

APPROXIMATE SOLUTIONS FOR SOME ADVANCED MULTISERVER RETRIAL QUEUES

Yang Woo Shin¹ and Dug Hee Moon²

¹Department of Statistics, Changwon National University, 641-773, Changwon, Republic of Korea

²Department of Industrial and Systems Engineering, Changwon National University,
641-773, Changwon, Republic of Korea

Keywords: Multiserver retrial queue, PH-retrial time, Impatient customers, Multiclass customers, Approximations.

Abstract: Retrial queues have been widely used for modelling many practical problems arising in computer and communication systems. It has been known to be difficult problems to develop a numerical algorithm or an approximate solution for advanced multiserver retrial queues such as the models with phase type distribution of retrial time, impatient customers governed by a general persistence function and multiclass of customers. Recently, we have developed an approximation method based on the approach in Fredericks and Reisner (1979) with some modifications for the advanced systems described above. In this paper, we introduce the approximation results developed recently.

1 INTRODUCTION

The retrial queueing system consists of a service facility with finite capacity and an orbit of an infinite size. A customer enters the service facility if the service facility is not full upon arrival. Otherwise, the customer joins orbit and repeats its request after random amount of time. The time interval between two consecutive attempts of each customer in orbit is called a retrial time.

Retrial queue has been widely used for modelling many practical problems arising in computer and communication systems. Even for the basic $M/M/c$ retrial queues with multiple servers, the exact results have not been obtained except for some special cases. Instead, attempts to develop algorithmic or approximation methods have been made for long time. For the Markovian retrial queues with multiple servers with Poisson arrival, exponential service and exponential retrial time, some algorithms and approximations are presented, e.g. (Neuts and Rao, 1990; Greenberg and Wolf, 1987; Fredericks and Reisner, 1979) and for more details of retrial queues, refer the monographs (Artalejo and Gómez-Corral, 2008; Falin and Templeton, 1997) and references therein. Neuts and Rao (1990) propose a generalized truncation method that uses the system in which only finite number of customers in orbit can retry for approximation of the basic $M/M/c$ retrial queue. Neuts and

Rao's method uses the matrix analytic method for computing the approximate system and its feasibility depends on the size of matrix components of the approximate system. Greenberg and Wolf (1987) present an approximation for the stationary distribution of the number of customers in service facility in the $M/M/c/K$ retrial queue under the assumption that retrials see time averages (RTA). The approximation using RTA assumption does not depend on the retrial rate and works well only for small value of retrial rate. Fredericks and Reisner (1979) develop an approximation for the number of customers at service facility in $M/M/c$ retrial queue with impatient customers. Fredericks and Reisner's method reflects the retrial rate and works well when the retrial rate γ is small, but it becomes worse as γ increases (Artalejo and Gómez-Corral, 2008).

Kulkarni and Liang (1997) suggest some open problems which include developments of analytical, numerical, and approximate solutions for advanced models such as the retrial queue with general retrial time, impatient customers, general service time. Recently, we have developed an approximation method based on the approach in Fredericks and Reisner (1979) with some modifications for the advanced systems. Objective of this paper is to introduce the recent results developed by Shin and Moon (Shin and Moon, 2011a; Shin and Moon, 2011b; Shin and Moon, 2011c).

Fredericks and Reisner's method with modification is described briefly in Section 2. Approximations of the $M/M/c$ retrial queue with phase type retrial time, impatient customers and multiclass customers are briefly described in Sections 3-5.

2 APPROXIMATION OF THE BASIC MODEL

Consider an $M/M/c$ retrial queue with no waiting space in service facility. Customers arrive from outside according to a Poisson process with rare λ and the service time and inter retrial time distributions are exponential with rates μ and γ , respectively. Let $C(t)$ be the number of customers at service facility and $N(t)$ be the number of customers in orbit at time t in the $M/M/c$ retrial queue. Assume the stability condition $\rho = \lambda/(s\mu) < 1$. Then $\mathbf{X} = \{(C(t), N(t)) | t \geq 0\}$ is a Markov chain on the state space $\mathcal{S} = \{0, 1, 2, \dots, c\} \times \{0, 1, 2, \dots\}$.

