SIAM J. APPL. MATH.
Vol. 1, No. 1, pp. 000–000, January 1998

RANDOMLY TIMED GATED QUEUEING SYSTEMS *

IDDO ELIAZAR† AND URI YECHIALI†

Abstract. Timers are used as a control mechanism to constrain the service periods in various telecommunication networks. In this paper we introduce and analyze a general family of timer-type models, called Randomly Timed Gated (RTG) queueing systems, which, in addition, extends and generalizes the family of $M/G/1$-type queues with server intermissions. The control mechanism is such that, whenever the server re-enters the system, an exponential Timer T is activated. If the server empties the queue before the Timer's expiration, it immediately leaves for another intermission. Otherwise (if the Timer expires when there is still work in the system), the server obeys one of the following rules, each leading to a different model: (1) The server completes *all* the work accumulated up to time T, and leaves. (2) The server completes only the service of the job *currently* being served, and leaves. (3) The server leaves immediately. Consequently, the RTG queue forms a unified structure for a vast collection of timer-type models: Under rule (1) it spans an entire spectrum of regimes lying between the Gated and the Exhaustive. Under rule (2) it spans an entire spectrum between the 1-Limited and the Exhaustive disciplines.

The analysis is achieved through a general solution of an infinite set of linear equations where the unknowns are the state-dependent joint transforms of two key characteristics: the length of a busy period starting with r jobs ($r = 0, 1, 2, \ldots$), and the number of jobs left behind at the end of such a busy period. We show that, in steady state, a complete description of the RTG system can be obtained *explicitly* in terms of the probability generating function (PGF) of the queue size at a *start* of a busy period. A fixed point equation for this PGF can be constructed when the structure of the intermission interval is specified. We derive various performance measures for the three models in steady state.

Key words. $M/G/1$, Randomly Timed Gated, timers, decomposition, exhaustive, gated, 1-limited.

AMS subject classifications. 60K25, 68M20, 90B12, 90B22

1. Introduction.

Queueing models where a single server is shared by multiple users, performs several tasks, or attends several channels are widely used to describe computer networks, manufacturing and production lines, and communication systems. In these systems, service periods to each user, line or channel are commonly constrained by some mechanism. The most studied constraining mechanisms are the so-called Exhaustive, Gated and 1-Limited.

Under the Exhaustive regime, at each visit of the server to a station (user, line, channel) the server attends the queue until it becomes completely empty, and only then is it allowed to leave. Under the Gated regime, at each visit the server performs only the work present in the station at the moment of its arrival. Upon completion of that amount of work, the server is allowed to leave. Under the 1-Limited regime, at each visit the server renders service to at most one 'job' and leaves.

In various real-world systems the constraining mechanism is governed by 'Timers' that limit the server's sojourn duration in a queue (see e.g. Altman [1] and references there). Thus, it is important to develop a cohesive theory for systems of these types.

*Received by the editors January, 1996; accepted by the editors July 7, 1997. This work was supported by the Broadband Telecommunications R&D Consortium administered by the Chief Scientist of the Israeli Ministry of Industry & Trade.

†Department of Statistics and Operations Research, School of Mathematical Sciences, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel, Tel: +972-3-6409637, Fax: +972-3-6409357, Email: eliazar@math.tau.ac.il, uriy@math.tau.ac.il.

In this paper we introduce and analyze a general family of timer-type models, called Randomly Timed Gated (RTG) queueing systems, which, in addition, extends and generalizes the family of $M/G/1$-type queues with server intermissions, and spans the entire spectrum of service regimes lying between the Gated and the Exhaustive. The underlying process is an $M/G/1$-type queue with general server's Intermission Intervals (II), where the rules governing the server's departure for an II, and hence the duration of a busy period, may be of various types. When the server returns to the queue after an II, a random Timer with exponential length T is activated. If the server empties the queue before the expiration of the Timer, it immediately leaves for another II. Otherwise (i.e. if the Timer expires while there is still unfinished work in the system), the server obeys one of the following rules, each leading to a different model:

(1) it completes all the work accumulated up to time T, and leaves.

(2) it completes only the service of the job currently being served, and leaves.

(3) it leaves immediately. (The service of the preempted job will have to be repeated).

Although each of the above 3 rules determines a different version of the general model, much of the analysis will be carried out for all versions simultaneously.

The general analysis is achieved by the solution of an infinite set of state-dependent linear equations of the joint transforms of the duration of a busy period, starting with a *fixed* number of jobs, and the queue size at the end of this period. Using this solution we study the system in steady state and calculate various performance measures, including queue sizes and waiting times. We show that, in steady state, all these entities can be *explicitly* expressed in terms of the PGF of X, the queue size at a beginning of a busy period. We note, however, that this general result can further be exploited in the applied prospective when the structure of the II is specified (e.g. in multiple or single vacation models, polling systems, communication multiplexer problems, etc.). In a following work we show, among other things, that in such cases one can construct a fixed point equation of the form:

$$\widehat{X}(z) = F\Big(\widehat{X}\big(f_1(z)\big), \widehat{X}\big(f_2(z)\big), \ldots, \widehat{X}\big(f_k(z)\big)\Big)$$

for the probability generating function, $\widehat{X}(\cdot)$, of X. Finally, we obtain the Gated, Exhaustive and 1-Limited service regimes as limiting cases of the general RTG model. Specifically, the Exhaustive regime is achieved in each of the above three models when the length of the Timer, T, tends to infinity. The Gated and 1-Limited regimes are obtained, respectively, from models 1 and 2 when T shrinks to zero.

Single-server queueing systems, with server's 'vacations', governed by the Gated or Exhaustive service regimes have been studied extensively in the literature (see, for example Levy and Yechiali [10], Cooper [4], Fuhrmann and Cooper [7], Doshi [5], Kella and Yechiali [9], and references there). However, works related to the RTG family of queues are few, and we mention Coffman, Fayolle and Mitrani [3], Leung and Eisenberg [11], [12], Leung [13], and Leung and Lucantoni [14]. In Coffman et al. a single server is alternating between two $M/M/1$-type queues, where each queue is equipped with exponential timer and obeys rule 3 above. The model is transformed into a boundary value problem. Leung and Eisenberg [11] study an $M/G/1$ queue with multiple server vacations and *gated time-limited* regime, where only those jobs present at the server's return from a vacation are *candidates* for service during the server visit. The server obeys rule 3 above, where the Timer is *deterministic* and the discipline is preemptive-repeat. They derive a functional equation for the amount of work and approximate it by a weighted sum of Laguerre functions with unknown coefficients. The equation is then transformed into a set of linear equations from which the coefficients can be computed. In a complementary work [12] they used similar techniques to study a non-gated model with time-limited service. Leung [13] further investigated a polling system controlled by rule (2) above. By using discrete Fourier transforms, the queue-length and delay distributions are numerically solved. Two vacation models with constant time-limited service and vacation-dependent, time-limited service are proposed by Leung and Lucantoni [14] for the performance analysis of stations in timed-token networks. By the use of exponential time stages to approximate the constant time limits, they applied the matrix analytic method to solve the vacation models for the queue length distribution and the moments of the sojourn time.

Throughout this work we use the following notation and basic properties of the $M/G/1$ queue.

Notation.

1_E	The indicator function of an even E.
$\stackrel{d}{=}, \sim$	Equality in distribution for a random variable (r.v).
$\widehat{X}(z)$	The Probability Generating Function (PGF), $E[z^X]$, of a discrete r.v. $X \in \{0,1,2,\ldots\}$, where $\lvert z\rvert \le 1$.
$\widetilde{X}(w)$	The Laplace-Stieltjes Transform (LST), $E\big[e^{-wX}\big]$, of a nonnegative r.v. X, where $\mathrm{Re}(w) \ge 0$.
ℓ_∞	The Banach space of all bounded (complex) sequences $x = (x_n)_{n=0}^{\infty}$, equipped with the usual supremum norm: $\lVert x\rVert_{\ell_\infty} = \sup\limits_{n\ge 0} \lvert x_n\rvert$.
$L(\ell_\infty)$	The Banach space of all bounded linear operators from ℓ_∞ to itself, equipped with the usual operator norm: $\lVert\mathbb{A}\rVert_{L(\ell_\infty)} = \sup\{\lVert\mathbb{A}x\rVert_{\ell_\infty} : \lVert x\rVert_{\ell_\infty} \le 1\}$.

