Analysis of the $M/M/2$ Queueing System with Restricted Service Duration

A G Luwanda*and J J Namangale †

Abstract

This paper deals with a two-server queue in which the service duration is governed by an exponential timer. If servers enter the system and find that it has no waiting customer, they leave for another vacation. If they find one or more waiting customers, they start giving service. Servers leave the system if all customers have been served. However, if the timer expires while there are customers waiting for service, then service continues until none of those accumulated up to the timer's life is left before servers go on vacation. We obtain expressions for the size of a queue at time t, and the number of services in a busy period. We also present expressions for the expected waiting times.

1 Introduction

Consider a production line where an operator controls machines. Machines are started and stopped by a single switch. After working for some time which is preassigned, the machines must be serviced. If they have worked on all fragments available before the preassigned time is exceeded, then the operator switches off the system and machines are serviced. However, if the time expires while there are fragments to be worked on, then machines are switched off after all fragments that enter the gate are processed. In this example, the machines represent servers and customers are the fragments.

In this paper we consider a Markovian queueing system whose service duration is controlled by a an exponential timer. When servers enter the system, a timer is activated. If servers find no job in the system upon arrival , they go on another vacation of exponential duration. Otherwise, they give service until (i) no customers remain or (ii) the timer expires (while the system has customers). In the former case, servers go for a vacation immediately; and in the latter case, work continues on customers accumulated up to T, timer's life, until all get served after which servers leave for a vacation. Thus a customer that arrives in the system after the timer's expiration waits until servers return from a vacation.

*Mathematics Department, Mzuzu University

†Mathematical Sciences Department, University of Malawi

Queueing systems with servers' vacation have received considerable attention. Alfa [1] has analyzed a single-server queue ($MAP/PH/1$ vacation queue with gated time-limited service) in which the server goes through periods of vacation and work. In each busy period, the server attends to the queue for no more than a fixed length of time, T. The server does not attend to customers that were not in the system before its visit. Esenberg and Leung [6] have studied a single server queue with gated time-limited service. They have obtained expressions for performance measures of the system. Eliazar and Yechali [5] in their paper, *Randomly Timed Gated Queueing Systems*, have studied an $M/G/1$ queue in steady state and calculated various performance measures including queue sizes and waiting times. The server goes into the system and its stay in the system has an exponential distribution and is governed by a timer. If it finds no jobs upon entry into the system, it goes for another vacation. If the server finishes the work before the timer's expiration, it proceeds on vacation. However, if the timer goes while the system has customers, then the server obeys one of the following rules: 1) all work accumulated up to T, the timer's life, is completed before it leaves; 2) the server completes only the work in service, then it goes on vacation; 3) the server obeys the timer and goes for leave immediately.
We have considered an $M/M/2$ queue with service duration determined by a timer. Our study follows the approach in Eliazar and Yechiali [5]. We have obtained performance measures of the system that include the expected number of services in busy period and length of a queue at time t. We have also found an expression for the waiting times.
The rest of the paper is organized as follows. In Section 2, we have introduced some results from stochastic processes that will be used in the analysis. Parameters and variables are also introduced in this section. In Section 3 we have found expressions for the expected number of services in a busy period and the length of a queue at *any* time t. Section 4 is devoted to the study of waiting times. Conclusions are given in Section 5.

2 Preliminary Concepts

In this section we introduce some results from stochastic processes that will be needed in the sequel. We also present system parameters and assumptions.

2.1 Renewal Process

Let $\{T_n : n = 1, 2, 3, \ldots\}$ be a sequence of non-negative independent random variables with a common distribution F. T_n is the time between the $(n-1)$st event and nth event. Let the mean time between successive events be denoted by μ. This mean is given by

$$\begin{aligned}\mu &= E[T_n] \\ &= \int_0^\infty t dF(t).\end{aligned}$$

Let S_n be such that $S_0 = 0$, and $S_n = \sum_{i=1}^{n} T_i, \quad n \geq 1$. Then S_n is the time of the nth event. Let $N(t_1)$ be the number of events by time t_1. Then

$$N(t_1) = \sup n : S_n \leq t_1. \tag{1}$$

This leads us to the following.

Definition 2.1
The counting process $\{N(t_1) : t_1 \geq 0\}$ is called a *renewal process.*

Let $\{T_1, T_2, \cdots\}$ denote a sequence of independent random variables. Then an integer-valued random variable N is said to be a *stopping time* for the sequence if the event $[N = n]$ is independent of $T_{n+1}, T_{n+2}, \ldots$ for $n = 1, 2, 3, \ldots$ We state one of the most important theorems in renewal theory;it is called *Wald's Equation* and will be used in the analysis.