Let (C, N) be the stationary version of $(C(t), N(t))$ and $P(i, j) = P(C = i, N = j)$, $(i, j) \in \mathcal{S}$. The balance equations are given as follows:

$$\begin{aligned} (\lambda + i\mu + j\gamma)P(i, j) &= \lambda P(i - 1, j) \\ + (i + 1)\mu P(i + 1, j) + (j + 1)\gamma P(i - 1, j + 1), \\ 0 \leq i < c, j \geq 0 \end{aligned} \tag{1}$$

$$\begin{aligned} (\lambda + c\mu)P(c, j) &= \lambda P(c - 1, j) + \lambda P(c, j - 1) \\ + (j + 1)\gamma P(c - 1, j + 1), j \geq 0, \end{aligned} \tag{2}$$

where $P(i, j) = 0$ for $(i, j) \notin \mathcal{S}$. Define

$$P_i = P(C = i), N_i = \mathbb{E}[N | C = i], 0 \leq i \leq c.$$

Summing over j in (1) yields

$$(\lambda_i + \mu_i)P_i = \lambda_{i-1}P_{i-1} + \mu_{i+1}P_{i+1}, 0 \leq i < c, \tag{3}$$

where $\mu_i = i\mu, P_{-1} = 0$ and

$$\lambda_i = \lambda + \gamma N_i, 0 \leq i \leq c - 1. \tag{4}$$

Thus the stationary distribution of the number of busy servers is identical to that of finite birth-and-death process with birth rates $\{\lambda_j\}_{j=0}^{c-1}$ and death rates $\{\mu_j\}_{j=1}^c$ and is given by

$$P_i = P_0 \prod_{k=1}^i \left(\frac{\lambda_{k-1}}{\mu_k} \right), i = 1, 2, \dots, c \tag{5}$$

with

$$P_0 = \left[1 + \sum_{i=1}^c \prod_{k=1}^i \left(\frac{\lambda_{k-1}}{\mu_k} \right) \right]^{-1}.$$

Let $u_j = P(N = j), j \geq 0$. Summing over i in (1) and (2), we have that

$$\begin{aligned} (j + 1)\gamma u_{j+1} - j\gamma u_j &= \lambda(P(c, j) - P(c, j - 1)) \\ + \gamma((j + 1)P(c, j + 1) - jP(c, j)), j \geq 0, \end{aligned} \tag{6}$$

and hence the mean $L = \mathbb{E}[N]$ is given by

$$L = \frac{\lambda_c P_c}{\gamma}, \tag{7}$$

where $\lambda_c = \lambda + \gamma N_c$. Let R_i be the proportion of returning customers from orbit who find the service facility in state i , that is,

$$R_i = \frac{\gamma N_i P_i}{\sum_{j=0}^c \gamma N_j P_j} = \frac{N_i P_i}{L}, 0 \leq i \leq c. \tag{8}$$

It can be seen from (7) and (8) that

$$\lambda_c = \frac{\lambda}{1 - \gamma R_c}.$$

We have from (4) and (8) that

$$\lambda_i = \lambda + \frac{\lambda_i \dots \lambda_c}{\mu_{i+1} \dots \mu_c} R_i, 0 \leq i \leq c - 1. \tag{9}$$

From the observations above, we adopt the following approximation assumption.

Assumption A. The service facility behaves like a birth-and-death process with birth rates $\{\lambda_i\}_{i=0}^{c-1}$ and death rates $\{\mu_i\}_{i=1}^c$ and is independent of the retrials.

Let Q be the generator of the birth-and-death process with birth rates $\{\lambda_i\}_{i=0}^{c-1}$ and death rates $\{\mu_i\}_{i=1}^c$ and $q_{ij}(t)$ be the transition probability of Q and $\tilde{q}_{ij}(\theta)$ the Laplace transform of $q_{ij}(t)$. To determine R_i , suppose that at time 0, a customer attempts to get service and it finds that all the servers are busy and joins orbit. This customer returns after an exponential time with parameter γ . Then, R_i is approximated by the fraction of returning customers from orbit that find the service facility in state i , that is,

$$R_i \approx \int_0^\infty q_{ci}(t) \gamma e^{-\gamma t} dt = \gamma \tilde{q}_{ci}(\gamma), 0 \leq i \leq c. \tag{10}$$

Remark. The approximation of R_i in (10) is the same as that in Fredericks and Reisner (1979).

