*Applied Probability Trust* (30 July 2002)

# THE WAITING TIME ANALYSIS OF A DISCRETE TIME QUEUE WITH ARRIVALS AS A DISCRETE AUTOREGRESSIVE PROCESS OF ORDER 1

GANG UK HWANG,* *Korea Advanced Institute of Science and Technology*

BONG DAE CHOI,** *Korea University*

JAE-KYOON KIM,*** *Korea Advanced Institute of Science and Technology*

## Abstract

We consider a discrete time queueing system with the discrete autoregressive process of order 1 (shortly, the DAR(1)) as an input process and obtain the actual waiting time distribution and the virtual waiting time distribution. As shown in the analysis, our approach provides natural numerical algorithm to compute the waiting time distributions, based on the theory of the GI/G/1 queue, and consequently we can easily investigate the effect of the parameters of the DAR(1) on the waiting time distributions. We also derive a simple approximation of the asymptotic decay rate of the tail probabilities for the virtual waiting time in the heavy traffic case.

*Keywords:* Discrete Autoregressive Process of Order 1; Waiting Time; Heavy Traffic Analysis; Decay Rate; Queue

AMS 2000 Subject Classification: Primary 60K25

Secondary 68M20

THIS WORK WAS SUPPORTED BY GRANT NO. 98-0101-02-01-3 FROM THE BASIC RESEARCH PROGRAM OF THE KOREA SCIENCE & ENGINEERING FOUNDATION.

* Postal address: Division of Applied Mathematics, Korea Advanced Institute of Science and Technology (KAIST), 373-1 Kusong-Dong, Yusong-Gu, Taejon 305-701, South Korea

** Postal address: Department of Mathematics and Telecommunication Mathematics Research Center, Korea University, Anam-dong, Sungbuk-ku Seoul, 136-701, South Korea

*** Postal address: Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 373-1 Kusong-Dong, Yusong-Gu, Taejon 305-701, South Korea

## 1. Introduction

Various Markovian processes have been found for traffics arising in telecommunication networks, especially when the traffics exhibit high autocorrelations. Many authors modeled correlated arrival processes as MAPs (Markovian Arrival Processes) which generalize MMPP (Markov Modulated Poisson Process) in a continuous time framework and MMBP (Markov Modulated Bernoulli Process) in a discrete time framework. However, the drawback of these models is to estimate many parameters of MAP which should be extracted from the marginal distribution and the correlation structure of the measured data, and such estimation requires time consuming works. Consequently, in order to reduce the number of parameters, we use the 2-state MMPP/MMBP simply because they have only 4 parameters or less to estimate and their autocorrelations are exponentially/geometrically decaying, which is one of the salient features of the traffics in telecommunication networks such as ATM (Asynchronous Transfer Mode). Although the 2-state MMPP and MMBP have many applications, marginal distributions of both processes are fixed in the sense that, the arrival processes at each state of underlying Markov chain are Poisson and Bernoulli, respectively. In contrast, there are some works based on time series such as Gaussian-based Time Series (GTS) [5], discrete autoregressive process [6, 10] and MA (Moving Average) [7, 13]. Refer to [10] for details. Among those time series, the discrete autoregressive process of order 1 (shortly, the DAR(1)) is a Markov process of which autocorrelation function is geometrically decaying, and it can exhibit general distribution (Refer to section 2 for details). In addition, the DAR(1) is much simpler than the BMAP (Batch Markovian Arrival Process) which can exhibit general distribution. In spite of such merits of the DAR(1), relatively little attention has been paid to the DAR(1) in performance analysis of telecommunication networks because the analytical method of the queue with the DAR(1) is not easy to develop. Elwalid, Heyman, Laksman, Mitra and Wiss [6] analyzed the queueing system with the DAR(1) based on large deviation theory, but they assumed that the distribution of the size of an arriving batch has a finite support. Recently, Hwang and Sohraby [10] developed an exact analysis of a queue fed by the DAR(1) where the size of an arriving batch has a general distribution with infinite support. They obtained closed form expressions for the PGF (Probability Generating

Function) of the queue length and the mean queue length of the system in a discrete time framework.

Despite their success in the analysis of the queue with the DAR(1), the PGF of the queue length in [10] is of quite complex form, which is not convenient for numerical computation of queue length distribution. The aim of this paper is focused on how to compute and approximate the waiting time distribution for the DAR(1)/D/1 queue.

