Model and Performance Evaluation for Multiservice Network Link Supporting ABR and CBR Services

Tae-Jin Lee and G. de Veciana, *Member, IEEE*

***Abstract*—We consider a "dynamic" model for a link jointly supporting two types of service: rate-adaptive and constant-bit rate services, e.g., ABR and CBR service in ATM networks. We construct and solve the associated two dimensional Markov chain via matrix-geometric techniques. This permits an analytical evaluation of performance when such services share resources. Moreover, when constant-bit rate connections are long-lived relative to rate-adaptive connections (e.g., file transfers), we prove a separation of time scale result. This leads to a useful approximation that closely matches the performance in this regime.**

***Index Terms*—ABR/CBR services, ATM networks, Markov chain, matrix-geometric equation, network model, QBD process, random environment.**

I. Introduction

RATE-ADAPTIVE service, e.g., available bit rate (ABR) service in ATM networks [2], is aimed at utilizing the remaining bandwidth after other service classes, e.g., variable bit rate (VBR) or constant bit rate (CBR) service [4], have been assigned their requested bandwidth. Networks supporting rate-adaptive service can provide more efficient network utilization from the network providers' point of view, and let users economically share available resources when they have relaxed quality of service (QoS) requirements. In multiservice networks encompassing rate-adaptive/constant-bit rate services, connections are setup and terminated arbitrarily, so the connection arrivals/departures are "random" in nature. To study this *dynamic* setting, previous research has focused on performance analysis using simulations [3], [9]. However, little attention has been paid to analytical study of performance for such environments.[1]

Our purpose in this letter is to model a single link (switch) supporting both rate-adaptive and constant-bit rate traffic and analyze its performance. In this letter, we consider ABR/CBR traffic as representatives of rate-adaptive/constant-bit rate service. The ABR connections we are considering are the ones for file transfers with random amount of data (bits) to transmit while there may be other applications using ABR service, for example, LAN-to-LAN connections. In a dynamic network environment, the number of connections and thus bandwidth occupied by CBR connections changes dynamically over time leading to varying available (remaining) bandwidth for ABR service. So it can be considered that ABR connections are operating under randomly varying bandwidth environment due to the dynamic CBR traffic.

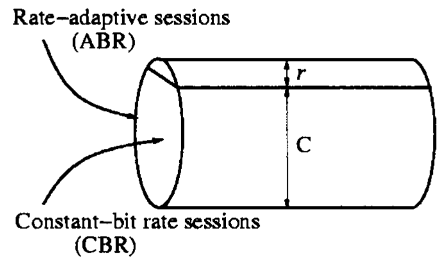

Fig. 1. A model of dynamic ABR and CBR connections to a link.

II. Model

Consider a single link with capacity C shared by CBR and ABR connections (see Fig. 1). CBR connections arrive at the link as a Poisson process with rate ν and have a connection holding time with exponential distribution and mean μ^{-1}. CBR connections are assumed to require unit bandwidth. So the number of CBR connections that the link can admit is no more than C.[2] If the link is full of CBR connections, further CBR connections are blocked until there is enough bandwidth to accommodate them.

In addition, ABR sessions enter the link as a Poisson process with rate λ, and the amount of work (bits) to be transmitted is exponentially distributed with parameter m. Thus the average volume of bits per ABR session will be $1/m$. We assume that ABR sessions share the available resources fairly in the max-min sense [1], and the available resources are the remaining bandwidth among total capacity once CBR connections are allocated their bandwidth. Since the available bandwidth is allocated fairly among the ABR connections, they may experience connection delays when the link is congested, i.e., ABR connections are not allocated enough bandwidth. We

Manuscript received February 28, 2000. The associate editor coordinating the review of this letter and approving it for publication was Prof. C. Douligeris. This work was supported by a National Science Foundation NSF Career Awards NCR 96-24230, by Southwestern Bell Company, and by an Intel Technology for Education 2000 equipment grant.

T.-J. Lee was with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA. He is now with the Corporate R&D Center, Samsung Electronics, Korea (e-mail: tjlee@samsung.com).

G. de Vecina is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail: gustavo@ece.utexas.edu).

Publisher Item Identifier S 1089-7798(00)11020-8.

[1]There has been previous research on links shared by two types of traffic, e.g., in particular in the ISDN context, see e.g., [5]. This work, however, takes a different view on the problem since in our context available bandwidth is shared via a processor sharing service discipline rather than a first come first served (FCFS) service discipline. In addition, connection-level performance is considered in ours.