Once an approximation of λ is given, the square matrix $\tilde{Q}(\theta) = (\tilde{q}_{ij}(\theta))$ can be calculated by the formula

$$\tilde{Q}(\theta) = (\theta I - Q)^{-1},$$

where I is the identity matrix of order $c + 1$.

Combining (9) and (10), we have the approximation formula for λ_i as follows

$$\lambda_i = \lambda + \lambda \frac{\lambda_i \dots \lambda_{c-1}}{\mu_{i+1} \dots \mu_c} \frac{\gamma \tilde{q}_{ci}(\gamma)}{1 - \gamma \tilde{q}_{cc}(\gamma)}, 0 \leq i \leq c - 1, \tag{11}$$

Now we describe the computational procedure for λ_j . Let $\boldsymbol{\lambda} = (\lambda_0, \dots, \lambda_{s-1})$ and denote by $g(\boldsymbol{\lambda})$ the right hand side of the equation (11) and define a sequence of $\boldsymbol{\lambda}^{(n)} = (\lambda_0^{(n)}, \dots, \lambda_{s-1}^{(n)})$ by

$$\boldsymbol{\lambda}^{(n)} = g(\boldsymbol{\lambda}^{(n-1)}), \quad n = 1, 2, \dots \quad (12)$$

with $\boldsymbol{\lambda}^{(0)} = (\lambda, \dots, \lambda)$. Repeating the successive substitution (12), we choose $\boldsymbol{\lambda}^{(n)}$ as an approximation of $\boldsymbol{\lambda}$ if it satisfies $\|\boldsymbol{\lambda}^{(n)} - \boldsymbol{\lambda}^{(n-1)}\|_\infty < \varepsilon$ for given tolerance $\varepsilon > 0$.

Numerical experiments provide that approximations for L_0 is the same as the exact result $L_0 = \frac{\lambda}{\mu}$. For improving the accuracy of approximation for $L = \mathbb{E}[N]$, a modification of approximation for L is provided as follows

$$\hat{L}(m) = L_{\text{Appr}}(m) + (L_{M/M/c} - L_{\text{Appr}}(m^*)), \quad (13)$$

where $L_{\text{Appr}}(m)$ is the approximation for L in the system with mean retrial time m and $L_{M/M/c}$ is the mean number of customers in queue for the ordinary $M/M/c$ queue and m^* is chosen to be large enough so that the variation of $L_{\text{Appr}}(m)$ for $m \geq m^*$ is negligible.

3 RETRIAL QUEUE WITH PH-RETRIAL TIME

Consider an $M/M/c$ retrial queue in which the retrial time is of phase type distribution $PH(\boldsymbol{\alpha}, \boldsymbol{\Gamma})$. For stability of the system $\rho = \frac{\lambda}{c\mu} < 1$ is assumed. Shin and Moon (2011a) proposed an approximation for this system as follows.

Let m be the number of phases of $PH(\boldsymbol{\alpha}, \boldsymbol{\Gamma})$. Let X_0 be the number of customers at service facility and X_i the number of customers in orbit whose retrial phase is of i in stationary state and $L_i = \mathbb{E}[X_i]$, $0 \leq i \leq m$. Then, it can be shown that $\pi_j = P(X_0 = j)$, $0 \leq j \leq c$ satisfies the balance equation of the a birth-and-death process with birth rates $\{\lambda_j\}_{j=0}^{c-1}$ and death rates $\{\mu_j\}_{j=1}^c$, where $\mu_j = j\mu$ and λ_j can be expressed in terms of R_j that is the proportion of returning customers from orbit who find the service facility in state j as

$$\lambda_j = \lambda + \lambda \left(\prod_{i=j+1}^c \frac{\lambda_{i-1}}{\mu_i} \right) \frac{R_j}{1 - R_c}, \quad 0 \leq j < c.$$

