ON THE QUEUE LENGTH DISTRIBUTION
FOR THE GI/G/1/K/V\textsubscript{M} QUEUE

Kyung C. Chae*, Nam K. Kim, Bong K. Yoon

Dept. of Industrial Engineering, Korea Advanced Institute of Science and Technology, Daejon, Korea 305-701
E-mail: kcchae@kaist.ac.kr

Dept. of Industrial Engineering, Korea Advanced Institute of Science and Technology, Daejon, Korea 305-701
E-mail: s.freedom@kaist.ac.kr

IBM Business Consulting Services, ASEM Building, 40th Floor, 159-1
Samsung-dong, Kangnam-Gu, Seoul, Korea 135-798
E-mail: bongkyu.yoon@kr.ibm.com

ABSTRACT

We present a transform-free distribution of the steady-state queue length for the GI/G/1/K queueing system with multiple vacations under exhaustive FIFO service discipline. The method we use is a modified supplementary variable technique and the result we obtain is expressed in terms of conditional expectations of the remaining service time, the remaining interarrival time, and the remaining vacation, conditional on the queue length at the embedded points. The case $K \rightarrow \infty$ is also considered.

*Corresponding author. Tel: +82-42-869-2915 Fax: +82-42-869-3110
1. INTRODUCTION

In this paper we derive a transform-free distribution of the steady-state queue length for the GI/G/1/K/V\textsubscript{M} queueing system under exhaustive FIFO (first-in first-out) service discipline. The case of unlimited queue capacity ($K \to \infty$) is also considered. \textit{V\textsubscript{M}} is the symbol used by Doshi [4] to represent the multiple vacation model: The server, when it becomes idle, keeps on taking vacations until, on return from a vacation, at least one customers is present. Let \(\{A_i\}_{i=1}^{\infty}, \{S_i\}_{i=1}^{\infty},\) and \(\{V_i\}_{i=1}^{\infty}\) denote the sequences of interarrival times, service times, and vacations, respectively. \(\{A_i\}, \{S_i\},\) and \(\{V_i\}\) form three independent renewal sequences with generic variables \(A, S,\) and \(V.\)

In his famous survey paper, Doshi [4] states that “Very little, if anything, is known about the queue length distribution for GI/G/1 queues with server vacations. This is no doubt due to the difficulty of dealing with supplementary variables to obtain a Markov process related to the queue length. ... Waiting times, on the other hand, are much easier to deal with.”

The supplementary variable technique (SVT) was firstly used by Kosten [9], and becomes one of the most popular approaches nowadays. Our approach is also basically SVT. However, we take an unconventional approach in the last step of SVT. The first step of SVT is to define a Markov process by including appropriate supplementary variables into the state vector. The second step is to set up system equations. Typically, the joint densities of the state vector at \(t + dt\) are related with those at \(t\) through mutually exclusive events which
can occur during \((t, t + dt)\). These relations are then converted into partial
differential equations which, in the limit \(t \to \infty\), become the steady-state
equations. The last step is to solve these equations. The conventional ap-
proach is to transform (such as Laplace transform) each equation and, in case
of infinite equations (e.g., when \(K \to \infty\)), combine the transformed equa-
tions into a form of probability generating function (see, e.g., Takagi [14]).
Our approach, on the other hand, is to directly integrate each equation after
multiplying a supplementary variate. An approach similar to ours can be
found in Kopocińska and Kopociński [8].

The result thus obtained is the steady-state queue length distribution ex-
pressed in terms of conditional expectations of the remaining (or residual)
\(A\), the remaining \(S\), and the remaining \(V\), conditional on respective numbers
of customers in the system as seen by an arriving customer, as left behind by
a departing customer, and as seen by the server on return from a vacation.

