A Novel Way of Treating the Finite-Buffer Queue *GI/M/c/N* Using Roots

James J. Kim and Mohan L. Chaudhry

Abstract**—**We present a new way of solving the model $GI/M/c/N$ using roots. By deriving and then solving the model's characteristic equation we are able to achieve this. The roots of the characteristic equation are quickly found and the solution is computed efficiently since it is entirely in terms of roots. The method presented embarks on the first application of the roots method in the finite-buffer multi-server queues and it remains robust even if the inter-arrival times follow heavy-tailed distributions. Some numerical results are provided.

Keywords**—**Finite-buffer, multi-server, roots, roots method, heavy-tailed.

I. Introduction

THE finite-buffer queues are characterized by the imposed limitation to the amount of waiting room such that when the line reaches a certain length, no further customers are allowed to enter until space becomes available.

The finite-buffer multi-server queues have been studied by various researchers (see e.g. Ferreira and Pacheco [2] and Laxmi [4]). In general, the finite-buffer multi-server queues can be solved in a few different ways. For instance, solving a set of balance equations leads to the steady-state queue-length distribution. In some cases the same result can be found through generating functions. The relations between various distributions can be developed through the use of a supplementary variable.

While the $GI/M/c/N$ queues are widely studied and are often extended to more sophisticated models in literature, the purpose of this paper is to introduce an alternative, but a novel solution procedure using the roots method. The robustness of the roots method, as demonstrated by Chaudhry et al. [1], remains effective even in the case of heavy-tailed inter-arrival times.

The heavy-tailed distributions constitute a class of probability distributions that are characterized by their slower decay than the light-tailed distributions. When considering the heavy-tailed distributions as an inter-arrival time distribution, the consensus among some researchers is that the roots method cannot be applied due to the unique probabilistic properties of the heavy-tailed random variables. In treating the $GI/M/c/\infty$ queues with heavy-tailed inter-arrival times, Harris et al. [3] state that "the standard root-finding problem gets complicated particularly when the inter-arrival time distribution possesses a complicated non-closed form or non-analytic Laplace-Stieltjes transform (L-S.T.)." This difficulty becomes even greater when applying the roots method to the $GI/M/c/N$ queues. This is because the model $GI/M/c/N$ has N roots while the model $GI/M/c/\infty$ has only one root (see later in Section 3 of this paper).

Nevertheless the heavy-tailed distributions are useful tools in modeling real life examples such as in telecommunications and financial engineering (see Willinger and Paxon [9], Leland et al. [5], Park et al. [6, 7], and Pitkow [8]). In particular, the heavy-tailed distributions (or synonymously referred to as the power, long or fat-tailed distribution) are useful when modeling the inter-arrival times of network packets and connection sizes under heavy traffic congestion (see Harris et al. [3]).

By expressing the entire queue-length distribution of the $GI/M/c/N$ queues in terms of the roots we are able to solve the model in an alternate, but a novel way that remains robust even if the inter-arrival times have a non-closed or a non-analytic form of L-ST.

II. MODEL DESCRIPTION

Consider the steady-state aspect of $GI/M/c/N$ queueing system where the service times and inter-arrival times are mutually independent. There are c parallel exponential servers, where each server has a service rate μ. Customers arrive at time epochs $T_1, T_2, \dots, T_n, \dots,$ and the inter-arrival times $t_{n+1} = T_{n+1} - T_n > 0, (n \geq 0)$ are identically independently distributed random variables (i.i.d.r.v.'s) with a probability density function (p.d.f.) $a(t)$, cumulative distribution function (c.d.f.) $A(t)$, mean $1/\lambda$ and L-S.T. $\bar{a}(s) = \int_0^\infty e^{-st} dA(t)$. Let $M(t)$ be the number of customers in the system at time t and put $M_n^- = M(T_n - 0), (n \geq 0)$. Thus M_n^- represents the number of customers in the system including the ones, if any, in service just before the arrival instant T_n. Its probability mass function (p.m.f.) or the queue-length-distribution is $p_j^- = \lim_{n\to\infty} P(M_n^- = j), (j \geq 0)$. Similarly, the queue-length-distribution at a random time epoch is $p_j = \lim_{n\to\infty} P(M_n = j), (j \geq 0)$. Let D_n be the total number of customer departures over the course of t_n with p.m.f. $k_l = \lim_{n\to\infty} \int_0^\infty P(D_n = l \mid t_n = t) dA(t), (l \geq 0)$, and probability generating function (p.g.f.) $K(z) = \sum_{l=0}^{\infty} k_l z^l$. The stochastic process $\{M_n^-, n \geq 1\}$ forms a homogenous Markov chain:

