INFINITE-SERVER QUEUES WITH SYSTEM'S ADDITIONAL TASKS AND IMPATIENT CUSTOMERS

Eitan Altman and Uri Yechiali

Probability in the Engineering and Informational Sciences / Volume 22 / Issue 04 / October 2008, pp 477 - 493
DOI: 10.1017/S0269964808000296, Published online: 25 September 2008

Probability in the Engineering and Informational Sciences, **22**, 2008, 477–493. Printed in the U.S.A.
doi:10.1017/S0269964808000296

INFINITE-SERVER QUEUES WITH SYSTEM'S ADDITIONAL TASKS AND IMPATIENT CUSTOMERS

Eitan Altman
INRIA
Sophia Antipolis, France
E-mail: altman@sophia.inria.fr

Uri Yechiali
Department of Statistics and Operations Research
School of Mathematical Sciences
Tel Aviv University, Tel Aviv 69978, Israel
E-mail: uriy@post.tau.ac.il

A system is operating as an $M/M/\infty$ queue. However, when it becomes empty, it is assigned to perform another task, the duration U of which is random. Customers arriving while the system is unavailable for service (i.e., occupied with a U-task) become impatient: Each individual activates an "impatience timer" having random duration T such that if the system does not become available by the time the timer expires, the customer leaves the system never to return. When the system completes a U-task and there are waiting customers, each one is taken immediately into service. We analyze both multiple and single U-task scenarios and consider both exponentially and generally distributed task and impatience times. We derive the (partial) probability generating functions of the number of customers present when the system is occupied with a U-task as well as when it acts as an $M/M/\infty$ queue and we obtain explicit expressions for the corresponding mean queue sizes. We further calculate the mean length of a busy period, the mean cycle time, and the quality of service measure: proportion of customers being served.

1. INTRODUCTION

Impatience of customers as well as server vacations are important features involved in the analysis of queuing models. Indeed, impatience of customers has lately been

 0269-9648/08 $25.00

studied in Bonald and Roberts [3] in order to describe networks' behavior in overload conditions. Queues with servers' vacations and their nice decomposition properties (cf. Levy and Yechiali [9,10], Fuhrmann [6], Fuhrmann and Cooper [7], Doshi [5], Takagi [16], Shomrony and Yechiali [13], Yechiali [17]) have been studied extensively in the literature and used for modeling various applications, including local area networks that are based on a Token Passing Ring (cf. Altman and Kofman [1]). Recently, it has been used for describing WLANs (see, e.g., Zussman, Segall, and Yechiali [18] and Zussman, Yechiali, and Segall [19] for vacations analysis in Bluetooth).

The common approach to studying impatience assigns a (possibly random) timer to each customer that arrives (cf. Takacs [15]). If the customer has not terminated service by the time its timer expires, then it abandons the queue (see, e.g., recent work on call centers by Gans, Koole, and Mandelbaum [8]), or it is lost. In contrast, we assume here that reneging occurs only when the system is unavailable to render service (i.e., it is "on vacation"), possibly working on another task. For example, one can observe such behavior of customers in various real-world public service systems. If, upon arrival, new customers find closed or unattended windows, each individual tends to wait for only a (random) limited time. If none of the closed windows is reopened within this limited time, the customer abbandons the system. This type of impatience has been recently introduced by Altman and Yechiali [2] and analyzed for the cases of $M/M/1$, $M/G/1$, and $M/M/c$ queues. Here, we extend the analysis further and study this impatience phenomenon in an infinite-server queuing setting.

Our analysis approach is based on the use of probability generating functions (PGFs), which quite often allows one to analyze queuing systems by transforming difference equations (which represent the balance equations for the steady-state probabilities) into algebraic ones. However, in the present work, the PGFs obtained from the balance equations transform into a pair of differential equations rather than into an algebraic set. By solving these equations, we are able to determine various important performance measures specified below.

We consider various models that differ mainly according to the vacation pattern: either a single-vacation or a multiple-vacation scenario. In the first case, if no customers are present at the end of a task, the system stays ready for the first customer to arrive in order to start a new busy period; however, if customers are present, working starts immediately. In the second case (multiple tasks), the system keeps taking vacations until it finds at least one waiting customer upon completion of a task. At this point, a new busy period starts. For each of these procedures we consider both exponentially and generally distributed customers' impatience T times and system's U-tasks. For each of the resulting four models, we derive several performance measures, such as the PGF of the stationary probabilities, the expected cycle time, the expected number of customers in the system, and the important "quality of service" performance measure: fraction of customers getting served without abandoning the system.

The article's sections are the following: Sections 2 and 3 deal with the multiple-task operating scheme. The former treats the case in which both T and U are exponentially distributed, whereas the latter considers the case in which each random variable has a generally distributed probability distribution function (PDF). Sections 4 and 5 study the single-task scenario, where in the former we assume exponentially distributed T and U, and in the latter, we let T and U have general PDFs. Section 6 concludes the work.

2. MULTIPLE TASKS, EXPONENTIALLY DISTRIBUTED TASKS, AND EXPONENTIALLY DISTRIBUTED IMPATIENCE TIMES

2.1. The Model

A system is operating as an $M/M/\infty$ queue with Poisson arrival rate λ and with exponentially distributed service times having mean $1/\mu$. However, when the system becomes empty, it is assigned to perform another (independent) task whose random duration U is exponentially distributed with mean $1/\gamma$.

Customers arriving while the system is unavailable to serve them (i.e., being occupied with a U-task) become impatient: Each individual customer activates an "impatience timer" T, exponentially distributed with parameter ξ, which is independent of the number of waiting jobs at that moment. If the system completes its U-task before the time T expires, the customer is immediately taken into service and leaves the system upon his service completion. If, however, T expires before the system becomes available, our customer abandons the queue never to return.

