ALTERNATIVE APPROACH FOR ANALYZING DISCRETE-TIME
FINITE-BUFFER QUEUES WITH SERVER VACATIONS

S. K. SAMANTA¹, U. C. GUPTA¹ AND R. K. SHARMA²

Abstract

In this paper, we present a discrete-time single-server finite-buffer queue in which the server takes single/multiple vacations whenever the system becomes empty. The interarrival times are geometrically distributed and service, vacation times are independently identically distributed random variables and their durations are integral multiples of a slot duration. We obtain the state probabilities at service completion, vacation termination and arbitrary epochs using supplementary variable and imbedded Markov chain techniques. The analysis of actual waiting time under the first-come-first-served (FCFS) queueing discipline is also carried out.

Key words: finite-buffer, queue, single-server, vacations.

1. Introduction

In recent years discrete time queueing systems with vacations have been studied by many researchers due to their wide applications in the performance analysis of computer data communication system and telecommunication network which is based on the Asynchronous Transfer Mode (ATM) environment. The ATM has been adopted as the transport mechanism for the implementation of Broadband Integrated Service Digital Networks (B-ISDN). In this paper, we consider Geo/G/1/N queueing system where the server works continuously until the system empties. When the server finishes serving a customer and finds the queue empty, he goes away for a length of time called a vacation. On return from a vacation if he finds one or more customers waiting, he takes them for service one by one until the system empties, after that he takes another vacation. However, on return from a vacation if he finds no customer waiting then in case of single vacation he remains dormant until at least one customer arrive in the queue, whereas in case of multiple vacations he immediately proceeds for another vacation and continues in this manner until he finds at least one waiting customer upon return from a vacation. It has been seen that a very little works have been done in this direction. Notice that the analysis of the models discussed here is also available in Takagi [4]. The method of analysis used in this paper and in Takagi is quite similar and resembles in some places. However, we obtain several new results which are not given in Takagi [4]. It may be remarked here that finite-buffer queues are more realistic and useful in many applications. Recently, Zhang and Tian [6] have carried out analysis of the Geo/G/1 queue with multiple adaptive vacations. Further they have discussed the discrete-time GI/Geo/1 queue with multiple vacations [5]. The discrete-time GeoX/G/1 queue with timed vacations is analyzed by Fiems and Bruneel [2]. Alfa [1] analyzed a class of discrete-time single-server vacation models using matrix-analytic method.

¹ Organized by Department of Applied Mathematics with Oceanology and Computer Programming, Vidyasagar University, Midnapore, West Bengal, India.
2. Models description and solutions

Let us consider a discrete-time single-server finite-buffer queue of size $N$ with single/multiple
vacations. The time axis is divided into constant length intervals called slots. We assume
that customers arrive at the system according to a Bernoulli process with parameter $\lambda$.
The service [vacation] times $S$ [V] are independently identically distributed random variables
with probability mass function $s_n = P(S = n), n \geq 1$ [$v_n = P(V = n), n \geq 1$], correspond-
ing probability generating function (p.g.f.) $S(z)$ [$V(z)$] and mean service [vacation] time is
$E(S)$ [$E(V)$]. Let $\rho'$ be defined as the carried load, i.e. the probability that the server is busy
at an arbitrary time, where the offered load $\rho$ is defined as usual to be $\rho = \lambda E(S)$. Here
we discuss the models for late arrival system with delayed access (LAS-DA) and therefore
a potential arrival takes place in $(t-, t)$ and a potential departure occurs in $(t, t+)$. These
points are depicted in Figure 1.

The state of the system just before a potential arrival (at $t-$) is described by the following
random variables: the number of customers in the queue ($N_t$); the remaining service time
of the customer in service ($U_t$); the remaining vacation time of the server ($V_t$) and the
state of the server ($\xi_t$) as $\xi_{t-} = \lfloor 2 \rfloor \{1\}$ (0) the server is [busy] {on vacation} (in dormancy).
Let us define the joint probabilities as

$$
\pi_n(u,t-) = P[N_{t-} = n, U_{t-} = u, \xi_{t-} = 2], \quad 0 \leq n \leq N, \quad u \geq 0,
\omega_n(u,t-) = P[N_{t-} = n, V_{t-} = u, \xi_{t-} = 1], \quad 0 \leq n \leq N, \quad u \geq 0,
\gamma_0(t-) = P[N_{t-} = 0, \xi_{t-} = 0].
$$

