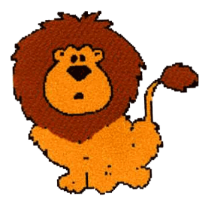
Computer Networks
More general queuing systems

Saad Mneimneh
Computer Science
Hunter College of CUNY
New York



$\mathbb{M}/\mathbb{G}/1$

1 Introduction

We now consider a queuing system where the customer service times have a general distribution - not necessarily exponential as in the M/M/1 system. We denote such a system by M/G/1. In M/G/1, customers still arrive according to a Poisson process with rate λ. We assume that customers are served in the order they arrive (FIFO) and that X_i is the service time of the i^{th} customer. We assume further that X_i are IID and independent of the interarrival times. Let

$$\overline{X} = E[X] = \frac{1}{\mu} = \text{ Average service time}$$

$$\overline{X^2} = E[X^2]$$

The goal is to derive the Pollaczek-Khinchin (P-K) formula:

$$W = \frac{\lambda \overline{X^2}}{2(1-\rho)}$$

where W is the expected waiting time in the queue (we assume here an infinite queue) and $\rho = \lambda/\mu = \lambda\overline{X}$.

2 The Pollaczek-Khinchin (P-K) formula

We will derive the P-K formula using the concept of mean residual time. While this derivation obtains only the system averages, it is simpler and more insightful than other derivations that can give a probability distribution of the system occupancy. Denote the following:

- W_i = waiting time in queue for customer i
- N_i = number of customers in queue seen by customer i upon arrival
- R_i = residual service time seen by customer i, i.e. the remaining time until the departure of the customer already being served upon the arrival of customer i

Then (FIFO),

$$W_i = R_i + \sum_{j=i-N_i}^{i-1} X_j$$

Assuming that $X_{i-N_i}, \ldots, X_{i-1}$ are independent of N_i,

$$E[W_i] = E[R_i] + E[N_i]E[X_j|N_i] = E[R_i] + E[N_i]E[X] = E[R_i] + \overline{X}E[N_i]$$

Let W, R, and N_Q be the corresponding limits as customer index $i \to \infty$ or time increases to infinity. We assume that these limits exist. Then, by the PASTA property of Poisson arrivals, these are also the averages seen by an outside observer:

$$W = R + \frac{1}{\mu}N_Q$$

By Little's theorem, $N_Q = \lambda W$, so

$$W = R + \rho W \Rightarrow W = \frac{R}{1-\rho}$$

We now compute R by a graphical argument.

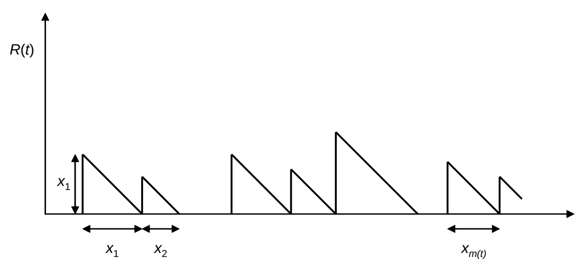


Figure 1: Residual service time

Let $R(t)$ be the residual service time as a function of time, and let t be a time for which $R(t) = 0$. As we can see from the above figure,

$$R_t = \frac{1}{t}\int_0^t R(\tau)d\tau = \frac{1}{t}\sum_{i=1}^{m(t)} \frac{1}{2}X_i^2 = \frac{m(t)}{t}\frac{\sum_{i=1}^{m(t)} \frac{1}{2}X_i^2}{m(t)}$$

Taking the limit as $t \to \infty$, and applying the law of large numbers (we also implicitly assume that time averages can be replaced by ensemble averages):

$$R = \frac{1}{2}\lambda\overline{X^2}$$

Therefore,

$$W = \frac{\lambda\overline{X^2}}{2(1-\rho)}$$

- M/G/1 (general): $W = \frac{\lambda\overline{X^2}}{2(1-\rho)}$
- M/M/1 (memoryless): $\overline{X^2} = 2/\mu^2 \Rightarrow W = \frac{\rho}{\mu(1-\rho)}$
- M/D/1 (deterministic): $\overline{X^2} = 1/\mu^2 \Rightarrow W = \frac{\rho}{2\mu(1-\rho)}$

Althoug we assumed FIFO, the P-K formula is valid for any order of sevice as long as the order is independent of the service times of individual customers (e.g. serving smaller jobs first or larger jobs first would make the order dependent on service times).

