Applied Mathematics, 2012, 3, 2174-2177
http://dx.doi.org/10.4236/am.2012.312A298 Published Online December 2012 (http://www.SciRP.org/journal/am)

Server Workload in an M/M/1 Queue with Bulk Arrivals and Special Delays

Percy H. Brill[1,2], Myron Hlynka[1]
[1]Department of Mathematics & Statistics, University of Windsor, Windsor, Canada
[2]Department of Management Science, University of Windsor, Windsor, Canada
Email: brill@uwindsor.ca, hlynka@uwindsor.ca

Received November 10, 2012; revised December 10, 2012; accepted December 17, 2012

ABSTRACT

We consider a variant of M/M/1 where customers arrive singly or in pairs. Each single and one member of each pair is called primary; the other member of each pair is called secondary. Each primary joins the queue upon arrival. Each secondary is delayed in a separate area, and joins the queue when "pushed" by the next arriving primary. Thus each secondary joins the queue followed immediately by the next primary. This arrival/delay mechanism appears to be new in queueing theory. Our goal is to obtain the steady-state probability density function (pdf) of the workload, and related quantities of interest. We utilize a typical sample path of the workload process as a physical guide, and simple level crossing theorems, to derive model equations for the steady-state pdf. A potential application is to the processing of electronic signals with error free components and components that require later confirmation before joining the queue. The confirmation is the arrival of the next signal.

Keywords: M/M/1 Queue; Bulk Arrivals; Delay before Joining; Workload; Integral Equations; Level crossing Method

1. Introduction

The M/M/1 arrival/delay mechanism considered in this paper was introduced by Hlynka [1], who derived the Laplace transform of the busy period of the server, using the probabilistic interpretation of the Laplace transform. The busy period in that analysis included the idle times of the server while a secondary is being delayed.

Here we analyze the model using a level crossing approach, and derive: the steady-state pdf of the server workload; probability that the system is empty; probability that the server is idle when a secondary is being delayed; expected busy period (as defined in Hlynka [1]); a stability condition; expected time the server is busy in a cycle between instants of system emptiness, or between instants the server becomes idle and a secondary is being delayed. An advantage of the level crossing method used here is that it focuses on the workload process in a concrete manner. That is, it uses physical properties of a typical sample path of the workload process as a guide, and simple level crossing theorems, to formulate the model equations for the key probability distributions of the model. Viewing the sample path in this concrete manner, makes the solution procedure intuitive, straightforward, and suggestive of future research ideas.

Section 2 specifies the M/M/1 variant and sample path structure. Section 3 derives the model equations for the steady-state pdf of the workload, and specifies related quantities. Section 4 uses the model equations to obtain relevant constant terms, and gives a numerical example of the steady-state pdf of workload.

2. The M/M/1 Variant

Singles (primaries) arrive at the system at Poisson rate λ_1 and pairs (pair = primary + secondary) arrive at the system at Poisson rate λ_2; let $\lambda = \lambda_1 + \lambda_2$. When a primary arrives at the system, it immediately joins the M/M/1 queue, either alone or just behind a secondary that was being delayed. When a secondary arrives at the system it splits from its primary and is delayed in a separate area outside the queue. The delayed secondary is "pushed" by the next arriving primary to join the queue, followed immediately by the new primary. (Thus the delay of each secondary before joining the queue is distributed as exponential-λ.) Customers in the queue are served one at a time at exponential rate μ in first-come-first-served order. When a primary joins the queue alone, the server workload is increased by exponential-μ. When a primary and secondary join the queue simultaneously, the server workload is increased by Erlang $(2, \mu)$, *i.e.*, the sum of two independent exponential-μ's. All secondaries join the queue simultaneously with (next) primaries. The number of secondaries being delayed in the system at any instant is either 0 or 1.

Define the state of the system as $\{W(t),\ J(t)\}_{t\geq 0}$ where $W(t)$ = server workload at time $t \geq 0$, $J(t)=0$ if *zero* secondaries are being delayed, $J(t)=1$ if *one* secondary is being delayed. The state with zero customers in the system is denoted by $\{0,0\}$. The state when the server is idle and one secondary is being delayed is denoted by $\{0,1\}$. Let E_1 denote an exponential-μ random variable, and E_2 denote an Erlang$(2,\mu)$ random variable.