Multiple Tasks: If the system completes a U-task and no customers are waiting, it is assigned a new U-task, independent of the previous U-tasks. If there are customers waiting at the end of a U-task, the system starts a busy period, whose duration depends on the number of customers still waiting.

2.2. Balance Equations

Let L denote the total number of customers in the system and let $J = 1$ if the system is operating regularly or $J = 0$ if it is "unavailable," being occupied with a U-task. Then the pair (J,L) defines a continuous-time Markov process with transition rate diagram as depicted in Figure 1.

Let $P_{jn} = P\{J = j, L = n\}(j = 0,1,; n = 0,1,2,\ldots)$ denote the system-state probabilities. Then, the set of balance equations is given by the following:

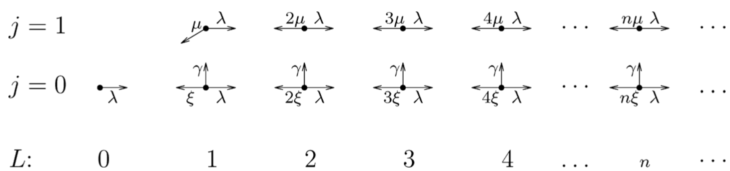

FIGURE 1. Transition rate diagram for the multiple-task scenario.

For $j = 0$:

$$\begin{cases} n = 0, & \lambda P_{00} = \xi P_{01} + \mu P_{11}, \\ n \geq 1, & (\lambda + n\xi + \gamma) P_{0n} = \lambda P_{0,n-1} + (n+1)\xi P_{0,n+1}; \end{cases} \tag{1}$$

for $j = 1$:

$$\begin{cases} n = 1, & (\lambda + \mu) P_{11} = 2\mu P_{12} + \gamma P_{01}, \\ n \geq 2, & (\lambda + n\mu) P_{1n} = \lambda P_{1,n-1} + (n+1)\mu P_{1,n+1} + \gamma P_{0n}. \end{cases} \tag{2}$$

Let $P_{10} = 0$ and define the (partial) PGFs

$$G_0(z) = \sum_{n=0}^{\infty} P_{0n} z^n, \qquad G_1(z) = \sum_{n=0}^{\infty} P_{1n} z^n.$$

Then by multiplying each equation by z^n and summing over n, we get from (1) the relation

$$\xi(1-z)G_0'(z) = [\lambda(1-z) + \gamma]G_0(z) - A, \tag{3}$$

where $A = \gamma P_{00} + \mu P_{11}$ and $G_j'(z) = dG_j(z)/dz,\ j = 0, 1$. Similarly, using (2), we obtain

$$\lambda(1-z)G_1(z) + \mu z G_1'(z) = \mu[G_1'(z) - P_{11}] + \gamma[G_0(z) - P_{00}]. \tag{4}$$

2.3. Solution of the Differential Equations

Equation (3) is similar to Eq. (2.4) in Altman and Yechiali [2] and its solution is given by

$$G_0(z) = P_{00} e(\lambda/\xi)^z \frac{\int_{s=z}^{1} (1-s)^{(\gamma/\xi)-1} e^{-(\lambda/\xi)s}\, ds}{K(1-z)^{\gamma/\xi}}, \tag{5}$$

where

$$K = \int_0^1 (1-s)^{(\gamma/\xi)-1} e^{-(\lambda/\xi)s}\, ds. \tag{6}$$

Also (see Eq. (2.11) in Altman and Yechiali [2]),

$$P_{00} = \frac{A}{\xi} K = \frac{\gamma P_{00} + \mu P_{11}}{\xi} K = \frac{\mu K}{\xi - \gamma K} P_{11}. \tag{7}$$

Define $P_{0\bullet} = \sum_{n=0}^{\infty} P_{0n}$ and $P_{1\bullet} = \sum_{n=1}^{\infty} P_{1n}$. Then, by applying L'Hospital's rule to (5) and using (7), we get

$$G_0(1) = P_{0\bullet} = \frac{\xi}{\gamma K} P_{00} = \frac{A}{\gamma}. \tag{8}$$

Now, (4) can be written as

$$G_1'(z) - \frac{\lambda}{\mu}G_1(z) = \frac{1}{\mu(1-z)}[-\gamma G_0(z) + A]. \tag{9}$$

Multiplying both sides of (9) by $e^{-(\lambda/\mu)z}$, we can write

$$\frac{d}{dz}\left[e^{(-\lambda/\mu)z}G_1(z)\right] = e^{(-\lambda/\mu)z}\frac{1}{\mu(1-z)}[-\gamma G_0(z) + A]. \tag{10}$$

Integrating, we have

$$e^{-(\lambda/\mu)z}G_1(z) - G_1(0) = \frac{1}{\mu}\int_{s=0}^{z}(1-s)^{-1}e^{-(\lambda/\mu)s}[-\gamma G_0(s) + A]\,ds. \tag{11}$$

Thus, finally, since $G_1(0) = P_{10} = 0$,

$$G_1(z) = \frac{1}{\mu}e^{(\lambda/\mu)z}\int_{s=0}^{z}(1-s)^{-1}e^{(-\lambda/\mu)s}[-\gamma G_0(s) + A]\,ds, \tag{12}$$

where $G_0(\cdot)$ is given by (5).

In order to solve completely for $G_0(z)$ [and $G_1(z)$], we need to calculate the value of P_{00}.