Under the approximation assumption that *the service facility behaves like a birth-and-death process with birth rates $\{\lambda_j\}_{j=0}^{c-1}$ and death rates $\{\mu_j\}_{j=1}^c$ and is independent of the retrials*, R_j is approximated as the

probability that a returning customer finds the service facility in state j , that is,

$$R_j \approx \int_0^\infty q_{cj}(t) \boldsymbol{\alpha} \exp(\boldsymbol{\Gamma}t) \boldsymbol{\Gamma}^0 dt, \quad 0 \leq j < c,$$

where $q_{ij}(t)$ is the transition probability of Q . Then, λ_j is calculated by successive substitution. Numerical experiments provide that approximations for L_0 is the same as the exact result $L_0 = \lambda/\mu$. For improving the accuracy of approximation for $L = \sum_{k=1}^m L_k$, the modification (13) for L is used. In Table 1, the \hat{L} and the simulation are listed for two retrial time distributions $\text{MER}_3(0.0740741; \frac{4}{3m}, \frac{10}{3m})$ and $\text{CE}_{3,1}(0.007773; \frac{0.146991}{m}, \frac{1.188568}{m})$, where $\text{MER}_k(p; \nu_1, \nu_2) = p\text{Erlang}(k, \nu_1) + (1 - p)\text{Erlang}(k, \nu_2)$ is the mixture of two Erlang distributions $\text{Erlang}(k, \nu_i)$, $i = 1, 2$ of order k and $\text{CE}_{k,j}(p; \nu_1, \nu_2)$ the distribution whose Laplace-Stieltjes transform is of the form

$$f^*(s) = p \left(\frac{\nu_1}{\nu_1 + s} \right)^k \left(\frac{\nu_2}{\nu_2 + s} \right)^j + (1 - p) \left(\frac{\nu_2}{\nu_2 + s} \right)^j.$$

Table 1: L in $M/M/5$ retrial queue ($\rho = 0.8$).

Retrial Time	m	\hat{L}	Sim(c.i.)
MER_3 $C_V^2 = 0.5$	0.1	2.622	2.523(± 0.051)
	1.0	4.922	4.961(± 0.084)
	5.0	16.01	16.06(± 0.25)
	10.0	30.16	30.24(± 0.40)
$\text{CE}_{3,1}$ $C_V^2 = 5.0$	0.1	2.785	2.673(± 0.054)
	1.0	5.439	5.534(± 0.090)
	5.0	16.79	17.00(± 0.26)
	10.0	30.97	31.15(± 0.46)

4 RETRIAL QUEUE WITH IMPATIENT CUSTOMERS

Consider an $M/M/c/K$ retrial queue with impatient customers. The impatience of customers is governed by the persistence function $\{b_k, k = 1, 2, \dots\}$, where b_k is the probability that after the k th attempt fails, a customer will make the $(k + 1)$ st one. The retrial time is exponential whose rate may depend on the number of failures to enter the service facility. Let γ_k be the retrial rate of the customer that has experienced blocking k times. For a technical reason, the number of retrials of a customer from orbit is limited by m , that is, $b_k = 0$ for $k \geq m + 1$. Since $b_k = 0$ for $k \geq m + 1$, it can be easily seen that the system is always stable. Shin and Moon (2011b) proposed an approximation for this system as follows.

Let X_0 be the number of customers in service facility and X_k the number of customers in orbit who

have failed k times to enter the service facility. It can be shown that $P_{0j} = P(X_0 = j)$, $0 \leq j \leq c$ satisfies the balance equation of the a birth-and-death process with birth rates $\{\lambda_j\}_{j=0}^{K-1}$ and death rates $\{\mu_j\}_{j=1}^K$, where $\mu_j = \min(j, c)\mu$ and λ_j can be expressed in terms of R_{kj} that is the proportion of returning customers of type k who find the service facility of state j as

$$\lambda_j = \lambda + \lambda \left(\prod_{i=j+1}^K \frac{\lambda_{i-1}}{\mu_i} \right) \sum_{k=1}^m \left(\prod_{i=1}^{k-1} b_i R_{iK} \right) b_k R_{kj},$$

and R_{kj} ($0 \leq j \leq K, 1 \leq k \leq m$) is approximated by

$$R_{kj} \approx \int_0^\infty q_{Kj}(t) \gamma_k e^{-\gamma_k t} dt = \gamma_k \tilde{q}_{Kj}(\gamma_k).$$

In Table 2, approximation results (App.) are compared with the simulation results (Sim.) with half length of 95% confidence interval (c.i.) for L_0 , mean number L_{Orbit} of customers in orbit and the blocking probability $P_B = P_{0c}$ in the system with $\gamma_k = \gamma$, $k = 1, 2, \dots, m$.