Theorem 2.1 (Wald's Equation)
If $T_1, T_2, \ldots$ are identically and independently distributed random variables having finite expectation, and if N is a stopping time for $T_1, T_2, \ldots,$ with $E[N]$ finite then

$$E\left[\sum_{n=1}^{N} T_n\right] = E[T]E[N]. \tag{2}$$

Its proof can be found in Ross [8].

2.2 System variables and parameters

2.2.1 System Parameters

The system parameters are as follows:

1. *Arrival process* $\{A(t) : t \geq 0\}$: This is a Poisson process with rate α, counting the number of arrivals in the interval $(0, t]$;

2. *Service times* $\{S_r : r \geq 1\}$: These are exponential random variables with parameter β, where S_r is the service time needed by the rth job;

3. *Vacation Intervals*: These are random time intervals at which servers are not available for service. A vacation interval starts each time servers leave the system and these intervals have a common exponential distribution with finite parameter, σ;

4. *Timer* T: This is an exponentially distributed random variable with parameter $\mu < \infty$. This is reactivated every time servers reenter the system;

5. *Discipline of queue*: Customers get service on first-come-first-served basis. A customer does not renege.

The following important variables are defined:

1. $X :\equiv$ queue size at a polling instant;
2. $B :\equiv$ the length of a busy period initiated by the presence of x jobs and
3. $N :\equiv$number of jobs served during a busy period.

3 The analysis

We look at the model when the systen is in steady state. An expression for the number of services in a busy period is derived along with the queue length at time t.

3.1 The number of services in a busy period

Let N be the number of services in a busy period. Let j be the number of arrivals that will depend on b and T, where b is a time point at the end of a busy period and T is the length of the timer's life. Thus the number of customers served is $(x+j)$ where $x \quad (\geq 1)$ is the number of customers at a polling instant. Thus $n = x + j$ which, implies that $j = n - x$. We note that the number of arrivals in $(0, t]$ is independent of the queue length at a polling instant.
Now, denote the arrival vector by $\underline{a} = \{(a_1, a_2) : a_1 + a_2 = j\}$ where a_1 is the number of arrivals when the system has one customer and a_2 is the number of arrivals when the system has two or more customers and let $\underline{d} = \{(d_1, d_2) : d_1 + d_2 = n\}$ be the corresponding departure vector. Necessarily, $d_1 = 1$ and $d_2 = n - 1$. Let $C(\underline{a}, \underline{d})$ be the number of sequences of arrivals and departures (exactly $(a_1 + 1)$ such events) when the system is in state one.

Proposition 3.1
Let $a(x)$ be the probability mass function of X. Then

$$P(N = n) = \sum_{n-x=0}^{\infty} \sum_{x=1}^{n} a(x) \left\{ \frac{(\alpha t)^{n-x} e^{-\alpha t}}{(n-x)!} C(\underline{a}, \underline{d}) \left(\frac{2\rho}{1+2\rho} \right)^{a_1} \times \left(\frac{\rho}{\rho+1} \right)^{n-x-a_1} \left(\frac{1}{\rho+1} \right)^{n-1} \left(\frac{1}{2\rho+1} \right) \right\} \quad (3)$$

where $\rho < 1$, $t = b$ for $b \leq T$ and $t = T$ when $b > T$.

Proof 1
a_1 jobs arrive in the system each with probability $\frac{2\rho}{2\rho+1}$ and one job departs with probability $\frac{1}{2\rho+1}$ (see Bunday [3]). The other events (arrivals and departures) occur when the system has two or more jobs. There are $(n - x - a_1)$ arrivals each with probability $\frac{\rho}{\rho+1}$ and $(n-1)$ departures each with probability $\frac{1}{\rho+1}$. Then the distribution of N is determined by conditioning on the number of

jobs at polling instant and the number of new arrivals that will also get service before the end of a busy period. Let $t \leq T$. Then

$$P(N = n \mid X = x, A(t) = n - x) = C(\underline{a}, \underline{d}) \left(\frac{2\rho}{1+2\rho}\right)^{a_1} \times \left(\frac{\rho}{\rho+1}\right)^{n-x-a_1} \left(\frac{1}{\rho+1}\right)^{n-1} \left(\frac{1}{2\rho+1}\right). \quad (4)$$