2.4. Calculation of P_{00}, $P_{0\bullet}$, and P_{11}

From (8), we have $P_{0\bullet} = (\xi/(\gamma K)P_{00}$, and (7) reads $P_{00} = [\mu K/(\xi - \gamma K)]P_{11}$. Using (12), we get, for $z = 1$,

$$G_1(1) = P_{1\bullet} = 1 - P_{0\bullet} = \frac{e^{(\lambda/\mu)}}{\mu}\int_{s=0}^{1}(1-s)^{-1}e^{-(\lambda/\mu)s}[-\gamma G_0(s) + A]\,ds. \tag{13}$$

Now, $G_0(s)$ is a function of P_{00} and $A = \gamma P_{00} + \mu P_{11}$. Hence, (7), (8), and (13) comprise a set of three independent equations in the three unknowns P_{00}, P_{11}, and $P_{0\bullet}$.

2.5. Calculation of $E[L_0]$ and $E[L_1]$

Define $G_j'(1) = E[L_j] = \sum_{n=0}^{\infty} nP_{jn}$, for $j = 0, 1$. Then, from (9) (since $\gamma G_0(1) = A$),

$$G_1'(1) = \frac{1}{\mu}\left[\lambda P_{1\bullet} + \frac{\lim_{z\to 1}\{d[-\gamma G_0(z) + A]/dz\}}{\lim_{z\to 1}\{d(1-z)/dz\}}\right];$$

that is,

$$E[L_1] = \frac{1}{\mu}[\lambda P_{1\bullet} + \gamma E[L_0]]. \tag{14}$$

Indeed, (14) written as

$$\mu E[L_1] = \lambda P_{1\bullet} + \gamma E[L_0]$$

states that the rate of arrivals to "level" $J = 1$ (viz. $\lambda P_{1\bullet} + \gamma E[L_0]$) is equal to the rate of departures from that level (i.e., $\mu E[L_1]$). Now, writing (3) as

$$G_0'(z) = \frac{[\lambda(1-z)+\gamma]G_0(z) - A}{\xi(1-z)},$$

we derive

$$G_0'(1) = E[L_0] = \frac{-G_0(1) + \gamma G_0'(1)}{-\xi} = \frac{-\lambda P_{0\bullet} + \gamma E[L_0]}{-\xi};$$

that is,

$$E[L_0] = \frac{\lambda}{\gamma+\xi} P_{0\bullet} \tag{15}$$

Again, the rate at which customers enter "level" $J = 0$ (viz. $\lambda P_{0\bullet}$) is equal to the rate of customers leaving that state, which is $(\gamma + \xi)E[L_0]$.

Substituting (15) in (14) and using $1 = P_{0\bullet} + P_{1\bullet}$ yields

$$E[L_1] = \frac{\lambda}{\mu}\left(1 - \frac{\xi}{\gamma+\xi} P_{0\bullet}\right). \tag{16}$$

3. MULTIPLE TASKS, GENERALLY DISTRIBUTED TASKS, AND GENERALLY DISTRIBUTED IMPATIENCE TIMES

3.1. The Model

The underlying process is, as earlier, the $M/M/\infty$ queue. The system is assigned a new U-task at the end of a busy period, or whenever it completes a task but no customers are present. U-tasks are generally distributed with mean $E[U]$, second moment $E[U^2]$, and Laplace–Stieltjes transform (LST) $U^*(s) = E[e^{-sU}]$. Impatience times of customers are independent and identically distributed (i.i.d.) random variables, all distributed as T, having mean $E[T]$, second moment $E[T^2]$, and LST $T^*(s)$. When the system is unavailable (level $J = 0$) due to a U-task, the time evolution of the queue size is that of an $M/G/\infty$ queue with arrival rate λ and service times distributed as T. It is well known (Takacs [14]) that for time $t \le U$, the number of customers present is distributed as a Poisson random variable with parameter

$$\Lambda(t) = \lambda \int_0^t [1 - P(T \le y)]\,dy, \qquad t \le U. \tag{17}$$

Let τ denote the duration of a *consecutive* series of U-tasks (from the end of a busy period until the start of the next one). Then it was shown in Altman and Yechiali [2] that the LST of τ, $\tau^*(s) := E\left[e^{-s\tau}\right]$, is given by

$$\tau^*(s) = \frac{U^*(s) - E_U\left[e^{-s(U+\Lambda(U))}\right]}{1 - E_U\left[e^{-s(U+\Lambda(U))}\right]}, \tag{18}$$

with mean

$$E[\tau] = \frac{E[U]}{1 - E_U\left[e^{-\Lambda(U)}\right]}. \tag{19}$$

Let N be the number of customers at the start of a busy period. Then (see Altman and Yechiali [2])

$$P(N = n) = \frac{(1/n!)E_U\left[e^{-\Lambda(U)}(\Lambda(U))^n\right]}{1 - E_U\left[e^{-\Lambda(U)}\right]} \qquad (n = 1, 2, 3, \ldots), \tag{20}$$

$$E[N] = \frac{E[\Lambda(U)]}{1 - E_U\left[e^{-\Lambda(U)}\right]}. \tag{21}$$

3.2. The Busy Period

Denote by $\overline{\Gamma}_n$ the mean of a busy period Γ_n starting with n customers in an $M(\lambda)/M(\mu)/\infty$ queue. We have, for $n \geq 1$,

$$\overline{\Gamma}_n = \frac{1}{\lambda + n\mu} + \frac{\lambda}{\lambda + n\mu}\overline{\Gamma}_{n+1} + \frac{n\mu}{\lambda + n\mu}\overline{\Gamma}_{n-1}, \tag{22}$$

where $\overline{\Gamma}_0 = 0$. Define $\overline{\delta}_n := \overline{\Gamma}_n - \overline{\Gamma}_{n-1}$. Then

$$\begin{aligned}\overline{\delta}_n &= \frac{1}{\lambda + n\mu} + \frac{\lambda}{\lambda + n\mu}\overline{\Gamma}_{n+1} - \frac{\lambda}{\lambda + n\mu}\overline{\Gamma}_{n-1} \\ &= \frac{1}{\lambda + n\mu} + \frac{\lambda}{\lambda + n\mu}\left(\overline{\delta}_{n+1} + \overline{\delta}_n\right).\end{aligned}$$