Some related references are as follows: Steady-state waiting time distribu-
tions are available for the G(I)/G/1/V_M queue (Doshi [5], Miyazawa [12])
((Doshi [3])). On the other hand, the steady-state queue length distribu-
tions are available for the MAP/G/1/(K)V_M queue (Lucantoni et al. [10])
((Blondia [2])) and the MGE/G/1/V_M queue (Bertsimas and Mourtzi-
nou [1]), where MAP stands for the Markovian arrival process and MGE
the mixed generalized Erlang.
2. MAIN RESULTS

The Markov process we consider is \( \{N(t), A_R(t), S_R(t), V_R(t); t > 0\} \), where \( N(t) \) denotes the number of customers in the system at \( t \) and the supplementary variables \( A_R(t) \), \( S_R(t) \), and \( V_R(t) \) respectively denote the remaining interarrival time, the remaining service time, and the remaining vacation, all at \( t \). By convention, \( S_R(t) = 0 \) if the server is on vacation at \( t \); and \( V_R(t) = 0 \) if the server is busy at \( t \).

For notational parsimony, let \( X \) denote \( \lim_{t \to \infty} X(t) \) for any variable \( X(t) \), and define the probability density functions as follows:

\[
\begin{align*}
a(x)dx &= \Pr\{x < A \leq x + dx\}, \\
s(y)dy &= \Pr\{y < S \leq y + dy\}, \\
v(y)dy &= \Pr\{y < V \leq y + dy\},
\end{align*}
\]

\[
\begin{align*}
f_n(x, y)dxdy &= \Pr\{N = n, x < A_R \leq x + dx, S_R = 0, y < V_R \leq y + dy\}, \\
&\quad \text{for } 0 \leq n \leq K, \\
g_n(x, y)dxdy &= \Pr\{N = n, x < A_R \leq x + dx, y < S_R \leq y + dy, V_R = 0\}, \\
&\quad \text{for } 1 \leq n \leq K.
\end{align*}
\]

By considering mutually exclusive events that can occur during \( dt \), we write the steady-state system equations as follows:

\[
\begin{align*}
f_0(x - dt, y - dt) &= f_0(x, y) + f_0(x, 0)v(y)dt + g_1(x, 0)v(y)dt,
\end{align*}
\]
\[ f_n(x - dt, y - dt) = f_n(x, y) + f_{n-1}(0, y)a(x)dt, \quad 1 \leq n \leq K - 1, \]
\[ f_K(x - dt, y - dt) = f_K(x, y) + f_{K-1}(0, y)a(x)dt + f_K(0, y)a(x)dt, \]
\[ g_n(x - dt, y - dt) = g_n(x, y) + g_{n-1}(0, y)a(x)dt + g_{n+1}(x, 0)s(y)dt + f_n(x, 0)s(y)dt, \quad 1 \leq n \leq K - 1, \]
\[ g_K(x - dt, y - dt) = g_K(x, y) + g_{K-1}(0, y)a(x)dt + g_K(0, y)a(x)dt + f_K(x, 0)s(y)dt, \]

where \( g_0(0, y) = 0 \), thereby we save an equation.

Converting these equations into partial differential equations, we have

\[
\begin{align*}
-(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})f_0(x, y) &= f_0(x, 0)v(y) + g_1(x, 0)v(y), \\
-(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})f_n(x, y) &= f_{n-1}(0, y)a(x), \quad 1 \leq n \leq K - 1, \\
-(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})f_K(x, y) &= f_{K-1}(0, y)a(x) + f_K(0, y)a(x),
\end{align*}
\]

\[
\begin{align*}
-(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})g_n(x, y) &= g_{n-1}(0, y)a(x) + g_{n+1}(x, 0)s(y) + f_n(x, 0)s(y), \quad 1 \leq n \leq K - 1, \\
-(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})g_K(x, y) &= g_{K-1}(0, y)a(x) + g_K(0, y)a(x) + f_K(x, 0)s(y).
\end{align*}
\]