The standard $M/G/1$ queue. Let V and λ denote the service time and arrival rate in the $M/G/1$ queue, and let $\rho = \lambda E[V]$. Defining $L_{M/G/1}$ and $W_{M/G/1}$ to be the queue size and waiting (sojourn) time in steady state, it is known (see Cooper [1981]) that

$$\widehat{L}_{M/G/1}(z) = \widetilde{W}_{M/G/1}(s) = (1-\rho)\frac{(1-z)\widetilde{V}(s)}{\widetilde{V}(s) - z} \qquad \begin{array}{c} 0 \le z \le 1 \\ s = \lambda(1-z) \end{array} \tag{1.1}$$

$$E\big[L_{M/G/1}\big] = \lambda E\big[W_{M/G/1}\big] = \rho + \frac{\lambda^2 E[V^2]}{2(1-\rho)} \ . \tag{1.2}$$

Furthermore, defining θ_r $(r \ge 0)$ to be the length of a busy period generated by r awaiting jobs, then

$$\widetilde{\theta}_r(w) = \big(\widetilde{\theta}(w)\big)^r \qquad \mathrm{Re}(w) \ge 0,\ r = 0,1,2,\ldots \tag{1.3}$$

where $\widetilde{\theta}(w) = \widetilde{\theta}_1(w)$ is the unique solution of the equation

$$\widetilde{\theta}(w) = \widetilde{V}\big(w + \lambda(1 - \widetilde{\theta}(w))\big) \ . \tag{1.4}$$

2. The model.

Formally, the model and the parameters of the system are as follows:

Arrival process $\{A(t)\}_{t\geq 0}$: A Poisson process with rate λ, counting the number of arrivals in $(0,t]$.

Service times $\{V_n\}_{n=1}^{\infty}$: A sequence of i.i.d random variables, where V_n is the service time required by the n^{th} job, and $V_n \sim V\ \forall n$.

Intermission intervals (II): Random time intervals at which the server is not available for service. A new II is initiated each time the server leaves the system. (Note that no assumptions are placed upon the intermission intervals.)

Timer T: An exponentially distributed random variable with mean μ^{-1}, ($T \sim \exp(\mu)$), which is re-activated at the moment the server re-enters the system (following an intermission interval).

Independence assumption: The arrival process $\{A(t)\}_{t\geq 0}$, service times $\{V_n\}_{n=1}^{\infty}$ and the timer T are mutually independent.

Queue discipline: Whenever the server re-enters the system (after an II) the Timer is re-activated. If the server empties the system before the expiration of the random time alloted by the Timer, the server leaves for another II. Otherwise, the server acts as follows:

Version 1: The server completes all the work accumulated up to time T and then leaves for an II.

Version 2: The server completes only the service of the job currently being served and then leaves for an II.

Version 3: The server leaves immediately for an II. In such a case, the service of the preempted job will be *repeated* during the next server's visit, where the job 'picks' up a *new*, independent service time.

Equivalently, the queue discipline dictates that the server leaves for a new II as soon as the *first* of the following two events occurs:

(a) the system becomes empty.
(b) Version 1: all jobs that arrived prior to time T have been served.
Version 2: the first service completion after time T has occurred.
Version 3: the Timer has expired.

An important role is played by the notion of a *busy period*, which is the time interval starting with the server's re-entrance to the system and ending when it next leaves for on II. If at re-entrance the queue size is $r \geq 0$, we say that the busy period is initiated by r awaiting jobs.

For Version i $(i = 1,2,3)$ we define the following key variables:
$B_r^{(i)}$ = the length of a busy period initiated by r awaiting jobs $(r \geq 0)$.
$Y_r^{(i)}$ = queue size at the end of $B_r^{(i)}$.
$N_r^{(i)}$ = number of jobs served during $B_r^{(i)}$.

For Version 3 it is also convenient to define $B_r^{(\text{eff})}$ and $Y_r^{(\text{eff})}$ as follows. $B_r^{(\text{eff})}$ = the 'effective' part of $B_r^{(3)}$, i.e. the time interval from the beginning of $B_r^{(3)}$ until the last service completion in $B_r^{(3)}$.
$Y_r^{(\text{eff})}$ = queue size at the end of $B_r^{(\text{eff})}$.
$N_r^{(\text{eff})}$ = number of jobs served during $B_r^{(\text{eff})}$.

Clearly, if $T > \theta_r$ then $B_r^{(\text{eff})} = B_r^{(3)} = \theta_r$. However, if $T < \theta_r$, then $B_r^{(3)} - B_r^{(\text{eff})} > 0$ and the difference is a wasted amount of work loaded on the server (and the system). Obviously, $N_r^{(\text{eff})} \equiv N_r^{(3)}$.

Note that, for $i = 1, 2, 3, \text{eff}$, $B_0^{(i)} = Y_0^{(i)} = N_0^{(i)} = 0$.

3. The joint distribution of (B_r, Y_r). Our aim is to derive the joint distribution of the random pair $\left(B_r^{(i)}, Y_r^{(i)}\right)$, for $i = 1, 2, 3, \text{eff}$. In what follows we suppress the superscript (i) and use the generic pair (B_r, Y_r) for all cases $i = 1, 2, 3, \text{eff}$. We define the joint transform

$$\Phi_r(w,z) = E\left[e^{-wB_r} z^{Y_r}\right], \ \operatorname{Re}(w) \geq 0, \ |z| \leq 1 \ . \tag{3.1}$$

Observing the process at the first service completion within B_r $(r \geq 1)$, enables one to analyze simultaneously all cases. Indeed, if $V_1 < T$, then at time V_1, the process regenerates itself with $r + A(V_1) - 1$ awaiting jobs. Therefore, for $r \geq 1$,

(3.2)

$$(B_r, Y_r) \stackrel{d}{=} 1_{[T>V_1]} \cdot \left(V_1 + B_{r+A(V_1)-1}, Y_{r+A(V_1)-1}\right) +$$

$$+ 1_{[T \leq V_1]} \cdot \begin{cases} \left(\sum\limits_{i=1}^{r+A(T)} V_i, A\left(\sum\limits_{i=1}^{r+A(T)} V_i \right) - A(T) \right) & i = 1 \quad \text{(Version 1)} \\ (V_1, r + A(V_1) - 1) & i = 2 \quad \text{(Version 2)} \\ (T, r + A(T)) & i = 3 \quad \text{(Version 3)} \\ (0, r) & i = \text{eff} \quad \text{(Version 3)} \end{cases}$$

When $T > V_1$, using the fact that T is exponential and $A(t)$ is Poisson, and utilizing the regeneration property, we have

$$\begin{aligned} &E\left[e^{-w(V_1 + B_{r+A(V_1)-1})} \cdot z^{Y_{r+A(V_1)-1}} \cdot 1_{[T>V_1]}\right] \\ &= \sum_{j=0}^{\infty} E\left[e^{-wV_1} \cdot 1_{[T>V_1, A(V_1)=j]}\right] \Phi_{r+j-1}(w,z) \\ &= \sum_{j=0}^{\infty} E\left[e^{-wV_1} \cdot P[T > V_1 | V_1] \cdot P[A(V_1) = j | V_1]\right] \Phi_{r+j-1}(w,z) \\ &= \sum_{j=0}^{\infty} E\left[\frac{(\lambda V_1)^j}{j!} e^{-(w+\mu+\lambda)V_1}\right] \Phi_{r+j-1}(w,z) \end{aligned}$$

From the above and equation (3.2) we obtain

$$E\left[e^{-wB_r} z^{Y_r} 1_{[T>V_1]}\right] = \sum_{j=0}^{\infty} a_j(w) \Phi_{r+j-1}(w,z) \tag{3.3}$$

where

$$a_j(w) = E\left[\frac{(\lambda V)^j}{j!} e^{-(w+\mu+\lambda)V}\right] \tag{3.4}$$

When $T \leq V_1$ the computation is carried out for each case separately.

Version 1 ($i = 1$).

Since the interval $(0, T]$ and the interval $(V_1, B_r]$ are disjoint and $B_r - V_1 = \sum_{i=2}^{r+A(T)} V_i$, we obtain

$$
\begin{aligned}
(3.5)\qquad & E\left[e^{-w(B_r-V_1)}z^{A(B_r)-A(V_1)}\Big|V_1,T\right] \\
=& E\left[E\left[E\left[e^{-w(B_r-V_1)}z^{A(B_r)-A(V_1)}\Big|V_1,B_r,A(T),T\right]\Big|A(T),T\right]\Big|T\right] \\
=& E\left[E\left[e^{-w(B_r-V_1)}e^{-\lambda(1-z)(B_r-V_1)}\big|A(T),T\right]\big|T\right] \\
=& E\left[(\widetilde{V}(w+\lambda(1-z)))^{r-1+A(T)}\big|T\right] \\
=& e^{-\lambda(1-\widetilde{V}(w+\lambda(1-z)))T}\left[\widetilde{V}(w+\lambda(1-z))\right]^{r-1} .
\end{aligned}
$$

