EPH – International Journal of Mathematics and Statistics
ISSN:2208-2212
https://eijms.com/index.php/ms/issue/archive

ON SINGLE SERVER RETRIAL QUEUES

[1]Abdul Rof V
[1]Department of Mathematics,
Korambayil Ahamed Haji Memorial Unity Women's College, Manjeri, Malappuram, Kerala–676122, India
Mail - rofabdul@gmail.com

ARTICLE INFO

Received: 02/11/2021
Revised: 04/12/2021
Accepted: 12/12/2021
Published:20/12/2021

DOI:https://doi.org/10.69980/nadwt698
Volume:7, Issue:2, December, 2021

Abstract

Retrial queues refer to a class of stochastic queueing models in which customers finding the server busy leave the service area and retry for service after a random period of time. This paper studies retrial queues of both M/M/1 and M/G/1 types, with particular emphasis on the joint distribution of the server state and the number of customers in the orbit. For the M/M/1 retrial queue, the joint stationary distribution of the server state and the orbit size is obtained, and the corresponding probability generating functions are derived. Several performance measures are also discussed.

Keywords: Retrial Queues, M/M/1 Queue, Single-Server Queue; Markov Process, Stationary Distribution, Orbit Size; Probability Generating Functions, Queueing Theory.

1. Introduction

Retrial queues constitute an important class of stochastic queueing models with applications in telephone systems, computer networks, communication systems, and other service systems. The basic idea of a retrial queue can be illustrated by a telephone subscriber who obtains a busy signal when attempting to make a call. Since the subscriber cannot wait for the service, the call is terminated and the subscriber makes another attempt after some time. Thus, the flow of calls in a telephone network consists of two components: primary calls and repeated calls. A customer who fails to obtain service on the first attempt becomes a source of repeated trials.

The characteristics of such systems can be investigated using retrial queueing models. A retrial queueing system is characterized by the requirement that customers finding the service facility busy do not remain in the service area. Instead, they leave the service area and join a group of unsatisfied customers, commonly referred to as the orbit. Customers in the orbit independently make repeated attempts to access the service facility. This feature distinguishes retrial queues from conventional waiting-line models.

Retrial queues have practical applications in several computer and communication systems. They also arise in other situations, such as aircraft landing systems and retrial shopping models, where customers who encounter a long queue may leave the system and return later when the congestion is expected to be lower. The analysis of retrial queues is generally more involved than that of ordinary queueing systems because the number of customers in the orbit must be considered together with the state of the server. In the M/M/1 case, the memoryless property of the exponential service distribution permits a direct continuous-time Markov chain analysis.

In this paper, we investigate the stationary behaviour of M/M/1 retrial queues.

2. Literature Review

Retrial queues constitute an important class of stochastic queueing models in which customers who cannot obtain immediate service leave the service area and make repeated attempts after a random amount of time. Such models are particularly appropriate for telephone systems, communication networks, call centres, computer systems and other applications in which waiting customers do not remain in a physical queue. The development of retrial queueing theory can be traced to the study of repeated attempts in telephone traffic and subsequently to a wide range of stochastic systems.

One of the early comprehensive treatments of retrial queues was provided by Yang and Templeton [1]. Their survey summarized the principal results available for retrial queueing systems and established the foundations for subsequent research. Falin [2] subsequently presented an extensive survey of retrial queues, covering Markovian single-channel and multichannel systems together with models involving batch arrivals, customer classes, impatience, finite buffers, two-way communication and control mechanisms. The survey also considered stationary and transient behaviour and asymptotic results for different retrial regimes.

During the 1990s, research on retrial queues expanded considerably. Artalejo [4] provided a classified bibliography of research conducted during 1990–1999, documenting the rapid development of the subject and the increasing variety of retrial models. The literature during this period included single-server and multiserver systems, finite-source models, nonpersistent customers, priority mechanisms, batch arrivals and models with different retrial disciplines.