We thus obtain

$$\overline{\delta}_{n+1} = \frac{n\mu}{\lambda}\overline{\delta}_n - \frac{1}{\lambda}.$$

Iterating with $\overline{\delta}_n$ and $\overline{\Gamma}_n$, the solution is

$$\overline{\delta}_n = (n-1)!\left(\frac{\mu}{\lambda}\right)^{n-1}\overline{\Gamma}_1 - \frac{1}{\lambda}\sum_{i=1}^{n-1}\frac{(n-1)!}{i!}\left(\frac{\mu}{\lambda}\right)^{n-1-i}. \tag{23}$$

Now, $\overline{\Gamma}_m = \sum_{n=1}^{m} \overline{\delta}_n$. Thus,

$$\overline{\Gamma}_m = \sum_{n=1}^{m} \left[(n-1)! \left(\frac{\mu}{\lambda}\right)^{n-1} \overline{\Gamma}_1 - \frac{1}{\lambda} \sum_{i=1}^{n-1} \frac{(n-1)!}{i!} \left(\frac{\mu}{\lambda}\right)^{n-1-i} \right]. \tag{24}$$

Hence, the mean duration of an arbitrary busy period Γ is given by

$$\begin{aligned} E[\Gamma] &= \sum_{m=1}^{\infty} P(N=m)\overline{\Gamma}_m \\ &= \sum_{m=1}^{\infty} \frac{E_U\left[e^{-\Lambda(U)}(\Lambda(U))^m\right]}{m!\left(1 - E_U\left[e^{-\Lambda(U)}\right]\right)} \\ &\quad \times \left(\sum_{n=1}^{m} (n-1)! \left[\left(\frac{\mu}{\lambda}\right)^{n-1} \overline{\Gamma}_1 - \frac{1}{\lambda} \sum_{i=1}^{n-1} \frac{1}{i!} \left(\frac{\mu}{\lambda}\right)^{n-1-i} \right] \right). \end{aligned} \tag{25}$$

The LST of the cycle time in an $M/M/\infty$ queue is given by Liu and Shi [11, p. 828] as

$$C^*_{M/M/\infty}(s) = 1 - \frac{1}{s+\lambda} \left(\sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!} \frac{1}{s+n\mu} \right)^{-1}, \tag{26}$$

whereas the mean length of an ordinary busy period (starting with a single customer) in an $M/G/\infty$ queue is given by (see Browne and Steele [4] and Miorandi and Altman [12])

$$\overline{\Gamma}_1 = \frac{1}{\lambda}\Big(\exp(\lambda E[B]) - 1 \Big), \tag{27}$$

where B denotes the service time of a single customer. For the $M/M/\infty$ queue, $E[B] = 1/\mu$.

3.3. Calculation of P_{00}

The proportion of time the system is empty, denoted as P_{00}, is calculated by

$$P_{00} = \frac{E[D]}{E[\Gamma] + E[\tau]}, \tag{28}$$

where D is the total time during τ in which there are no customers in the system. $E[D]$ was calculated in Altman and Yechiali [2] as

$$E[D] = \frac{E_U\left[\int_0^U e^{-\Lambda(t)}\, dt\right]}{1 - E_U[e^{-\Lambda(U)}]}. \tag{29}$$

Combining (19), (25), and (29), P_{00} can now be calculated.

3.4. Proportion of Customers Served

If $N = m \geq 1$, the expected number of customers served during the immediate following busy period is $m + \lambda\overline{\Gamma}_m$. Thus, the expected number of customers served during an arbitrary busy period Γ is calculated as

$$\sum_{m=1}^{\infty} P(N = m)[m + \lambda\overline{\Gamma}_m] = E[N] + \lambda E[\Gamma]. \tag{30}$$

Hence, the proportion of customers served without abandoning the system is given by

$$P(\text{served}) = \frac{E[N] + \lambda E[\Gamma]}{\lambda(E[\tau] + E[\Gamma])}, \tag{31}$$

where $E[\tau]$, $E[N]$, and $E[\Gamma]$ are given by (19), (21), and (25), respectively.

3.5. Back to the Fully Markovian Case of Section 2

In the case where U and T are exponentially distributed with parameters γ and ξ, respectively, we get

$$\begin{aligned} E_U\left[e^{-\Lambda(U)}\right] &= \int_{u=0}^{\infty} \left(e^{-\lambda \int_0^u e^{-\xi y}\, dy}\right) \gamma e^{-\gamma u}\, du \\ &= \int_{u=0}^{\infty} e^{-(\lambda/\xi)\left(1-e^{-\xi u}\right)} \gamma e^{-\gamma u}\, du. \end{aligned}$$

By letting $s = 1 - e^{-\xi u}$ we have $u = -(1/\xi)\ln(1-s)$, $du = ds/[\xi(1-s)]$, and $e^{-\gamma u} = (1-s)^{\gamma/s}$. Hence,

$$E_U\left[e^{-\Lambda(U)}\right] = \int_{s=0}^{1} e^{-(\lambda/\xi)s} \frac{\gamma}{\xi}(1-s)^{\gamma/\xi-1}\, ds = \frac{\gamma}{\xi}K. \tag{32}$$

