J. KSIAM Vol.12, No.1, 33–39, 2008

QUEUE RESPONSE APPROXIMATION WITH DISCRETE AUTOREGRESSIVE PROCESSES OF ORDER 1

YOORA KIM AND GANG UK HWANG†

DEPARTMENT OF MATHEMATICAL SCIENCES, KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (KAIST), KOREA
E-mail address: guhwang@kaist.edu

ABSTRACT. We consider a queueing system fed by a superposition of multiple discrete autoregressive processes of order 1, and propose an approximation method to estimate the overflow probability of the system. Numerical examples are provided to validate the proposed method.

1. INTRODUCTION

Video teleconference services are expected to be substantial portion of the traffic in the future broadband networks [1], and the queue response of video services is one of important performance metrics in the design of the future broadband networks supporting good quality of services. Recently, it has been shown that discrete autoregressive processes of order 1 (DAR(1) processes) provide good performance models for video teleconference traffic [1],[2],[3],[4]. For example, Heyman [1] studied the validity of the DAR process by investigating the system performance such as the packet loss rate through simulations.

While most studies on the queue response of video teleconference traffic are based on simulations, Hwang *et al.* [5] and Hwang and Sohraby [6] developed new analytic methods to analyze a queueing system fed by a single DAR(1) process. Elwalid *et al.* [4] considered a queueing system fed by a superposition of multiple DAR(1) processes and estimated the overflow probability of the queue length distribution based on the mathematical analysis. Their method is based on a single exponential form which is called the CDE (Chernoff-Dominant Eigenvalue) method.

In this paper, we consider a queueing system fed by a superposition of multiple video teleconference traffic, each of which is modelled by the DAR(1) process as in Elwalid *et al.* [4], and propose an approximation method to estimate the overflow probability of the queue length distribution based on our previous analytic results in Hwang and Sohraby [7]. While the CDE method in [4] is based on a single exponential form, our method is based on a linear sum of

2000 *Mathematics Subject Classification.* 60K25, 68M20.
Key words and phrases. Overflow probability, DAR(1), Video Teleconference Traffic.
This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. R01-2007-000-20053-0).

† Corresponding author.

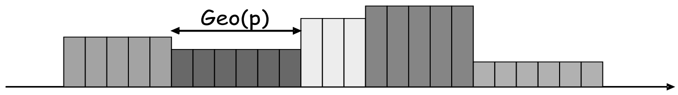

FIGURE 1. A discrete autoregressive process of order 1

two exponential forms. The reason why we use the linear sum of two exponential forms comes from the fact that two different time scale components - a burst scale component and a packet scale component - exist in the overflow probability of the queue length distribution for a general queueing model fed by bursty input traffic. Even though our method needs more parameters than Elwalid *et al.* [4], all the parameters are easily computed as shown later. Using our linear sum of two exponential forms we compute the overflow probability of the queue length distribution, and compare it with the simulation result. Our comparison study shows that our approximation method performs well.

The remainder of this paper is organized as follows. In section 2, we introduce the DAR(1) process and describe the system model considered in this paper. We also give some known results which will be used in our approximation method. In section 3, we propose a new approximation method to estimate the overflow probability of the queue length distribution. In section 4, we provide some numerical results to show the validity of our approximation method. In section 5, we give our conclusions.

2. MATHEMATICAL MODELLING AND KNOWN RESULTS

We first introduce a DAR(1) process. To define a DAR(1) process, we consider a sequence of independent and identically distributed (i.i.d.) random variables $\{B_n\}_{n\geq 1}$. We assume that B_n takes its values on $N = \{0, 1, 2, \cdots\}$. A DAR(1) process $\{A_n\}$ is then defined by the following equations:

$$\begin{aligned} A_1 &= B_1, \\ A_{n+1} &= (1-\alpha_n)A_n + \alpha_n B_n, \quad n \geq 1, \end{aligned}$$

where $\{\alpha_n\}_{n\geq 1}$ are i.i.d. Bernoulli random variables with $P\{\alpha_n = 1\} = p$ $(0 < p \leq 1)$, and independent of the sequence $\{B_n\}$. The stochastic properties of the DAR(1) process can be found in Hwang and Sohraby [7] and references therein.

