Performance Analysis, Autumn 2010

Bengt Jonsson

November 16, 2010

Kendall Notation

Queueing process described by $A/B/X/Y/Z$, where

- A is the arrival distribution
- B is the service pattern
- X the number of parallel service channels (servers)
- Y the restriction on system capacity
- Z the ququing discipline

Example

- $M/D/2/\infty/\mathrm{FCFS}$ is a process with exponential interarrival times, deterministic service times, two parallel servers, unbounded queueing capacity, FCFS queue discpline.

Little's Formula

Holds for general queueing models.

- Relates steady-state mean system size to steady-state waiting times.
- Let
  - L be average number of jobs in system (average queue length)
  - W be average time that a job spends in system (average waiting time)
  - λ is the average arrival rate (arriving jobs per time unit)

  Then

$$L = \lambda W$$

Also, if we restrict to time in and size of queue (not in server), then

$$L_q = \lambda W_q$$

Application of Little's Formula

Average waiting time for $M/M/1$ queue:

- Expected number of jobs: $L = \dfrac{\lambda}{\mu - \lambda}$
- Hence expected total time in system: $W = L/\lambda = \dfrac{1}{\mu - \lambda}$
- Expected number of jobs in "pure" queue

$$L_q = \sum_{n=1}^{\infty}(n-1)\pi_n = \sum_{n=1}^{\infty} n\pi_n - \sum_{n=1}^{\infty}\pi_n =$$

$$L - (1 - \pi_0) = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}$$

- i.e.,

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

- Hence expected waiting time

$$W_q = L_q/\lambda = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{\rho}{\mu - \lambda}$$

PASTA Property

PASTA = **P**oisson **A**rrivals **S**ee **T**ime **A**verages

- Consider a stochastic system with arrivals according to a Poisson process with intensity λ
- (service times can be arbitrarily distributed)
- With a state i we may associate
  - long-term probability π_i
  - π_i^*: Probability that an arrival will find system in state i.
- In general $\pi_i \neq \pi_i^*$.
- But for Poisson arrivals, we have $\pi_i = \pi_i^*$.

(* Supply counterexample, e.g., deterministic arrivals *)

Proof of PASTA Property

Sketch:

- Consider system state to be a stochastic process $(X_t, t \geq 0)$.
- Consider a specific small interval $(t-h, t]$.
- The events "$X_{t-h} = i$" and "some job arrives in $(t-h, t]$" are independent (by memoryless property):

$$P(X_{t-h} = i \;\cap\; N(h) \geq 1) = P(X_{t-h} = i)P(N(h) \geq 1)$$

which implies

$$P[X_{t-h} = i | N(h) \geq 1] = P(X_{t-h} = i)$$

By letting $h \to 0$, the LHS expresses probability that system is in state i when job arrives at time t. The equality says that this is independent of whether a job arrives.

Hitchhiker's paradox

Setting:

- Cars are passing a point of a road according to a Poisson process with rate $1/10$ (i.e., mean interarrival time is 10 (minutes))
- A hitchhiker arrives at a random instant.
- What sis the mean waiting time W until next car?
- According to memoryless property of exponential distribution, it should be 10 minutes?
- But: the hitchhiker arrives between two cars, such that the mean time between the two cars is 10 minutes. Therefore, the mean waiting time should be 5 minutes.
- Resolve the paradox

Multiserver queues

i.e., $M/M/c$ queues for integer $c \geq 1$.