In this paper, we consider a discrete time queueing system with the DAR(1) and obtain the actual waiting time distribution and the virtual waiting time distribution. We assume the time is divided into slots of equal size and one slot is needed to serve a customer. In this paper we will use the term "*cell*"instead of the term "customer"because our model has applications in ATM networks. We consider *the early arrival model* [16] which means that the cells arrive just after the beginning of each slot, so that one of the cells newly arriving in the slot may be served if there is no cell waiting in the queue at the arrival epoch. We analyze the system by constructing an embedded Markov chain for the waiting time of the GI/G/1 queue, and obtain the PGFs of the actual waiting time distribution and the virtual waiting time distribution. As will be shown in the analysis, our approach provides natural numerical algorithm to compute both waiting time distributions, based on the theory of the GI/G/1 queue, and consequently we can easily investigate the effect of the parameters of the DAR(1) on the waiting time distributions. We also present a simple approximation of the asymptotic decay rate of the tail probabilities for the virtual waiting time in the heavy traffic case.

The organization of the paper is as follows: In section 2 we provide a mathematical description of the system. In section 3 we investigate the queue with the DAR(1) and derive the PGFs of the actual waiting time distribution and the virtual waiting time distribution. We also provide some numerical results on the waiting time distribution. In section 4 we present an approximation of the asymptotic decay rate for the virtual waiting time distribution in terms of $1 - \rho$ where $\rho$ denotes the offered load of the system.

## 2. Mathematical Description

In this paper, we consider a discrete time single server queueing system with the DAR(1). To introduce the DAR(1), we consider a sequence of independent and identically distributed (i.i.d.) random variables $\{B_n\}_{n\geq 0}$ with PGF $B(z) = E[z^{B_n}]$. We assume that $B_n$ takes its values on $\{0, 1, 2, \cdots\}$ and denote $b_k = P\{B_n = k\}, k \geq 0$. We further assume that the mean and the second moment of the random variable $B_0$ exist. We define the DAR(1) $\{A_n\}_{n\geq 0}$ by the following regression equation, for $n \geq 0$

$$\begin{aligned} A_0 &=^d B_0, \\ A_{n+1} &= (1-\alpha_n)A_n + \alpha_n B_n, \end{aligned} \tag{1}$$

where $\{\alpha_n\}_{n\geq 0}$ are i.i.d. Bernoulli random variables with $P\{\alpha_n = 1\} = p$ $(0 < p \leq 1)$, and independent of the sequence $\{B_n\}$. We assume that $A_0$ is independent of $B_n$ as well as $\alpha_n$ for $n \geq 0$.

Note that the DAR(1) $\{A_n\}_{n\geq 0}$ is a Markov chain and determined by the parameter $p$ and the probability mass function $\{b_k\}_{k\geq 0}$ of $B_n$. The state space of $\{A_n\}_{n\geq 0}$ is $\{0, 1, 2, \cdots\}$ and the one step transition probability matrix for $\{A_n\}_{n\geq 0}$ is given by

$$\mathcal{P} = (1-p)\begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} + p\begin{pmatrix} b_0 & b_1 & b_2 & \cdots \\ b_0 & b_1 & b_2 & \cdots \\ b_0 & b_1 & b_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

so that the smaller $p$ generates the stronger correlation of consecutive arrivals, and we have

$$A_{n+1} = \begin{cases} A_n & \text{with probability } 1-p, \\ B_n & \text{with probability } p. \end{cases}$$

We can show that $A_n$ and $B_n$ have the *same* distribution, i.e., $P\{A_n = k\} = b_k$. This statement implies that $A_n$ can have any given distribution. Let $A$ and $B$ be the generic random variables for $A_n$ and $B_n$, respectively.

In order to find the autocorrelation function of $\{A_n\}_{n\geq 0}$, observe that

$$\begin{aligned} E[A_{m+n}A_m] &= \int_{\alpha_{m+n-1}=1} B_{m+n-1}A_m dP + \int_{\alpha_{m+n-1}=0} A_{m+n-1}A_m dP \\ &= p\ E[B]E[A] + (1-p)E[A_{m+n-1}A_m]. \end{aligned} \tag{2}$$

Now we apply above relation (2) inductively

$$\begin{aligned}
E[A_{m+n}A_m] &= p(E[A])^2 + (1-p)\left\{p(E[A])^2 + (1-p)E[A_{m+n-2}A_m]\right\} \\
&= p(E[A])^2 + p(1-p)(E[A])^2 + (1-p)^2 E[A_{m+n-2}A_m] \\
&= \cdots \\
&= p(E[A])^2 + p(1-p)(E[A])^2 + p(1-p)^2(E[A])^2 \\
&\quad + \cdots + p(1-p)^{n-1}(E[A])^2 + (1-p)^n E[A^2] \\
&= p(E[A])^2 \left\{\frac{1-(1-p)^n}{1-(1-p)}\right\} + (1-p)^n E[A^2] \\
&= (1-p)^n(E[A^2] - (E[A])^2) + (E[A])^2 \qquad (3)
\end{aligned}$$

Hence the autocorrelation $\gamma_n$ of the sequence $\{A_n\}$ can be obtained from (3) as follows:

$$\begin{aligned}
\gamma_n &=^{def} \frac{E[(A_0 - E[A_0])(A_n - E[A_n])]}{E[(A - E[A])(A - E[A])]} \\
&= (1-p)^n, \qquad n \geq 0. \qquad (4)
\end{aligned}$$

Equation (3) shows that the DAR(1) is a short range dependent process and the smaller $p$ gives the stronger correlation of $\{A_n\}$. When $p = 1$, $A_n$ is the same as $B_n$ and so the DAR(1) generates i.i.d. arrivals.