3 Example 1: M/G/1 system

- What is the probability that the system is empty?
  - By Little's theorem, the average number of customers in service is $\lambda\overline{X}$
  - $P(system is empty) = 1 - \lambda\overline{X}$
- What is the average time I (for idle) between busy periods?
  - consider the end of a busy period, since arrivals are Poisson (memoryless), the next arrival is exponentially distributed (which signals the beginning of next busy period)
  - $I = 1/\lambda$
- What is the average time B of a busy period?
  - $\frac{B}{B+I} = \lambda\overline{X}$
  - $B = \frac{\overline{X}}{1-\lambda\overline{X}}$
- What is the average number of customers served during a busy period
  - from above $\frac{1}{1-\lambda\overline{X}}$

4 Example 2: sliding window ARQ

- Assume Go back n with one sided error
  - probability of error is p
  - ACKs always arrive
  - timeout $= n$ (for retransmissions)
- When packet i is successfully transmitted, packet $i+1$ is successfully transmitted $1+kn$ time units later with probability $(1-p)p^k$
- Transmitter's queue behaves like M/G/1
  - $P(X = 1 + kn) = (1-p)p^k$
  - $\overline{X} = 1 + \frac{np}{1-p}$, $\overline{X^2} = 1 + \frac{2np}{1-p} + \frac{n^2(p+p^2)}{(1-p)^2}$ (after some calculation)
  - A packet waits in window on average $W = \frac{\lambda \overline{X^2}}{2(1-\lambda\overline{X})}$

5 M/G/1 with vacations

Suppose that at the end of each busy period, the M/G/1 server goes on "vacation" for some random interval of time. If the system is still idle at the completion of a vacation, a new vacation start immediately. This is useful for modeling slotted systems as we will see in the following section. Let V_1, V_2, ..., $V_{l(t)}$ be the durations of vacations at up to time t. The following formula is still valid:

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

where R is now the residual time for completion of the service *or* vacation in process when the customer arrives.

We can compute R using a similar graphical argument as before.

$$R_t = \int_0^t R(\tau)d\tau = \frac{1}{t}\sum_{i=1}^{m(t)} \frac{1}{2}X_i^2 + \frac{1}{t}\sum_{i=1}^{l(t)} \frac{1}{2}V_i^2$$

$$R_t = \frac{m(t)}{t}\frac{\sum_{i=1}^{m(t)} \frac{1}{2}X_i^2}{m(t)} + \frac{l(t)}{t}\frac{\sum_{i=1}^{l(t)} \frac{1}{2}V_i^2}{l(t)}$$

Taking the limit as $t \to \infty$ (assuming V_i are IID and applying the law of large numbers):

$$R = \frac{1}{2}\lambda\overline{X^2} + \frac{1}{2}\overline{V^2}\lim_{t\to\infty}\frac{l(t)}{t}$$

But

$$\lim_{t\to\infty}\frac{t(1-\rho)}{l(t)} = \overline{V} \quad \text{(why?)}$$

Therefore,

$$R = \frac{1}{2}\lambda\overline{X^2} + \frac{1}{2}\frac{(1-\rho)\overline{V^2}}{\overline{V}}$$

$$W = \frac{R}{1-\rho} = \frac{\lambda\overline{X^2}}{2(1-\rho)} + \frac{\overline{V^2}}{2\overline{V}}$$

6 Example 3: slotted FDM

Consider m sessions, each a Poisson process with rate λ/m, frequency division multiplexed on a channel. The transmission time per packet is m time units on each subchannel. M/D/1 $\Rightarrow W = \frac{\lambda m}{2(1-\lambda)}$ ($\overline{X} = 1/\mu = m$, $\rho = \frac{\lambda/m}{\mu} = \lambda$). If the system is slotted, i.e. packets can only leave at times m, $2m$, $3m$, etc..., then we can view this as M/D/1 with vacations. If no packet is waiting, server takes a vacation of m units, $\overline{V} = m$, $\overline{V^2} = m^2$. M/D/1 with vacations $\Rightarrow$ $W = \frac{\lambda m}{2(1-\lambda)} + \frac{m}{2} = \frac{m}{2(1-\lambda)}$.