Hence

$$P(N = n, X = x, A(t) = n - x) = a(x)\frac{(\alpha t)^{n-x}e^{-\alpha t}}{(n-x)!}C(\underline{a}, \underline{d}) \times \left(\frac{2\rho}{1+2\rho}\right)^{a_1} \left(\frac{\rho}{\rho+1}\right)^{n-x-a_1} \left(\frac{1}{\rho+1}\right)^{n-1} \left(\frac{1}{2\rho+1}\right). \quad (5)$$

Thus, the probability that $N = n$ jobs have been served in a busy period is given by

$$P(N = n) = \sum_{n-x=0}^{\infty} \sum_{x=1}^{n} a(x)\frac{(\alpha t)^{n-x}e^{-\alpha t}}{(n-x)!}C(\underline{a}, \underline{d}) \left(\frac{2\rho}{1+2\rho}\right)^{a_1} \times \left(\frac{\rho}{\rho+1}\right)^{n-x-a_1} \left(\frac{1}{\rho+1}\right)^{n-1} \left(\frac{1}{2\rho+1}\right). \quad (6)$$

This finishes the proof.

Now using Equation 3, the expected number of jobs served in a busy period is given by

$$E[N] = \sum_{n=1}^{\infty} n \left\{ \sum_{n-x=0}^{\infty} \sum_{x=1}^{n} a(x)\frac{(\alpha t)^{n-x}e^{-\alpha t}}{(n-x)!}C(\underline{a}, \underline{d}) \left(\frac{2\rho}{1+2\rho}\right)^{a_1} \times \left(\frac{\rho}{\rho+1}\right)^{n-x-a_1} \left(\frac{1}{\rho+1}\right)^{n-1} \left(\frac{1}{2\rho+1}\right) \right\}. \quad (7)$$

This result can be used in determining the expected length of a busy period. Let B be the busy period. Then

$$B = \sum_{r=1}^{N} S_r \quad (8)$$

where S_r is the service time needed by the rth customer and N stands for the number of all customers served in a busy period. Since service times have finite means and also $E[N] < \infty$ (since $\mu < \infty$ (see Section 2.2)) , we can now employ Theorem 2.1, Wald's Equation, to get

$$\begin{aligned} E[N]E[S] &= E\left[\sum_{r=1}^{N} S_r\right] \\ &= E[B] \end{aligned} \tag{9}$$

where $E[S]$ is the expected service time of a customer and is given by $[S] = \frac{1}{2\beta}$. Hence,

$$E[B] = \frac{E[N]}{2\beta}. \tag{10}$$

Thus from Equation (7) we have

$$\begin{aligned} E[B] = \frac{1}{2\beta}\sum_{n=1}^{\infty} n \Bigg\{ & \sum_{n-x=0}^{\infty}\sum_{x=1}^{n} a(x)\frac{(\alpha t)^{n-x}e^{-\alpha t}}{(n-x)!}C(\underline{a},\underline{d})\left(\frac{2\rho}{1+2\rho}\right)^{a_1} \\ & \times \frac{1}{2\beta}\left(\frac{\rho}{\rho+1}\right)^{n-x-a_1}\left(\frac{1}{\rho+1}\right)^{n-1}\left(\frac{1}{2\rho+1}\right)\Bigg\}. \end{aligned} \tag{11}$$

We note that this parameter depends on the number of jobs in the system at a polling instant.

3.2 Queue length

Let Q_0 be the queue length for an $M/M/2$ queueing system without vacation and Q_v be the size of the queue at some time point within a vacation period. If we let Q denote the queue size at an arbitrary inspection time instant in a cycle, then $E[Q]$ is given in the following.

Proposition 3.2
Let X be the size of a queue at polling instant and B be length of a busy period. Then

$$\begin{aligned} E[Q] &= \frac{2\rho}{1-\rho} - \sigma\alpha\left(\frac{E[X]-1}{2\beta} - \frac{1-\rho}{\mu}\right)^2 \\ &\quad + \frac{\sigma}{\alpha}E[X(X-1)] \end{aligned} \tag{12}$$

where $\mu < \infty$ and $\sigma < \infty$.