Since we consider a discrete time system, we assume that time is divided into slots of equal size. We also assume that a batch of size $A_n$ arrives during $(n-1, n)$ (which is called the $n$th slot). Let $q_n$ be the queue length at the beginning of the $n$th slot. Then $q_n$ satisfies

$$q_{n+1} = (q_n - 1 + A_n)^+,$$

where $(X)^+ = \max(0, X)$. Even though the above evolution equation for $\{q_n\}$ seems to be the same as the Lindley equation of the GI/G/1 queue, the standard GI/G/1 queueing theory cannot be applied directly to $\{q_n\}$ because there is correlation in the sequence $\{A_n\}$ when $0 < p < 1$. Hence, it seems to be inevitable to consider a sequence of supplementary variables, say, $\{s_n\}$ to capture the correlation in the sequence $\{A_n\}$ and construct a two-dimensional Markov chain $\{(q_n, s_n)\}$ in the analysis. For example, one candidate for $\{s_n\}$ is the sequence $\{A_{n-1}\}$, and in this case we have $\{(q_n, A_{n-1})\}$

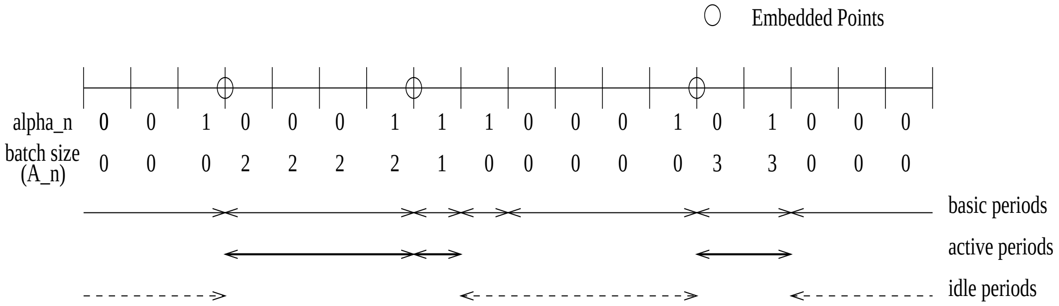

Figure 1. Basic Periods and Embedded Points

as the two-dimensional Markov chain. Another candidate for $\{s_n\}$ is that $s_n$ denotes the discrete time since the sequence $\{\alpha_k\}$ last changed its value. However, it seems to be hard or at least complex to obtain the steady state joint PGF of $(q_n, s_n)$ or the steady state PGF of $q_n$ in general. In this paper, by choosing suitable time epochs as our embedded points and constructing a *one-dimentional* embedded Markov chain we show that the analysis to obtain the actual and virtual waiting time distributions resembles the analysis of the standard GI/G/1 queue and consequently, the analysis would be relatively simple and efficient.

Let $\rho$ be the offered load of the system, defined by $\rho = E[A]$. Since $A$ and $B$ have the same distribution, we also have $\rho = E[B]$. We assume that $\rho$ is less than 1 for stability.

## 3. Waiting Time Distributions

In this section we derive the PGFs of the actual waiting time distribution and the virtual waiting time distribution. To do this, it is quite helpful to introduce *a basic period*. A basic period is defined by the time period which begins with the slot, say $n$, satisfying $\alpha_n = 0$ and $\alpha_{n-1} = 1$, and ends with the slot, say $k(> n)$, satisfying $\alpha_k = 1$. If we have consecutive slots, say $n$ and $n+1$, satisfying $\alpha_n = 1$ and $\alpha_{n+1} = 1$, then the basic period is defined by the last slot of the two consecutive slots. Refer to figure 1.