Now we consider a discrete time queueing system and time axis is divided into slots of equal size. The queueing system is fed by a superposition of N identical video teleconference traffic where source i is modelled by a DAR(1) process with $\{A_n^{(i)}, B_n^{(i)}\}$. That is, the number of packets arriving at the system in slot n is given by $\sum_{i=1}^{N} A_n^{(i)}$. The service time of a packet is assumed to be one slot.

For later use, let ρ be the offered load of the system which is given by [7]

$$\rho = \sum_{i=1}^{N} E[B_n^{(i)}] = N \cdot E[B_n^{(1)}].$$

Let q_n denote the queue length of the system at time n. Then the evolution equation for $\{q_n\}$ is given by

$$q_{n+1} = (q_n - 1)^+ + \sum_{i=1}^{N} A_{n+1}^{(i)}.$$

Let q denote the queue length at an arbitrary time in the steady state. Recently, Hwang and Sohraby [7] computed the expectation $E[q]$ and the second moment $E[q^2]$ of the queue length distribution in the steady state as follows.

Theorem 2.1. [7] *In the steady state, the expectation $E[q]$ of the queue length distribution is given by*

$$E[q] = \frac{\rho}{2} + \frac{v}{2(1-\rho)} + \frac{(1-p)v}{p(1-\rho)} - \frac{(1-p)\rho}{p}.$$

Here, v denotes the variance of $\sum_{i=1}^{N} B^{(i)}$, and $B^{(i)}$ is a generic random variable for $\{B_n^{(i)}\}$.

Theorem 2.2. [7] *In the steady state, the second moment $E[q^2]$ is given by*

$$E[q^2] = E[\tilde{q}^2] - \frac{(1-p)\rho^2(N-1)}{Np^2}$$

where $E[\tilde{q}^2]$ is given by

$$\begin{aligned} E[\tilde{q}^2](1-\rho) = E[q] &\left\{1 - \frac{2\rho}{p} + \frac{2-p}{p}(v+\rho^2)\right\} \\ &+E[(\sum_{i=1}^{N} B^{(i)})^3]\left\{\frac{2(1-p)}{p^2} + \frac{1}{3}\right\} \\ &+\frac{(1-p)^2}{p^2}\left\{2\rho(1-\rho) - 2(v+\rho^2)\right\} \\ &+\frac{1-p}{p^2}\left\{-2\rho p(v+\rho^2) - (2+p)v + p\rho(1-\rho)\right\} \\ &+\rho^2 - \rho(v+\rho^2) - \frac{1}{3}\rho. \end{aligned}$$

Then, since we know the expectation $E[q]$ and the second moment $E[q^2]$, the variance $Var[q]$ of the queue length distribution in the steady state is easily computed by $Var[q] = E[q^2] - E[q]^2$. Therefore, from now on we assume that we know $E[q]$ and $Var[q]$ in the steady state.

3. Overflow Probability Approximation

First, we presume that the overflow probability of the queue length is well approximated by a linear sum of two exponential forms as follows:

$$P\{q > x\} \approx ae^{-\delta_1 x} + (\rho - a)e^{-\delta_2 x}, \tag{1}$$

where a, δ_1, and δ_2 are constants to be determined and $\delta_1 > \delta_2 > 0$.

If the approximation (1) is good enough to estimate the overflow probability of the queue length distribution, then the expectation and variance of the queue length distribution are well approximated as follows.

Lemma 3.1. *The expectation and variance of the queue length distribution are well approximated by*

$$E[q] \approx \frac{a}{\delta_1} + \frac{\rho - a}{\delta_2}, \tag{2}$$

$$Var[q] \approx \frac{2a}{\delta_1^2} + \frac{2(\rho - a)}{\delta_2^2} - (E[q])^2. \tag{3}$$

Proof. From equation (1), we see that the distribution of the queue length q can be derived from a mixture of two exponential random variables with respective parameters δ_1 and δ_2 as follows. Let Z be a random variable defined by

$$Z = \begin{cases} 0 & \text{with probability } 1 - \rho, \\ 1 & \text{with probability } a, \\ 2 & \text{with probability } \rho - a. \end{cases}$$

Then we get

$$\begin{aligned} E[q] &\approx E[E[q|Z]] \\ &= aE[X_1] + (\rho - a)E[X_2] \end{aligned}$$

where X_1 and X_2 are exponential random variables with parameters δ_1 and δ_2, respectively. Since $E[X_1] = \frac{1}{\delta_1}$ and $E[X_2] = \frac{1}{\delta_2}$ we get

$$E[q] \approx \frac{a}{\delta_1} + \frac{\rho - a}{\delta_2}.$$

Similarly, we get

$$\begin{aligned} E[q^2] &\approx E[E[q^2|Z]] \\ &= aE[X_1^2] + (\rho - a)E[X_2^2] \\ &= \frac{2a}{\delta_1^2} + \frac{2(\rho - a)}{\delta_2^2}. \end{aligned}$$

This completes the proof. □

Then, we can obtain the following theorem which shows the relations between the expectation and variance of the queue length distribution and the parameters a and δ_1.