- Arrival poisson w. rate λ
- Each server serves with rate μ.
- From balance equations, we get:

$$\pi_n = \begin{cases} \dfrac{\lambda^n}{n!\mu^n}\pi_0 & (0 \leq n < c), \\ \frac{\lambda^n}{c^{n-c}c!\mu^n}\pi_0 & (n \geq c). \end{cases}$$

- Let $r = \lambda/\mu$, let $\rho = r/c = \lambda/c\mu$. Then

$$\pi_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \sum_{n=c}^{\infty} \frac{r^n}{c^{n-c}c!} \right)^{-1} = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c}{c!(1-\rho)} \right)^{-1}$$

Average “pure” queue length

i.e., $M/M/c$ queues for integer $c \geq 1$.

$$L_q = \sum_{n=c+1}^{\infty}(n-c)\pi_n = \sum_{n=c+1}^{\infty}(n-c)\frac{r^n}{c^{n-c}c!}\pi_0$$

$$= \frac{r^c\pi_0}{c!}\sum_{n=c+1}^{\infty}(n-c)\rho^{n-c} = \frac{r^c\pi_0}{c!}\sum_{m=1}^{\infty}m\rho^m$$

$$= \frac{r^c\pi_0}{c!}\frac{\rho}{(1-\rho)^2}$$

Other queue types

$M/G/1$ queues.

- Arrival poisson w. rate λ
- Services with arbitrary distributions.
- In general, this is not a Markov process.
- Expression for expected waiting time, by PASTA property:

$$E[W_q] = \underbrace{E[L_q]}_{\text{waiting jobs}} \cdot \underbrace{E[S]}_{\text{service time}} + \underbrace{E[R]}_{\text{remaining service}}$$

- Using Little's law:

$$E[L_q] = \lambda E[W_q] \quad \text{i.e.,} \quad E[W_q] = \frac{E[R]}{1 - \lambda E[S]} = \frac{E[R]}{1 - \rho}$$

- It remains to compute residual service time.

$M/G/1$ Residual Service time

- Find $E[R]$:
- Over a long period T, we have λT jobs.
- Average remaining time for job is $\frac{1}{2}E[S^2]$
- $E[R] = \frac{\lambda}{2}E[S^2]$
- $E[W_q] = \frac{\lambda E[S^2]}{2(1-\rho)}$
- Hence $E[L_q] = \lambda E[W_q] = \frac{\lambda^2 E[S^2]}{2(1-\rho)}$
- and $E[W] = E[W_q] + E[S] = \frac{\lambda E[S^2]}{2(1-\rho)} + \frac{1}{\mu}$
- and $E[L] = \lambda E[W] = \frac{\lambda^2 E[S^2]}{2(1-\rho)} + \rho$

Observations on PK formula

Consider $E[W_q] = \frac{\lambda E[S^2]}{2(1-\rho)}$

- Now, $E[S^2] = (E[S])^2 + \mathrm{Var}(S) = (E[S])^2(1 + C_v^2)$.
- For same mean, Mean values increase linearly by variance.

Consider $M/M/1$:

- $Var(S) = (E[S])^2 = 1/\mu^2$:
- $$E[L] = \frac{\lambda^2 E[S^2]}{2(1-\rho)} + \rho = \frac{\rho^2}{(1-\rho)} + \rho = \frac{\rho}{(1-\rho)}$$

Consider $M/D/1$:

- $Var(S) = 0$
- $$E[L] = \frac{\lambda^2 E[S^2]}{2(1-\rho)} + \rho = \frac{\rho^2}{2(1-\rho)} + \rho$$

Queueing Networks

Modeling a queue of jobs

- Helpful fact: Burke's theorem:

  The departure process from a stable $M/M/c$ queue with arrival and service rates λ and μ, respectively, is a Poisson process with rate λ.

- This allows to analyze acyclic QNs, since each queue will behave like an $M/M/1$ queue
- Let r_{ij} be routing probabilities, i.e., probability that job leaving i will go to j.
- Each queue i will behave like $M/M/1$, with arrival rate $\lambda_0 r_{0i}$ and service rate μ_i,
- Assume no bottlenecks, i.e., $\frac{\lambda_0 r_{0i}}{\mu_i} = \rho_i < 1$,

Jackson Queueing Networks

Take the preceding slide, but allow cycles.

- Now arrival streams may not be Poisson.
- But steady-state probabilities are as if each queue is independent $M/M/1$ queue
- Solve the *traffic equations*:

$$\lambda_i = \lambda_0 r_{0i} = \sum_{j=1}^{k} \lambda_j r_{ji}$$

- Example (BH206)