James J.Kim is with the Royal Canadian Air Force, Ottawa, Canada and Mohan L.Chaudhry is with Royal Military College of Canada, Ontario, Canada

$$M_{n+1}^- = \begin{cases} \min(M_n^- + 1 - D_n, N), & (M_n^- + 1 - D_n \geq 0) \\ 0, & (M_n^- + 1 - D_n < 0) \end{cases}$$

Since we deal with the steady-state solution of the problem under consideration, we assume that the traffic intensity of the system is $\rho = \frac{\lambda}{c\mu} > 0$. The model $GI/M/c/N$ has a finite-buffer $N, (N \geq c)$ such that an incoming customer is rejected if there are $(N - c)$ customers in queue at an arrival point.

III. QUEUE-LENGTH DISTRIBUTION AT PRE-ARRIVAL TIME EPOCH

To compute the queue-length distribution of the $GI/M/c/N$ queues at a pre-arrival time epoch we first define the transition probabilities of the model. Let $P_{i,j}(n) = P[M_n^- = j \mid M_0^- = i]$, $(i, j \geq 0, n \geq 1)$ be the n-step transition probabilities of $\{M_n^-, n \geq 0\}$. Thus, the one-step transition probabilities are defined as $P_{i,j} \equiv P_{i,j}(1)$. The $P_{i,j}$ are defined as

$$P_{i,j} = \begin{cases} k_{i+1-j}, & (j > i \geq 0, j \geq c) \\ k_{i+1-j} + k_{N-j}, & (c \leq j \leq i) \\ V_{i+1,j} + V_{N,j}, & (i \geq 0, 1 \leq j \leq c-1) \end{cases}$$

where $P_{i,0} = 1 - \sum_{j=1}^{N} P_{i,j}$, $k_n = \int_0^\infty \frac{(c\mu t)^n}{n!} e^{-c\mu t} dA(t)$, $(n \geq 0 \text{ and } k_n = 0 \text{ for } n < 0)$, and $V_{h,j}$ is defined as

$$V_{h,j} = \begin{cases} 0, & (h < j) \\ \int_0^\infty \binom{h}{j} (1 - e^{-\mu t})^{h-j} (e^{-\mu t})^j dA(t), & (1 \leq j \leq h \leq c) \\ \int_0^\infty \int_0^t \frac{e^{-c\mu u}(c\mu u)^{h-c-1}}{(h-c-1)!} c\mu \binom{c}{j} e^{-\mu j(t-u)} \left[1 - e^{-\mu(t-u)}\right]^{c-j} du\, dA(t), & (1 \leq j < c < h) \end{cases}$$

We define the Chapman-Kolmogorov equation for the $GI^X/M/c/N$ queues as

$$p_j^- = \sum_{i=0}^{N} p_i^- P_{i,j}, (0 \leq j \leq N) \quad (1)$$

which is a set of j first order linear difference equations. As a remark, $N = 0$ indicates that no customers are allowed in the system (i.e. $p_0^- = 1$). In presenting a novel way of treating the $GI/M/c/N$ queues using roots, we assume the solution of a general form $p_j^- = Cz^j, (1 \leq j \leq N, C \neq 0)$. By substituting the general solution into (1), we have

$$Cz^j = \sum_{i=0}^{N} Cz^i P_{i,j}, (1 \leq j \leq N)$$

$$0 = \sum_{i=0}^{N} z^i P_{i,j} - z^j$$

By summing both sides of the above over $1 \leq j \leq N$, we have the characteristic equation of the $GI/M/c/N$ queues as

$$0 = \sum_{j=1}^{N} \left(\sum_{i=0}^{N} z^i P_{i,j} - z^j \right) \quad (2)$$

Since the characteristic equation of the $GI/M/c/N$ queues is an N-th degree polynomial, solving it gives N roots. As a remark, In case (2) has repeated roots, modern computational software packages can find them. The MAPLE script below demonstrates the finding of repeated roots for the equation

$$g(x) = (x-1)(x-2)(x-3)^2(x-4)$$

```
restart : Digits := 10 :with(RootFinding) :
g := (x − 1)·(x − 2)·(x − 3)^2·(x − 4);
Analytic(g, x, re =−1 ..10, im =−2 ..10);
3.00000000000000, 3.00000000000000, 4.00000000000000,
    2.00000000000000, 1.00000000000000
```