Now, by letting $s = 1 - e^{-\xi t}$, we get

$$\begin{aligned} E_U\left[\int_0^U e^{-\Lambda(t)}\, dt\right] &= \int_{u=0}^{\infty} \left(\int_{t=0}^{u} e^{-(\lambda/\xi)(1-e^{-\xi t})}\, dt\right) \gamma e^{-\gamma u}\, du \\ &= \int_{t=0}^{\infty} e^{-(\lambda/\xi)(1-e^{-\xi t})} \left(\int_{u=t}^{\infty} \gamma e^{-\gamma u}\, du\right) dt \\ &= \int_{t=0}^{\infty} e^{-(\lambda/\xi)(1-e^{-\xi t})} e^{-\gamma t}\, dt \\ &= \int_{s=0}^{1} \frac{1}{\xi} e^{-(\lambda/\xi)s}(1-s)^{(\gamma/\xi)-1}\, ds \\ &= \frac{K}{\xi}. \end{aligned} \tag{33}$$

Thus, from (29),

$$E[D] = \frac{K/\xi}{1-(\gamma/\xi)K} = \frac{K}{\xi - \gamma K}. \tag{34}$$

Also, from (19),

$$E[\tau] = \frac{1/\gamma}{1-(\gamma/\xi)K} = \frac{\xi/\gamma}{\xi - \gamma K}. \tag{35}$$

Applying (28) we obtain, for the Markovian case,

$$P_{00} = \frac{K/(\xi-\gamma K)}{E[\Gamma] + (\xi/\gamma)/(\xi-\gamma K)} = \frac{\gamma K}{\gamma(\xi-\gamma K)E[\Gamma]+\xi}. \tag{36}$$

In order to get the corresponding expression for $E[N]$, we calculate

$$\begin{aligned} E[\Lambda(U)] &= \int_{u=0}^{\infty} \left(\lambda \int_{y=0}^{u} e^{-\xi y}\,dy\right) \gamma e^{-\gamma u}\,du \\ &= \int_{u=0}^{\infty} \frac{\lambda}{\xi}\left(1-e^{-\xi u}\right)\gamma e^{-\gamma u}\,du \\ &= \frac{\lambda}{\xi}\left(1-\frac{\gamma}{\xi+\gamma}\right) \\ &= \frac{\lambda}{\xi+\gamma}. \end{aligned} \tag{37}$$

Thus, from (21),

$$E[N] = \frac{\lambda/(\xi+\gamma)}{1-(\gamma/\xi)K} = \frac{\lambda\xi}{(\xi+\gamma)(\xi-\gamma K)} = \lambda\frac{\gamma}{\xi+\gamma}E[\tau], \tag{38}$$

where, for each individual customer, $\gamma/(\xi+\gamma)$ is the probability that the task U is completed before T expires, implying that the customer will not abandon and will be served. Thus, the expected number of customers present at the end of τ is

$$E[N] = \lambda P(\text{not abandoning})E[\tau].$$

Finally, substituting result (35) in (31) gives

$$\begin{aligned} P(\text{served}) &= \left(\frac{\lambda\xi}{(\xi+\gamma)(\xi-\gamma K)} + \lambda E[\Gamma]\right)\left[\lambda\left(\frac{\xi}{\gamma(\xi-\gamma K)} + E[\Gamma]\right)\right]^{-1} \\ &= \left(\frac{\xi}{\xi+\gamma} + (\xi-\gamma K)E[\Gamma]\right)\left(\frac{\xi}{\gamma} + (\xi-\gamma K)E[\Gamma]\right)^{-1}. \end{aligned} \tag{39}$$

Remark 3.1: We can use (26) to obtain an alternative expression for K. From [Liu and Shi [11] p. 828] we have

$$\int_0^\infty e^{-\gamma u} e^{-(\lambda/\xi)(1-e^{-\xi u})}\, du = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\xi}\right)^n \frac{1}{n!} \frac{1}{\gamma + n\xi}.$$

Using the derivation of (32), it follows that the left-hand side of the above equation equals K/ξ. Hence,

$$K = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\xi}\right)^n \frac{1}{n!} \frac{\xi}{\gamma + n\xi}.$$

In the special case when $\xi = \gamma$ (i.e., each individual customer is willing to wait, on the average, the mean duration of U), we get a simpler expression for K:

$$K = \frac{\xi}{\lambda}\left(e^{\lambda/\xi} - 1\right).$$

4. SINGLE TASK, EXPONENTIALLY DISTRIBUTED TASKS, AND EXPONENTIALLY DISTRIBUTED IMPATIENCE TIMES

4.1. The Model

We consider now the case where as soon as the system becomes empty of customers, it is assigned a *single* U-task, exponentially distributed with mean $1/\gamma$.

Customers arriving while the system is occupied with a U-task are impatient: Each individual customer, upon arrival, activates his impatience timer T, exponentially distributed with parameter ξ. When the system completes its U-task and there are $n \geq 1$ waiting customers, a busy period starts. However, if there are no customers waiting, the system stays *dormant* until the first customer arrives. This customer initiates a busy period, at the end of which the system is assigned a new U-task.

The transition-rate diagram is depicted in Figure 2.

4.2. Balance Equations and Generating Functions

As previously, let $P_{jn} = P\{J = j, L = n\}$ $(j = 0, 1; n = 0, 1, 2, \ldots)$ denote the system-state probabilities, where L denotes the total number of customers in the system and

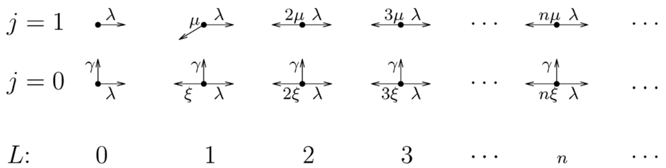

Figure 2. Transition rate diagram for the single-task scenario.