Proof 2
Let Q_0 and Q_v be independent random variables and suppose that $\rho < 1$. Then (see Tian et al [9])

$$Q = Q_0 + Q_v. \tag{13}$$

Now the probability generating function of Q is given by

$$\widetilde{Q}(z) = \widetilde{Q}_0(z)\widetilde{Q}_v(z) \tag{14}$$

where $\widetilde{Q}(z)$, $\widetilde{Q}_0(z)$ and $\widetilde{Q}_v(z)$ are, respectively, generating functions of the random variables Q, Q_0 and Q_v. Now the probability generating function of Q_0 is given by (see Luwanda and Namangale [7])

$$\begin{aligned} \widetilde{Q}_0(z) &= \sum_{l=0}^{\infty} P(Q_0 = l)z^l, \quad \mid z \mid \leq 1 \\ &= (1-\rho)\sum_{l=0}^{\infty} (\rho z)^l \\ &= \frac{1-\rho}{1-\rho z}. \end{aligned} \tag{15}$$

Also applying the Strong Law of Large Numbers (see Eliazar and Yechiali [5]) on Y gives

$$\widetilde{Q}_v(z) = \frac{1}{z-1}\frac{\widetilde{X}(z) - \widetilde{Y}(z)}{E[X] - E[Y]} \tag{16}$$

where X is the number of jobs at a polling instant and Y is the number of jobs left at the end of a busy period. Now, $E[X] - E[Y] = \frac{\alpha}{\sigma}$ ([7]) Thus

$$\widetilde{Q}(z) = \frac{1-\rho}{1-\rho z}\frac{\widetilde{X}(z) - \widetilde{Y}(z)}{z-1}\frac{\sigma}{\alpha}. \tag{17}$$

The expected queue length is found by differentiating Equation 17 with respect to z and evaluating the derivative of the resulting expression at $z = 1$.

$$\begin{aligned} E[Q] &= \frac{d}{dz}\widetilde{Q}(z) \mid_{z=1} \\ &= \frac{\sigma(1-\rho)}{\alpha}\frac{d}{dz}\left(\frac{\widetilde{X}'(z) - \widetilde{Y}'(z)}{1+\rho(1-2z)}\right) \mid_{z=1} \\ &= \frac{\sigma}{\alpha}\frac{\left[\widetilde{X}''(z) - \widetilde{Y}''(z)\right](1+\rho(1-2z))}{1-\rho} \mid_{z=1} \\ &\quad + \frac{2\rho\sigma}{\alpha(1-\rho)}\left[\widetilde{X}'(z) - \widetilde{Y}'(z)\right] \mid_{z=1} \end{aligned} \tag{18}$$

Now, $E[Y(Y-1)] = \left(\alpha(E[B] - \frac{1}{\mu})\right)^2$.
Thus the expression in Equation 18 simplifies to

$$\begin{aligned} E[Q] &= \frac{\sigma(1-\rho)}{\alpha}\left\{\frac{E[X(X-1)]-E[Y(Y-1)]}{1-\rho}+\frac{2\alpha\rho}{\sigma(1-\rho)^2}\right\} \\ &= \frac{\sigma}{\alpha}E[X(X-1)]-\sigma\alpha\left(E[B]-\frac{1}{\mu}\right)^2+\frac{2\rho}{1-\rho} \end{aligned} \tag{19}$$

as required.

4 Waiting times

We look at the waiting time of a customer who arrives in the system at a time instant t_1. There are three possibilities. But we present results for only two in Propositons 4.1 and 4.2. The cases are (i) $t_1 \leq T$, (ii) $T < t_1 < B$ and (iii) $t_1 > B$.
Let w_k denote the waiting time for a customer who arrives in the system in case k ($k = 1, 2$), $S(t_1)$ be the number of services completed by time t_1 and R_d and R_v denote residual service time and residual vacation time respectively, where d is the number of arrivals in the interval $(0, t_1]$.

Proposition 4.1
Let $\rho < 0$ and $t_1 < B < T$. Then the waiting time w_1 is given by

$$E[w_1] = \frac{E[X]}{2\beta} + \frac{1}{\beta}\frac{\rho^2}{1-\rho^2}. \tag{20}$$

Proof 3
The dth arrival will wait for $X + (d-1) - S(t_1)$ customers to be served and these will need an amount of time given by $E[S]E[X + (d-1) - S(t_1))]$. Besides this time, there is the residual service time for jobs found in service. Now, in the interval $(0, t_1]$, we have $A(t_1)$ arrivals and $S(t_1)$ services. So

$$E[w_1] = E[S]E[X] + E[S][A(t_1) - S(t_1)] + E[R_d] \tag{21}$$

Let w_d be the amount of time the dth customer (excluding those present at polling instant) has to wait before getting service in order for the $(d-1)$ jobs to be served. This time includes the residual service time R_d. Both servers are busy, thus the inter-departure times are independent exponential random variables with parameter 2β. Thus the total time until the $(d-1)$st departure is the $(d-1)$-fold convolution of this distribution. Thus the expected value of w_d is given by

$$\begin{aligned} E[w_d] &= \frac{1-\rho}{1+\rho}\frac{\alpha^2}{\beta}\int_0^\infty te^{-t(2\beta-\alpha)}dt \\ &= \frac{1-\rho}{1+\rho}\frac{\alpha^2}{\beta}\frac{1}{(2\beta-\alpha)^2} \\ &= \frac{1}{\beta}\left(\frac{\rho^2}{1-\rho^2}\right). \end{aligned} \tag{22}$$

Hence, Equation 21 simplifies to Eq. 20.