Table 2: $M/M/5/7$ retrial queue with $\mu = 1.0, \rho = 1.2$.

		$(b_k = 0.85^{k-1}, k = 1, 2, 3, b_k = 0, k \geq 4)$		
γ		L_0	L_{Orbit}	P_B
0.1	App.	6.345	61.42	0.593
	Sim.	6.333	60.97	0.589
	c.i.	± 0.003	± 0.207	± 0.001
1.0	App.	6.252	5.997	0.567
	Sim.	6.141	5.835	0.541
	c.i.	± 0.005	± 0.024	± 0.002
5.0	App.	5.881	1.108	0.481
	Sim.	5.744	1.065	0.448
	c.i.	± 0.005	± 0.003	± 0.001

5 RETRIAL QUEUE WITH MULTICLASS OF CUSTOMERS

Consider an $M/M/c$ retrial queue with m different types customers. Let λ_i, μ_i and γ_i denote the arrival rate, service rate and retrial rate of type i customers (i -customers), respectively. Assume $\rho = \sum_{i=1}^m \frac{\lambda_i}{c\mu_i} < 1$ for stability of the system. Shin and Moon (2011c) proposed an approximation for this system as follows. Let C_i and N_i be the number of i -customers being served and in orbit in stationary state, respectively and set $\mathbf{C} = (C_1, \dots, C_m), \mathbf{N} = (N_1, \dots, N_m)$. Let $\mathcal{X} = \cup_{i=0}^c \mathcal{X}(i)$, where $\mathcal{X}(i) = \{(k_1, \dots, k_m) \in \mathbb{Z}_+^m : \sum_{j=1}^m k_j = i\}$ and $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$. It can be shown that $\pi(\mathbf{k}) = P(\mathbf{C} = \mathbf{k}), \mathbf{k} \in \mathcal{X}$ satisfies $\pi Q = 0$, where $\pi = (\pi(\mathbf{k}), \mathbf{k} \in \mathcal{X})$ and Q has same form as the generator of level dependent quasi-birth-and-death (LDQBD) process with finite level which describes

Table 3: $M/M/3$ retrial queue with $m = 2(\rho = 0.8)$.

Γ		P_B	L_1	L_2
0.5	App	0.5584	15.301	9.5116
	Sim	0.5630	15.361	9.5189
	c.i.	± 0.0025	± 0.1562	± 0.1027
1.0	App	0.5603	8.3544	5.4528
	Sim	0.5673	8.3597	5.4162
	c.i.	± 0.0025	± 0.1332	± 0.0693
5.0	App	0.5725	2.7856	2.1833
	Sim	0.5923	2.7696	2.1402
	c.i.	± 0.0036	± 0.0673	± 0.0424
10.0	App	0.5832	2.0811	1.7624
	Sim	0.6055	2.0321	1.7041
	c.i.	± 0.0034	± 0.0470	± 0.0324

the ordinary $M/M/c/c$ queue with m classes of customers in which the arrival rate $a_i(\mathbf{k})$ of i -customer depends on the system state $\mathbf{k} \in \mathcal{X}$ and service rate is $\mu_i, 1 \leq i \leq m$. The arrival rate $a_i(\mathbf{k})$ is given by

$$a_i(\mathbf{k}) = \lambda_i + \frac{\gamma_i L_i R_i(\mathbf{k})}{\pi(\mathbf{k})}, i = 1, 2, \dots, m,$$

where $L_i = \mathbb{E}[N_i], P_B = P(\mathbf{C} \in \mathcal{X}(c)), R_B(i) = \sum_{\mathbf{k} \in \mathcal{X}(c)} R_i(\mathbf{k})$ and $R_i(\mathbf{k})$ is the proportion of returning customers of i -customers from orbit who find the service facility in state \mathbf{k} . It can be shown that