J indicates whether the system is occupied with a U-task ($J = 0$) or it is available for regular service ($J = 1$).

The set of balance equations is given by the following:
For $j = 0$:

$$\begin{cases} n = 0, & (\lambda + \gamma)P_{00} = \xi P_{01} + \mu P_{11}, \\ n \geq 1, & (\lambda + n\xi + \gamma)P_{0n} = \lambda P_{0,n-1} + (n+1)\xi P_{0,n+1}; \end{cases} \tag{40}$$

for $j = 1$:

$$\begin{cases} n = 0, & \lambda P_{10} = \gamma P_{00}, \\ n \geq 1, & (\lambda + n\mu)P_{1n} = \lambda P_{1,n-1} + (n+1)\mu P_{1,n+1} + \gamma P_{0n}. \end{cases} \tag{41}$$

Note that the difference between this model and the multiple U-tasks is the positive existence of the state $(1, 0)$.

Define the (partial) PGFs as

$$G_j(z) = \sum_{n=0}^{\infty} P_{jn} z^n \qquad (j = 0, 1).$$

Then, similar to the derivation of (3) and (4), we obtain

$$\xi(1 - z)G_0'(z) = [\lambda(1 - z) + \gamma]G_0(z) - \mu P_{11} \tag{42}$$

and

$$\lambda(1 - z)G_1(z) + \mu z G_1'(z) = \mu[G_1'(z) - P_{11}] + \gamma G_0(z). \tag{43}$$

The solution of the differential equation (DF) (42) is given, similar to (5), by

$$G_0(z) = P_{00} e^{(\lambda/\xi)z} \left(1 - \frac{\int_0^z (1 - s)^{(\gamma/\xi)-1} e^{-(\lambda/\xi)s}\, ds}{K}\right)(1 - z)^{-(\gamma/\xi)}, \tag{44}$$

where K is given in (6), and now

$$P_{00} = \frac{\mu P_{11}}{\xi} K. \tag{45}$$

Define $P_{j\bullet} = \sum_{n=0}^{\infty} P_{jn} \quad (j = 0, 1)$. Then, from (44), we have

$$G_0(1) = P_{0\bullet} = \frac{\xi}{\gamma K} P_{00} = \frac{\mu}{\gamma} P_{11}\,. \tag{46}$$

Note that, in the multiple-task case,

$$P_{0\bullet} = \frac{\xi}{\xi - \mu K} \frac{\mu}{\gamma} P_{11}\,.$$

Clearly, (46) can be directly obtained by considering a horizontal "cut" in Figure 2 between "level" $j = 0$ and "level" $j = 1$.

Similar to (12), the solution of the DF (43) is given by

$$G_1(z) = \frac{1}{\mu} e^{(\lambda/\mu)z} \int_{s=0}^{z} (1-s)^{-1} e^{-(\mu/\lambda)s} [-\gamma G_0(s) + \mu P_{11}] \, ds. \tag{47}$$

Also, similar to (13),

$$\begin{aligned} G_1(1) &= P_{1\bullet} \\ &= 1 - P_{0\bullet} \\ &= \frac{1}{\mu} e^{(\lambda/\mu)} \int_{0}^{1} (1-s)^{-1} e^{-(\mu/\lambda)s} [-\gamma G_0(s) + \mu P_{11}] \, ds. \end{aligned} \tag{48}$$

The set (45), (46), and (48) uniquely determines the values of P_{00}, P_{11}, and $P_{0\bullet}$.

Comparing the multiple- and single-task models, it is seen that

$$P_{00}(\text{single}) + P_{10}(\text{single}) = P_{00}(\text{multiple}). \tag{49}$$

The mean queue lengths $E[L_0]$ and $E[L_1]$ are readily calculated, similar to Section 2.5. Indeed, they have the same form-namely $E[L_0] = [\lambda/(\gamma+\xi)]P_{0\bullet}$ and $E[L_1] = \frac{\lambda}{\mu}(1 - [\xi/(\gamma+\xi)]P_{0\bullet})$. The difference from the multiple-task case is expressed by the different value of $P_{0\bullet}$.

In addition, the result for $E[L_1]$ can be explained as follows: The probability of abandonment by an arbitrary customer is $[\xi/(\gamma+\xi)]P_{0\bullet}$. Thus, the "effective" arrival rate is $\lambda_{\text{eff}} = \lambda\,(1 - [\xi/(\gamma+\xi)]P_{0\bullet})$. Now, the mean number of customers in an $M(\lambda_{\text{eff}})/M(\mu)/\infty$ queue is λ_{eff}/μ.

5. SINGLE TASK, GENERALLY DISTRIBUTED TASKS, AND GENERALLY DISTRIBUTED IMPATIENCE TIMES

5.1. The Model

Again, we study the case of a *single* U-task assignment after a completion of a busy period, but where U has a general probability distribution function, and impatience times T are generally distributed as well.

5.2. The Busy Period

Due to the impatience process, the number of customers, $N(U)$, present at the end of a single U-task is given by ($M/G/\infty$ queue)

$$P(N(U) = m) = e^{-\Lambda(U)} \frac{(\Lambda(U))^m}{m!} \qquad (m = 0, 1, 2, \ldots). \tag{50}$$

If $N(U) = 0$, the server stays dormant until the first arrival, when a regular $M/M/\infty$ busy period is initiated. If $N(U) = m \geq 1$, the server starts serving immediately, where the mean duration of the resulting busy period is $\overline{\Gamma}_m$.