Based on the concept of the basic period, we define an active period by a basic period during which there arrives a batch of size greater than 0, and define an idle period by the time period between two consecutive active periods. Note that we might have an active period immediately followed by the next active period. In this case, we assume that an idle period between two consecutive active periods is of length 0. Introduce two random variables $X_n$ and $Y_n$ which denote the lengths of the $n$-th active period and the $n$-th idle period, respectively. Then, from the definitions of two periods, we know

$$\begin{aligned} P\{X_n = l\} &= (1-p)^{l-1}p, \qquad l \geq 1, \\ P\{Y_n = 0\} &= 1 - b_0, \\ P\{Y_n = m\} &= b_0[1-p(1-b_0)]^{m-1}p(1-b_0), m \geq 1. \end{aligned}$$

For the analysis we consider the beginnings of active periods as our embedded points, and refer to the beginning epoch of the $n$-th active priod as the $n$-th embedded point. We assume that a batch of size greater than 0 arrives to an empty queue at slot 1. Since the batches arriving in a sigle active period are of the same size, we let $\bar{B}_n$ to denote the batch size during the $n$-th active period. It is known that

$$P\{\bar{B}_n = k\} = \frac{b_k}{1-b_0}, \qquad k \geq 1.$$

Note that $\{\bar{B}_n\}$ are i.i.d.. In the analysis we exclude the trivial case where $b_0 + b_1 = 1$, so that we have $b_k > 0$ for some $k \geq 2$.

Let $\bar{q}_n$ be the queue length at the $n$-th embedded point. Then $\bar{q}_n$ satisfies the following evolution equation:

$$\bar{q}_{n+1} = (\bar{q}_n + X_n \times (\bar{B}_n - 1) - Y_n)^+, \tag{5}$$

which is equivalent to the evolution equation for the waiting time of the GI/G/1 queue (called the induced GI/G/1 queue in this paper) where the interarrival times are according to $\{Y_n\}$ and the service times are $\{X_n \times (\bar{B}_n - 1)\}$. Let $\bar{\rho}$ be the offered load of the induced GI/G/1 queue. Then, from the fact that $\rho = E[B] > P\{B \geq 1\}$ it is easy to show that

$$\begin{aligned} \bar{\rho} &= \frac{E[X](E[\bar{B}]-1)}{E[Y]} \\ &= \frac{\rho - (1-b_0)}{b_0}, \end{aligned}$$

and $0 < \bar{\rho} < 1$. Here, $X$ and $Y$ denote the generic random variables for $\{X_n\}$ and $\{Y_n\}$, respectively. So, the induced GI/G/1 queue is stable and hence the limiting steady state distribution of $\{\bar{q}_n\}$ exists and its PGF $\bar{q}(z)$ satisfies [4, 14]

$$\bar{q}(z) = \frac{1 - R^+(1)}{1 - R^+(z)}, \tag{6}$$

where $R^+(z) = \sum_{i=1}^{\infty} r_i^+ z^i$ is the generating function of a strictly ascending ladder height random variable of the random walk generated by $\{X_n \times (\bar{B}_n - 1) - Y_n\}$.[1]

Now we are ready to derive the PGF of the actual waiting time distribution. For doing this, let's tag an arbitrary cell in the arrivals in the steady state. Let $\bar{B}_e$ be the size of the batch to which our tagged cell belongs. From Renewal theory we know the distribution of $\bar{B}_e$ is given by

$$\begin{aligned} P\{\bar{B}_e = n\} &= \frac{nb_n}{E[\bar{B}](1 - b_0)} \\ &= \frac{nb_n}{E[A]}, \qquad n \geq 1, \end{aligned}$$

where $\bar{B}$ denotes the generic random variable for $\{\bar{B}_n\}$. Consider the active period in which our tagged cell arrives. Let $T$ denote the elapsed time in the active period until the slot in the active period where our tagged cell arrives. Using Renewal theory again, we know the distribution of $T$ is given by

$$\begin{aligned} P\{T = m\} &= \frac{P\{X > m\}}{E[X]} \\ &= p(1-p)^m, \qquad m \geq 0. \end{aligned}$$

Observe the following facts:

1. The lengths between two consecutive embedded points are i.i.d.. Hence, by Markov renewal theory, the number of cells in the queue, seen by the first batch arriving in the active period, has the same distribution as the random variable $\bar{q}$.

2. The increase in the number of cells in the queue during the elapsed time $T$ is $(n-1)m$, given that the elapsed time $T$ is $m$ and the batch size arriving at the slot in the active period is $n$. We denote this number of cells by $K_1$.

[1] For a random walk $\{S_n\}$ generated by $\{Z_n\}$, i.e., $S_0 = 0$ and $S_n = Z_1 + \cdots + Z_n$, the strictly ascending ladder epoch $\tau_+ = \inf\{n \geq 1 : S_n > 0\}$ and the strictly ascending ladder height (defined only on $\{\tau_+ < \infty\}$ only) is defined by $S_{\tau_+}$.

3. The number of cells prior to our tagged cell in the batch to which our tagged cell belongs is according to a uniform distribution depending on the size of the batch. We denote the number of cells prior to our tagged cell in the batch by $K_2$.

From the above observations and the fact that a single slot is needed to serve a cell, it is easy to show that the actual waiting time $W_a$ of the system satisfies

$$W_a \quad =^d \quad \bar{q} + K_1 + K_2. \tag{7}$$

Consequently, the PGF $W_a(z)$ of the actual waiting time distribution is derived from (6) and (7) as given in the following theorem.