In addition,

$$
(3.6)\quad E\left[e^{-xV_1-yT}1_{[T\le V_1]}\right] = E\left[e^{-xV_1}E\left[e^{-yT}1_{[T\le V_1]}\mid V_1\right]\right] = E\left[e^{-xV_1}\int_0^{V_1} e^{-yt}\mu e^{-\mu t}dt\right]
$$

$$
= E\left[e^{-xV_1}\frac{\mu}{y+\mu}\left(1-e^{-(y+\mu)V_1}\right)\right] = \frac{\mu}{y+\mu}\left(\widetilde{V}(x)-\widetilde{V}(x+y+\mu)\right) .
$$

Hence,

$$
\begin{aligned}
(3.7)\qquad & E\left[e^{-wB_r}z^{Y_r}1_{[T\le V_1]}\right] = E\left[e^{-w(V_1+(B_r-V_1))}z^{A(V_1)-A(T)}z^{A(B_r)-A(V_1)}1_{[T\le V_1]}\right] \\
=& E\left[e^{-wV_1}1_{[T\le V_1]}E\left[z^{A(V_1-T)}|V_1,T\right]E\left[e^{-w(B_r-V_1)}z^{A(B_r)-A(V_1)}|V_1,T\right]\right] \\
=& \text{ (by (3.5)) } E\left[e^{-wV_1}1_{[T\le V_1]}e^{-\lambda(1-z)(V_1-T)}e^{-\lambda(1-\widetilde{V}(w+\lambda(1-z)))T}\right][\widetilde{V}(w+\lambda(1-z))]^{r-1} \\
=& E\left[e^{-(w+\lambda(1-z))V_1-\lambda(z-\widetilde{V}(w+\lambda(1-z)))T}1_{[T\le V_1]}\right][\widetilde{V}(w+\lambda(1-z))]^{r-1} \\
=& \text{ (by (3.6)) } \frac{\mu}{\lambda\Big(z-\widetilde{V}(w+\lambda(1-z))\Big)+\mu}\cdot \\
&\cdot\left\{\widetilde{V}\Big(w+\lambda(1-z)\Big)-\widetilde{V}\Big(w+\lambda(1-\widetilde{V}(w+\lambda(1-z)))+\mu\Big)\right\}[\widetilde{V}(w+\lambda(1-z))]^{r-1}
\end{aligned}
$$

Version 2 ($i = 2$).

$$
\begin{aligned}
(3.8)\qquad & E\left[e^{-wV_1}z^{r+A(V_1)-1}1_{[T\le V_1]}\right] \\
=& E\left[e^{-wV_1}E\left[z^{A(V_1)}|V_1\right]P[T\le V_1|V_1]\right]z^{r-1} \\
=& E\left[e^{-wV_1}e^{-\lambda(1-z)V_1}(1-e^{-\mu V_1})\right]z^{r-1}
\end{aligned}
$$

Hence,

$$
E\left[e^{-wB_r}z^{Y_r}1_{[T\le V_1]}\right] = \Big(\widetilde{V}(w+\lambda(1-z))-\widetilde{V}(w+\lambda(1-z)+\mu)\Big)z^{r-1}
$$

Version 3 ($i = 3$).

$$E\left[e^{-wT}z^{r+A(T)}1_{[T\le V_1]}\right] = E\left[e^{-wT}1_{[T\le V_1]}E\left[z^{A(T)}|T,V_1\right]\right]z^r$$
$$= E\left[e^{-(w+\lambda(1-z))T}1_{[T\le V_1]}\right]z^r = E\left[\int_0^{V_1} e^{-w+\lambda(1-z))t}\mu e^{\mu t}dt\right]z^r$$
$$= E\left[\frac{\mu}{w+\lambda(1-z)+\mu}\left(1-e^{-(w+\lambda(1-z)+\mu)V_1}\right)\right]z^r$$

Hence,

$$E\left[e^{-wB_r}z^{Y_r}1_{[T\le V_1]}\right] = \mu\frac{1-\widetilde{V}(w+\lambda(1-z)+\mu)}{w+\lambda(1-z)+\mu}z^r \tag{3.9}$$

Version 3 ($i = \text{eff}$).

$$E\left[e^{-w0}z^r1_{[T\le V_1]}\right] = E\left[P[T\le V_1|V_1]\right]z^r = E\left[1-e^{-\mu V_1}\right]z^r$$

Hence,

$$E\left[e^{-wB_r}z^{Y_r}1_{[T\le V_1]}\right] = (1-\widetilde{V}(\mu))z^r \tag{3.10}$$

To conclude the analysis we substitute results (3.3), (3.7), (3.8), (3.9) and (3.10) into (3.11)

$$\Phi_r(w,z) = E\left[e^{-wB_r}z^{Y_r}1_{[T>V_1]}\right] + E\left[e^{-wB_r}z^{Y_r}1_{[T\le V_1]}\right] \tag{3.11}$$

and obtain an infinite set of equations

$$\begin{cases} \Phi_0(w,z) = 1 \\ \Phi_r(w,z) = \sum_{j=0}^{\infty} a_j(w)\Phi_{r+j-1}(w,z) + c(w,z)\xi(w,z)^r, \quad r\ge 1 \end{cases} \tag{3.12}$$

where $a_j(w)$ is given by (3.4). The terms $c(w,z)$ and $\xi(w,z)$ are specified (for each case) in Table 1 below.

TABLE 1
Values of $c(w,z)$ and $\xi(w,z)$ for equation (3.12)

case	$c(w,z)$	$\xi(w,z)$
$i=1$ (Version 1)	$\frac{\mu}{\widetilde{V}(w+\lambda(1-z))}\cdot\frac{\widetilde{V}(w+\lambda(1-z))-\widetilde{V}(w+\lambda(1-\widetilde{V}(w+\lambda(1-z)))+\mu)}{\lambda(z-\widetilde{V}(w+\lambda(1-z)))+\mu}$	$\widetilde{V}(w+\lambda(1-z))$
$i=2$ (Version 2)	$\frac{1}{z}\left(\widetilde{V}(w+\lambda(1-z))\right)-\widetilde{V}(w+\lambda(1-z)+\mu)\big)$	z
$i=3$ (Version 3)	$\mu\frac{1-\widetilde{V}(w+\lambda(1-z)+\mu)}{w+\lambda(1-z)+\mu}$	z
$i=\text{eff}$ (Version 3)	$1-\widetilde{V}(\mu)$	z

Solution of Equation (3.12).

For fixed values of w and z we rewrite equation (3.12) in a matrix form

$$\vec{\Phi} = \mathbb{A}\vec{\Phi} + \vec{b} \tag{3.13}$$

where

$$\left(\vec{\Phi}\right)_r = \Phi_r(w,z) \ , \ \left(\vec{b}\right)_r = \begin{cases} 1, & r=0 \\ c(w,z)\xi(w,z)^r, & r \geq 1 \end{cases}$$

and

$$\mathbb{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ a_0(w) & a_1(w) & a_2(w) & a_3(w) & \cdots \\ 0 & a_0(w) & a_1(w) & a_2(w) & \cdots \\ 0 & 0 & a_0(w) & a_1(w) & \cdots \\ & & \ddots & & \end{bmatrix}$$

We claim

Theorem 1. *Equation (3.12) has a unique solution given by*

$$\Phi_r(w,z) = \varphi(w,z)[\xi(w,z)]^r + (1-\varphi(w,z))[\widetilde{\theta}(w+\mu)]^r \tag{3.14}$$

where $\varphi(w,z)$ and $\xi(w,z)$ are given, for each case, by Table 2 below. (We repeat $\xi(w,z)$ for convenience).

Table 2

Values of $\varphi(w,z)$ and $\xi(w,z)$ for equation (3.14).

	$\varphi(w,z)$	$\xi(w,z)$
$i=1$ (Version 1)	$\frac{\mu}{\mu+\lambda(z-\widetilde{V}(w+\lambda(1-z)))}$	$\widetilde{V}(w+\lambda(1-z))$
$i=2$ (Version 2)	$\frac{\widetilde{V}(w+\lambda(1-z))-\widetilde{V}(\mu+w+\lambda(1-z))}{z-\widetilde{V}(\mu+w+\lambda(1-z))}$	z
$i=3$ (Version 3)	$\frac{\mu z}{\mu+w+\lambda(1-z)} \frac{1-\widetilde{V}(\mu+w+\lambda(1-z))}{z-\widetilde{V}(\mu+w+\lambda(1-z))}$	z
$i=\text{eff}$ (Version 3)	$\frac{z(1-\widetilde{V}(\mu))}{z-\widetilde{V}(\mu+w+\lambda(1-z))}$	z