A major milestone was the monograph of Falin and Templeton [3], which provided a systematic treatment of retrial queues. The book covered stationary and transient distributions, busy-period analysis, waiting-time characteristics, limit theorems and multiserver retrial systems. It also considered advanced models such as finite-source systems, priority customers and nonpersistent subscribers. This work became one of the principal references for the mathematical analysis of retrial queueing systems.

Artalejo and Gomez-Corral [5] compared retrial queues with conventional waiting-line and loss systems. They emphasized that, although retrial queues are closely related to standard queueing models, the orbit of blocked customers produces distinct stochastic characteristics. Their analysis covered single-server M/G/1-type and multiserver M/M/c-type retrial systems and highlighted important similarities and differences between retrial and conventional queues.

The development of matrix-analytic and computational methods became particularly important in the analysis of retrial systems. Artalejo and GomezCorral [6] provided a comprehensive computational treatment of retrial queueing systems, including numerical algorithms and matrix-analytic techniques for models for which closed-form solutions are difficult or unavailable. The book also addressed approximation methods and computational evaluation of important performance measures.

The literature also expanded toward discrete-time retrial systems. Nobel [9] presented a survey of one-server discrete-time retrial models under late-arrival assumptions. Generating-function methods were used to study steady-state behaviour and performance measures such as orbit size and busy-period characteristics. The survey also emphasized the distinction between early-arrival and late-arrival conventions and discussed the relative tractability of generating-function and matrix-analytic approaches.

A further important development was the systematic survey of Kim and Kim [8]. They reviewed a broad class of retrial queueing systems with particular emphasis on queue-length distributions, waiting-time distributions, tail asymptotics and stability conditions. Their survey demonstrated that retrial models had developed far beyond the classical M/M/1 and M/M/c settings and had become relevant to a wide range of applied probability problems.

Artalejo [7] provided an updated bibliography covering the progress in retrial queues during 2000–2009. The bibliography documented the increasing diversity of models studied during this period, including discrete-time systems, finite-source models, multiserver systems, unreliable servers, vacation models and communication-network applications. During the 2010s, research increasingly focused on complex retrial systems incorporating server vacations, breakdowns and repairs, impatience, collisions, priority classes, working vacations and non-Poisson input processes. These extensions were motivated by the need to represent realistic communication and service systems more accurately. In particular, models with unreliable servers and repair mechanisms became an active research area because service interruptions can have a significant influence on both the orbit size and system performance.

Phung-Duc [10] presented a comprehensive survey of retrial queueing models and their applications. The survey reviewed exact analytical solutions, stability conditions, asymptotic analysis and multidimensional retrial models. It also discussed applications in call centers, cellular networks and random-access communication protocols. Particular attention was given to the development of tractable models and to open problems involving higherdimensional and application-oriented retrial systems.

Another direction of research concerned the behaviour of individual customers and their decisions to retry. Su and Veeraraghavan [11] introduced a framework for rational retrials in queues, in which customers explicitly account for the cost associated with returning to the system and the delay associated with service. This line of research connected retrial queueing theory with strategic and economic models of customer behaviour.

Research immediately preceding 2021 also considered more complicated finite-source and unreliable-server retrial systems. Such models incorporated collisions, server breakdowns and repairs and were motivated by applications in computer and communication systems. These developments illustrate the continuing movement from classical analytical retrial models toward multidimensional stochastic systems that capture several operational features simultaneously.