**Theorem 3.1.**

$$\begin{aligned} W_a(z) &= \bar{q}(z) \sum_{m=0}^{\infty} p(1-p)^m \sum_{n=1}^{\infty} \frac{nb_n}{E[A]} \sum_{l=0}^{n-1} \frac{1}{n} z^{(n-1)m+l} \\ &= \frac{1-R^+(1)}{1-R^+(z)} \sum_{m=0}^{\infty} p(1-p)^m \sum_{n=1}^{\infty} \frac{b_n}{E[A]} \sum_{l=0}^{n-1} z^{(n-1)m+l}. \end{aligned} \tag{8}$$

Next, we derive the PGF of the virtual waiting time distribution. For doing this, consider an arbitrary slot and refer to it as our tagged slot. Introduce a random variable $I_a$ which is defined by 1 if our tagged slot is in an active period and by 0 otherwise. Then, we know

$$\begin{aligned} P\{I_a = 1\} &= \frac{E[X]}{E[X+Y]} = 1 - b_0, \\ P\{I_a = 0\} &= \frac{E[Y]}{E[X+Y]} = b_0, \end{aligned} \tag{9}$$

Consider the case where our tagged slot is in an active period. In this case, let $T_a$ denote the elapsed time (including our tagged slot) in the active period until our tagged slot. Then, we know

$$P\{T_a = m\} = \frac{P\{X \geq m\}}{E[X]} = p(1-p)^{m-1}, m \geq 1. \tag{10}$$

Next, consider the case where our tagged slot is in an idle period. In this case, let $T_i$ denote the elapsed time (including our tagged slot) in the idle period until our tagged slot. Then we know

$$P\{T_i = n\} = \frac{P\{Y \geq n\}}{E[Y]} = p(1-b_0)[1-p(1-b_0)]^{n-1}, n \geq 1. \tag{11}$$

From the above notations, it is easy to show that the virtual waiting time $W$ at the end of our tagged slot, satisfies

$$W =^d I_a \times [\bar{q} + T_a \times (\bar{B} - 1)] + (1 - I_a) \times [\bar{q}^* + X \times (\bar{B}^* - 1) - T_i]^+, \tag{12}$$

where the random variables with superscript $*$ have the same distributions as the random variables without superscript $*$, respectively. Note that all the random variables involved are mutually independent. Hence, from (9), (10), (11) and (12) we obatin the following theorem.

**Theorem 3.2.** *Let $W(z)$ be the PGF of the virtual waiting time distribution. Then we have $W(z) = \bar{q}(z)$, that is, the distribution of the virtual waiting time $W$ is the same as that of the queue length $\bar{q}$ at an embedded point.*

*Proof.* Note that the random variable $T_a$ has the same distribution as $X$, which is a well-known property of a geometric random variable. In addition, note that the distribution of $T_i$ is the same as the conditional distribution of $Y$, given that $Y \geq 1$. We use $Y_c$ to denote the random variable having the conditional distribution of $Y$, given $Y \geq 1$. So, from (12) we have

$$\begin{aligned}
W &=^d I_a \times [\bar{q} + T_a \times (\bar{B} - 1)] + (1 - I_a) \times [\bar{q}^* + X \times (\bar{B}^* - 1) - T_i]^+ \\
&=^d I\{Y = 0\} \times [\bar{q} + X \times (\bar{B} - 1)] + I\{Y \geq 1\} \times [\bar{q}^* + X^* \times (\bar{B}^* - 1) - Y_c]^+ \\
&=^d [\bar{q} + X \times (\bar{B} - 1) - Y]^+ \\
&=^d \bar{q},
\end{aligned}$$

where $I\{\cdot\}$ denotes the indicate function. Therefore, the proof is completed.

In the rest of this section, we mention the numerical algorithm that our approach naturally provides. To obtain the waiting time distributions of our system, as seen in (8) and Theorem 3.2 it is quite inevitable to compute the waiting time distribution of the GI/G/1 queue induced from our embedded method. A number of authors proposed the numerical algorithms to compute the waiting time distribution of the GI/G/1 queue. Ackroyd [3], Fryer and Winsten [8], Konheim [12] and Grassmann and Jain [9] considered a discrete time GI/G/1 queue. The first three works used root finding algorithms to compute the waiting time distribution, while, the work in

[9] used iterative schemes to compute the waiting time distribution. Here we use the iterative scheme proposed in [9] to obtain our numerical results simply because the task of finding all roots of a complex polynomial is very difficult, in general. Once we compute the waiting time distribution of the induced GI/G/1 queue, it is quite easy to obtain the actual/virtual waiting time distribution of our system from Theorem 3.1 and Theorem 3.2.