Proof. We first show existence and uniqueness of the solution of equation (3.13). Indeed, as $\text{Re}(w) \geq 0$ and $|z| \leq 1$, then clearly $|\Phi_r(w,z)| \leq 1 \ \ \forall r \geq 0$, and $|\xi(w,z)^r| \leq 1 \ \ \forall r \geq 1$. Hence $\|\vec{\Phi}\|_\infty \leq 1$ and $\|\vec{b}\|_\infty \leq 1$. Furthermore,

$$\begin{aligned} \|\mathbb{A}\|_{L(\ell_\infty)} &\leq \sum_{j=0}^{\infty} |a_j(w)| = \sum_{j=0}^{\infty} \left| E\left[\frac{(\lambda V)^j}{j!} e^{-(w+\mu+\lambda)V}\right] \right| \\ &\leq \sum_{j=0}^{\infty} E\left[\frac{(\lambda V)^j}{j!} e^{-(\text{Re}(w)+\mu+\lambda)V}\right] = E\left[e^{-(\text{Re}(w)+\mu)V}\right] < 1 \ . \end{aligned}$$

Hence, $[I-\mathbb{A}]^{-1} \in L(\ell_\infty)$ and the unique solution of (3.13) is

$$\vec{\Phi} = [I-\mathbb{A}]^{-1}\vec{b}\ .$$

Next we show that the unique solution of (3.12) (or (3.13)) is given by

$$\vec{\Phi} = \varphi\vec{\xi} + (1-\varphi)\vec{\theta} \tag{3.15}$$

where $(\vec{\xi})_r = \xi(w,z)^r$, $(\vec{\theta})_r = \widetilde{\theta}(w+\mu)^r$ and

$$\varphi = \varphi(w,z) = \frac{c(w,z)\xi(w,z)}{\xi(w,z) - \widetilde{V}(w+\mu+\lambda(1-\xi(w,z))}\ . \tag{3.16}$$

The above solution is suggested by the similarity between the matrix $\mathbb{A}$ and the one-step transition matrix of the $M/G/1$ queue and the corresponding geometric solution (see Neuts [15]). For an intuitive derivation see remark below.

In order to show the above we note that, for $|t| \le 1$

$$\begin{aligned}\sum_{j=0}^{\infty} a_j(w)t^{r+j-1} &= E\left[\sum_{j=0} t^{r+j-1}\frac{(\lambda V)^j}{j!}e^{-(w+\mu+\lambda)V}\right] = \\ =& t^{r-1}E\left[e^{\lambda tV}e^{-(w+\mu+\lambda)V}\right] = t^{r-1}\widetilde{V}(w+\mu+\lambda(1-t))\end{aligned}$$

Furthermore, if $t=\widetilde{\theta}(w+\mu)$ then $\widetilde{V}((w+\mu)+\lambda(1-t)) = t$. Now, $\varphi(w,z)\xi(w,z)^0 + (1-\varphi(w,z))\widetilde{\theta}(w+\mu)^0 = 1$, and by the above, for all $r \ge 1$

$$\begin{aligned}&\sum_{j=0}^{\infty} a_j(w)\cdot\left\{\varphi(w,z)\xi(w,z)^{r+j-1} + (1-\varphi(w,z))\widetilde{\theta}(w+\mu)^{r+j-1}\right\} + c(w,z)\xi(w,z)^r \\ =&\varphi(w,z)\left[\xi(w,z)^{r-1}\widetilde{V}(w+\mu+\lambda(1-\xi(w,z))\right] \\ &\qquad + (1-\varphi(w,z))\left[\widetilde{\theta}(w+\mu)^{r-1}\widetilde{\theta}(w+\mu)\right] + c(w,z)\xi(w,z)^r \\ =&\left(\frac{\varphi(w,z)}{\xi(w,z)}\widetilde{V}(w+\mu+\lambda(1-\xi(w,z)) + c(w,z)\right)\xi(w,z)^r + (1-\varphi(w,z))\widetilde{\theta}(w+\mu)^r \\ =&\varphi(w,z)\xi(w,z)^r + (1-\varphi(w,z))\widetilde{\theta}(w+\mu)^r\end{aligned}$$

where the last step follows by substituting $\varphi(w,z)$ from (3.16). Hence, (3.15) is indeed the solution of (3.12). $\square$

Remark (An intuitive derivation of Theorem 1). We start from the infinite linear algebraic system (3.12). Set $\Phi_r = \Phi_r^H + \Phi_r^P$, where Φ_r^H is the homogeneous solution and Φ_r^P is a particular one. Clearly the particular solution is proportional to ξ^r so we set $\Phi_r^P = \varphi\xi^r$. Substituting Φ_r^P in (3.12) yields

$$\varphi = \varphi\sum_{j=0}^{\infty} a_j\xi^{j-1} + c = \varphi\xi^{-1}\widetilde{V}\big(w+\mu+\lambda(1-\xi)\big) + c\ .$$

Hence,

$$\varphi = c\xi/\Big[\xi - \widetilde{V}\big(w+\mu+\lambda(1-\xi)\big)\Big]\ .$$

The (pseudo) homogeneous solution satisfies

$$\Phi_r^H = \sum_{j=0}^{\infty} a_j \Phi_{r+j-1}^H , \quad \Phi_0^H = 1 \ .$$

This equation admits a geometric solution $\Phi_r^H(w) = \alpha \widetilde{\theta}(w+\mu)^r$, where $\widetilde{\theta}(\cdot)$ is given in (1.4).

Finally, $\alpha = 1 - \varphi$ since $1 = \Phi_0 = \Phi_0^H + \Phi_0^P = \alpha + \varphi$.

4. The system in steady state. For the system in steady state we define the following:

Polling instant: A moment where the server enters the system, following an Intermission Interval (II).

Cycle: Time interval between two consecutive polling instants.

C = Cycle length.

B = Length of the busy period during a cycle.

I = Length of the idle period (i.e. II) during a cycle.

N = Number of jobs served during a cycle.

X = Queue size at a polling instant.

Y = Queue size at the end of a busy period.

Clearly,

$$\begin{aligned} (B, Y, N) &= \sum_{r=0}^{\infty} (B_r, Y_r, N_r) \cdot 1_{[X=r]} \\ X &= Y + A(I) \\ C &= B + I \end{aligned} \tag{4.1}$$

Now, define the joint transform of (B, Y) as

$$\Phi(w, z) = E\,[e^{-wB} z^Y] \tag{4.2}$$

Then, by (4.1) and (3.16)

$$\begin{aligned} \Phi(w, z) &= E\,[E\,[e^{-wB} z^Y | X]] = E\,[\Phi_X(w, z)] \\ &= \varphi(w, z) \widehat{X}\,(\xi(w, z)) + (1 - \varphi(w, z)) \widehat{X}(\widetilde{\theta}(w+\mu)) \ . \end{aligned} \tag{4.3}$$

Equation (4.3) implies that the joint transform of (B, Y) (and hence, $\widetilde{B}(\cdot)$ and $\widehat{Y}(\cdot)$) is *explicitly* given in terms of $\widehat{X}(\cdot)$, the PGF of X.

As mentioned in the Introduction, we stress once again the fact that a fixed point equation for $\widehat{X}(\cdot)$ can be constructed when the structure of the II is specified. In particular, in a following work we construct a fixed-point equation of the form

$$\widehat{X}(z) = F\Big(\widehat{X}\big(f_1(z)\big), \widehat{X}\big(f_2(z)\big), \ldots, \widehat{X}\big(f_k(z)\big)\Big)$$

for applications like polling systems, vacation models, multiplexer problems, etc, where F and $f_1, \ldots, f_k$ are known functions depending on the underlying model.

In the sequel, we calculate various performance measures (Section 4.1), queue size and waiting times (Section 4.2), and study the Exhaustive, Gated and 1-Limited regimes as limiting cases of the Timer's expiration time (Section 4.3). All the results will be derived from equation (4.3), implying that a complete description of the system is given in terms of $\widehat{X}(\cdot)$.

4.1. Performance measures. The system's performance measures are presented in Table 3 below. We show how these measures are derived.

First, the means of B and Y, are formally given by

$$E[B] = -\frac{d}{dw}\widetilde{B}(w)\Big|_{w=0} = -\frac{\partial}{\partial w}\Phi(w,1)\Big|_{w=0} \tag{4.4}$$

$$E[Y] = \frac{d}{dz}\widehat{Y}(z)\Big|_{z=1} = \frac{\partial}{\partial z}\Phi(0,z)\Big|_{z=1} \tag{4.5}$$

Using (4.3) and Table 2 while noting that, for all cases, $\varphi(0,1) = 1$, the differentiation above leads to

***i* = 1 (Version 1):**

$$E[B^{(1)}] = E[V]E[X] - \frac{d}{dw}\varphi(w,1)\Big|_{w=0} \cdot (1 - \widehat{X}(\widetilde{\theta}(\mu))) \tag{4.6}$$

$$E[Y^{(1)}] = \rho E[X] + \frac{d}{dz}\varphi(0,z)\Big|_{z=1} \cdot (1 - \widehat{X}(\widetilde{\theta}(\mu))) \tag{4.7}$$

***i* = 2, 3, eff (Versions 2 and 3):**

$$E[B^{(i)}] = -\frac{d}{dw}\varphi(w,1)\Big|_{w=0} \cdot (1 - \widehat{X}(\widetilde{\theta}(\mu))) \tag{4.8}$$

$$E[Y^{(i)}] = E[X] + \frac{d}{dz}\varphi(0,z)\Big|_{z=1} \cdot (1 - \widehat{X}(\widetilde{\theta}(\mu))) \tag{4.9}$$

Differentiating $\varphi(w,z)$ appropriately, and substituting the results in (4.6), (4.7) and in (4.8), (4.9), yields $E[B]$ and $E[Y]$ in Table 3.

