A SURVEY OF RETRIAL QUEUES

BEENA P.

Abstract: Mean value analysis is an elegant tool for determining mean performance measures in queueing models. Although retrial queues are in general much harder to analyze than ordinary queues mean performance measures are sometimes simple. Concentration is mainly on single server queueing models with batch arrivals, priority subscribers, two phase service, impatience subscribers, network blocking. Their performance analysis and mean value analysis are discussed. The retrial models for which mean value analysis works and does not works are discussed.

Keywords: retrial queues, batch arrivals, priority subscribers, two phase service, impatience subscribers, network blocking, performance analysis, mean value analysis.

Introduction: Queueing systems in which arriving customers find all servers and waiting positions (if any) occupied, may retry for service after a period of time. Such queues are called retrial queues or queues with repeated attempts. The goal of the mean value analysis technique is to provide an elegant alternative for obtaining the expected number of customers in orbit (and the expected waiting time) by avoiding the use of generating functions. This significantly reduces the algebra. Next, we show how the mean value relations can be adapted to obtain mean performance measures in more advanced M/G/1-type retrial queues. We use the Mean value analysis to analyze the M/G/1 retrial queue with exponential retrial times.

The main M/G/1 retrial queue: In M/G/1 retrial queue, customers arrive according to a Poisson process with rate λ. These customers are identified as primary calls. Service times are generally distributed with mean E(B) and mean residual service time E(R) = $(E(B^2)/2E(B))$. Customers in orbit retry after an exponentially distributed time with parameter μ. To ensure stability, we assume that $\rho = \lambda\ E(B) < 1$. With W we denote the steady-state waiting time of customers and with L we denote the steady-state orbit size.

The mean value relations for the system are given by

$$E(W) = E(L)E(B) + \rho\ (E(R) + \frac{1}{\mu})$$

$$E(L) = \lambda\ E(W)$$

Combination of the two relations gives the expression for the mean waiting time

$$E(W) = \frac{\rho}{1-\rho}\ (E(R) + \frac{1}{\mu})$$

The advanced M/G/1-type retrial queues: The mean value analysis technique also leads to the expected values for the waiting times and orbit sizes in more advanced M/G/1-type retrial queues. This fact is illustrated in the sequel for a variety of systems including batch arrivals, priorities, impatience, network blocking, etc.

The model with batch arrivals: In the batch arrival retrial queue it is assumed that at every arrival epoch a batch of K primary calls arrive with probability x_k . Every such customer produces a Poisson flow of repeated calls with rate μ. The batch size of the arrivals is given by the random variable X with probability distribution

$x_k = P\ (X = k\ ), k \geq 1$

The mean value relations for the systems are given by

$$E(W) = \left(E(L) + \sum_{k=1}^{\infty} r_k\,(k-1) \right) E(B) + \rho E(R) + \frac{1 - r_1\,(1-\rho)}{\mu}$$

Where the server utilization factor is $\rho = \lambda\ E(X)E(B)$

The model with priority subscribers: In this section the subscript 1 is associated with the high priority customers and the subscript 2 is associated with the low priority customers. The stability condition is given by $\rho_1 + \rho_2 < 1$, where $\rho_i = \lambda_i\ E\ (B_i\ )$, for i = 1, 2. The mean value relations for the high priority customers are now given by

$$E(W_1) = E(L_1)E(B_1) + \sum_{i=1}^{2} \rho_i\, E(R_i)$$

$E(L_1) = \lambda_1\ \ E(W_1)$ leading to

$$E(W_1) = \frac{\sum_{i=1}^{2} \rho_i\, E(R_i)}{(1-\rho_1)}$$

and the mean value relations for the low priority customers are given by

$$E(W_2) = \sum_{i=1}^{2} E(L_i)\, E(B_i) + \sum_{i=1}^{2} \rho_i\, E(R_i) + \rho_1\, E(W_2) + \frac{\rho_1 + \rho_2}{\mu}$$

$E(L_2) = \lambda_2\ E(W_2)$ leading to

$$E(W_2) = \frac{\sum_{i=1}^{2} \rho_i\, E(R_i)}{(1-\rho_1)(1-\rho_1-\rho_2)} + \frac{\rho_1 + \rho_2}{(1-\rho_1-\rho_2)\mu}$$

The model with impatient subscribers: In this model, the probability that a customer retries after the first attempt equals H_1. The system is stable when $\rho H_1 < 1$

The blocking probability $P_b = \frac{\rho}{1+\rho\,(1-H_1)}$ The mean value relations are given by

$$E(W_0) = \frac{H_1\, P_b}{\mu}$$

$$E(W_1) = E(L_0)\, E(B) + H_1\ E(L_1)\, E(B) + H_1\, P_b\, E(R)$$

$$E(L_0) = \lambda E(W_0)\,,\ E(L_1) = \lambda E(W_1)\,.$$

Combining formulas we get

$$E(W) = \frac{H_1\,\rho}{1-H_1\,\rho}\left(\frac{E(R)}{1+\rho(1-H_1)} + \frac{1}{\mu}\right)$$

The model with two-phase service: In this model, the server provides an essential service B_1 (with residual service time R_1) to all customers. Some customers are provided with a second optional service B_2 (with residual service time R_2) with probability p. Thus, we have

$$E(W_1^1) = \frac{\rho_1\,\rho}{1-\rho}\left(E(R) + \frac{1}{\mu}\right) + \rho_1\, E(R_1)$$

$$E(W_1^2) = \frac{\rho_2\,\rho}{1-\rho}\left(E(R) + \frac{1}{\mu}\right) + \rho_2\, E(R_2) + \frac{\rho_1\,\rho_2}{\lambda}$$

where

W_1^1: the busy time of the server providing essential service during the waiting time of the tagged customer,

W_1^2 : the busy time of the server providing optional service during the waiting time of the tagged customer.

The model with network blocking: Incoming customers who find a free server can receive an engaged signal with probability p. In that case, the customer will join the orbit. With probability q = 1 – p customers who find a free server really enter the service station. After each exponential period with parameter μ, the idle time ends with probability q and continues with probability p. So, we conclude

$$E(W_0) = \frac{p + q\rho}{\mu q}$$

Retrial models for which mean value analysis does not seem to work are discrete-time models, models with general retrial times, multi-server models and models with finite buffers and retrials.

References:

1. I.J.B.F. Adan and J.A.C. Resing (2001). Lecture Notes Queueing Theory, Eindhoven University of technology.
2. J.R. Artalejo and G. Choudhury (2004). Steady state analysis of an M/G/1 queue with repeated attempts and two-phase service. Quality Technology & Quantitive Management 1, 189–199.
3. G.I. Falin (1986). Single-line repeated orders queueing systems. Optimization 17, 649–667.
4. G.I. Falin and J.G.C. Templeton (1997). Retrial Queues, Chapman and Hall,London.
5. S.W. Fuhrmann and R.B. Cooper (1985). Stochastic decompositions in the M/G/1 queue with generalized vacations. Operations Research 33, 1117–1129.

A. G´omez-Corral (1999). Stochastic analysis of a single server retrial queue with general retrial times. Naval Research Logistics 46, 561–581.

6. E.M.M. Winands, I.J.B.F. Adan and G.J. van Houtum (2006). Mean value analysis for polling systems. Queueing Systems 54, 35–44.

Artalejo, J. R., “Accessible Bibliography on Retrial Queues,” *Mathematical and Computer Modeling*

***

Beena P., Assistant Professor in Mathematics
Sree Neelakanta Govt. Sanskrit College ,Pattambi