Now, we give a numerical example to see the effect of the change in the actual waiting time distribution as $p$ varies. In our numerical example, we use $B(z) = (qz + 1 - q)^N$, i.e., the marginal distribution is according to a Binomial distribution. We put $q = 0.15$ and $N = 5$, which means the offered load $\rho = 0.75$. Figure 2 displays the change of the actual waiting time distribution as $p$ varies from 0.1 to 1.0. Figure 2 shows that the value of $p$ significantly affects the tail probabilities and the effect due to the change in the value of $p$ is getting more significantly when the value of $p$ is getting smaller.

## 4. Decay Rate Approximation

In most discrete time queueing systems with short range dependent process as arrival process, it is quite remarkable to approximate the tail probability of the waiting time distribution by a simply geometric form

$$\lim_{k\to\infty} z_0^k P\{W \geq k\} = \beta.$$

Here, $\beta$ and $1/z_0$ are called the asymptotic decay constant and the asymptotic decay rate, respectively. Such exponential asymptotic result for the GI/G/1 queue was first established by Smith [15] and has been recently extended by many authors. Refer to [1, 2] and the references therein.

Heavy traffic expansion was also first proposed by Smith [15]. For further extension, see [1]. In this section we first derive the upper and lower bounds of asymptotic tail probability of the virtual waiting time distribution and next we approximate the parameter $1/z_0$ in terms of $1 - \rho$ when the offered load $\rho$ is near to 1.

Let's begin with the equation (5) for the waiting time of the induced GI/G/1 queue. From the theory of the GI/G/1 queue we know there are three constants $c_u$, $c_l$ and $\eta$

which satisfy the inequalities: (for example, see [11])

$$c_l e^{-\eta k} \le P\{\bar{q} \ge k\} \le c_u e^{-\eta k}, \quad k \ge 0, \tag{13}$$

where $\eta(>0)$ satisfies

$$E[e^{\eta(X\times(\bar{B}-1)-Y)}] = 1. \tag{14}$$

For the existence of such $\eta$ we further assume that $X \times (\bar{B}-1)$ has a finite moment generating function around the origin in this section.

From Theorem 3.2 and (13) we know the asymptotic decay rate $1/z_0$ of tail probabilities for the virtual waiting time is $e^{-\eta}$. Consequently, our next step is to approximate $\eta$ in terms of $1-\rho$ when $\rho$ is near 1.

Observe that

$$\begin{aligned} E[e^{\eta(X\times(\bar{B}-1))}] &= 1 + \eta E[X]\left(-1 + \frac{\rho}{1-b_0}\right) + \eta^2 E[X^2]\left(\frac{1}{2} - \frac{\rho}{1-b_0} + \frac{E[\bar{B}^2]}{2}\right) \\ &\quad + \eta^3 E[X^3]\left(\frac{-1}{6} + \frac{\rho}{2(1-b_0)} - \frac{E[\bar{B}^2]}{2} + \frac{E[\bar{B}^3]}{6}\right) + o(\eta^3), \end{aligned} \tag{15}$$

and that

$$E[e^{-\eta Y}] = 1 - \eta E[Y] + \frac{\eta^2 E[Y^2]}{2} - \frac{\eta^3 E[Y^3]}{6} + o(\eta^3). \tag{16}$$

Further, from the definition of the random variables $X$ and $Y$ we know

$$\begin{aligned} E[X] &= \frac{1}{p}, \qquad E[Y] = \frac{b_0}{p(1-b_0)}, \\ E[X^2] &= \frac{2-p}{p^2}, \qquad E[Y^2] = b_0\frac{2-p(1-b_0)}{p^2(1-b_0)^2}, \\ E[X^3] &= \frac{6-6p+p^2}{p^3}, \\ E[Y^3] &= b_0\frac{6-6p(1-b_0)+p^2(1-b_0)^2}{p^3(1-b_0)^3}. \end{aligned}$$

Substituting (15), (16) and the above equations into (14) we have, after some manipulations

$$1-\rho = \eta\{h_1(1-\rho)+h_2\} + \eta^2\{h_3(1-\rho)+h_4\} + o(\eta^2), \tag{17}$$

where

$$\begin{aligned}
h_1 &= \frac{-2+b_0+p-pb_0}{(-1+b_0)p} \\
h_2 &= \frac{-2+p+(2-p)(1-b_0)E[\bar{B}^2]}{2p} \\
h_3 &= -\frac{6-6p+p^2+b_0(-6+9p-2p^2)+b_0^2(2-3p+p^2)}{2(1-b_0)^2p^2} \\
h_4 &= \frac{(-2+3E[\bar{B}^2]-E[\bar{B}^3])(6-6p+p)}{6(-1+b_0)p^2} \\
&\quad +\frac{b_0[6-9p+2p^2+E[\bar{B}^2](-30+33p-6p^2)+2E[\bar{B}^3](6-6p+p^2)]}{6(-1+b_0)p^2} \\
&\quad +\frac{b_0^2[-E[\bar{B}^3](6-6p+p^2)+3E[\bar{B}^2](4-5p+p^2)]}{6(-1+b_0)p^2}. \qquad (18)
\end{aligned}$$