Thus, for the single-task (ST) scenario, the mean length of an arbitrary busy period Γ_{ST} is obtained by

$$E[\Gamma_{ST}] = E_U[P(N(U)=0)]\overline{\Gamma}_1 + \sum_{m=1}^{\infty} E_U[P(N(U)=m)]\overline{\Gamma}_m$$
$$= E_U\left[e^{-\Lambda(U)}\right]\overline{\Gamma}_1 + \sum_{m=1}^{\infty} \frac{1}{m!} E_U\left[e^{-\Lambda(U)}(\Lambda(U))^m\right]\overline{\Gamma}_m, \qquad \textbf{(51)}$$

where $\overline{\Gamma}_m$ is given by (24).

5.3. Mean Cycle Time and P_{00}

A cycle C consists of a single vacation U followed by either the sum of an interarrival time plus Γ_1 (with probability $P(N(U)=0)$) or, with probability $P(N(U)=m$, by Γ_m ($m \geq 1$). Thus,

$$E[C] = E[U] + E_U\left[e^{-\Lambda(U)}\right]\left(\frac{1}{\lambda} + \overline{\Gamma}_1\right) + \sum_{m=1}^{\infty} \frac{1}{m!} E_U\left[e^{-\Lambda(U)}(\Lambda(U))^m\right]\overline{\Gamma}_m$$
$$= E[U] + \frac{1}{\lambda} E_U\left[e^{-\Lambda(U)}\right] + E[\Gamma_{ST}], \qquad \textbf{(52)}$$

where $E[\Gamma_{ST}]$ is given by (51).

The time D, within U, where there are no customers in the system, is

$$D = \int_0^U I\{N(t)=0\}\, dt, \qquad \textbf{(53)}$$

where $N(t)$ is the number of customers present at time t (t in $[0, U]$) and $I\{N(t)\}$ is its indicator function.

Thus, P_{00}, the fraction of time that the system is *unavailable* and there are no customers present, is calculated as

$$P_{00} = \frac{E_U[D]}{E[C]} = \frac{E_U\left[\int_0^U e^{-\Lambda(t)}\, dt\right]}{E[C]}. \qquad \textbf{(54)}$$

The fraction of time the system is *available* but no customers are present is given by

$$P_{10} = \frac{E_U\left[e^{-\Lambda(U)}\right]/\lambda}{E[C]}. \qquad \textbf{(55)}$$

Finally,

$$P_{0\bullet} = \frac{E[U]}{E[C]}. \qquad \textbf{(56)}$$

5.4. Proportion of Customers Served

The proportion of customers *not* abandoning the system is calculated as

$$P(\text{served}) = \frac{E_U[P(N(U)=0)](1+\lambda\overline{\Gamma}_1) + \sum_{m=1}^{\infty} E_U[P(N(U)=m)]\cdot(m+\lambda\overline{\Gamma}_m)}{\lambda E[C]}$$

$$= \frac{E_U\left[e^{-\Lambda(U)}\right] + \lambda E[\Gamma_{\text{ST}}] + E_U[\Lambda(U)]}{\lambda\left(E[U] + \frac{1}{\lambda}E_U\left[e^{-\Lambda(U)}\right] + E[\Gamma_{\text{ST}}]\right)}. \tag{57}$$

This performance measure of effectiveness is important when evaluating the (indirect) "cost" to the system, in terms of lost customers, due to vacations.

5.5. Back to the Fully Markovian Case of Section 4

When U and T are exponentially distributed with parameters γ and ξ, respectively, we get, from (33),

$$E[D] = E_U\left[\int_0^U e^{-\Lambda(t)}\,dt\right] = \frac{K}{\xi} \tag{58}$$

and, from (32),

$$E\left[e^{-\Lambda(U)}\right] = \frac{\gamma K}{\xi}.$$

Thus, using (54), (58), and (52), we have

$$P_{00} = (K/\xi)\left(\frac{1}{\gamma} + \frac{\gamma K}{\lambda\xi} + E[\Gamma_{\text{ST}}]\right)^{-1} \tag{59}$$

and, by (51),

$$P_{10} = \frac{\gamma K/(\lambda\xi)}{E[C]}$$

$$= \left(\frac{\gamma K}{\lambda\xi}\right)\left(\frac{1}{\gamma} + \frac{\gamma K}{\lambda\xi} + E[\Gamma_{\text{ST}}]\right)^{-1}. \tag{60}$$

Equations (59) and (60) imply that $P_{10} = (\gamma/\lambda)P_{00}$, which is exactly the balance equation for $n = 0$.

Now, by using (37), (57) is simplified

$$P(\text{served}) = \left(\frac{\gamma K}{\xi} + \lambda E[\Gamma_{\text{ST}}] + \frac{\lambda}{\xi+\gamma}\right)\left(\frac{\lambda}{\gamma} + \frac{\gamma K}{\xi} + \lambda E[\Gamma_{\text{ST}}]\right)^{-1}. \tag{61}$$

In order to get, in this case, a more explicit expression for $E[\Gamma_{\text{ST}}]$, we write

$$E_U\left[e^{-\Lambda(U)}\Lambda(U)^m\right] = \int_{u=0}^{\infty} e^{-(\lambda/\xi)}\left(1-e^{-\xi u}\right)\left[\frac{\lambda}{\xi}\left(1-e^{-\xi u}\right)\right]^m \gamma e^{-\gamma u}\,du, \tag{62}$$