Theorem 3.2.

$$\frac{1}{\delta_1} = \frac{Var[q] + (E[q])^2 - 2\rho/\delta_2^2}{2(E[q] - \rho/\delta_2)} - \frac{1}{\delta_2}. \tag{4}$$

$$a = \frac{E[q] - \rho/\delta_2}{\frac{1}{\delta_1} - \frac{1}{\delta_2}}. \tag{5}$$

Proof. From (2) and (3) we obtain

$$E[q] - \frac{\rho}{\delta_2} = a\left(\frac{1}{\delta_1} - \frac{1}{\delta_2}\right) \quad (6)$$

$$\begin{aligned} Var[q] + (E[q])^2 &= \frac{2a}{\delta_1^2} + \frac{2(\rho - a)}{\delta_2^2} = 2a\left(\frac{1}{\delta_1^2} - \frac{1}{\delta_2^2}\right) + \frac{2\rho}{\delta_2^2} \\ &= 2a\left(\frac{1}{\delta_1} + \frac{1}{\delta_2}\right)\cdot\left(\frac{1}{\delta_1} - \frac{1}{\delta_2}\right) + \frac{2\rho}{\delta_2^2}. \end{aligned} \quad (7)$$

Substituting (6) into (7) yields equation (4). In addition, from (6) we obtain equation (5). □

If we know the constant δ_2, then we can obtain two constants δ_1 and a from (4) and (5), respectively. Accordingly, it remains to compute δ_2. To do it, we first truncate the distribution of the random variable B_n by a constant number M as follows:

$$b_i = P\{B_n = i\} \text{ for } 0 \le i \le M - 1, \quad b_M = \sum_{i \ge M} P\{B_n = i\}. \quad (8)$$

Then it was shown that δ_2 is the largest solution of the following equation [4], [8]:

$$\sum_{i=0}^{M} \frac{pb_i}{e^{x(1/N - i)} - (1 - p)} = 1. \quad (9)$$

By solving the above equation (9) the constant δ_2 can be obtained numerically.

Remark: When we use equation (4) to compute the constant δ_1, we have $\delta_1 < \delta_2$ in some cases, which violates our assumption that $\delta_1 > \delta_2$. In these cases we propose to use either the following simple approximation

$$P\{q > x\} \approx \rho e^{-\delta_2 x}$$

or the CDE method proposed by Elwalid *et al.* [4] instead of equation (1). In fact, a number of numerical studies reveal that the condition $\delta_1 > \delta_2$ is valid when $Var[q]$ is *not* significantly large.

4. Numerical Examples

In this section, we give some numerical examples to check the validity of our method. In numerical examples we assume that the random variables $\{B_n\}$ are according to a negative binomial distribution because previous empirical studies showed that the negative binomial distribution is a good mathematical model for the random variables $\{B_n\}$ when the DAR(1) process is used to model a VBR-coded video teleconference traffic [1], [4]. We simulate our system and compare the resulting overflow probability with the proposed formula (1). In all figures, the results obtained from the approximation formula (1) are denoted by *Approximation* and those obtained from simulation are denoted by *Simulation*. For simplicity, we omit the confidence intervals in the resulting figures.