First, we integrate (1) and (2), both over \( x \) and \( y \), \( 0 \leq x, y < \infty \). Simplifying
the results, we have
\[
\int_0^\infty f_n(0,y)\,dy + \int_0^\infty g_n(0,y)\,dy = \int_0^\infty g_{n+1}(x,0)\,dx, \quad 0 \leq n \leq K - 1, \quad (3)
\]
\[
\begin{align*}
\int_0^\infty f_n(x,0)\,dx &= \int_0^\infty f_{n-1}(0,y)\,dy - \int_0^\infty f_n(0,y)\,dy, \quad 1 \leq n \leq K - 1, \\
\int_0^\infty f_K(x,0)\,dx &= \int_0^\infty f_{K-1}(0,y)\,dy.
\end{align*}
\quad (4)
\]

In order to express (3) and (4) in terms of meaningful quantities, we define
\[
P_n^A(V) = \Pr\{N = n, \text{ Server is on vacation} \mid 0 < A_R \leq dx\}, \quad 0 \leq n \leq K,
\]
\[
P_n^A(B) = \Pr\{N = n, \text{ Server is busy} \mid 0 < A_R \leq dx\}, \quad 1 \leq n \leq K,
\]
\[
P_n^A = P_n^A(V) + P_n^A(B), \quad 0 \leq n \leq K; (P_0^A(B) = 0)
\]
\[
P_n^D = \Pr\{N = n+1 \mid 0 < S_R \leq dx \cup \{N = K, 0 < A_R \leq dx\}\}, \quad 0 \leq n \leq K - 1,
\]
\[
P_K^D = \Pr\{N = K, 0 < A_R \leq dx \mid 0 < S_R \leq dx \cup \{N = K, 0 < A_R \leq dx\}\}.
\]

Note that $P_n^A$ ($P_n^D$) is the probability that an arriving customer finds (a departing customer leaves behind) $n$ customers in the system. Also for convenience, we assume that blocked customers, who see $K$ customers in the systems upon arrival, immediately depart the system leaving behind $K$ customers. That is, we take into consideration not only unblocked customers but also blocked customers. Consequently, both the arrival rate and the departure rate are $\lambda = 1/E[A]$. Note that, according to our assumption, $P_K^D = P_K^A = \Pr\{N = K \mid 0 < A_R \leq dx\}$. When $K \to \infty$, we require $\rho = \lambda E[S] < 1$ to ensure stability of the GI/G/1/V_M queue.
Based on the assumption mentioned above, we identify (3) as \( \lambda P_n^A(V) + \lambda P_n^A(B) = \lambda P_n^D \). Thus, together with \( \lambda P_K^A(V) + \lambda P_K^A(B) = \lambda P_K^D \), we have

\[
P_n^A(V) + P_n^A(B) = P_n^D, \quad 0 \leq n \leq K,
\]

which is known as the Burke’s theorem (see Takagi [14, p. 7]). Likewise, we rewrite (4) as

\[
\begin{align*}
\xi_n & = \lambda P_{n-1}^A(V) - \lambda P_n^A(V), \quad 1 \leq n \leq K - 1, \\
\xi_K & = \lambda P_{K-1}^A(V),
\end{align*}
\]

where \( \xi_n \) denotes \( \int_0^\infty f_n(x,0)dx \), \( 0 \leq n \leq K \), and is interpreted as the expected frequency per unit time that the server returns from a vacation finding \( n \) customers waiting in the system.

Now, in order to express the results of the next procedure, we define the following probabilities and conditional expectations:

\[
\begin{align*}
P_n(V) & = \Pr\{N = n, \text{ Server is on vacation}\}, \quad 0 \leq n \leq K, \\
P_n(B) & = \Pr\{N = n, \text{ Server is busy}\}, \quad 1 \leq n \leq K, \\
P_n & = P_n(V) + P_n(B), \quad 0 \leq n \leq K, \quad (P_0(B) = 0)
\end{align*}
\]