By using the transformation $s = 1 - e^{-\xi u}$, we have

$$\begin{aligned} E_U\left[e^{-\Lambda(U)}\Lambda(U)^m\right] &= \frac{\gamma}{\xi}\int_{s=0}^{1} e^{-(\lambda/\xi)s}\left(\frac{\lambda}{\xi}s\right)^m (1-s)^{(\gamma/\xi)-1}\,ds \\ &= \frac{\gamma}{\xi}\left(\frac{\lambda}{\xi}\right)^m \int_{s=0}^{1} e^{-(\lambda/\xi)s}s^m(1-s)^{(\gamma/\xi)-1}\,ds. \end{aligned} \tag{63}$$

Hence, from (51),

$$E[\Gamma_{\mathrm{ST}}] = \frac{\gamma K}{\xi}\overline{\Gamma}_1 + \sum_{m=1}^{\infty}\frac{1}{m!}\left[\frac{\gamma}{\xi}\left(\frac{\lambda}{\xi}\right)^m \int_{s=0}^{1} e^{-(\lambda/\xi)s}s^m(1-s)^{(\gamma/\xi)-1}\,ds\right]\overline{\Gamma}_m. \tag{64}$$

6. CONCLUSION

Impatience of customers is a phenomenon affecting the performance of many real-life queuing systems. For example, its affect has been demonstrated and extensively analyzed in recent works on the so-called "call centers" (see Gans et al. [8]). In almost all previous works in the literature it has been assumed that customers' abandonment occurs as a result of exceeding waiting time already experienced or due to an anticipation of a too long waiting time. In contrast, in this work we analyze the situation in which customers' impatience is due to absence (or unavailability) of the servers or of the system. In a previous work we studied the $M/M/1$, $M/G/1$ and $M/M/c$ queues. Here, we further extend the scope of the analysis by examining the $M/M/\infty$ case. It turns out that various results from the analysis of the single-server or the multiserver queues can be utilized in the study of infinitely many servers systems. We have derived the (conditional) PGF of the number of customers in the system when it is functioning or when it is unavailable and we have calculated values of key performance measures such as mean queue sizes, mean length of a busy period, and mean duration of a cycle. We have considered both multiple and single system's task scenarios, for both exponentially and generally distributed impatience and task durations, and for each model we calculated the "quality of service" index: the proportion of customers being served without abandoning the system.

Acknowledgment

This work was partially supported by the Euro NGI network of excellence.

References

1. Altman, E. & Kofman, D. (1996). Bounds for performance measures of token rings. *IEEE/ACM Transactions on Networking* 4: 292–299.
2. Altman, E. & Yechiali, U. (2006). Analysis of customers' impatience in queues with server vacations. Queueing Systems 52: 261–279.

3. Bonald, T. & Roberts, J. (2001). Performance modeling of elastic traffic in overload. *ACM Sigmetrics Performance Evoluation Review* 29: 342–343.
4. Browne, S. & Steele, J.M. (1993). Transient behavior of coverage processes with applications to the infinite-server queue. *Journal of Applied Probability* 30: 589–602.
5. Doshi, B.T. (1985). A note on stochastic decomposition in a $GI/G/1$ queue with vacations or set-up times. *Journal of Applied Probability* 22: 419–428.
6. Fuhrmann, S.W. (1984). A note on the $M/G/1$ queue with server vacations. *Operations Research* 32: 1368–1373.
7. Fuhrmann, S.W. & Cooper, R.B. (1985). Stochastic decomposition in the $M/G/1$ queue with generalized vacations. *Operations Research* 33: 1117–1129.
8. Gans, N., Koole, G. & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management* 5: 79–141.
9. Levy, Y.J. & Yechiali, U. (1975). Utilization of the idle time in an $M/G/1$ queuing system. *Management Science* 22: 202–211.
10. Levy, Y. & Yechiali, U. (1976). An $M/M/s$ queue with servers' vacations. *Canadian Journal of Operational Research and Information Processing* 14: 153–163.
11. Liu, L. & Shi, D. (1996). Busy period in $GI^X/G/\infty$. *Journal of Applied Probability* 33: 815–829.
12. Miorandi, D. & Altman, E. (2004). Connectivity in ad–hoc networks: A queueing theoretical approach. In *Proceedings of Workshop on Modeling and Optimization in Mobile, Ad–Hoc and Wireless Networks (WiOPT 2004)*, pp. 132–141.
13. Shomrony, M. & Yechiali, U. (2001). Burst arrival queues with server vacations and random timers. *Mathematical Methods in Operations Research* 53: 117–146.
14. Takacs, L. (1962). *Introduction to the theory of queues*. New York: Oxford University Press.
15. Takacs, L. (1974). A single-server queue with limited virtual waiting time. *Journal of Applied Probability* 11: 612–617.
16. Takagi, H. (1991). *Queueing analysis* Vol 1: *Vacations and Priority Systems*. Amsterdam: North-Holland.
17. Yechiali, U. (2004). On the $M^X/G/1$ queue with a waiting server and vacations. *Sankhya* 66, (Part 1): 1–17.
18. Zussman, G., Segall, A., & Yechiali, U. (2004). Bluetooth time division duplex: Analysis as a polling system. In *Proceedings of the 1st IEEE Conference on Sensor and Ad Hoc Communications and Networks (SECON'04)*, pp. 547–556.
19. Zussman, G., Yechiali, U., & Segall, A. (2003). Exact probabilistic analysis of the limited scheduling algorithm for symmetrical bluetooth piconets. In *Proceedings of IFIP TC6 Personal Wireless Communications (PWC'03)* M. Conti, S. Giordano, E. Gregori, & S. Olaria (Eds.). Lecture Notes in Computer Science Vol. 2775. Berlin: Springer-Verlag, pp. 276–290.