2.1. Sample Path of $\{W(t),J(t)\}_{t\geq 0}$

Technique of "Lines and Sheets (or Pages)" We utilize a technique of "lines and sheets (or pages)" to picture the state space and a sample path in it (e.g., see Brill [2] Section 4.6). This technique partitions the state space into mutually exclusive and exhaustive physical lines and sheets corresponding to the states of $\{W(t),J(t)\}_{t\geq 0}$. We select an arbitrary continuous subset in each sheet, having one boundary as a fixed level x in the state space of $W(t)$, e.g., (x,∞), $x>0$. We use this concrete physical picture as a guide to balance the sample-path exit and entrance rates of the selected state-space subsets. Simple level-crossing theorems (e.g., Brill [2]) guarantee that the partial steady-state pdf of $W(t)=x$ for each sheet is a unique term, or linear factor in a term, of the corresponding balance equation. The balance equations are generally Volterra integral equations of the second kind with parameter. Thus there is an isomorphism between the physical sample path structure and the model equations.

Consider the sample path of $\{W(t),J(t)\}_{t\geq 0}$ in **Figure 1**. All jumps due to an arrival on page 0 are distributed as E_1 because only primaries join the queue when arrivals find zero delayed secondaries present. All jumps due to an arrival on page 1 are distributed as E_2 because both the delayed secondary and the arriving primary join the queue simultaneously.

2.2. Description of the Sample Path of $\{W(t),J(t)\}_{t\geq 0}$ in Figure 1

At time 0 the system is empty (state $\{0,0\}$). A single arrives and the SP (sample path) jumps to level E_1 on page 0; the arrival immediately starts service. The SP decreases at rate 1 $\left(\mathrm{d}W(t)/\mathrm{d}t=-1\right)$. A pair arrives, the primary joins the queue and the secondary is delayed; the SP jumps E_1 since the server workload includes only those customers in the queue, and transits to page 1 (delayed secondary present).

A single arrives and pushes the delayed secondary to join the queue; the single joins just after it. The SP jumps E_2 and transits to page 0 (no delayed secondary present).

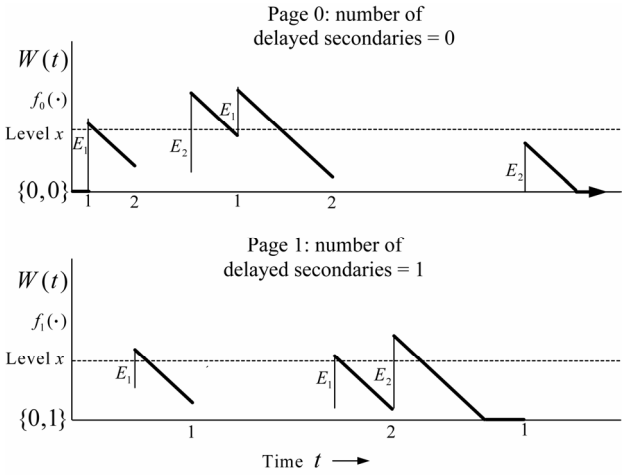

Figure 1. A sample path of $\{W(t),J(t)\}_{t\geq 0}$.

A single arrives and joins the queue; the SP jumps E_1 and remains on page 0. A pair arrives; the primary joins the queue and the split-off secondary becomes delayed. The SP jumps E_1 and transits to page 1. A pair arrives, the primary pushes the delayed secondary into the queue and joins just after it; the new secondary becomes delayed. The SP jumps E_2 and remains on page 1. The SP hits level 0 from above, and remains in state $\{0,1\}$ for a time distributed as exponentia-λ (server idle). A single arrives, pushes the delayed secondary into the queue and follows just after it. The SP jumps by E_2 and transits to page 0. Finally the SP hits level 0 from above and enters $\{0,0\}$ (system empty).

Figure 1 illustrates a cycle starting and ending in state $\{0,0\}$. A cycle starting and ending in state $\{0,1\}$ would be produced similarly.

3. Equations for Probability Distribution of Server Workload

Let $P_{0,0}=P\left(\{W(t),J(t)\}_{t\to\infty}=\{0,0\}\right)$,

$P_{0,1}=P\left(\{W(t),J(t)\}_{t\to\infty}=\{0,1\}\right)$,

$\overline{E_2}(x)=P(E_2>x)=\mathrm{e}^{-\mu x}(1+\mu x), x>0$,

$f_j(x), x>0$ = mixed joint steady-state pdf of $\{W(t),J(t)\}_{t\to\infty}$, $j=0,1$.