The mean of N is obtained by using Wald's identity. For Versions 1 and 2

$$B^{(i)} = \sum_{n=1}^{N^{(i)}} V_n \qquad i = 1,2 \tag{4.10}$$

and therefore

$$E[N^{(i)}] = \frac{E[B^{(i)}]}{E[V]} \qquad i = 1,2 \tag{4.11}$$

For version 3 a similar argumentation holds:

$$B^{(\mathrm{eff})} = \sum_{n=1}^{N^{(3)}} S_k \tag{4.12}$$

where S_k are i.i.d as S, the duration of a successful service attempt, and

$$S \sim V\Big|_{V<T} \tag{4.13}$$

Since

$$E[V|V<T]=\frac{E[V1_{[V<T]}]}{P[V<T]}=\frac{E[VP[V<T|V]]}{E[P[V<T|V]]}=\frac{E[Ve^{-\mu V}]}{E[e^{-\mu V}]}$$

it follows that

$$E[S]=-\frac{\widetilde{V}'(\mu)}{\widetilde{V}(\mu)} \tag{4.14}$$

Hence, from (4.12) and (4.14)

$$E[N^{(3)}]=\frac{E\big[B^{(\mathrm{eff})}\big]}{E[S]}=-\frac{\widetilde{V}(\mu)}{\widetilde{V}'(\mu)}E\big[B^{(\mathrm{eff})}\big] \tag{4.15}$$

The mean cycle time is now readily calculated since, in steady state, $E[A(C)]=E[N]$. Hence, for all versions,

$$E[C]=\frac{E[N]}{\lambda} \tag{4.16}$$

We define p_{busy} as the probability of finding the server busy. Clearly, for all versions

$$p^{(i)}_{\mathrm{busy}}=\frac{E[B^{(i)}]}{E[C^{(i)}]} \qquad i=1,2,3 \tag{4.17}$$

For versions 3 we also define p_{eff} as the proportion of time the server is working 'effectively', i.e.

$$p_{\mathrm{eff}}=\frac{E\big[B^{(\mathrm{eff})}\big]}{E[C^{(3)}]} \tag{4.18}$$

Finally, the last row in Table 3 shows the connection between $E[X]-E[Y]$ and $E[N]$. For technical convenience, in Table 3 and in the sequel, we write B_{eff} and Y_{eff} instead of $B^{(\mathrm{eff})}$ and $Y^{(\mathrm{eff})}$, respectively.

TABLE 3
Performance Measures

	Version 1	Version 2	Version 3	
$E[B]$	$E[V]E[X]+\rho\gamma_1$	$E[V]\gamma_2$	γ_1	$E[B_{\mathrm{eff}}]=-\widetilde{V}'(\mu)\gamma_2$
$E[Y]$	$\rho E[X]-(1-\rho)\lambda\gamma_1$	$E[X]-(1-\rho)\gamma_2$	$E[X]+\lambda\gamma_1-\widetilde{V}(\mu)\gamma_2$	$E[Y_{\mathrm{eff}}]=E[X]-(\widetilde{V}(\mu)+\lambda\widetilde{V}'(\mu))\gamma_2$
$E[N]$	$E[X]+\lambda\gamma_1$	γ_2	$\widetilde{V}(\mu)\gamma_2$	
$E[C]$	$\frac{1}{\lambda}E[X]+\gamma_1$	$\frac{1}{\lambda}\gamma_2$	$\frac{1}{\lambda}\widetilde{V}(\mu)\gamma_2$	
p_{busy}	ρ	ρ	$\lambda\frac{1-\widetilde{V}(\mu)}{\mu\widetilde{V}(\mu)}$	$p_{\mathrm{eff}}=\widehat{\rho}=\left(-\lambda\frac{\widetilde{V}'(\mu)}{\widetilde{V}(\mu)}\right)$
$E[X]-E[Y]$	$(1-\rho)E[N]$	$(1-\rho)E[N]$	$(1-p_{\mathrm{busy}})E[N]$	$E[X]-E[Y_{\mathrm{eff}}]=(1-p_{\mathrm{eff}})E[N]$

$$\gamma_1=\frac{1-\widehat{X}(\widetilde{\theta}(\mu))}{\mu}=E[\mathrm{Min}(T,\theta_X)] \quad \gamma_2=\frac{1-\widehat{X}(\widetilde{\theta}(\mu))}{1-\widetilde{V}(\mu)}=\frac{E[\mathrm{Min}(T,\theta_X)]}{E[\mathrm{Min}(T,V)]}$$

We conclude the section with three interpreting remarks.

Remark 4.1. Stationary conditions.

From Table 3 it is clear that necessary conditions for steady-state are $p_{\text{busy}} < 1$, i.e. $\rho < 1$ for Versions 1 and 2, and $\lambda \cdot \frac{1-\widetilde{V}(\mu)}{\mu \widetilde{V}(\mu)} < 1$ for Version 3. Sufficient conditions, however, depend on the nature of the II's. If, for example, the II's form a sequence of i.i.d random variables with duration D and mean $E[D] < \infty$, we can use Fricker & Jaïbi [6] result for non-preemptive queues stating that a necessary and sufficient condition for stability is that $\rho + \lambda E[D]/E[M] < 1$, where $E[M]$ is the mean number of jobs that would have been served during a busy period starting with an infinite number of awaiting jobs. Hence, for Versions 1 and 2, we have:

Version 1: $E[M] = \infty$. Thus, $\rho < 1$ is also sufficient.

Version 2: M is distributed geometrically with parameter $P[T \leq V] = 1 - \widetilde{V}(\mu)$. Thus, $E[M] = 1/(1 - \widetilde{V}(\mu))$. Hence, the necessary and sufficient condition is $\rho + \lambda E[D](1 - \widetilde{V}(\mu)) < 1$. That is, $\lambda \left[E[V] + E[D]\Big(1 - \widetilde{V}(\mu)\Big)\right] < 1$. Indeed, every job, beyond its own service time, loads the system with an II if and only if $T \leq V$.

If the II's are independent of the other parameters of the system then, both Versions 2 and 3 can be looked upon as variations of the standard $M/G/1$ queue, where each job loads the system with an *actual* service time of length $\overline{S}$, given by:

$$
\begin{aligned}
\overline{S} &= V + D1_{[V>T]} && \text{(Version 2)}\\
\overline{S} &= \text{Min}(V,T) + (D + \overline{S}')1_{[V>T]} && \text{(Version 3)}
\end{aligned}
$$

where $\overline{S}\,'$ is independent of V, T and D, and is distributed as $\overline{S}$. Hence (see also (4.19)),

$$
\begin{aligned}
E[\overline{S}] &= E[V] + E[D]\Big(1 - \widetilde{V}(\mu)\Big) && \text{(Version 2)}\\
E[\overline{S}] &= \tfrac{1-\widetilde{V}(\mu)}{\widetilde{V}(\mu)}\Big(\tfrac{1}{\mu} + E[D]\Big) && \text{(Version 3)}
\end{aligned}
$$

Clearly, a necessary and sufficient condition for stationarity is $\lambda E[\overline{S}] < 1$.

Remark 4.2. Interpretation of γ_1 and γ_2 in Table 3.

Note that

$$
\begin{aligned}
E[\text{Min}(V,T)] &= E\big[V1_{[V\leq T]}\big] + E\big[T1_{[V>T]}\big]\\
&= E[VP[V \leq T|V]] + E\left[\int_0^V t\mu e^{-\mu t}dt\right]\\
&= E[Ve^{-\mu V}] + E\left[-Ve^{-\mu V} + \frac{1}{\mu}\big(1 - e^{-\mu V}\big)\right]
\end{aligned}
$$

Hence,

$$
E[\text{Min}(V,T)] = \frac{1 - \widetilde{V}(\mu)}{\mu} \tag{4.19}
$$

Similarly,

$$
E[\text{Min}(\theta_X, T)] = \frac{1 - \widehat{X}(\widetilde{\theta}(\mu))}{\mu}\ . \tag{4.20}
$$

Result (4.20) follows since $\big(\widetilde{\theta}_X\big)(w) = E\big[e^{-w\theta_X}\big] = E\big[E\big[e^{-w\theta_X}|X\big]\big] = E\Big[\widetilde{\theta}(w)^X\Big] = \widehat{X}(\widetilde{\theta}(w))$. Therefore,

$$\gamma_1 = \frac{1-\widehat{X}\big(\widetilde{\theta}(\mu)\big)}{\mu} = E[\mathrm{Min}(\theta_X, T)] \tag{4.21}$$

$$\gamma_2 = \frac{1-\widehat{X}\big(\widetilde{\theta}(\mu)\big)}{1-\widetilde{V}(\mu)} = \frac{E\big[\mathrm{Min}(\theta_X, T)\big]}{E\big[\mathrm{Min}(V, T)\big]} \tag{4.22}$$

Remark 4.3. Interpretation of p_{busy} and p_{eff} in Version 3.