[2]In case CBR requires bandwidth b instead, the number of possible CBR connections to the link will be C/b.

assume a bandwidth r has been reserved for ABR connections to ensure the stability of the system.

III. Analysis

We can formulate a two dimensional Markov chain for the model, in which a state (i, j) denotes a number i of CBR connections and j of ABR connections, see Fig. 2. Let $\pi(i, j)$ denote the stationary distribution for the numbers of CBR and ABR connections. The transition rate from (i, j) to $(i+1, j)$ is ν, the rate from $(i+1, j)$ to (i, j) is $(i+1)\mu$, the rate from (i, j) to $(i, j+1)$ is λ, and the rate from $(i, j+1)$ to (i, j) is η_i, where $\eta_i = ((C-i)+r)/(1/m)$ is the effective service rate of ABR sessions when i CBR connections are present. We let the reserved bandwidth $r > \rho$ and $\rho = \lambda/m$ to guarantee the stability of ABR connections. Let $\Pi_j = [\pi(0, j)\, \pi(1, j) \cdots \pi(C, j)]$ be a row vector of the stationary distribution when j ABR connections are present.

The model can be studied using a QBD process [7] and the following balance equation is formulated [8]:

$$\Pi \mathbb{Q} = \vec{0} \tag{1}$$

where $\Pi = [\Pi_0\, \Pi_1 \cdots]$ and

$$\mathbb{Q} = \begin{bmatrix} A-\Delta(\lambda) & \Delta(\lambda) & 0 & \ddots \\ \Delta(\eta) & A-\Delta(\lambda+\eta) & \Delta(\lambda) & \ddots \\ 0 & \Delta(\eta) & A-\Delta(\lambda+\eta) & \ddots \\ \ddots & 0 & \Delta(\eta) & \ddots \\ \ddots & \ddots & \ddots & \ddots \end{bmatrix},$$

and where

$$A = \begin{bmatrix} -\nu & \nu & 0 & \ddots & \ddots \\ \mu & -(\mu+\nu) & \nu & 0 & \ddots \\ 0 & 2\mu & -(2\mu+\nu) & \nu & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & 0 & C\mu & -(C\mu+\nu) & \nu \end{bmatrix}$$

$$\begin{aligned} \Delta(\lambda) &= \mathrm{diag}[\lambda\, \lambda \cdots \lambda], \\ \Delta(\eta) &= \mathrm{diag}[\eta_0\ \eta_1 \cdots \eta_C], \\ \Delta(\lambda+\eta) &= \mathrm{diag}[\lambda+\eta_0\ \lambda+\eta_1 \cdots \lambda+\eta_C]. \end{aligned}$$

Note that the diagonal structure of the matrix $\mathbb{Q}$ corresponds to a QBD process [8]. Thus (1) has a matrix-geometric solution

$$\Pi_k = \Pi_0 R^k = \vec{\pi}(I-R)R^k, \tag{2}$$

where $\vec{\pi}A = \vec{0}$ with $\vec{\pi} = [\pi(0)\, \pi(1) \cdots \pi(C)]$, which is a balance equation of M/M/C/C queue (CBR connections). Also R is the minimum nonnegative solution to the following equation

$$R^2\Delta(\eta) + R(A - \Delta(\lambda+\eta)) + \Delta(\lambda) = \vec{0} \tag{3}$$

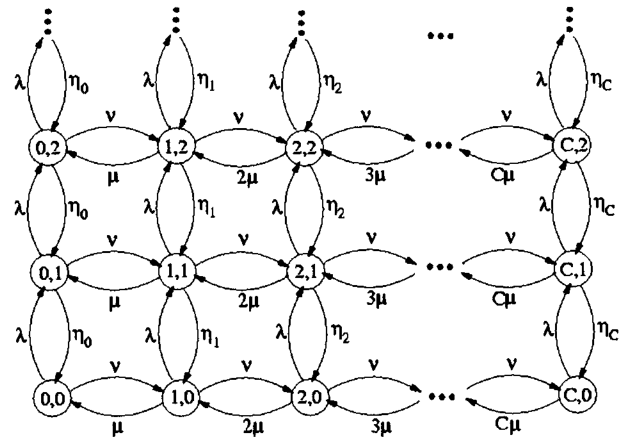

Fig. 2. A Markov chain for the model.

with boundary conditions for Π_0

$$\begin{aligned} \Pi_0(R\Delta(\eta) + (A - \Delta(\lambda))) &= \vec{0}, \\ \Pi_0(I-R)^{-1}\vec{e}^T &= 1 \end{aligned}$$

where $\vec{e} = [1\,1 \cdots 1]$.