Guided by the sample path (**Figure 1**) we write a system of balance equations for the singleton states $\{0,0\}$ and $\{0,1\}$, and for sets of continuous states $\{(x,\infty),j\}$, $j=0,1$. Explanations of the equations are given immediately after Equation (5).

State $\{0,0\}$

$$\lambda P_{0,0}=f_0\left(0^+\right) \tag{1}$$

State $\{0,1\}$

$$\lambda P_{0,1} = f_1\left(0^+\right) \tag{2}$$

States $\{(x,\infty),0\}$

$$\begin{aligned} & f_0(x) + \lambda_2 \int_x^\infty f_0(y)\,\mathrm{d}y \\ &= \lambda_1 P_{0,0} \mathrm{e}^{-\mu x} + \lambda_1 \int_0^x \mathrm{e}^{-\mu(x-y)} f_0(y)\,\mathrm{d}y \\ &+ \lambda_1 \int_x^\infty f_1(y)\,\mathrm{d}y + \lambda_1 \int_0^x \overline{E_2}(x-y)\, f_1(y)\,\mathrm{d}y \\ &+ \lambda_1 P_{0,1} \overline{E_2}(x). \end{aligned} \tag{3}$$

States $\{(x,\infty),1\}$

$$\begin{aligned} & f_1(x) + \lambda_1 \int_x^\infty f_1(y)\,\mathrm{d}y \\ &= \lambda_2 P_{0,0} \mathrm{e}^{-\mu x} + \lambda_2 \int_0^x \overline{E_2}(x-y)\, f_1(y)\,\mathrm{d}y \\ &+ \lambda_2 \int_x^\infty f_0(y)\,\mathrm{d}y + \lambda_2 \int_0^x \mathrm{e}^{-\mu(x-y)} f_0(y)\,\mathrm{d}y \\ &+ \lambda_2 P_{0,1} \overline{E_2}(x) \end{aligned} \tag{4}$$

Total Probability = 1

$$P_{0,0} + P_{0,1} + \int_0^\infty f_0(x)\,\mathrm{d}x + \int_0^\infty f_1(x)\,\mathrm{d}x = 1. \tag{5}$$

Explanation of Equations (1)-(5)

Equation (1): Left side = (*exit rate of* $\{0,0\}$). Right side = (*downcrossing rate of level* 0 *on page* 0) = (*rate at which system is emptied*).

Equation (2): Left side = (*exit rate of* $\{0,1\}$). Right side = (*downcrossing rate of level* 0 *on page* 1).

Equation (3): Left side = (*exit rate of* $\{(x,\infty),0\}$) due to: i) (*downcrossings of level* x *on page* 0) + ii) (*arrivals of pairs*). Right side = (*entrance rate of* $\{(x,\infty),0\}$) due to: i) (*singles arriving to an empty system bringing a workload* $>x$) + ii) (*singles arriving when the state is* $(y,0), y\in(0,x)$ *bringing a workload* $>(x-y)$, *summed for* $y\in(0,x)$) + iii) (*singles arriving when the state is in* $\{(x,\infty),1\}$) + iv) (*singles arriving when the state is* $(y,1), y\in(0,x)$ *adding a workload* $>(x-y)$ *summed on* $y\in(0,x)$) + v) (*singles arriving when the state is* $\{0,1\}$ *adding a workload* $>x$).

Equation (4): Similar reasoning as for Equation (3).

Equation (5): Sum of probabilities = 1.

Related Quantities

We now consider related quantities obtainable from the solution of Equations (1)-(5): $P_{0,0}$ = proportion of time the system is empty; $P_{0,1}$ = proportion of time the server is idle and a delayed secondary is present;

$F_0 \equiv \int_0^\infty f_0(x)\,\mathrm{d}x$ = proportion of time the server is busy and no delayed secondary is present; $F_1 \equiv \int_0^\infty f_1(x)\,\mathrm{d}x$ = proportion of time the server is busy and a delayed secondary is present.