Recall that at the beginning of each service the exponential timer is reactivated and the service is successful only if $V \leq T$. Thus, each job loads the system with a geometric number of service attempts, from which only the last is successful.

Let R be the total service time required by a job and recall (see (4.13)) that S is the duration of a successful attempt. Then,

$$R = V1_{[V\leq T]} + (T + R_1)1_{[V>T]} = \mathrm{Min}(V,T) + R_1 1_{[V>T]} \tag{4.23}$$

where $R_1 \sim R$, and R_1 is independent of V and T. From (4.23) it is easily verified that

$$E[R] = \frac{1-\widetilde{V}(\mu)}{\mu\widetilde{V}(\mu)} \tag{4.24}$$

Hence, from (4.24) and (4.14) it follows that

$$p_{\text{busy}} = \lambda E[R] = E[A(R)] \tag{4.25}$$

$$p_{\text{eff}} = \lambda E[S] = E[A(S)] \equiv \widehat{\rho} \tag{4.26}$$

Indeed, (4.25) and (4.26) are the correct analogs of $\rho = \lambda E[V] = E[A(V)]$.

4.2. Queue sizes and waiting times. There are several possible ways to derive $\widehat{L}(\cdot)$, the PGF of the queue size L at steady state. We demonstrate two. The first, applicable to the non-preemption cases Versions 1 and 2, uses the Fuhrmann-Cooper decomposition result [1985]. The other, applicable to all three versions, is based on a result from renewal-reward processes. We start with the first approach.

The Fuhrmann-Cooper decomposition result states that

$$\widehat{L}(z) = \widehat{L}_{M|G|1}(z)\cdot\chi(z) \tag{4.27}$$

where $\chi(z)$ is the PGF of the queue size at an arbitrary instant within an II. Furthermore, Borst [1994], by using SLLN arguments, has shown that

$$\chi(z) = \frac{1}{z-1}\,\frac{E[z^X]-E[z^Y]}{E[X]-E[Y]} \tag{4.28}$$

Substituting (1.1) and (4.28) in (4.27) yields

$$\widehat{L}(z) = (1-\rho)\frac{\widetilde{V}(\lambda(1-z))}{z-\widetilde{V}(\lambda(1-z))}\,\frac{E[z^X]-E[z^Y]}{E[X]-E[Y]} \tag{4.29}$$

By using (4.3), and since $E[z^Y] = \Phi(0,z)$, we can write

$$\begin{aligned} E[z^X] - E[z^Y] &= \widehat{X}(z) - \Phi(0,z) \\ &= \varphi(0,z)\Big(\widehat{X}(z) - \widehat{X}\big(\xi(0,z)\big)\Big) + \big(1-\varphi(0,z)\big)\Big(\widehat{X}(z) - \widehat{X}\big(\widetilde{\theta}(\mu)\big)\Big) \end{aligned} \tag{4.30}$$

From Table 3, $E[X] - E[Y] = (1-\rho)E[N]$. Hence,

$$\begin{aligned} \widehat{L}(z) = \frac{1}{E[N]} \frac{\widetilde{V}(\lambda(1-z))}{z - \widetilde{V}(\lambda(1-z))} &\Big\{\varphi(0,z)\Big(\widehat{X}(z) - \widehat{X}\big(\xi(0,z)\big)\Big) \\ &+ \big(1-\varphi(0,z)\big)\Big(\widehat{X}(z) - \widehat{X}\big(\widetilde{\theta}(\mu)\big)\Big)\Big\} \end{aligned} \tag{4.31}$$

Using Table 3 again (see also (4.21) and (4.22)), we have, for Versions 1 and 2,

$$\begin{aligned} \widehat{X}(z) - \widehat{X}\big(\widetilde{\theta}(\mu)\big) &= \big(\widehat{X}(z) - 1\big) + \Big(1 - \widehat{X}\big(\widetilde{\theta}(\mu)\big)\Big) \\ &= \big(\widehat{X}(z) - 1\big) + \mu\gamma_1 = \big(\widehat{X}(z) - 1\big) + \big(1 - \widetilde{V}(\mu)\big)\gamma_2 \end{aligned} \tag{4.32}$$

Version 1. From Table 2, $\xi(0,z) = \widetilde{V}\big(\lambda(1-z)\big)$,

$$\varphi(0,z) = \frac{\mu}{\mu + \lambda(z - \widetilde{V}(\lambda(1-z)))}, \quad 1 - \varphi(0,z) = \frac{\lambda(z - \widetilde{V}(\lambda(1-z)))}{\mu + \lambda(z - \widetilde{V}(\lambda(1-z)))} .$$

From Table 3, $E[N] = E[X] + \lambda\gamma_1$. Substituting the above and (4.32) in (4.31) yields (while setting $s = \lambda(1-z)$)

$$\begin{aligned} \widehat{L}(z) = \frac{\widetilde{V}(s)}{\mu + \lambda(z - \widetilde{V}(s))} &\left\{\mu \frac{\lambda}{E[N]}\big(\widehat{X}(z) - 1\big)\right. \\ &\left. + \frac{\mu}{E[N]}\left(\frac{\widehat{X}(z) - \widehat{X}(\widetilde{V}(s))}{z - \widetilde{V}(s)} - E[X]\right)\right\} \end{aligned} \tag{4.33}$$

Version 2. Here $\xi(0,z) = z$, while $1 - \varphi(0,z) = \frac{z - \widetilde{V}(s)}{z - \widetilde{V}(\mu+s)}$. Hence from (4.33),

$$\widehat{L}(z) = \frac{1}{E[N]} \frac{\widetilde{V}(s)}{z - \widetilde{V}(\mu+s)}\Big(\widehat{X}(z) - \widehat{X}\big(\widetilde{\theta}(\mu)\big)\Big) \tag{4.34}$$

Using (from Table 3) $E[N] = \gamma_2 = \frac{1 - \widehat{X}(\widetilde{\theta}(\mu))}{1 - \widetilde{V}(\mu)}$ yields

$$\begin{aligned} \widehat{L}(z) &= \big(1 - \widetilde{V}(\mu)\big)\frac{\widetilde{V}(s)}{z - \widetilde{V}(\mu+s)}\left(1 - \frac{1 - \widehat{X}(z)}{1 - \widehat{X}(\widetilde{\theta}(\mu))}\right) \\ &= \frac{\widetilde{V}(s)}{z - \widetilde{V}(s)}\left(\big(1 - \widetilde{V}(\mu)\big) - \frac{1 - \widehat{X}(z)}{E[N]}\right) \end{aligned} \tag{4.35}$$

As mentioned above, the Fuhrmann-Cooper decomposition does *not* hold for Version 3 where preemptions *do* occur. We therefore use a second approach (also used in

Yechiali [16]), which is applicable to *all* versions and is based on the following result, obtained from the theory of renewal-reward processes (see Heyman & Sobel [8], Chap. 6).

$$\widehat{L}(z) = \frac{E\Big[\sum\limits_{n=1}^{N} z^{Q_n}\Big]}{E[N]} \tag{4.36}$$

where Q_n = Queue size left behind by the n^{th} departing job within a busy period $(1 \leq n \leq N)$. Now, for a single busy period generated by r awaiting jobs, define $F_r(z) = \sum\limits_{n=1}^{N_r} z^{Q_n}$ and $f_r(z) = E[F_r(z)]$.

The key observation, just as in Section 3, is that, under the event $\{T > V_1\}$, the process regenerates at time V_1 with $r + A(V_1) - 1$ awaiting jobs. On the other hand, when $\{T < V_1\}$,

$$E[F_r(z) \cdot 1_{[T<V_1]}] = \begin{cases} E\left[\sum_{n=1}^{r+A(T)} z^{Q_n}\right] & \text{Version 1} \\ E[z^{Q_1}] & \text{Version 2} \\ 0 & \text{Version 3} \end{cases} \tag{4.37}$$

(The expression for Version 3 results from the fact that, when $T < V_1$, there are no departures during the busy period.)