In general it is difficult to find a closed form solution to (3). One can however solve for matrix R numerically. In this case the average number of ABR connections is given by

$$\mathbb{E}[N_{\mathrm{ABR}}] = \sum_{k=1}^{\infty} k\Pi_k \vec{e}^T = \vec{\pi}R(I-R)^{-1}\vec{e}^T, \tag{4}$$

and by Little's law, the average delay of ABR connections

$$\mathbb{E}[D_{\mathrm{ABR}}] = \frac{1}{\lambda}\,\mathbb{E}[N_{\mathrm{ABR}}]. \tag{5}$$

IV. Approximation

If the set-up and tear-down of ABR connections is fast relative to that of CBR connections, i.e., ABR is operating in faster time scale than CBR[3] then we expect the system might exhibit a simpler quasi-static behavior. This is formalized by the following theorem (see [6] for proof).

Theorem IV.1: Suppose $\lambda, m \to \infty$ with $\rho = \lambda/m$ fixed, then

$$\pi(i, j) = \pi^1(i)\pi^2(j|i)$$

where $\pi^2(j|i) = \rho_2(i)^j(1-\rho_2(i))$ with $\rho_2(i) = \lambda/\eta_i$, and $\pi^1(i) = G^{-1}\rho_1^i/i!$ with $G = \sum_{k=0}^{C}(\rho_1^k/k!)$ and $\rho_1 = \nu/\mu$. The theorem implies that as λ increases and $1/m$ decreases, the joint stationary distribution of ABR and CBR connections can be expressed as product of two simple distributions—$\pi^1(i)$ is a stationary distribution of M/M/C/C queue and $\pi^2(j|i)$ is a stationary distribution of M/M/1 queue with service rate η_i.

Letting $\Delta(\varrho) := \mathrm{diag}[\lambda\eta_0^{-1}\ \lambda\eta_1^{-1} \cdots \lambda\eta_C^{-1}]$, we can approximate the matrix R by $R \approx \Delta(\lambda)\Delta(\eta)^{-1} = \Delta(\varrho)$, which is

[3]For example, CBR connections for video transmission are long-lived while ABR connections for small data transmission are short-lived.

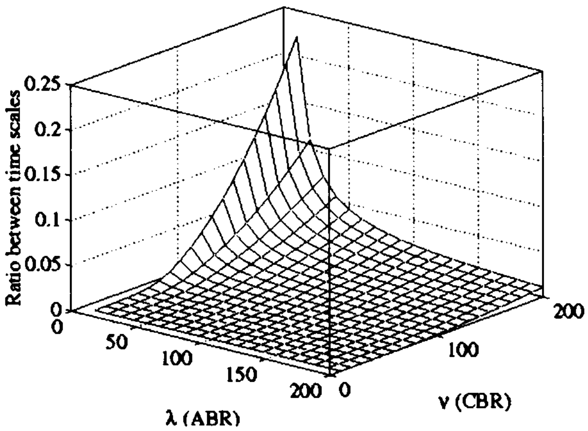

Fig. 3. Ratio between time scales when $r = 9$ kbps.

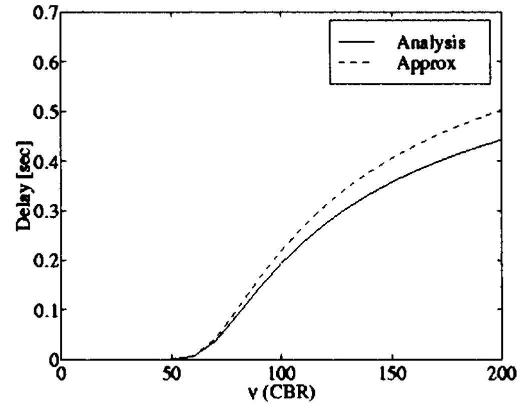

Fig. 4. Average delay as ν (CBR) changes for a given $\lambda = 10$ (ABR) with $r = 9$ kbps.