Recursive substituting (17) into itself yields, after some manipulations

$$1-\rho = \eta h_2 + \eta^2(h_1h_2+h_4)+o(\eta^2). \qquad (19)$$

From (19) and using the same technique given in section 4 of [1] we obtain the expression of $\eta$ in terms of $1-\rho$ as follows:

$$\eta = \frac{1}{h_2}(1-\rho) - \frac{h_1h_2+h_4}{h_2^3}(1-\rho)^2 + o((1-\rho)^2). \qquad (20)$$

In addition, from (18) the coefficients of $1-\rho$ and $(1-\rho)^2$ can be derived as follows:

$$\begin{aligned}
\frac{1}{h_2} &= \frac{2p}{(2-p)(E[B^2]-1)} \\
\frac{h_1h_2+h_4}{h_2^3} &= \frac{4p[-p^2-6(1-p)E[B^2]+(6-6p+p^2)E[A^3]]}{3(2-p)^3(E[B^2]-1)^3}. \qquad (21)
\end{aligned}$$

Here we use $E[B^k]=(1-b_0)E[\bar{B}^k], k\geq 1$.

Hence, we finally obtain the following approximation of the asymptotic decay rate of the tail probabilities for the virtual waiting time when $\rho$ is near 1:

**Theorem 4.1.** *Under heavy traffic condition, the asymptotic decay rate $1/z_0$ of the tail probabilities for the virtual waiting time, can be approximated as follows:*

$$1/z_0 \quad = \quad 1+c_1(1-\rho)+c_2(1-\rho)^2+o((1-\rho)^2),$$

*where the coefficients $c_1$ and $c_2$ are given as*

$$\begin{aligned}
c_1 &= -\frac{2p}{(2-p)(E[B^2]-1)} \\
c_2 &= \frac{2p[p(p-6)-3(4-6p+p^2)E[B^2]+2(6-6p+p^2)E[B^3]]}{3(2-p)^3[E[B^2]-1]^3}.
\end{aligned}$$

*Proof.* From the fact that $1/z_0 = e^{-\eta}$, we have

$$\begin{aligned} 1/z_0 &= 1 - \eta + \frac{\eta^2}{2} + o(\eta^2) \\ &= 1 - \frac{1}{h_2}(1-\rho) + \left(\frac{1}{2h_2^2} + \frac{h_1 h_2 + h_4}{h_2^3}\right)(1-\rho)^2 + o((1-\rho)^2). \end{aligned}$$

The coefficients of $(1-\rho)$ and $(1-\rho)^2$ can be computed from (21) as given in the theorem, which completes the proof.

**Remark:** When we have found an approximation for the decay rate, we can obtain an approximation for the asymptotic decay constant $\beta$ as follows: [1]

$$\beta = \eta E[W] + O((1-\rho)^2) \text{ as } \rho \to 1,$$

where $\eta$ is given by (20) and (21) and an approximation for the mean virtual waiting time $E[W]$ can be obtained from, for example, Tijms [17]. Interesting reader may refer to the discussion below Theorem 4 of [1].

## Acknowledgements

The first author would like to thank Prof. Khosrow Sohraby for helpful advise by discussing the DAR(1). The authors would like to thank the reviewer for his valuable comments which have improved the presentation of the paper.

## References

[1] J. Abate, G.L. Choudhury and W. Whitt (1994). Exponential Approximations for Tail Probabilities in Queues, I: Waiting Times, *Operations Research*, Vol. 43, No. 5, pp. 885–901.

[2] J. Abate and W. Whitt (1994). A Heavy-traffic Expasion for Asymptotic Decay Rates of Tail Probabilities in Multichannel Queues, *Operations Research Letters*, Vol. 15, pp. 223–230.

[3] M.H. Ackroyd (1980). Computing the Waiting Time Distribution for the G/G/1 Queue by Signal Processing Methods, *IEEE Transactions on Communications*, Vol. 28, pp. 52–58.

[4] S. Asmussen (1987). *Applied Probability and Queues*, John Wiley and Sons.

[5] C.E.P. Box and G.M. Jenkins (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day.

[6] A. Elwalid, D. Heyman, T.V. Laksman, D. Mitra, and A. Weiss (1995). Fundamental Bounds and Approximations for ATM Multiplexers with Applications to Video Teleconferencing, *IEEE Journal of Selected Areas in Communications*, Vol. 13, No. 6, pp. 1004–1016.