Define the generating functions $\pi_n(z) = \sum_{u=0}^{\infty} \pi_n(u) z^u$ and $\omega_n(z) = \sum_{u=0}^{\infty} \omega_n(u) z^u, |z| \leq 1, 0 \leq n \leq N$. One may note that $\pi_n = \pi_n(1)$ [$\omega_n = \omega_n(1)$] denotes the probability of
$n$ customers in the queue, when the server is busy [on vacation] at arbitrary epoch. The
normalization condition is $\sum_{n=0}^{N} (\pi_n + \omega_n) + (1 - \delta_m) \gamma_0 = 1$.
Let $f_j$ and $h_j$ denote the probabilities that $j$ customers enter into the system during a
service time $S$ of a customer and a vacation time $V$ respectively. Hence for $j \geq 0$, we have

$$
f_j = \sum_{k=1}^{\infty} s_k \binom{k}{j} \lambda^j (1 - \lambda)^{k-j}, \quad h_j = \sum_{k=1}^{\infty} v_k \binom{k}{j} \lambda^j (1 - \lambda)^{k-j}
$$

with $f_j = h_j = 0$, for $j > k$ and let $\widehat{f}_i = \sum_{n=1}^{N} f_n, \ i \geq 1, \widehat{h}_N = \sum_{n=N}^{\infty} h_n$. 

![Figure 1. Various time epochs in late arrival system with delayed access.](image-url)
Now observing the state of the system at two consecutive epochs \( t-\) and \( (t+1)-\), in steady-state we have the following equations for \( u \geq 1 \)

\[
\begin{align*}
\pi_0(u - 1) &= (1 - \lambda)\pi_0(u) + (1 - \lambda)\pi_1(0)s_u + \lambda\pi_0(0)s_u + \lambda\omega_0(0)s_u \\
&\quad + (1 - \lambda)\omega_1(0)s_u + (1 - \delta_m)\lambda\gamma_0s_u, \\
\pi_n(u - 1) &= (1 - \lambda)\pi_n(u) + (1 - \lambda)\pi_{n+1}(0)s_u + \lambda\pi_n(0)s_u + \lambda\pi_{n-1}(u) \\
&\quad + \lambda\omega_{n-1}(0)s_u + (1 - \lambda)\omega_{n+1}(0)s_u, \quad 1 \leq n \leq N - 2, \\
\pi_N(u - 1) &= (1 - \lambda)\pi_N(u) + \pi_N(0)s_u + \lambda\pi_{N-1}(0)s_u + \lambda\pi_{N-2}(u) \\
&\quad + \lambda\omega_{N-1}(0)s_u + \omega_N(0)s_u, \\
\pi_N(u - 1) &= \pi_N(u) + \lambda\pi_{N-1}(u), \\
\omega_0(u - 1) &= (1 - \lambda)\omega_0(u) + \delta_m(1 - \lambda)\omega_0(0)v_u + (1 - \lambda)\pi_0(0)v_u, \\
\omega_n(u - 1) &= (1 - \lambda)\omega_n(u) + \lambda\omega_{n-1}(u), \quad 1 \leq n \leq N - 1, \\
\omega_N(u - 1) &= \omega_N(u) + \lambda\omega_{N-1}(u), \\
\gamma_0 &= (1 - \lambda)\gamma_0 + (1 - \lambda)\omega_0(0).
\end{align*}
\]

where \( \delta_m = 0 \) for single vacation and \( \delta_m = 1 \) for multiple vacations. Notice that the equation (8) will not appear in case of multiple vacations due to the absence of dormant state. Multiplying equations (1) - (7) by \( z^u \) and summing over \( u \), we get