independent of CBR transition matrix A. Based on this approximation and by (4), the approximate average number of ABR connections will be

$$\mathbb{E}[N_{\mathrm{ABR}\,approx}] = \sum_{i=1}^{\infty} \frac{\rho_2(i)}{1-\rho_2(i)}\,\pi^1(i) = \sum_{i=1}^{\infty} \frac{\lambda}{\eta_i - \lambda}\,\pi^1(i).$$

Note that, in the approximation, from the ABR connections' point of view, the link corresponds to M/M/1 queue conditioned on that the number of CBR connections i being fixed. The approximate average delay experienced by ABR connections is

$$\mathbb{E}[D_{\mathrm{ABR}\,approx}] = \frac{1}{\lambda}\,\mathbb{E}[N_{\mathrm{ABR}\,approx}]. \tag{6}$$

V. Example

In this section, we present an example applying the performance results. We consider an OC3 link with capacity $C = 150$ Mbps and CBR sessions requiring $b = 2$ Mbps bandwidth corresponding to MPEG sources. The parameters $\nu = 10-200$, $\mu = 1$, $\lambda = 10 - 200$ conn./hour, $1/m = 2.88$ M–144 kbits/conn., and $r = 9$, 10 kbps are used.

The proposed approximation is based on the idea that ABR connections might come and go much faster than CBR connections. Thus conditioning on a given number of CBR connections, we might assume the distribution of ABR connections has reached "steady state." In order for this to be the case, the time scale on which the number of CBR connections in the system changes should be slow relative to that on which the ABR "queue" reaches steady state. To determine when this is the case, we compute a ratio between these time scales.

The ratio between the time scales is given by [6]

$$\mathrm{Ratio} = \frac{\mathrm{Time-scale}_{\mathrm{ABR}}}{\mathrm{Time-scale}_{\mathrm{CBR}}} = \frac{\nu + \kappa\mu}{\lambda + \sigma - 2\sqrt{\lambda\sigma}}.$$

where $k = E[N_{\mathrm{CBR}}]$ and $\alpha = ((c - kb) + \gamma)/(1/m)$.

Based on this ratio we can approximately see when our approximation might hold. For example, if the ratio is small enough, say 0.05, then we might conclude the time scales are separated. Fig. 3 illustrates how is ratio varies as CBR and ABR arrival rates change for the operating condition. Fig. 4 shows the average delay as ν changes. We note that the approximation yields *conservative* average delay estimate (5) compared with that of exact analysis (6), which can be used for the design, see [6]. Further result on blocking probability of constant-bit rate connections, the effect of r, design of link's capacity are also addressed in [6].

VI. Conclusion

We have considered a single link *dynamic* model supporting rate-adaptive service and constant-bit rate service. By formulating a two-dimensional Markov chain and solving a matrix-geometric equation, we have analytically determined the average connection delay for rate-adaptive connections. The joint distribution of ABR and CBR sessions has been shown to be appropriately captured by a quasi-static approximation when both services operate at different time scales. We believe the result provides a first step toward performance analysis and design of multiservice networks including adaptive services.

References

[1] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice Hall, 1992.

[2] F. Bonomi and K. W. Fendick, "The rate-based flow control framework for the available Bit Rate ATM service," *IEEE Network Mug.*, pp. 25–39, Mar./Apr. 1995.

[3] C. Fulton, S.-Q. Li, and C. S. Lim, "An ABR feedback control scheme with tracking," in *Proc. IEEE INFOCOM*, 1997, pp. 806–815.

[4] ATM Forum's Traffic Management Working Group, "ATM forum traffic management specification version 4.0.,", Tech. Rep., 1995.

[5] B. Kraimeche and M. Schwartz, "Analysis of traffic access control strategies in integrated service networks," *IEEE Trans. Commun.*, vol. 33, pp. 1085–1093, Oct. 1985.

[6] T.-J. Lee, "Traffic management and design of multiservice networks: The Internet and ATM networks," Ph.D. Dissertation, E.C.E. Dept., Univ. Texas, Austin, 1999.

[7] W. A. Massey and R. Srinivasan, "A packet delay analysis for cellular digital packet data," *IEEE J.Select. Areas Commun.*, vol. 15, no. 7, pp. 1364–1372, Sept. 1997.

[8] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*. Baltimore, MD: Johns Hopkins Univ. Press, 1981.

[9] H. Ohsaki, M. Murata, and H. Miyahara, "Robustness of rate-based congestion control algorithm for ABR service class in ATM networks," in *Proc. IEEE GLOBECOM*, 1996.