By 2021, retrial queueing theory had therefore developed into a broad research area encompassing classical Markovian retrial queues, multiserver systems, finite-source models, discrete-time systems, nonpersistent and impatient customers, unreliable servers, vacation models, priority systems, randomaccess protocols and strategic retrial behaviour. Despite substantial progress, analytical difficulties remain for models involving multiple interacting components, general service and retrial distributions, state dependent retrial rates and multidimensional state spaces. These limitations provide motivation for developing efficient matrix-analytic, generating-function and numerical approaches for new classes of retrial queueing systems. "'

3. Model Description

Consider a single server queueing system in which customers arrive in a Poisson process with rate λ.These customers are called primary customers.If the server is free at the arrival of a primary customer, the arriving customer begins to be served immediately and leaves the system after service completion. Otherwise,if the server is busy, the arriving customer becomes a customer in orbit, or a customer in pool.Each customer in orbit repeat the attempt,which produces a Poisson process with intensity μ.If one such customer find the server free,he is served and leaves the system forever after service.Otherwise he return to orbit.

The service time distribution is exponential for both primary and retrial customers, and the service rate is ν.

The input flow of primary calls, intervals between repetitions, and service times are mutually independent.

The queueing process evolves in the following manner:

Sppose the (i − 1)th customer completes his service at the epoch ηi−1 (the customers are numbered in the order of service) and the server become free.Evenif there are some customers in the system who want to get service,they cannot occupy the server immediately, because of their ignorance of the server state.Therefore the next, ith customer enters service only after some time interval Ri,during which the server is free,while there may be waiting customers.If the number of orbital customers at time, ηi−1, Ni−1 is n,then the random variable Ri has an exponential distribution with parameter (λ + nμ).The ith customer is a primary customer with probability, λ/(λ+nμ),and he is a retrial customer with probability, nμ/(λ+nμ).At the epoch ξi = ηi−1 +Ri, the ith customer,starts his service and continues during a time Si.All primary calls arriving during Si,goes to the orbit.They do not affect the service.At the epoch ηi = ξi + Ri,the ith customer completes his service and the server becomes free again.

at time t let N(t) be the number of orbital customers, C(t), the number of busy servers (server status).

i.e.C(t) = 1,if the server is busy, and C(t) = 0,if the server is idle. The process (C(t),N(t)) describes the number of customers in the system,in the simplest form.Since the service time distribution is exponential,the process (C(t),N(t)) is a Markov process.The rate of transition is given by the following infinitesimal generating matrix,

$$\begin{bmatrix} A_{00} & A_0 & 0 & . & . \\ A_{21} & A_{01} & A_0 & 0 & . \\ 0 & A_{22} & A_{02} & 0 & . \\ 0 & 0 & \ddots & \ddots & . \\ \vdots & . & . & \ddots & \ddots \end{bmatrix}$$

where,

$$A_{0n} = \begin{bmatrix} -(\lambda + n\mu) & \lambda \\ \nu & -(\lambda + \nu) \end{bmatrix}$$

$$A_0 = \begin{bmatrix} 0 & 0 \\ 0 & \lambda \end{bmatrix}$$

$$A_{2n} = \begin{bmatrix} 0 & n\mu \\ 0 & 0 \end{bmatrix}$$

4. Joint distribution of the server state and queue length in steady state

For an M/M/1 queue in the steady state,the joint distribution of the server state C(t) and queue length N(t), pin = P(C(t) = i,N(t) = n) is given by,

$$p_{0n} = \frac{\rho^n}{n!\mu^n} \prod_{i=0}^{n-1} (\lambda + i\mu)(1-\rho)^{\frac{\lambda}{\mu}+1}$$

$$p_{1n} = \frac{\rho^{n+1}}{n!\mu^n} \prod_{i=1}^{n} (\lambda + i\mu)(1-\rho)^{\frac{\lambda}{\mu}+1}$$

The corresponding partial generating function is given by,

$$P_0(z) := \sum_{n=0}^{\infty} z^n p_{0n} = (1-\rho) \left[\frac{1-\rho}{1-\rho z}\right]^{\frac{\lambda}{\mu}}$$

$$P_1(z) := \sum_{n=0}^{\infty} z^n p_{1n} = \rho \left[\frac{1-\rho}{1-\rho z}\right]^{\frac{\lambda}{\mu}+1}$$