$$L_i = \frac{\lambda_i P_B}{\gamma_i (1 - R_B(i))}, i = 1, 2, \dots, m.$$

Under the approximation assumption that the service facility behaves like a LDQBD process with generator Q and is independent of the retrials, $R_i(\mathbf{k})$ is approximated by

$$R_i(\mathbf{k}) \approx \frac{\gamma_i}{P_B} \sum_{\mathbf{j} \in \mathcal{X}(c)} \pi(\mathbf{j}) [(\gamma_i I - Q)^{-1}]_{\mathbf{j}\mathbf{k}}, i = 1, 2, \dots, m$$

and $a_i(\mathbf{k})$ is calculated by iteration.

Let $\Gamma = \sum_{i=1}^m \gamma_i, \beta_i = \frac{\gamma_i}{\Gamma}, 1 \leq i \leq m$ and $\beta = (\beta_1, \dots, \beta_m)$. It can be seen that multiclass $M/M/c$ retrial queue converges to the ordinary $M/M/c$ queue with discriminatory random order service (DROS) with the selecting probability $\beta = (\beta_1, \dots, \beta_m)$ denoted by $M/M/c/DROS(\beta)$ as Γ tends to infinity for fixed β . For the queue with DROS discipline, see (Kim et al., 2011). From this observations, a refinement of $L_{i,App}(\Gamma)$ is proposed by

$$\hat{L}_i(\Gamma) = L_{i,App}(\Gamma) + (L_{i,M/M/c/DROS(\beta)} - L_{i,App}(\Gamma^*)),$$

where $L_{i,M/M/c/DROS(\beta)}$ is the mean number of i -customers in queue for $M/M/c/DROS(\beta)$ and Γ^* is large enough so that the variation of $L_{i,App}(\Gamma)$ is negligible for $\Gamma \geq \Gamma^*$.

Table 3 lists the approximation results and simulation ones with 95% confidence interval (c.i.) for P_B

and L_i in $M/M/3$ retrial queue with $m = 2$ classes of customers, $\boldsymbol{\mu} = (1.0, 2.0)$, $\boldsymbol{\beta} = (0.2, 0.8)$ and the ratio of arrival rates $\boldsymbol{\alpha} = (0.3, 0.7)$, where $\alpha_i = \frac{\lambda_i}{\sum_{j=1}^m \lambda_j}$, $1 \leq j \leq m$.

ACKNOWLEDGEMENTS

The first and second authors were supported by Basic Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology, Grant Number 2009-0072282 and 2010-008957, respectively.

REFERENCES

- Artalejo, J. R. and Gómez-Corral, A. (2008). *Retrial Queueing Systems, A Computational Approach*. Springer-Verlag, Berlin.
- Falin, G. I. and Templeton, J. G. C. (1997). *Retrial Queues*. Chapman and Hall, London.
- Fredericks, A. A. and Reisner, G. A. (1979). Approximations to stochastic service systems with an application to a retrial model. *Bell Systems Technical Journal* 58, 557-576.
- Greenberg, B. S. and Wolf, R. W. (1987). An upper bound on the performance of queues with returning customer, *J. Appl. Probab.* 24, 466-475.
- Kim, J., Kim, J. and Kim, B. (2011). Analysis of the $M/G/1$ queue with discriminatory random order service policy, *Performance Evaluation*. 68, 256-270.
- Kulkarni, V. G. and Liang, H. M. (1997). Retrial queues revisited. In *Frontiers in Queueing, Models and Applications in Science and Engineering*, Chapter 2, pages 19-34, CRC.
- Neuts, M. F. and Rao, B. M. (1990). Numerical investigation of a multiserver retrial model. *Queueing Systems*. 7, 169-190.
- Shin, Y. W. and Moon, D. H. (2011a). Approximations of $M/M/c$ retrial queue with PH -retrial times. *European Journal of Operational Research*. 213:205-209.
- Shin, Y. W. and Moon, D. H. (2011b). Approximation of $M/M/s/K$ retrial queue with nonpersistent customers, submitted for publication.
- Shin, Y. W. and Moon, D. H. (2011c). Approximation of $M/M/c$ Retrial queue with multiclass of customers, submitted for publication.