\[
\begin{align*}
v_n & = E[V_R|N = n, 0 < A_R \leq dx, \text{ Server is on vacation}], \quad 0 \leq n \leq K, \\
\sigma_n & = E[S_R|N = n, 0 < A_R \leq dx, \text{ Server is busy}], \quad 1 \leq n \leq K, \\
\alpha_n^V & = E[A_R|N = n, 0 < V_R \leq dx, \text{ Server is on vacation}], \quad 0 \leq n \leq K, \\
\alpha_n^D & = E[A_R|N = n + 1, 0 < S_R \leq dx, \text{ Server is busy}], \quad 0 \leq n \leq K - 1.
\end{align*}
\]
Note that $v_n$ ($\sigma_n$) is the expected remaining vacation (service time) at the arrival point of a customer who finds $n$ customers in the system. Likewise, $\alpha_n^V$ and $\alpha_n^D$ are expected remaining interarrival times respectively at the point the server returns from a vacation finding $n$ customers in the system and at the departure point of a customer who leaves $n$ customers behind in the system.

Next, we multiply $x$ to both sides of (1) and then integrate over $x$ and $y$, $0 \leq x, y < \infty$. The procedure, which is omitted due to space consideration, is long but straightforward. It is mostly integration by parts under the assumption of finite first moments. Expressing the results using the identities $\lambda P_n^A(V)v_n = \int_0^\infty y f_n(0,y)dy$, $\xi_n\alpha_n^V = \int_0^\infty x f_n(x,0)dx$, $\lambda P_n^A(B)\sigma_n = \int_0^\infty y g_n(0,y)dy$, and $\lambda P_n^D\alpha_n^D = \int_0^\infty x g_{n+1}(x,0)dx$, we have

$$P_0(V) = \lambda P_0^D \alpha_0^D;$$

$$P_n(V) = P_{n-1}^A(V) - \xi_n \alpha_n^V, \quad 1 \leq n \leq K-1,$$

$$P_K(V) = P_{K-1}^A(V) + P_K^A(V) - \xi_K \alpha_K^V.$$  \hspace{1cm} (8)

On the other hand, multiplying $y$ to both sides of (1) and integrating over $x$ and $y$, $0 \leq x, y < \infty$, we have

$$\{\lambda P_0^A(V)v_0 + P_0(V)\}/E[V] = \lambda P_0^D + \xi_0;$$

$$P_n(V) = \lambda P_{n-1}^A(v_{n-1} - \lambda P_n^A(v_n), \quad 1 \leq n \leq K-1,$$

$$P_K(V) = \lambda P_{K-1}^A(v_{K-1}).$$  \hspace{1cm} (10)
Then we obtain $P_n^A(V)$ and $P_n(V)$, $1 \leq n \leq K$, by solving (6), (8) and (10) simultaneously as follows: ($P_0^A(V)$ and $P_0(V)$ will be determined later.)

\[
P_n^A(V) = P_0^A(V) \prod_{i=1}^{n} \beta_i, \quad 1 \leq n \leq K, \tag{11}
\]

\[
P_n(V) = \lambda P_n^A(V) \gamma_n, \quad 1 \leq n \leq K, \tag{12}
\]

where

\[
\beta_n = \frac{\{v_{n-1} + \alpha_n^V - E[A]\}}{(v_n + \alpha_n^V)}, \quad 1 \leq n \leq K - 1, \tag{13}
\]

\[
\gamma_n = \frac{\{v_{n-1}\alpha_n^V - v_n\alpha_n^V + v_n E[A]\}}{\{v_{n-1} + \alpha_n^V - E[A]\}}, \quad 1 \leq n \leq K - 1, \tag{14}
\]

\[
\beta_K = \frac{\{v_{K-1} + \alpha_K^V - E[A]\}}{E[A]},
\]

\[
\gamma_K = \frac{v_{K-1} E[A]}{\{v_{K-1} + \alpha_K^V - E[A]\}}.
\]

