Queueing Theory

For a single queue:

$N(t)$ = number of customers in the queue at time t.

$\alpha(t)$ = number of customers who arrived in the interval $[0, t]$.

T_i = time spent in the system by customer i.

Assume that the following three limits exist:

$N = \lim_{t\to\infty} N(t)$

$\lambda = \lim_{t\to\infty} \frac{\alpha(t)}{t}$

$T = \lim_{t\to\infty} \frac{\sum_{i=1}^{\alpha(t)} T_i}{\alpha(t)}$

Then $N = \lambda T$ (Little's Lemma).

The M/M/1 system

A single queue with a single server.

Customers arrive according to a Poisson process with rate λ.

Service time has exponential distribution with mean $1/\mu$.

Poisson Process

Let $A(t)$ be the number of arrivals in the interval $[0, t]$.

$A(t)$ is a Poisson process if for all $\tau > 0$:

- $\Pr(A(t + \tau) - A(t) = n) = e^{-\lambda\tau}\frac{(\lambda\tau)^n}{n!}$
- Number of arrivals in disjoint intervals are independent.

Properties of a Poisson process

1. The expected number of arrivals in an interval of t steps is λt.

2. Let τ_i be the interval between the ith and $i+1$th arrivals.

   τ_i has an exponential distribution with parameter λ:

   $$\Pr(\tau_i \leq s) = 1 - \Pr(A(t+s) - A(t) = 0) = 1 - e^{-\lambda s}$$

3. $\Pr(A(t+\tau) - A(t) = 1) = \lambda\tau + o(\tau)$.

4. $\Pr(A(t+\tau) - A(t) \geq 2) = o(\tau)$.

5. Sum of Poisson processes is a Poisson process with sum of rates.

6. If a Poisson process is split randomly, the two processes are Poisson.

Exponential Distribution

$X \sim Exp(\mu) \Rightarrow \Pr(X \leq s) = 1 - e^{-\mu s}$

$E[X] = 1/\mu$

The exponential distribution is **memoryless**:

$$\Pr(X > t + \tau | X > t) = \frac{\Pr(X > t + \tau)}{\Pr(X > t)}$$

$$= \frac{e^{-\mu(t+\tau)}}{e^{-\mu t}} = e^{-\mu\tau} = \Pr(X > \tau)$$

Interarrival and Waiting time distributions

Let X_n ($n \geq 1$) denote the time from the $(n-1)$th to the nth arrival.

The sequence $\{X_n, n = 1, 2, \ldots\}$ is the **sequence of interarrival times**.

If the arrival process is Poisson with rate λ then each random variable in the above sequence is independently and identically distributed exponential random variable with mean $1/\lambda$.

Markov Chain Analysis

Let $N(t)$ be the number of customers at time t.

$\{N(t)|t \geq 0\}$ is a continuous-time Markov chain.

However, for simplicity we will analyze a discrete-time Markov chain.

Let N_k = number of customers in the system at time $k\delta$, $k = 0, 1, \ldots$, and δ is a small positive number.

Let $P_{ij} = \Pr(N_{k+1} = j | N_k = i)$ denote the transition probabilities.

We can show that:

$P_{00} = 1 - \lambda\delta + o(\delta)$

$P_{ii} = 1 - \lambda\delta - \mu\delta + o(\delta)$, $i \geq 1$

$P_{i,i+1} = \lambda\delta + o(\delta)$, $i \geq 0$

$P_{i,i-1} = \mu\delta + o(\delta)$, $i \geq 1$

$P_{i,j} = o(\delta)$, $j \neq i, i+1, i-1$

Steady State

$\{N_k | k = 0, 1, \ldots\}$ is a discrete Markov chain with steady state probabilities:

$$p_n = \lim_{t \to \infty} \Pr(N(t) = n) = \lim_{k \to \infty} \Pr(N_k = n)$$

In steady state, we have the global balance equations:

$p_n \lambda \delta + o(\delta) = p_{n+1} \mu \delta + o(\delta)$

As $\delta \to \infty$,

$p_n \lambda = p_{n+1} \mu$

Thus,

$p_{n+1} = \rho^{n+1} p_0$ where $\rho = \lambda / \mu$.

If $\rho < 1$, $\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} \rho^n p_0 = \frac{p_0}{1-\rho}$

Thus $p_n = \rho^n (1 - \rho)$, $n = 0, 1, \ldots$.

The expected number of customers in the system at steady state:

$$N = \sum_{n=0}^{\infty} n p_n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$$

The average delay per customer is

$$T = N/\lambda = \frac{1}{\mu-\lambda}$$

The average number of customers in the queue is

$$N_Q = \lambda W = \frac{\rho^2}{1-\rho}$$

Applications

1. Increasing arrival and transmission rate by the same factor $K > 1$ does not change the utilization factor and hence the average number of packets in the system.

   However, the average delay per packet is now $T = N/(K\lambda)$ (!)

2. Statistical multiplexing verses time-division multiplexing:

   Transmitting m identical Poisson process each with arrival rate λ/m:

   Mixing in one channel: $T = \frac{1}{\mu - \lambda}$

   Transmitting through m separate channels:

   $T' = \frac{m}{\mu - \lambda}$