Mean Value Analysis For Polling Systems

Ivo Adan, Geert-Jan van Houtum, Erik Winands

20th June 2006

Polling System:

System consisting of a number of queues served by a single server in fixed order

Applications in:

- Manufacturing
- Communication
- Maintenance
- . . .

Service disciplines:

- *Exhaustive*
  A queue must be empty before the server moves on
- *Gated*
  Only those customers are served present on arrival of the server at the queue
- k-limited
  At most k customers are served
- Time-limited
  The time of the server spent at the queue is limited
- ...

Performance measures:

- Utilization rate
- Cycle length
- Queue length
- Waiting time
- . . .

Queue length and waiting time distribution:

If the service discipline satisfies a certain *branching property*, then an exact generating function analysis of the queue length distribution is possible (*Resing* 1993)

Mean queue length and mean waiting time:

- Buffer occupancy method (*Cooper* 1970, *Eisenberg* 1972)
- Descendant set method (*Konhein et al.* 1994)
- Individual station method (*Srinivasan et al.* 1996)
- Station time method (*Ferguson and Aminetzah* 1985)
- . . .

Mean Value Analysis:

Approach to determine mean queue lengths and waiting times based on

- The **PASTA** property: **P**oisson **A**rrivals **S**ee **T**ime **A**verages
- Little's law: $E(L) = \lambda E(W)$

Advantage:
Probabilistic intuitive approach to determine mean values, resulting in linear equations only involving first moments

Disadvantage:
Computational complexity

Model Description:

N queues, numbered $1, 2, \ldots, N$, served by 1 server in a fixed cyclic order; the service discipline is exhaustive or gated and within a queue customers are served in order of arrival. In queue i, arrivals are Poisson with rate λ_i, the service time and setup time are denoted by B_i and S_i, and the *residual service and setup* time are denoted by R_{B_i} and R_{S_i}.

Mean total setup time in a cycle and mean residual service time are

$$E(S) = \sum_{i=1}^{N} E(S_i), \qquad E(R_{B_i}) = \frac{E(B_i^2)}{2E(B_i)}$$

Utilization rate of queue i and the total utilization rate are

$$\rho_i = \lambda_i E(B_i), \qquad \rho = \sum_{i=1}^{N} \rho_i < 1$$

Preliminaries:

Cycle length of queue i is the time between two successive arrivals of the server to this queue; mean cycle time is

$$E(C) = E(S)/(1-\rho)$$

Visit time θ_i of queue i is the service period of queue i plus *preceding* setup time if exhaustive service or plus *succeeding* setup time if gated service; thus

$$E(\theta_i) = E(S_i) + \rho_i E(C) \text{ (exhaustive)},$$

$$E(\theta_i) = \rho_i E(C) + E(S_{i+1}) \text{ (gated)}$$

Let L_i denote the length of queue i (excluding customer possibly in service), let W_i denote the waiting time of a type-i customer

Goal:

Determine $E(L_i)$ and $E(W_i)$

Mean Value Analysis:

Derive (i) *arrival relation* for the mean waiting time and use (ii) Little's law

First $N = 1$: ($M/G/1$ queue)

Arrival relation

$$E(W_1) = E(L_1)E(B_1) + \rho_1 E(R_{B_1})$$

Little's law

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

yields

$$E(W_1) = \frac{1}{1-\rho_1} \cdot \rho_1 E(R_{B_1}) = \frac{\rho_1}{1-\rho_1} \cdot \frac{\rho_1}{2}(1 + c_{B_1}^2)E(B_1)$$

Now $N = 2$ for exhaustive service:

Arrival relation

$$E(W_1) = E(L_1)E(B_1) + \rho_1 E(R_{B_1}) + \frac{E(S_1)}{E(C)}E(R_{S_1}) + \frac{E(\theta_2)}{E(C)}(E(R_{\theta_2}) + E(S_1))$$

Little's law

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

yields

$$E(W_1) = \frac{1}{1-\rho_1} \cdot \left[\rho_1 E(R_{B_1}) + \frac{E(S_1)}{E(C)}E(R_{S_1}) + \frac{E(\theta_2)}{E(C)}(E(R_{\theta_2}) + E(S_1))\right]$$

But what is $E(R_{\theta_2})$?

