 1988

[5] A. A. Nilsson F.-Y. Lai and H. G. Perros, "An approximate analysis of a bufferless N × N synchronous Clos ATM switch", *Proceedings, Canadian Conference on Electrical and Computer Engineering*, pp.39.1.1–39.1.4, 1990.

[6] L. Kleinrock, *Queueing System*, Vol. I, John Wiley & Sons, Inc, New York, NY, pp.134-136, 1975.

[7] A. A. Nilsson and Zhi Cui, "ATM Adaptation Layer Issues", *Proceedings, The Twenty-Fourth Southeastern Symposium on System Theory*, pp.434–437, 1992.