Let these roots be $z_1, z_2, \ldots, z_N$ such that the solution becomes

$$p_j^- = \sum_{h=1}^{N} C_h z_h^j, (1 \leq j \leq N) \quad (3)$$

where $C_h, (1 \leq h \leq N)$ are the unknown non-zero constant coefficients. To determine these unknowns, we substitute (3) into (1) such that it leads to

$$\sum_{h=1}^{N} C_h z_h^j = \sum_{i=0}^{N} \sum_{h=1}^{N} C_h z_h^i P_{i,j}, (1 \leq j \leq N)$$

The above expression can be rearranged to

$$0 = \sum_{h=1}^{N} C_h \left(\sum_{i=j-h}^{N} z_h^i P_{i,j} - z_h^j \right), (1 \leq j \leq N) \quad (4)$$

To make (3) also true for the case when $j = 0$, we establish the normalizing condition as

$$1 = \sum_{j=0}^{N} \sum_{h=1}^{N} C_h z_h^j$$

The normalizing condition, in conjunction with letting $j = 1, 2, \ldots, N-1$ in (4), gives N equations. Solving these equations give the solution to the $GI/M/c/N$ queues in terms of roots as

$$p_j^- = \sum_{h=1}^{N} C_h z_h^j, (0 \leq j \leq N) \quad (5)$$

IV. QUEUE-LENGTH DISTRIBUTION AT RANDOM TIME EPOCH

The queue-length distribution of the $GI/M/c/N$ queues at a random time epoch (say p_j for $0 \leq j \leq N$) can be explicitly expressed as $p_j = \sum_{i=0}^{N} p_i^- P_{i,j}^*, (1 \leq j \leq N)$ where $p_0 = 1 - \sum_{j=1}^{N} p_j$ and $P_{i,j}^*$ is $P_{i,j}$ except $A(t)$ is replaced with $A_R(t)$ where $A_R(t) = \frac{1}{a} \int_0^t [1 - A(w)] dw, (0 < w \leq t)$. This way of computing the $p_j, (0 \leq j \leq N)$ is simpler than the proposed method by Laxmi [4]. As a remark, as $a \to \infty$ the $p_0 \to 1$ since $\sum_{j=1}^{N} \sum_{i=0}^{N} p_i^- P_{i,j}^* = 0$ (this phenomenon is numerically demonstrated later in Table 2 under Section 6). This property is unique to heavy-tailed inter-arrival times (see next Section) and can be understood as follows: An infinite mean inter-arrival time indicates that no customers are arriving while existing customers continue to get served. Hence, in this case, at a random time epoch, the system is empty.

V. The GI/M/c/N Queues involving heavy-tailed inter-arrival times

The use of heavy-tailed distributions in queueing theory allows niche applications to areas that resemble a system with extremities that cannot be estimated with light-tailed distributions (e.g. large income disparity and extreme delays of any sort). In treating heavy-tailed distributions as either an inter-arrival or service time distribution, previous methods offered workaround solutions that may consume large time and computing resources. Though such methods can be extended to solve the $GI/M/c/N$ queues involving heavy-tailed inter-arrival times, it is identified that our solution procedure remains robust and whether the inter-arrival time distribution is light or heavy-tailed.

VI. Numerical Results

Some numerical results of our analytical method are provided in Tables 1 (light-tail) and 2 (heavy-tail). All roots were accurately found and successfully used to compute the queue-length distribution at the pre-arrival and random time epochs. All results were computed to nine decimal places and in presenting our results they were rounded to four decimal places.

Table 1: $E_m/M/4/6$ with $a(t) = \frac{(m\lambda)^m t^{m-1} e^{-mt\lambda}}{(m-1)!}, (t > 0), \mu = 5, m = 2, \rho = 0.4, 1,$ and 1.2. This gives $\lambda = \rho c \mu m$ $(16, 40,$ and $48)$

	p_j^-		
j	$\rho = 0.4$	$\rho = 1$	$\rho = 1.2$
0	0.0320	0.0003	0.0001
1	0.1276	0.0031	0.0012
2	0.2250	0.0164	0.0082
3	0.2388	0.0549	0.0338
4	0.1753	0.1294	0.0989
5	0.1250	0.2866	0.2703
6	0.0762	0.5095	0.5876
Sum	1.0000	1.0000	1.0000

	p_j		
j	$\rho = 0.4$	$\rho = 1$	$\rho = 1.2$
0	0.0220	0.0002	0.0001
1	0.1024	0.0022	0.0009
2	0.2042	0.0123	0.0059
3	0.2400	0.0438	0.0261
4	0.1911	0.1097	0.0810
5	0.1403	0.2587	0.2373
6	0.1000	0.5731	0.6488
Sum	1.0000	1.0000	1.0000

*As a remark, when $m = 1$, then $p_j^- = p_j$ for $0 \leq j \leq N$ due to PASTA property.