\[
\begin{align*}
z\pi_0^s(z) &= (1 - \lambda)\{\pi_0^s(z) - \pi_0(0)\} + (1 - \lambda)\{\pi_1^s(0) + \omega_0(z)\}S(z) \\
&\quad + \lambda\{\pi_0(0) + \omega_0(0)\}S(z) + (1 - \delta_m)\lambda\gamma_0S(z), \\
z\pi_n^s(z) &= (1 - \lambda)\{\pi_n^s(z) - \pi_n(0)\} + (1 - \lambda)\{\pi_{n+1}(0) + \omega_{n}(0)\}S(z) \\
&\quad + \lambda\{\pi_{n-1}(z) - \pi_{n-1}(0)\} + \lambda\{\pi_n(0) + \omega_n(z)\}S(z), \\
&\quad 1 \leq n \leq N - 2, \\
z\pi_{N-1}^s(z) &= (1 - \lambda)\{\pi_{N-1}^s(z) - \pi_{N-1}(0)\} + \lambda\{\pi_{N-2}(z) - \pi_{N-2}(0)\} \\
&\quad + \lambda\{\pi_{N-1}(0) + \omega_{N-1}(0)\}S(z) + \{\pi_N(0) + \omega_N(z)\}S(z), \\
z\pi_N^s(z) &= \pi_N^s(z) - \pi_N(0) + \lambda\{\pi_{N-1}(z) - \pi_{N-1}(0)\}, \\
z\omega_0^s(z) &= (1 - \lambda)\{\omega_0^s(z) - \omega_0(0)\} + (1 - \lambda)\{\delta_m\omega_0(0) + \pi_0(0)\}V(z), \\
z\omega_n^s(z) &= (1 - \lambda)\{\omega_n^s(z) - \omega_n(0)\} + \lambda\{\omega_{n-1}(z) - \omega_{n-1}(0)\}, \\
&\quad 1 \leq n \leq N - 1, \\
z\omega_N^s(z) &= \omega_N^s(z) - \omega_N(0) + \lambda\{\omega_{N-1}(z) - \omega_{N-1}(0)\}.
\end{align*}
\]

**Lemma 1.** Setting \( z = 1 \) in equations (13) - (15) and adding them, after simplification we get

\[
(1 - \lambda)\left\{\pi_0(0) + \delta_m\omega_0(0)\right\} = \sum_{n=0}^{N} \omega_n(0).
\]

**Lemma 2.**

\[
\sum_{n=0}^{N} \pi_n = E(S)\sum_{n=0}^{N} \pi_n(0) = \rho' \quad \text{and} \quad \sum_{n=0}^{N} \omega_n = E(V)\sum_{n=0}^{N} \omega_n(0) = 1 - \rho' - (1 - \delta_m)\gamma_0.
\]

**Proof.** Adding equations (9) - (12) and using Lemma 1, we get

\[
(z - 1)\sum_{n=0}^{N} \pi_n^s(z) = \left\{S(z) - 1\right\}\sum_{n=0}^{N} \pi_n(0).
\]

Taking limit \( z \to 1 \), yields the first result. Applying similar arguments to the equations (13) - (15), which leads to the second result.
2.1. Queue length distribution at service completion and vacation termination epochs. Let \( \pi_n^+ \) \( [\omega_n^+] \) be the probability that there are \( n \) \((0 \leq n \leq N) \) customers in the queue at service completion [vacation termination] epoch. Therefore, we have

\[
\pi_n^+ = P\{n - 1 \geq 0 \text{ or } n \text{ customers in the queue just prior to service completion epoch} \mid \leq N \text{ customers in the queue just prior to service completion or vacation termination epoch}\} = \begin{cases} \frac{1}{\sigma}((1 - \lambda)\pi_0(0)) & : n = 0, \\ \frac{1}{\sigma}(1 - \lambda)\pi_n(0) + \lambda\pi_{n-1}(0) & : 1 \leq n \leq N - 1, \\ \frac{1}{\sigma}(\pi_N(0) + \lambda\pi_{N-1}(0)) & : n = N. \end{cases}
\]

Similarly, the expression of \( \omega_n^+ \) is given by

\[
\omega_n^+ = \begin{cases} \frac{1}{\sigma}((1 - \lambda)\omega_0(0)) & : n = 0, \\ \frac{1}{\sigma}(1 - \lambda)\omega_n(0) + \lambda\omega_{n-1}(0) & : 1 \leq n \leq N - 1, \\ \frac{1}{\sigma}(\omega_N(0) + \lambda\omega_{N-1}(0)) & : n = N. \end{cases}
\]

where

\[
s = P\{\leq N \text{ customers in the queue just prior to service completion or vacation termination epoch}\} = \sum_{n=0}^{N} \{\pi_n(0) + \omega_n(0)\}.
\]