Let $C_{0,0}$ = time between successive $\{W(t), J(t)\}_{t\ge 0}$-entrances into state $\{0,0\}$ (system becomes empty). Let $C_{0,1}$ = time between successive $\{W(t), J(t)\}_{t\ge 0}$-entrances into state $\{0,1\}$ (server becomes idle and a delayed secondary is present). Then $C_{0,0}$ and $C_{0,1}$ are regenerative cycles of regenerative processes. From the elementary renewal theorem and (1) and (2)

$$\begin{aligned} E\left(C_{0,j}\right) &= 1/(\text{downcrossing rate of level 0 on page } j) \\ &= 1/f_j\left(0^+\right) = 1/\left(\lambda P_{0,j}\right),\ j = 0,1. \end{aligned} \tag{6}$$

Let $B_{0,j}$ be the total time that the server is busy during a cycle $C_{0,j}$. (Note that $B_{0,j}$ is composed generally of non-contiguous time segments.) By the theory of regenerative processes (e.g., Cohen [3]) and (6)

$$\begin{aligned} E\left(B_{0,j}\right)/E\left(C_{0,j}\right) &= P(\text{server is busy}) = F_0 + F_1, \\ E\left(B_{0,j}\right) &= \left(F_0 + F_1\right)/\left(\lambda P_{0,j}\right),\ j = 0,1. \end{aligned} \tag{7}$$

If we define a "busy period" as the proportion of $C_{0,0}$ such that there is at least one customer in the system as in Hlynka [1] (which includes the time that the server is idle while a delayed secondary is present), then

$$E(\text{busy period}) = \left(1 - P_{0,0}\right)/\left(\lambda P_{0,0}\right). \tag{8}$$

4. Relevant Constants and Numerical Example of Steady-State PDF of Workload

4.1. Relevant Constants

Before solving for $f_j(x), j = 0,1$ and $f(x) = f_0(x) + f_1(x), x > 0$, we solve six linearly independent equations for the constants $f_j\left(0^+\right), F_j, P_{0,j}, j = 0,1$. The first two equations are (1) and (2). The second two equations are obtained by letting $x \downarrow 0$ in (3) and (4). The fifth equation is (5).

The sixth equation is obtained by considering the system from the server's point of view. The arrival rate to the queue is $1\cdot\lambda\left(P_{0,0} + F_0\right) + 2\cdot\lambda\left(P_{0,1} + F_1\right)$ since only the arriving primary joins when zero delayed secondaries are present, and two join (arriving primary + present delayed secondary) when one secondary is pre-

sent. The steady-state probability that the server is idle is $1-$ (long-run traffic intensity) = $1-\lambda\left(P_{0,0}+F_0+2\left(P_{0,1}+F_1\right)\right)/\mu$. The server does not distinguish between idle periods when a delayed secondary is present or is not present. Thus P (server is idle) $=P_{0,0}+P_{0,1}$. This yields the sixth equation

$$P_{0,0}+P_{0,1}=1-\lambda\left(P_{0,0}+F_0+2\left(P_{0,1}+F_1\right)\right)/\mu. \quad (9)$$

Solving the six equations gives the four quantities (as well as $f_j\left(0^+\right), j=0,1$):

$$\left.\begin{aligned} P_{0,0}&=\left(\lambda_1\left(\mu-\lambda_1-2\lambda_2\right)\right)/\left(\mu\lambda\right);\\ P_{0,1}&=\left(\lambda_2\left(\mu-\lambda_1-2\lambda_2\right)\right)/\left(\mu\lambda\right);\\ F_0&=\left(\lambda_1\left(\lambda_1+2\lambda_2\right)\right)/\left(\mu\lambda\right);\\ F_1&=\left(\lambda_2\left(\lambda_1+2\lambda_2\right)\right)/\left(\mu\lambda\right). \end{aligned}\right\} \quad (10)$$

From (7)

$$\left.\begin{aligned} E\left(B_{0,0}\right)&=\left(\lambda_1+2\lambda_2\right)/\left(\lambda_1\left(\mu-\lambda_1-2\lambda_2\right)\right);\\ E\left(B_{0,1}\right)&=\left(\lambda_1+2\lambda_2\right)/\left(\lambda_2\left(\mu-\lambda_1-2\lambda_2\right)\right). \end{aligned}\right\} \quad (11)$$

From (8)

$$\begin{aligned} &E\left(\text{busy period}\right)\\ &=\left(\lambda_1^2+2\lambda_2\lambda_1+\lambda_2\mu\right)/\left(\lambda\lambda_1\left(\mu-\lambda_1-2\lambda_2\right)\right). \end{aligned} \quad (12)$$

At least one of $P_{0,j}, j=0,1$ must be positive for a steady state distribution to exist. This implies that $\lambda_1+2\lambda_2<\mu$ is the condition for stability, which agrees with intuition.