[7] P.D. Finch and C. Pearse (1965). A Second Look at a Queueing System with Moving Average Input Process, *Journal of Australian Math. Soc.*, Vol. 5, pp. 100–106.

[8] M.J. Fryer and cB. Winsten (1986). An Algorithm to Compute the Equilibrium Distribution of A One-dimensional Bounded Random Walk, *Operations Research*, Vol. 34, pp. 449–454.

[9] W.K. Grassmann and J.L. Jain (1989). Numerical Solutions of the Waiting Time Distribution and Idle Time Distribution of the Arithmatic GI/G/1 Queue, *Operations Research*, Vol. 37, pp. 141–150.

[10] G.U. Hwang and K. Sohraby (2001). An Exact Analysis of A Queueing System with An Autoregressive Model of Order 1, *Submitted for Publication.*

[11] J.F.C. Kingman (1970). Inequalities in the Theory of Queues, *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 32, No. 1, pp. 102–110.

[12] A.G. Konheim (1975). An Elementary solution of the Queueing System GI/G/1, *SIAM Journal of Computing*, Vol. 4, No. 4, pp. 540–545.

[13] A.J. Lawrance and P.A.W. Lewis (1977). An Exponential Moving Average Sequence and Point Process EMA1 Process, *Journal of Applied Probability*, Vol. 14, pp. 98–113.

[14] N.U. Prabhu (1998). *Stochastic Storage Processes: Queues, Insurance Risk, Dams and Data Communications*, 2nd ed., Springer.

[15] W.L. Smith (1953). On the Distribution of Queueing Times, *Proc. Camb. Phil. soc.*, Vol. 49, pp. 449–461.

[16] H. Takagi (1993). *Queueing Analysis : A Foundation of Performance Evaluation*, Vol. 1,2,3, Elservier Science, Amsterdam.

[17] H.C. Tijms (1986). *Stochastic Modelling and Analysis: A Computational Approach*, John Wiley & Sons, New York.

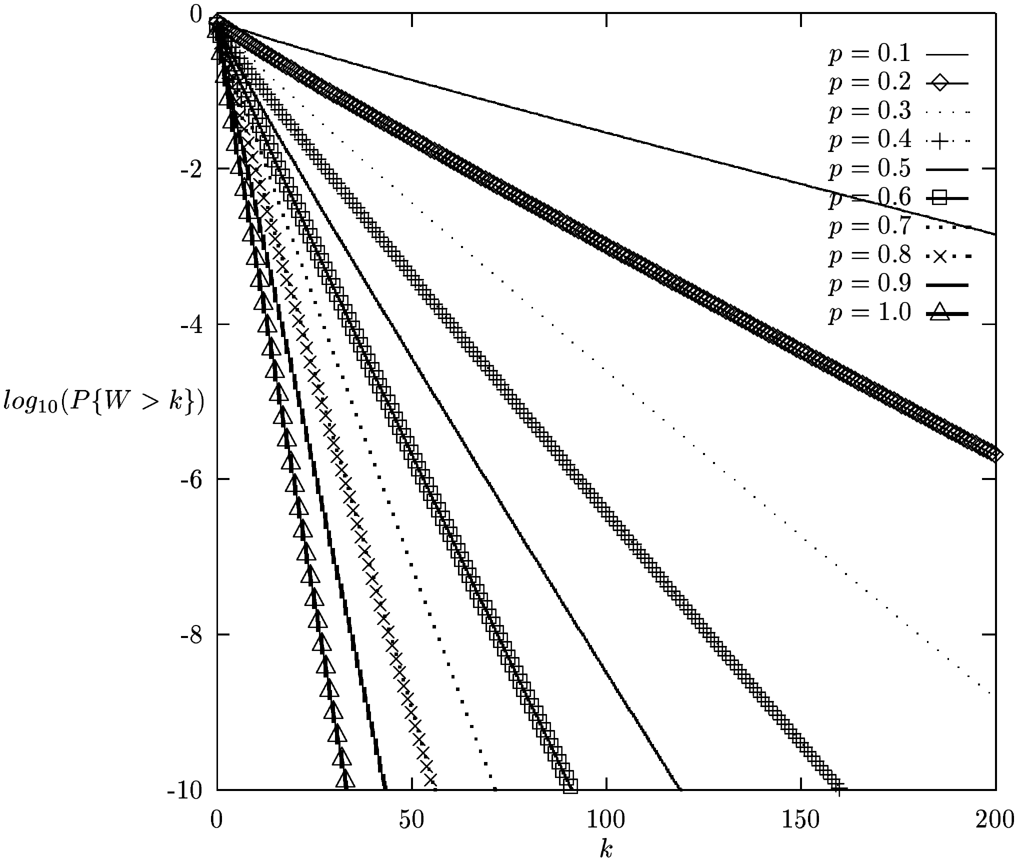

Figure 2. The effect of $p$ on the waiting time distribution