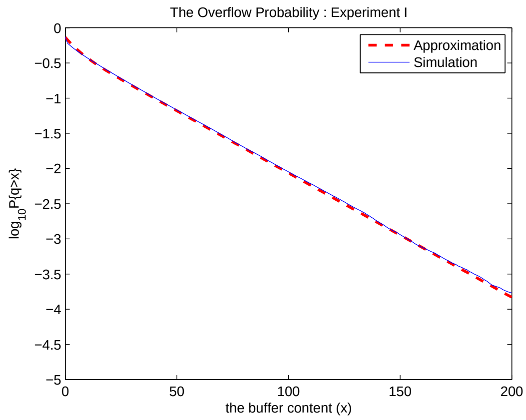

FIGURE 2. The overflow probability in experiment I

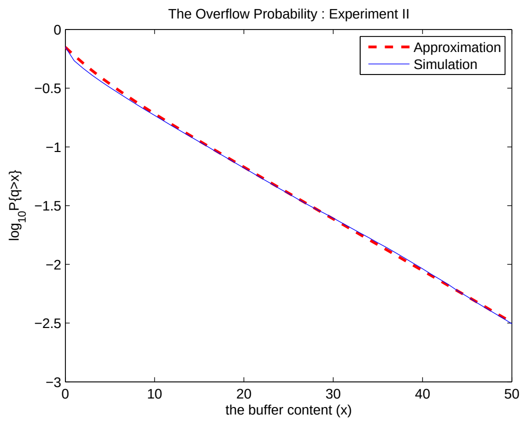

FIGURE 3. The overflow probability in experiment II

In the first example, called experiment I, we consider 12 homogeneous DAR(1) processes where $p = P\{\alpha_n = 1\} = 0.15$ and B_n is according to a negative binomial distribution with parameters $(3, 0.98)$, i.e.,

$$P\{B_n = k\} = \binom{k+3-1}{k} (0.98)^3 (0.02)^k.$$

The truncation value M in equation (8) is 5. The results from experiment I are plotted in Figure 2. As shown in Figure 2 the proposed formula (1) predicts the queue response well.

In the second experiment, called experiment II, we consider 35 homogeneous DAR(1) process where $p = P\{\alpha_n = 1\} = 0.3$ and B_n is according to a negative binomial distribution with parameters $(2, 0.99)$. The truncation value M in equation (8) is 4. The results from experiment II are plotted in Figure 3 which also shows that the proposed formula (1) predicts the queue response well.

We summarize parameters used in experiments I and II in Table 1. The resulting values of the constants such as a, ρ, δ_1, and δ_2 for the proposed formula (1) are summarized in Table 2.

TABLE 1. Parameters used in experiments I and II

	N	M	p	Parameters for B_n
Experiment I	12	5	0.15	$(3, 0.98)$
Experiment II	35	4	0.3	$(2, 0.99)$

TABLE 2. Resulting values of the constants for the proposed formula (1)

	a	ρ	δ_1	δ_2
Experiment I	0.23026	0.73469	0.18673	0.04069
Experiment II	0.19818	0.70707	0.33639	0.10132

5. CONCLUSIONS

In this paper, we propose an approximation method to estimate the queue length distribution of our system. Our method is to use a linear combination of two exponential forms based on some analytic results and the observation that the overflow probability has two time scale components. We also give some numerical examples to show the validity of our proposed method.

REFERENCES

[1] D.P. Heyman, and T.V. Lakshman, *Long-range dependence and queueing effects for VBR video*, in: K. Park and W. Willinger, (Ed.), Self-similar Network Traffic and Performance Evaluation, John Wiley & Sons, INC., 285–318, 2000.

[2] D.P. Heyman, A. Tabatabai, and T.V. Lakshman, *Statistical analysis and simulaton study of video teleconference traffic in ATM networks*, IEEE Trans. Circuits Syst. Video Technology, **2(1)** (1992), 49–59.

[3] D.P. Heyman, A. Tabatabai, T.V. Lakshman, and H. Heeke, *Modeling teleconference traffic from VBR video coders*, Proceedings of ICC '94, 1744–1748, 1994.

[4] A. Elwalid, D. Heyman, T.V. Laksman, D. Mitra, and A. Weiss, *Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing*, IEEE Journal of Selected Areas in Communications, **13** (1995), 1004–1016.

[5] G.U. Hwang, B.D. Choi and J. Kim, *The waiting time analysis of a discrete time queue with arrivals as a discrete autoregressive process of order 1*, Journal of Applied Probability, **39(3)** (2002), 619–629.

[6] G.U. Hwang and K. Shoraby, *On the exact analysis of a discrete-time queueing system with autoregressive inputs*, Queueing Systems, **43(1-2)** (2003), 29–41.

[7] G.U. Hwang and K. Shoraby, *On the queueing behavior of multiple first-order autoregressive sources*, Globecom 2004, Dallas, USA, December, 1187–1191, 2004.

[8] W. Ganesh, N. O'Connell and D. Wischik, *Big queues*, Springer, 2004.