The above results have been obtained by observing the events at epochs \( t^- \) and \( t^+ \) of Figure 1. It can be seen from (16) - (17) that to get \( \pi_n^+ \) and \( \omega_n^+ \) we need to find out \( \pi_n(0) \) and \( \omega_n(0) \). As \( \pi_n(0) \) and \( \omega_n(0) \) are cumbersome to evaluate directly from (9) - (15), we obtain them using imbedded Markov chain technique. The unknown quantities \( \{\pi_n^+\}_0^N \) and \( \{\omega_n^+\}_0^N \) can be obtained by solving the system of equations \((\pi^+, \omega^+) = (\pi^+, \omega^+)(\mathbf{P}) \) with \((\pi^+ + \omega^+)e = 1\), where \( e \) is a column vector of ones with an appropriate dimension. Here \( \pi^+ = [\pi_0^+, \pi_1^+, \ldots, \pi_N^+] \) and \( \omega^+ = [\omega_0^+, \omega_1^+, \ldots, \omega_N^+] \) are the stationary probability vectors of the one-step transition probability matrix \( \mathbf{P} \) of order \((2N + 2) \times (2N + 2)\) as given below. We solved the system of equations \((\pi^+, \omega^+) = (\pi^+, \omega^+)(\mathbf{P}) \) with \((\pi^+ + \omega^+)e = 1\) using the GTH (Grassmann, Taksar and Heyman) algorithm given in Latouche and Ramaswami ([3], pg 123). The transition probability matrix \( \mathbf{P} \) is given by

\[
\mathbf{P} = \begin{pmatrix}
0 & 0 & 0 & 0 & h_0 & h_1 & \cdots & h_{N-1} & \hat{h}_N \\
0 & f_0 & f_1 & f_{N-1} & \hat{f}_N & 0 & 0 & \cdots & 0 & 0 \\
0 & f_0 & f_1 & f_{N-1} & \hat{f}_N & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\end{pmatrix},
\]

where \( c_n = (1 - \delta_n)f_n, \) \((0 \leq n \leq N - 1)\) and \( \hat{c}_N = (1 - \delta_n)\hat{f}_N. \) In the transition probability matrix \( \mathbf{P} \), the top left corner refers to a transition from service completion to service completion, the top right corner refers to a transition from service completion to vacation termination, the bottom left corner refers to a transition from vacation termination to service completion and the bottom right corner refers to a transition from vacation termination to vacation termination.
Lemma 3. Let $\Theta_B$ [$\Theta_I$] be the random variable denoting the busy [server unavailable] period and $E(\Theta_B)$ [$E(\Theta_I)$] be the corresponding mean. From the definition of the carried load $\rho’$ (the fraction of time that the server is in a busy period), we have

$$\rho’ = \frac{E(\Theta_B)}{E(\Theta_B) + E(\Theta_I)}.$$  \hfill (18)

Using Lemma 2 and equation (18), we have

$$E(\Theta_I) \frac{E(\Theta_B)}{E(\Theta_I)} = 1 - \frac{\rho’}{\rho’} = (1 - \delta_m)\gamma_0 + \lambda E(V) \sum_{n=0}^{N} \omega_n(0) \frac{E(S)}{\sum_{n=0}^{N} \pi_n(0)},$$  \hfill (19)

Applying equations (8) and (16) - (17) in above, we obtain

$$E(\Theta_I) \frac{E(\Theta_B)}{E(\Theta_B)} = (1 - \delta_m)\gamma_0 + \lambda E(V) \sum_{n=0}^{N} \omega_n^+ \frac{\pi_0^+}{\lambda E(S) \sum_{n=0}^{N} \pi_n^+}.$$  \hfill (19)

Now dividing the numerator and denominator of the right side expression of equation (18) by $E(\Theta_B)$ and then using equation (19), we get the value of $\rho’$.

Lemma 4. Adding all the terms in equation (16) and using this in the first identity of Lemma 2, immediately yields

$$\sigma = \frac{\rho’}{E(S) \sum_{n=0}^{N} \pi_n^+}.$$  

2.2. Queue length distribution at arbitrary epoch. The arbitrary epoch probabilities are obtained from (8) and setting $z = 1$ in equations (9) - (11), (13) - (14) and then using (16) - (17), we get

$$\gamma_0 = \frac{\sigma}{\lambda} \omega_0^+,$$

$$\pi_0 = (1 - \delta_m)\gamma_0 + \frac{\sigma}{\lambda} \left\{ \pi_1^+ + \omega_1^+ - \pi_0^+ \right\},$$

$$\pi_n = \pi_{n-1} + \frac{\sigma}{\lambda} \left\{ \pi_{n+1}^+ + \omega_{n+1}^+ - \pi_n^+ \right\}, \quad 1 \leq n \leq N - 1,$$

$$\omega_0 = \frac{\sigma}{\lambda} \left\{ \pi_0^+ - (1 - \delta_m)\omega_0^+ \right\},$$

$$\omega_n = \omega_{n-1} - \frac{\sigma}{\lambda} \omega_n^+, \quad 1 \leq n \leq N - 1.$$  