Now,

$$\begin{aligned} E[F_r(z)1_{[T>V_1]}] &= E\Big[E\big[F_r(z)1_{[T>V_1]}|T, V_1, A(V_1)\big]\Big] \\ &= E[1_{T>V_1]}E\big[F_r(z)|T, V_1, A(V_1)\big]\Big] = E\Big[1_{[T>V_1]}\big(z^{r+A(V_1)-1} + f_{r+A(V_1)-1}(z)\big)\Big] \\ &= z^{r-1}E[z^{A(V_1)}1_{[T>V_1]}] + E[f_{r+A(V_1)-1}(z)1_{[T>V_1]}] \end{aligned} \tag{4.38}$$

We therefore calculate

$$\begin{aligned} E[z^{A(V_1)}1_{[T>V_1]}] &= E\big[E[z^{A(V_1)}1_{[T>V_1]}|V_1]\big] \\ &= E\big[E[z^{A(V_1)}|V_1]P(T > V_1|V_1)\big] = E[e^{-\lambda(1-z)V_1}e^{-\mu V_1}] \\ &= \widetilde{V}\big(\mu + \lambda(1-z)\big) \end{aligned} \tag{4.39}$$

Also

$$\begin{aligned} E[f_{r+A(V_1)-1}(z)1_{[T>V_1]}] &= E\big[E[f_{r+A(V_1)-1}(z)1_{[T>V_1]}|V_1]\big] \\ &= E\big[E[f_{r+A(V_1)-1}(z)|V_1] \cdot P(T > V_1|V_1)\big] \\ &= E\left[\sum_{j=0}^{\infty} f_{r+j-1}(z)P\big(A(V_1) = j|V_1\big) \cdot P(T > V_1|V_1)\right] \\ &= E\left[\sum_{j=0}^{\infty} f_{r+j-1}(z)\frac{(\lambda V_1)^j}{j!}e^{-\lambda V_1} \cdot e^{-\mu V_1}\right] \\ &= \sum_{j=0}^{\infty} E\left[\frac{(\lambda V_1)^j}{j!}e^{-(\lambda+\mu)V_1}\right] f_{r+j-1}(z) \\ &= \sum_{j=0}^{\infty} a_j(0)f_{r+j-1}(z) \triangleq \sum_{j=0}^{\infty} a_j f_{r+j-1}(z) \end{aligned} \tag{4.40}$$

Finally, substituting (4.39) and (4.40) in (4.38) yields,

$$E\big[F_r(z)1_{[T>V_1]}\big] = \sum_{j=0}^{\infty} a_j f_{r+j-1}(z) + \widetilde{V}\big(\mu+\lambda(1-z)\big)z^{r-1} \tag{4.41}$$

As for (4.37), we skip in what follows the derivations for Version 1 and present the calculations for Versions 2 and 3.

Version 2.

$$\begin{aligned} E[F_r(z)1_{[T<V_1]}] &= E[z^{Q_1}1_{[T<V_1]}] \\ &= E[z^{r+A(V_1)-1}1_{[T<V_1]}] = z^{r-1}E[z^{A(V_1)}1_{[T<V_1]}] \\ &= z^{r-1}\big(E[z^{A(V_1)}] - E[z^{A(V_1)}1_{[T>V_1]}]\big) \\ &= z^{r-1}\Big(\widetilde{V}\big(\lambda(1-z)\big) - \widetilde{V}\big(\mu+\lambda(1-z)\big)\Big) \end{aligned} \tag{4.42}$$

where the last term follows from (4.39).

Version 3.
$E[F_r(z)1_{[T<V_1]}] = 0$

Combining (4.37) with (4.41) leads to

$$\begin{cases} f_0(z) = 0 \\ f_r(z) = \sum_{j=0}^{\infty} a_j f_{r+j-1}(z) + g(z)z^r \quad (r \geq 1) \end{cases} \tag{4.43}$$

where

$$g(z) = \begin{cases} \frac{1}{z}\widetilde{V}\big(\lambda(1-z)\big) & \text{Version 2} \\ \frac{1}{z}\widetilde{V}\big(\mu+\lambda(1-z)\big) & \text{Version 3} \end{cases} \tag{4.44}$$

Theorem 2. *The unique solution of (4.43) is given by*

$$f_r(z) = \psi(z)\big(z^r - \widetilde{\theta}(\mu)^r\big) \quad r = 0, 1, 2, \cdots \tag{4.45}$$

where

$$\psi(z) = \frac{zg(z)}{z - \widetilde{V}\big(\mu+\lambda(1-z)\big)} \tag{4.46}$$

Proof. Existence and uniqueness of the solution follows similarly to Theorem 1. As for the explicit solution, we have

$$\begin{aligned} \sum_{j=0}^{\infty} a_j t^{r+j-1} &= \widetilde{V}\big(\mu+\lambda(1-t)\big)t^{r-1} \\ &= \begin{cases} \frac{\widetilde{V}(\mu+\lambda(1-z))}{z} z^r & t = z \\ \widetilde{\theta}(\mu)^r & t = \widetilde{\theta}(\mu) \end{cases} \end{aligned} \tag{4.47}$$

Therefore, for $r \geq 1$, substituting (4.45) in (4.43), we write

$$\begin{aligned} f_r(z) &= \sum_{j=0}^{\infty} a_j f_{r+j-1}(z) + g(z) z^r \\ &= \psi(z) \left(\sum_{j=0}^{\infty} a_j z^{r+j-1} - \sum_{j=0}^{\infty} a_j \big(\widetilde{\theta}(\mu)\big)^{r+j-1} \right) + g(z) z^r \\ &= \left(\psi(z) \frac{\widetilde{V}(\mu+\lambda(1-z))}{z} + g(z) \right) z^r - \psi(z) \big(\theta(\mu)\big)^r \end{aligned}$$

By using (4.46) we get

$$\psi(z) = \frac{\widetilde{V}(\mu+\lambda(1-z))}{z} + g(z) = \frac{z g(z)}{z - \widetilde{V}(\mu+\lambda(1-z))} = \psi(z)$$

Thus, $f_r(z) = \psi(z)\big(z^r - \theta(\mu)^r\big)$, as claimed. $\square$

It readily follows that

Corollary 1.

$$f(z) \equiv E\big[f_X(z)\big] = E\left[\sum_{n=1}^{N} z^{Q_n}\right] = \psi(z)\Big(\widehat{X}(z) - \widehat{X}\big(\widetilde{\theta}(\mu)\big)\Big) \tag{4.48}$$

Corollary 2. *For Versions 2 and 3,*

$$\widehat{L}(z) = \frac{\psi(z)}{E[N]}\Big(\widehat{X}(z) - \widehat{X}\big(\widetilde{\theta}(\mu)\big)\Big) \tag{4.49}$$

where $\psi(z)$ is given by (4.46) and $g(z)$ by (4.44).

Version 2: Substituting $\psi(z) = \frac{\widetilde{V}(\lambda(1-z))}{z-\widetilde{V}(\mu+\lambda(1-z))}$ in (4.49) results, again, in (4.34).

Version 3: From (4.44) and (4.46), $\psi(z) = \frac{\widetilde{V}(\mu+\lambda(1-z))}{z-\widetilde{V}(\mu+\lambda(1-z))}$. Substituting in (4.49) yields

$$\widehat{L}(z) = \frac{1}{E[N]} \, \frac{\widetilde{V}(\mu+\lambda(1-z))}{z - \widetilde{V}(\mu+\lambda(1-z))}\Big(\widehat{X}(z) - \widehat{X}\big(\widetilde{\theta}(\mu)\big)\Big) \tag{4.50}$$

Since $E[N] = \widetilde{V}(\mu)\gamma_2 = \frac{\widetilde{V}(\mu)}{1-\widetilde{V}(\mu)}\big(1 - \widehat{X}(\widetilde{\theta}(\mu)\big)$ and $\widehat{X}(z) - \widehat{X}(\widetilde{\theta}(\mu)) = \big(1 - \widehat{X}(\widetilde{\theta}(\mu)) - (1 - \widehat{X}(z))\big)$, we finally have, similarly to (4.35),

$$\begin{aligned} \widehat{L}(z) &= \frac{1-\widetilde{V}(\mu)}{\widetilde{V}(\mu)} \, \frac{\widetilde{V}(\mu+\lambda(1-z))}{z-\widetilde{V}(\mu+\lambda(1-z))} \left(1 - \frac{1-\widehat{X}(z)}{1-\widehat{X}(\widetilde{\theta}(\mu))}\right) \\ &= \frac{\widetilde{V}(\mu+\lambda(1-z))}{z-\widetilde{V}(\mu+\lambda(1-z))} \left(\frac{1-\widetilde{V}(\mu)}{\widetilde{V}(\mu)} - \frac{1-\widehat{X}(z)}{E[N]}\right) \end{aligned} \tag{4.51}$$