Consider *conditional* queue lengths:

Let $L_{i,n}$ denote the length of queue i at an arbitrary point in time during a visit time of queue n

Then, since each of the $L_{2,2}$ customers initiates a busy period,

$$E(R_{\theta_2}) = E(L_{2,2})\frac{E(B_2)}{1-\rho_2} + \frac{\rho_2 E(C)}{E(\theta_2)}\frac{E(R_{B_2})}{1-\rho_2} + \frac{E(S_2)}{E(\theta_2)}\frac{E(R_{S_2})}{1-\rho_2}$$

Further

$$E(L_2) = \frac{E(\theta_1)}{E(C)}E(L_{2,1}) + \frac{E(\theta_2)}{E(C)}E(L_{2,2})$$

Finally $L_{2,1}$ is equal to the number of type-2 arrivals during the age of θ_1, and the age has the same distribution as the residual time of θ_1. So

$$E(L_{2,1}) = \lambda_2 E(R_{\theta_1})$$

Now $N = 2$ for gated service:

The visit time θ_i is the service period of queue i plus succeeding setup time,

$$E(\theta_i) = \rho_i E(C) + E(S_{i+1})$$

An (i, j)-period is defined as the sum of j consecutive visit times starting in queue i. Then

$$\theta_{1,1} = \theta_1$$

and $\theta_{1,2}$ is a cycle starting at queue 1,

$$E(\theta_{1,2}) = E(\theta_1) + E(\theta_2)$$

Further distinguish between customers waiting before and after the gate.

Let $\tilde{L}_{i,n}$ denote the length of queue i before the gate and $\bar{L}_{i,n}$ denotes the length of queue i after the gate at an arbitrary point in time during a visit time of queue n. Then

$$\bar{L}_{i,n} = 0, \qquad i \neq n,$$

and

$$L_{1,1} = \tilde{L}_{1,1} + \bar{L}_{1,1},$$

$$L_{1,2} = \tilde{L}_{1,2}$$

Arrival relation

$$E(W_1) = E(\tilde{L}_1)E(B_1) + E(R_{\theta_{1,2}})$$

Little's law

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

and

$$E(L_1) = E(\tilde{L}_1) + \frac{E(\theta_1)}{E(C)} E(\bar{L}_{1,1}),$$

$$E(\tilde{L}_1) = \frac{E(\theta_1)}{E(C)} E(\tilde{L}_{1,1}) + \frac{E(\theta_2)}{E(C)} E(\tilde{L}_{1,2})$$

But what are $E(\tilde{L}_{1,1})$, $E(\tilde{L}_{1,2})$ and $E(R_{\theta_{1,2}})$?

We have again

$$E(\tilde{L}_{1,1}) = \lambda_1 E(R_{\theta_{1,1}}),$$

$$\frac{E(\theta_1)}{E(C)} E(\tilde{L}_{1,1}) + \frac{E(\theta_2)}{E(C)} E(\tilde{L}_{1,2}) = \lambda_1 E(R_{\theta_{1,2}})$$

and

$$E(R_{\theta_{1,1}}) = E(\bar{L}_{1,1}) E(B_1) + \frac{\rho_1 E(C)}{E(\theta_1)} [E(R_{B_1}) + E(S_2)] + \frac{E(S_2)}{E(\theta_1)} E(R_{S_2})$$

$$\begin{aligned} E(R_{\theta_{1,2}}) &= \frac{E(\theta_2)}{E(C)} E(R_{\theta_{2,1}}) + \frac{E(\theta_1)}{E(C)} \\ &\times \left[E(R_{\theta_{1,1}}) + (\lambda_2 E(R_{\theta_{1,1}}) + E(\tilde{L}_{2,1})) E(B_2) + E(S_2) \right] \end{aligned}$$

Model variations:

- *Mixed service*: Queue 1 has exhaustive and queue 2 has gated service
  Then further condition on queue lengths:
  $L_{i,n}$ is the length of queue i at an arbitrary point in time during a service period of queue i and $L'_{i,n}$ is the length of queue i at an arbitrary point in time during a setup time of queue i

- *Zero setup times*
  If all setup times are zero, then replace probabilities like

$$\frac{E(\theta_1)}{E(C)}$$

  by ρ_1

Extensions:

- Systems with Poisson *Batch* arrivals
- Systems with fixed polling tables
- Discrete time polling systems
- Systems with *branching property*?