The procedure of obtaining $P_n^A(B)$ and $P_n(B)$ is the same. Multiplying $x$ to (2) and integrating over $x$ and $y$, $0 \leq x, y < \infty$, we have

\[
\begin{align*}
P_n(B) &= P_{n-1}^A(B) + \lambda P_n^D \alpha_n^D + \xi_n \alpha_n^V - \lambda P_{n-1}^D \alpha_{n-1}^D, \\
& \quad 1 \leq n \leq K - 1, \\
P_K(B) &= P_{K-1}^A(B) + P_K^A(B) + \xi_K \alpha_K^V - \lambda P_{K-1}^D \alpha_{K-1}^D.
\end{align*}
\tag{15}
\]

Multiplying $y$ to (2) and integrating over $x$ and $y$, $0 \leq x, y < \infty$, we have

\[
\begin{align*}
P_n(B) &= \lambda P_{n-1}^A(B) \sigma_{n-1} + \rho P_n^D + \xi_n E[S] - \lambda P_n^A(B) \sigma_n, \\
& \quad 1 \leq n \leq K - 1, \\
P_K(B) &= \lambda P_{K-1}^A(B) \sigma_{K-1} + \xi_K E[S].
\end{align*}
\tag{16}
\]

9
Solving (5), (6), (15) and (16) simultaneously, we obtain

\[ P_n^A(B) = P_{n-1}^A(B)\delta_n + P_{n-1}^A(V)\varepsilon_n, \quad 1 \leq n \leq K; \quad (17) \]
\[ P_n(B) = \lambda P_n^A(B)\theta_n + \lambda P_n^A(V)\phi_n, \quad 1 \leq n \leq K; \quad (18) \]

where

\[ \delta_n = \{\alpha_n^D + \sigma_n - E[A]/\{\alpha_n^D + \sigma_n - E[S]\}, \quad 1 \leq n \leq K - 1, \quad (19) \]
\[ \varepsilon_n = \{\alpha_n^D - \alpha_n^V + E[S]/\{\alpha_n^D - \alpha_n^V + E[S]\}/\{\alpha_n^D + \sigma_n - E[S]\}, \quad 1 \leq n \leq K - 1, \quad (20) \]
\[ \theta_n = E[S] - \sigma_n + \sigma_{n-1}/\delta_n, \quad 1 \leq n \leq K - 1, \quad (21) \]
\[ \phi_n = \{E[S] - \sigma_n - \varepsilon_n/\delta_n\}/\beta_n, \quad 1 \leq n \leq K - 1, \quad (22) \]
\[ \delta_K = \{\alpha_{K-1}^D + \sigma_{K-1} - E[A]/E[A], \quad (23) \]
\[ \varepsilon_K = \{\alpha_{K-1}^D - \alpha_K^V + E[S]/E[A], \]
\[ \theta_K = \sigma_{K-1}/\delta_K \]
\[ \phi_K = \{E[S] - \varepsilon_K\theta_K\}/\beta_K. \quad (24) \]

For the stable GI/G/1/V_M queue, equations (11) through (14) and (17) through (22) are valid for \( n \geq 1 \).

Finally, we need \( P_0^A(V) \) and \( P_0(V) \). In principle, \( P_0^A(V) \) can be determined by normalization \( 1 = \sum_{n=0}^K P_n^A(V) + \sum_{n=1}^K P_n^A(B) \). \( P_0(V) \) is then obtained by (5) and (7). Note that \( P_0^A(V) = P_0^D \) since \( P_0^A(B) = 0 \). A more efficient way to normalize by utilizing (9) is given in the Appendix.
3. CONCLUDING REMARKS

For GI/G/1/K/V_M queue under exhaustive FIFO discipline, we derived a steady-state (time-average) queue length distribution, \( P_n(V) \) and \( P_n(B) \), as given in (12) and (18). As a byproduct, we also obtained an arrival time queue length distribution, \( P_n^A(V) \) and \( P_n^A(B) \), as given in (11) and (17). Queue length distributions for other vacation models, such as the single vacation model, could be obtained by the same approach as in this paper, once the state vector of the Markov process is modified appropriately. For the GI/G/1/K queue, the results of Franken et al. [6, p. 115–124] and Kim and Chae [7] turn out to be \( P_n^A = P_{n-1}^A \delta_n \) and \( P_n = \lambda P_n^A \theta_n \), \( 1 \leq n \leq K \), where \( \delta_n \) and \( \theta_n \) are as given in (19), (21), (23), and (24).