But $\pi_N$ and $\omega_N$ can not be obtained from equations (12) and (15), respectively by setting $z = 1$. We can obtain them using Lemma 2 in the sequel

$$\pi_N = \rho’ - \sum_{n=0}^{N-1} \pi_n$$

and $\omega_N = 1 - \rho’ - \sum_{n=0}^{N-1} \omega_n - (1 - \delta_m)\gamma_0$.

Let $p_n$ denotes the probability that there are $n$ customers in the queue at arbitrary epoch. Then

$$p_n = \begin{cases} \pi_0 + \omega_0 + (1 - \delta_m)\gamma_0, & n = 0, \\ \pi_n + \omega_n, & 1 \leq n \leq N. \end{cases}$$

Notice that the probability of blocking ($PBL$) is given by $PBL = p_N$.

2.3. Waiting time distribution. Let $W_q(z)$ be the probability generating function of actual waiting time of an arrived customer. Note that an arrived customer may be either

(i) served immediately if he sees the server in dormant state, or

(ii) served after the completion of the customer being served and all the waiting customers in front of him depart from the system if he sees the server in busy state, or

(iii) served after a vacation period ends and all the waiting customers in front of him depart from the system if he sees the server in vacation state.
Therefore, the actual waiting time p.g.f. is given by
\[ W_q(z) = \frac{1}{1 - pN} \left\{ (1 - \delta m) \gamma_0 + \sum_{n=0}^{N-1} \pi_n^*(z) \{S(z)\}^n + \sum_{n=0}^{N-1} \omega_n^*(z) \{S(z)\}^n \right\}. \]

Thus, the expected actual waiting time is given by
\[ W_q = \frac{1}{1 - pN} \left\{ \sum_{n=0}^{N-1} n \{\pi_n + \omega_n\} E(S) + \sum_{n=0}^{N-1} \{\pi_n^*(1) + \omega_n^*(1)\} \right\}, \]

where \(\pi_n^*(1)\) and \(\omega_n^*(1)\), \(0 \leq n \leq N - 1\) are obtained by differentiating equations (9) - (11) and (13) - (14) w.r.t. \(z\) at \(z = 1\). These are given as
\[ \pi_0^*(1) = \left\{ (1 - \delta m) \gamma_0 + \frac{\sigma}{\lambda} \{\pi_1^* + \omega_1^*\} \right\} E(S) - \frac{\pi_0}{\lambda}, \]
\[ \pi_n^*(1) = \frac{\sigma}{\lambda} \{\pi_{n+1} + \omega_{n+1}\} E(S) - \frac{\pi_n}{\lambda}, \quad 1 \leq n \leq N - 1, \]
\[ \omega_0^*(1) = \frac{\sigma}{\lambda} \{\pi_0^* + \delta m \omega_0^*\} E(V) - \frac{\omega_0}{\lambda}, \]
\[ \omega_n^*(1) = \omega_{n-1}^*(1) - \frac{\omega_n}{\lambda}, \quad 1 \leq n \leq N - 1. \]

3. Performance measures

In this section, we obtain some key performance measures such as the average number of customers in the queue at an arbitrary epoch \((L_q)\), average waiting time in the queue \((W_q)\), using Little’s rule, are given by
\[ L_q = \sum_{n=0}^{N} np_n, \quad W_q = \frac{L_q}{\lambda'}, \]

where \(\lambda' = \lambda(1 - P\text{BL})\) is the effective arrival rate.

We have checked numerically that in case of infinite buffer queues \(\rho\) and \(\rho'\) are equal where as in finite buffer queues they are different. The expected waiting time in the queue obtained through p.g.f. of actual waiting time in the queue is exactly same as the one obtained using Little’s rule. Also the blocking probability \((P\text{BL})\) is exactly matches with \((1 - \rho'/\rho)\). We have not presented any numerical result due to lack of space.

References


Address of the authors.

1Department of Mathematics, Indian Institute of Technology Kharagpur-721302, India.
2Department of Mathematics, Indian Institute of Technology Delhi-110016, India.