Table 2: *Standard-Cauchy* $/M/7/10$ with $a(t) = \frac{2}{\pi(1+t^2)}, (t > 0), \lambda = 0, \rho = 0,$ and $\mu = 0.001,\ 0.01,$ and 0.1

	p_j^-		
j	$\mu = 0.001$	$\mu = 0.01$	$\mu = 0.1$
0	0.0003	0.0027	0.0458
1	0.0004	0.0041	0.0806
2	0.0005	0.0056	0.1123
3	0.0006	0.0074	0.1339
4	0.0008	0.0097	0.1403
5	0.0011	0.0128	0.1309
6	0.0015	0.0174	0.1097
7	0.0023	0.0246	0.0828
8	0.0045	0.0430	0.0641
9	0.0237	0.1219	0.0525
10	0.9643	0.7509	0.0473
Sum	1.0000	1.0000	1.0000

	p_j		
j	$\mu = 0.001$	$\mu = 0.01$	$\mu = 0.1$
0	1.0000	1.0000	1.0000
1	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000

6	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000
Sum	1.0000	1.0000	1.0000

*As a remark, when $\lambda = 0$ (i.e. an infinite mean inter-arrival time), the $p_0 = 1$ (see Section 4 for the analytical explanation).

VII. Conclusion

We have presented a new and simple approach to explicitly express the queue-length distribution in terms of the roots of the model's characteristic equation. The method not only gives an alternative solution but also unifies the theory in the sense that both the models $GI/M/c/\infty$ and $GI/M/c/N$ can be solved through roots. While our analytical method embarks on the first application of the roots method in the finite-buffer queues of the $GI/M/c$ type, it is identified that the roots method works well even if the inter-arrival times follow a heavy-tailed distribution. In the future, it is possible that the roots method can be extended to bulk-arrival or bulk-service multi-server finite-buffer queues, as well, in the discrete-time finite-buffer queues.

References

[1] Chaudhry, M.L., Harris, C.M., and Marchal, W.G.: Robustness of root finding in single-server queueing models. INFORMS Journal on Computing. 2, 273-286 (1990)

[2] Ferreira, F., Pacheco, A. Analysis of $GI^X/M/s/c$ systems via uniformisation and stochastic ordering. In T. Czachórski and N.Pekergin, editors, Procs.First Workship on New Trends in Modelling. Quantitative Methods and Measurement, Zakopane, Poland, 27/06/2004-29/06/2004, pp. 97-133 (2004)

[3] Harris, C.M., Brill, P.H., and Fischer, M.J.: Internet-Type Queues with Power-Tailed Interarrival Times and Computational Methods for Their Analysis. Informs Journal on Computing. 12(4), 261-271 (2000).

[4] Laxmi, P.V.: Modelling and Analysis of Some Finite Buffer Bulk-Arrival/Service Queues. Ph.D. Thesis, India Institute of Technology Kharagpur (2001)

[5] Leland, W., Taqqu, M., Willinger, W., and Wilson, D.: On the Self-Similar Nature of Ethernet Traffic (extended version). IEEE/ACM Trans. Networking. 2, 1-13 (1994).

[6] Park, K., Kim, G., and Crovella, M.: On the Relationship Between File Sizes, Transport Protocols, and Self-Similar Network Traffic. Proceedings 1996 International Conference on Network Protocols. Columbus, OH. 171-180 (1996)

[7] Park, K., Kim, G., and Crovella, M.: On the Effect of Traffic Self-Similarity on Network Performance. Proceedings of the SPIE Conference on Performance and Control of Network Systems. Boston, M.A. 296-310 (1997)

[8] Pitkow, J.: Summary of WWW Characterizations. World Wide Web. 2, 3-13 (1999)

[9] Willinger, W. and Paxon, M.S.: Where Mathematics Meets the Internet. Notices of the AMS. 45, 961-970 (1998)