Note that if $\lambda_2=0$ the system behaves like a standard $M_{\lambda_1}/M_\mu/1$ queue. If $\lambda_1=0$ it behaves like a standard $M_{\lambda_2}/E_{2,\mu}/1$ queue with regard to the workload.

4.2. Numerical Example of Steady-State PDF of Server Workload

We obtain formulas for $f_j\left(x\right), j=0,1$ and $f\left(x\right), x>0$ by computing Laplace transforms and inverting them, with the aid of Maple software. Let $\tilde{f}_j\left(s\right)$ denote the Laplace transform of $f_j\left(x\right), j=0,1$. We take Laplace transforms on both sides of (3) and (4) and solve two linear equations for $\tilde{f}_j\left(s\right), j=0,1$. Taking the inverse Laplace transforms and adding gives the steady-state pdf $f\left(x\right)=f_0\left(x\right)+f_1\left(x\right)$. The analytical expressions obtained are lengthy, and are not shown here.

We present a numerical example with arbitrary parameter values $\lambda_1=1.5, \lambda_2=0.9, \mu=3.8$, which illustrates typical results.

The steady-state pdf of workload is (see **Figure 2**)

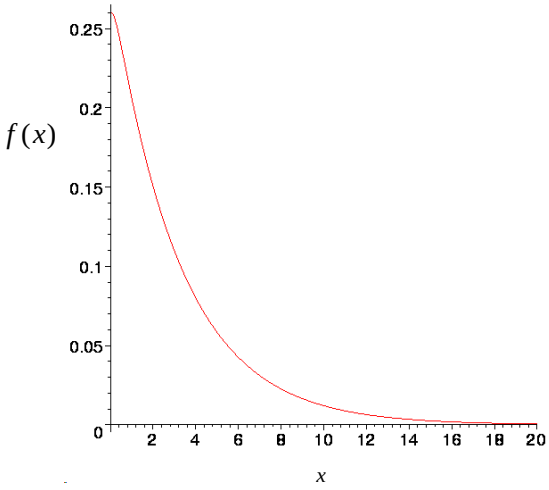

Figure 2. Plot of steady-state pdf f(x) using the parameter values in Subsection 4.2.

$$\begin{aligned} f\left(x\right)&=0.2834250634\times e^{-.315967x}\\ &\quad -0.02396560398\times e^{-4.684033x}, \quad (13)\\ &\quad x>0. \end{aligned}$$

Also

$$\begin{aligned} &P_{0,0}=0.06756756758; P_{0,1}=0.04054054054;\\ &F_0=0.5574324325; F_1=0.3344594595;\\ &E\left(C_{0,0}\right)=6.166667; E\left(C_{0,1}\right)=10.277778,\\ &E\left(B_{0,0}\right)=5.5; E\left(B_{0,1}\right)=9.166667;\\ &E\left(\text{busy period}\right)=5.75. \end{aligned}$$

Note that $\int_0^\infty f\left(x\right)\mathrm{d}x=F_0+F_1=0.89189189$;

$f\left(0^+\right)=\lambda\left(P_{0,0}+P_{0,1}\right)=0.2594594594$;

$P_{0,0}+P_{0,1}+\int_0^\infty f\left(x\right)\mathrm{d}x=1.$

5. Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] M. Hlynka, "An M/M/1 Queue with Bulk Arrivals and Delays," *Canadian Operational Research Society Conference Presentation*, Niagara Falls, June 2012.

[2] P. H. Brill, "Level Crossing Methods in Stochastic Models," Springer, New York, 2008. doi:10.1007/978-0-387-09421-2

[3] J. W. Cohen, "On Regenerative Processes in Queueing Theory," *Lecture Notes in Economics and Mathematical Systems*, Spring-Verlag, New York, 1976.