We have (C(t),N(t)) is a Markov process whose state space is,
{0, 1}× Z+,where, Z+ is the set of non negative integers.
from the state (0,n) transitions into the folowing states are possible.
1. (1,n), with rate, λ
2.(1,n-1),with rate, nμ
The first is due to a primary arrival, and second is due to retrial of an orbital customer.
Reaching the state (0,n) is possible only from state (1,n)with rate, ν
From the state (1,n) the possible transitions are,
1.(1,n+1),with rate,λ 2.(0,n),with rate, ν the first is due to a primary arrival, and second is due to a service completion.
Reaching (1,n) is possible from
1.(0,n) with rate λ
In the infinitesimal interval (t,t+h),we have

$$p'_{0n} = -(\lambda + n\mu)p_{0n} + \nu p_{1n}$$

$$p'_{1n} = -(\lambda + \nu)p_{1n} + \lambda p_{0n} + (n+1)\mu p_{0,(n+1)} + \lambda p_{1,(n-1)}$$

In the long run, 0,
∴ we have,

$$(\lambda + n\mu)p_{0n} = \nu p_{1n} \text{..}(1)$$

$$(\lambda + \nu)p_{1n} = \lambda p_{0n} + (n+1)\mu p_{0,(n+1)} + \lambda p_{1,(n-1)} \text{................}(2)$$

Now consider the partial generating functions, $P_0(z) := \sum_{n=0}^{\infty} z^n p_{0n}$,and, $P_1(z) := \sum_{n=0}^{\infty} z^n p_{1n}$

Now from (1) and (2) we have,

$$\lambda P_0(z) + \mu z P'_0(z) = \nu P_1(z) \text{..}(3)$$

$$(\lambda + \nu)P_1(z) = \lambda P_0(z) + \mu P'_0(z) + \lambda z P_1(z) \text{...................}(4)$$

Eliminating P1(z) between (3) and (4),

$$P'_0(z) = \frac{\lambda\rho}{\mu(1-\rho z)} P_0(z)$$

where ρ = λ ÷ ν

Solving,

$$\log P_0(z) = \frac{C}{(1-\rho z)^{\frac{\lambda}{\mu}}}$$

where C is a constant.Again,

$$P_1(z) = \frac{C\rho}{(1-\rho z)^{\frac{\lambda}{\mu}+1}}$$

To find C we have,

$$\sum_{n=0}^{\infty}(p_{0n}+p_{1n}) = 1$$

i.e.

$$P_0(1) + P_1(1) = 1$$

So that,

$$\frac{C}{(1-\rho)^{\frac{\lambda}{\mu}}} + \frac{C\rho}{(1-\rho)^{\frac{\lambda}{\mu}+1}} = 1$$

$$\Rightarrow C = (1-\rho)^{\frac{\lambda}{\mu}+1}$$

Hence,

$$P_0(z) = (1-\rho)\left[\frac{1-\rho}{1-\rho z}\right]^{\frac{\lambda}{\mu}}$$

$$P_1(z) = \rho\left[\frac{1-\rho}{1-\rho z}\right]^{\frac{\lambda}{\mu}+1}$$

Expanding P0(z) and P1(z) using binomial series and equating coefficient of zn, we get,

$$p_{0n} = \frac{\rho^n}{n!\mu^n}\prod_{i=0}^{n-1}(\lambda+i\mu)(1-\rho)^{\frac{\lambda}{\mu}+1}$$

$$p_{1n} = \frac{\rho^{n+1}}{n!\mu^n}\prod_{i=1}^{n}(\lambda+i\mu)(1-\rho)^{\frac{\lambda}{\mu}+1}$$