Mean queue size, $E[L]$. By differentiating equations (4.33), (4.35) and (4.52) we obtain the mean queue size, $E[L]$, for each version, respectively:

Version 1.

$$E[L] = \rho - \frac{\lambda}{\mu}(1-\rho) + \frac{\frac{1+\rho}{2}E[X(X-1)] + \frac{\lambda}{\mu}E[X]}{E[N]} \tag{4.52}$$

Version 2.

$$E[L] = \frac{1}{1-\widetilde{V}(\mu)}\left(\frac{E[X]}{\gamma_2} - \widetilde{V}(\mu) - \lambda\widetilde{V}'(\mu)\right) - (1-\rho) \tag{4.53}$$

Version 3.

$$E[L] = \frac{1}{1-\widetilde{V}(\mu)}\left(\frac{E[X]}{\gamma_2} - 1 - \lambda\frac{\widetilde{V}'(\mu)}{\widetilde{V}(\mu)}\right) \tag{4.54}$$

Waiting times. For all three versions, being $M/G/1$-type systems, and assuming First Come First Served order of service,

$$\widetilde{W}(\lambda(1-z)) = \widehat{L}(z)\ , \quad E[W] = E[L]/\lambda\ .$$

4.3. Special cases. As was indicated in the Introduction, the Exhaustive, Gated and 1-Limited regimes (each with server's II) are limiting cases of the general family of RTG regimes. We show how each case is obtained from the general model.

Exhaustive regime. In all three versions, taking limits of the general results when $\mu \searrow 0$ (i.e. $T \nearrow \infty$) yields the exhaustive regime, as follows.

From Table 2, $\varphi(w,z) \to 0$ when $\mu \searrow 0$. This implies that $\Phi(w,x) \to \widehat{X}(\widetilde{\theta}(w))$. Thus, since $\widetilde{B}(w) = \Phi(w,1)$, $\widetilde{B}(w) \to \widehat{X}(\widetilde{\theta}(w))$. However, $\widehat{X}(\widetilde{\theta}(w))$ is the LST of the duration of a busy period in an $M/G/1$ queue starting with X jobs. Hence $B \to \theta_X$ in distribution. Similarly, since $\widehat{Y}(z) = \Phi(0,z)$, we get that $\widehat{Y}(z) \to 1$. That is, $Y \to 0$ in distribution. The above pair (B,Y) characterizes the system under the *pure* Exhaustive regime.

As for the queue size, letting $\mu \searrow 0$ in equation (4.29) for Versions 1 and 2, and using $Y \to 0$ leads to (see (4.27))

$$\widehat{L}(z) \longrightarrow \widehat{L}_{M|G|1}(z) \cdot \frac{1-\widehat{X}(z)}{E[X](1-z)} \tag{4.55}$$

which is the PGF of the queue size in a regular $M/G/1$ queue with Intermission Intervals and Exhaustive service regime (see Levy and Yechiali [10]). For Version 3, equation (4.55) is obtained as the limit of equation (4.51).

Gated regime. The Gated regime is obtained as a limiting case of Version 1 when $\mu \nearrow \infty$ (i.e. $T \searrow 0$). Indeed, when $\mu \nearrow \infty$, $\varphi(w,z) \to 1$. This leads to $\Phi(w,z) \longrightarrow \widehat{X}(\widetilde{V}(w+\lambda(1-z)))$. But then $\widetilde{B}(w) = \widehat{X}(\widetilde{V}(w))$, which is the LST of a busy period (under the Gated regime) in an $M/G/1$ queue whose duration is

$B = \sum_{k=1}^{X} V_k$. Now, $\mu \nearrow \infty$ implies $\widehat{Y}(z) \longrightarrow \widehat{X}(\widetilde{V}(\lambda(1-z)))$, which is the PGF of the number of arrivals Y during the service of X jobs, as is the case under the Gated regime.

Since $E[Y] \to E[X]E[A(V)] = \rho E[X]$, we obtain from (4.27) and (4.28) the PGF of the queue size under the Gated regime, namely

$$\widehat{L}(z) \to \frac{\widetilde{V}(\lambda(1-z))}{E[X]} \cdot \frac{\widehat{X}(z) - \widehat{X}(\widetilde{V}(\lambda(1-z)))}{z - \widetilde{V}(\lambda(1-z))} \tag{4.56}$$

1-Limited regime. The 1-limited regime is obtained as a limiting case of Version 2 when $\mu \nearrow \infty$ (i.e. $T \searrow 0$). Indeed, $\mu \nearrow \infty$ leads to $\varphi(w,z) \longrightarrow \frac{\widetilde{V}(w+\lambda(1-z))}{z}$. This implies that $\Phi(w,z) \longrightarrow \frac{z-\widetilde{V}(w+\lambda(1-z))}{z}\widetilde{X}(0) + \frac{\widetilde{V}(w+\lambda(1-z))}{z}\widetilde{X}(z)$. Hence,

$$\widetilde{B}(w) \longrightarrow 1 \cdot \widetilde{X}(0) + \widetilde{V}(w) \cdot (1 - \widehat{X}(0))$$

which is equivalent to $B \longrightarrow 0 \cdot 1_{[X=0]} + V \cdot 1_{[X \geq 1]}$ in distribution. Also,

$$\widehat{Y}(z) \longrightarrow 1 \cdot \widehat{X}(0) + \frac{\widetilde{V}(\lambda(1-z))}{z} \cdot (\widehat{X}(z) - \widehat{X}(0)) \tag{4.57}$$

which is equivalent to $Y \longrightarrow 0 \cdot 1_{[X=0]} + (X + A(V) - 1) \cdot 1_{[X \geq 1]}$ in distribution. The above pair (B, Y) characterizes the 1-Limited regime.

Finally, using (4.57) and $E[Y] \to E[X] - (1-\rho)(1-\widehat{X}(0))$ in (4.29) yields

$$\widehat{L}(z) \longrightarrow \frac{\widetilde{V}(\lambda(1-z))}{z} \cdot \frac{\widehat{X}(z) - \widehat{X}(0)}{1 - \widehat{X}(0)} \tag{4.58}$$

REFERENCES

[1] E. Altman, *Analysing timed-token ring protocols using the power series algorithm*, in The Fundamental Role of Teletraffic in the Evolution of Telecommunications Networks, J. Labetoulle, J. W. Roberts, eds., Elsevier, 1994, pp. 961–971.

[2] S. C. Borst, *Polling systems with multiple coupled servers*, Queueing Systems, 20 (1995) pp. 369–393.

[3] E. G. Coffman, G. Fayolle and I. Mitrani, *Two queues with alternating service periods*, in Performance '87, P. J. Courtois, G. Latouche, eds., North-Holland, 1988, pp. 227–237.

[4] R. B. Cooper, *Introduction to Queueing Theory*, 2nd edition, North-Holland, 1981.

[5] B. T. Doshi, *Queueing systems with vacations – A survey*, Queueing Systems, 1 (1986), pp. 29–66.

[6] C. Fricker and M. R. Jaïbi, *Monotonicity and stability of periodic polling models*, Queueing Systems, 15 (1994) pp. 211–238.

[7] S. W. Fuhrmann and R. B. Cooper, *Stochastic decomposition in the $M/G/1$ queue with generalized vacations*, Operations Research, 33 (1985), pp. 1117–1129.

[8] D. P. Heyman and M. G. Sobel, *Stochastic Models in Operations Research*, Vol. 1: Stochastic Processes and Operating Characteristics, McGraw-Hill, 1982.

[9] O. Kella and U. Yechiali, *Priorities in $M/G/1$ queue with server vacations*, Naval Research Logistics, 35 (1988), pp. 23–34.

[10] Y. Levy and U. Yechiali, *Utilization of the idle time in an $M/G/1$ queueing system*, Management Science, 22 (1975), pp. 202–211.

[11] K. K. LEUNG AND M. EISENBERG, *A single-server queue with vacations and gated time-limited service*, IEEE Transactions on Communications, 38 (1990), pp. 1454–1462.

[12] K. K. LEUNG AND M. EISENBERG, *A single server queue with vacations and non-gated time-limited service*, Performance Evaluation, 12 (1991), pp. 115–125.

[13] K. K. LEUNG, *Cyclic-service systems with nonpreemptive, time-limited service*, IEEE Transactions on Communications, 42 (1994), pp. 2521–2524.

[14] K. K. LEUNG AND D. M. LUCANTONI, *Two vacation models for token-ring networks where service is controlled by timers*, Performance Evaluation, 20 (1994), pp. 165–184.

[15] M. F. NEUTS, *Matrix-Geometric Solutions in Stochastic Models*, The Johns Hopkins University Press, 1981.

[16] U. YECHIALI, *Analysis and control of polling systems*, in Performance Evaluation of Computer and Communication Systems, L. Donatiello, R. Nelson, eds., Springer-Verlag, 1993, pp. 630–650.