Our results are expressed in terms of \( v_n, \sigma_n, \alpha_n^V \), and \( \alpha_n^D \) which are conditional expectations of the supplementary variables. Note that \( \alpha_n^D = \alpha_n^V = E[A] \) for M/G/1/K/V_M, \( \sigma_n = E[S] \) for GI/M/1/K/V_M, and \( v_n = E[V] \) for GI/G/1/K/V_M with exponential \( V \), due to memoryless \( A, S \), and \( V \), respectively. In general, however, these conditional expectations are not easy to compute. On the other hand, approximation for \( P_n(V) \) and \( P_n(B) \) can be made via approximating \( v_n, \sigma_n, \alpha_n^V \), and \( \alpha_n^D \). The quickest approximation is to replace all the conditional expectations with the unconditional counterparts, e.g., \( v_n \approx E[V_R] = E[V^2]/2E[V] \) (for the GI/G/1/K queue, this simple approximation works fairly well (Kim and Chae [7])). Improvements over this quickest approximation are left to the interested readers.
The final remark is on our unconventional approach. Our approach is basically SVT but, in the last step of solving system equations, we multiply a supplementary variate \( x \) or \( y \) and then integrate over both \( x \) and \( y \). As a result, we obtain equations for queue length probabilities in terms of conditional expectations of the supplementary variables. (If we multiply \( x^2 \) or \( y^2 \) and then integrate over both \( x \) and \( y \), we would obtain equations in terms of conditional second moments of the supplementary variables.) We believe that our approach belongs to a more fundamental approach called the rate conservation law. See Miyazawa [11] for a survey on the rate conservation law. See also Sigman [13] and Bertsimas and Mourtzinou [1] for efforts to unify fundamental approaches including the generalized Little’s formula \( H = \lambda G \). Our approach, however, is simpler to handle and easier to understand.

APPENDIX: AN AID FOR NORMALIZATION

The right hand side of (9) is interpreted as the expected frequency per unit time that the server takes vacations. Thus, multiplying \( E[V] \) to the right hand side of (9), we get \( \sum_{n=0}^{K} P_n(V) \) which is the long-run proportion of time the server is on vacation. On the other hand, \( \sum_{n=0}^{K} P_n(V) \) is related to \( P^A_R(V) \) and \( P^A_R(B) \) through the Little’s formula as follows:

\[
\sum_{n=0}^{K} P_n(V) = 1 - \sum_{n=1}^{K} P_n(B) \\
= 1 - E[\text{number of customers being served}] \\
= 1 - \lambda_e E[S],
\]

(A.1)
where the effective arrival rate $\lambda_e$ equals $\lambda \{1 - P_K^A(V) - P_K^A(B)\}$. Consequently, from (5), (7), (9), and (A.1), we have

$$\lambda(v_0 + \alpha^{D}_0)P_0^A(V) = 1 - \rho \{1 - P_K^A(V) - P_K^A(B)\},$$

which is an aid for determining $P_0^A(V)$. For the stable GI/G/1/V_M queue, we have

$$\lambda(v_0 + \alpha^{D}_0)P_0^A(V) = 1 - \rho = \sum_{n=0}^{\infty} P_n(V).$$

ACKNOWLEDGEMENT

This work was supported by Korea Research Foundation Grant (KRF-2002-070-C00021).

REFERENCES

[5] Doshi, B.T. Generalizations of the Stochastic Decomposition Results for