5. Stationary distribution of the number of orbital customers

The stationary distribution of the number of orbital customers, is given by,

$$q_n = P\{N(t) = n\} = p_{0n} + p_{1n}$$

The corresponding generating function is,

$$P(z) = \sum_{n=0}^{\infty} q_n z^n = P_0(z) + P_1(z)$$

$$= (1+\rho-\rho z)\left[\frac{1-\rho}{1-\rho z}\right]^{\frac{\lambda}{\mu}+1}$$

$$E\left[N(t)\right] = \left.\frac{dP(z)}{dz}\right|_{z=1}$$

$$= \frac{\rho(\lambda+\rho\mu)}{\mu(1-\rho)}$$

$$E\left[N(t)^2\right] - E\left[N(t)\right] = \left.\frac{d^2P(z)}{dz^2}\right|_{z=1}$$

$$= \frac{\rho^2}{\mu^2(1-\rho)^2}(\lambda+\mu)(\lambda+2\rho\mu)$$

So,

$$V\left[N(t)\right] = E\left[N(t)^2\right] - \left[E\left(N(t)\right)\right]^2$$

$$= \frac{\rho}{\mu(1-\rho)^2}\left(\lambda+\rho\mu+\rho^2\mu-\rho^3\mu\right)$$

6. Stationary distribution of the system state

Let K(t) be the number of customers in the system at time t. Then,

$$Q_n = P\left(K(t)=n\right) = p_{0n} + p_{1,(n-1)}$$

The corresponding generating function is,

$$Q(z) = \sum_{n=0}^{\infty} Q_n z^n$$

$$= P_{0(z)} + zP_1(z)$$

$$= \left[\frac{1-\rho}{1-\rho z}\right]^{\frac{\lambda}{\mu}+1}$$

$$E\left(K(t)\right) = \left.\frac{dQ(z)}{dz}\right|_{z=1}$$

$$= \frac{\rho(\lambda+\mu)}{\mu(1-\rho)}$$

$$E\left(K(t)^2\right) - E\left(K(t)\right) = \left.\frac{d^2P(z)}{dz^2}\right|_{z=1}$$

$$= \frac{\rho^2(\lambda+\mu)(\lambda+2\mu)}{\mu^2(1-\rho)^2}$$

$$V\left(K(t)\right) = E\left(K(t)^2\right) - \left[E\left(K(t)\right)\right]^2$$

$$= \frac{\rho(\lambda+\mu)}{\mu(1-\rho)^2}$$

7. Blocking probability

Blocking probability,is the probability that the server is busy.Therefore,
Blocking probability

$$= P\{C(t)=1\}$$

$$= \sum_{n=0}^{\infty} p_{1n} = P_1(1) = \rho$$

Therefore,the probability that the server is busy=ρ
The probability that the server is idle= 1 − ρ

References

1. T. Yang and J. G. C. Templeton, A survey on retrial queues, Queueing Systems, 2(3), 201–233, 1987.
2. G. I. Falin, A survey of retrial queues, Queueing Systems, 7(2), 127–167, 1990.
3. G. I. Falin and J. G. C. Templeton, Retrial Queues, Chapman & Hall, London, 1997.
4. J. R. Artalejo, A classified bibliography of research on retrial queues: Progress in 1990–1999, TOP, 7, 187–211, 1999.

5. J. R. Artalejo and G. I. Falin, Standard and retrial queueing systems: A comparative analysis, Revista Matem´atica Complutense, 15(1), 101–129, 2002.
6. J. R. Artalejo and A. Gomez-Corral, Retrial Queueing Systems: A Computational Approach, Springer, Berlin, 2008.
7. J. R. Artalejo, Accessible bibliography on retrial queues: Progress in 2000–2009, Mathematical and Computer Modelling, 51(9–10), 1071– 1081, 2010.
8. J. Kim and B. Kim, A survey of retrial queueing systems, Annals of Operations Research, 247(1), 3–36, 2016.
9. R. Nobel, Retrial queueing models in discrete time: A short survey of some late arrival models, Annals of Operations Research, 247, 37–63, 2016.
10. T. Phung-Duc, Retrial queueing models: A survey on theory and applications, 2019.
11. X. Su and S. Veeraraghavan, A model of rational retrials in queues, Operations Research, 2019.