We note that Equation 20 is also true for the case where $t_1 < T < B$ because our interest is the waiting time for a job that arrives in the system before the timer expires.

Proposition 4.2
The waiting time of a customer that arrives in the system in vacation period is given by

$$E[w_3] = \frac{\rho^2\left(E[X]-1\right)}{\alpha} - \frac{\rho(1-\rho)}{\mu} + \frac{1}{\sigma}(1+\rho). \quad (23)$$

Proof 4
The customer arrives in the interval $(0, V]$. There are customers that servers left behind as they (servers) were going for a vacation and those that arrive in the system during a vacation but before the *tagged* customer. In addition to the time needed to serve earlier arrivals, there is residual vacation interval, R_v, which the customer has to endure. Thus the waiting time has three parts: the residual vacation interval, the amount of time needed to serve $A(t_1-B)$ arrivals and the time needed to serve $A(t_b)$ jobs. This leads us to Eq. 24

$$E[w_3] = E[R_v] + E[S]E[A(t_1-B)] + E[S]E[A(t_b)]. \quad (24)$$

Now, because of the memoryless property of an exponential random variable, the residual vacation period has the same distribution as the whole vacation. So

$$E[R_v] = \frac{1}{\sigma}. \quad (25)$$

$E[A(t_b)]$ is the length of the queue left in the system after a busy period. To determine $E[A(t_1-B)]$, we use the fact that these arrivals take place during an exponential vacation duration. The expected number of such arrivals is given in Eq. 26

$$\begin{aligned} E[A(t_1-B)] &= \alpha E[V] \\ &= \frac{\alpha}{\sigma}. \end{aligned} \quad (26)$$

Substituting Eq.25, Eq. 26, second case of Eq.**??** and the result in Proposition **??** in Eq. 24 gives us

$$E[w_3] = \frac{\rho^2\left(E[X]-1\right)}{\alpha} - \frac{\rho(1-\rho)}{\mu} + \frac{1}{\sigma}(1+\rho). \quad (27)$$

which is what we needed to show.

We further note that all the performance measures we have looked at depend on X, the number of jobs at a polling instant.

5 Summary of results

We have analysed the system and obtained expressions for the number of jobs served in this system and the corresponding length of a busy period.

Expressions for the expected length of a queue and waiting times for a customer who arrives in the system before and after the timer expires have been derived (see Proposition 3.2 and Equations 20 and 23).

References

[1] Alfa, A. *Discrete time analysis of $MAP/PH/1$ vacation queue with gated time-limited service*, **Queueing Systems** Vol. 29, 1998, pp 35–54.

[2] Boxma, O.; Yechiali, U. *An $M/G/1$ queue with multiple types of feedback and gated vacations*, Invited paper for the book **'Applied Probability (Liber Amicorum for J. Gani)'** Edited by C Hyde et al (Springer Verlag, Notes in Statistics), 1991.

[3] Bunday, B.D. **An Introduction to Queuing Theory**, Arnold, London, 1996.

[4] Chae, K.; Lee, H. *Number of services in an $M/M/2$ busy period*, **International Journal of Mathematics Education and Science Teaching** Vol. 23, 1991, pp. 786–789.

[5] Eliazar, I.; Yechiali, U. *Randomly timed gated queuing systems*, **SIAM Journal of Applied Mathematics** 59(2), 1997, pp. 423–441.

[6] Esenberg, M.; Leung, K. *A Single -server queue with vacations and gated time-limited service*, **IEEE Transactions On Communication** 38(9), 1990, pp. 1454–1462.

[7] Luwanda, A. G. and Namangale, J. J. *The $M/M/2$ Queueing System with Constrained Servers' Sojourn Time*, **MSc Thesis**, 2005.

[8] Ross, S. M. **Stochastic Processes**, 2nd Edition John Wiley and Sons, Inc., 1996.

[9] Tian, N.; Li,Quan-Lin; Cao, Q. *Conditional decomposition in the $M/M/c$ queue with servers vacation*, **Comminucation Statistics-Stochactic Models** Vol. 15, 1999, pp. 367–377.