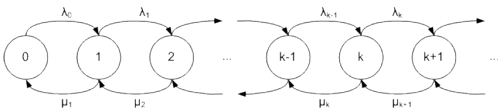
M / M / 1 Queue

1) Consider the $M/M/1$ queue with arrival rate λ and departure (or service) rate μ. Let the *state* of the system be the number of packets in the system (queue and service), and let $s(t)$ denote the state at time t. The state transition diagram of this queue is represented in the figure below (where $\lambda = \lambda_1 = \lambda_2 = \cdots$ and $\mu = \mu_1 = \mu_2 = \cdots$). Remember that there is no upper bound to the queue size, and assume that $\lambda < \mu$. We want to compute the *stationary distribution* of the



queue size, that tells us how likely it is to be in each state when the queue has reached a stationary condition (read "after a large number of steps"). To be precise the stationary distribution is a vector

$$\pi = (\pi_0, \pi_1, \pi_2, \ldots),$$

such that

$$\Pr(s(t) = k) = \pi_k$$

for all $k \geq 0$ and for all t suitably large. As the values in π constitute a probability distribution, we have that

$$\sum_{k=0}^{\infty} \pi_k = 1.$$

As the diagram above suggests, we have that

$$\lambda\pi_0 = \mu\pi_1$$
$$\lambda\pi_1 = \mu\pi_2$$
$$\ldots$$
$$\lambda\pi_i = \mu\pi_{i+1}$$
$$\ldots$$

1) Use the fact above to express π_k, $k > 0$, as a function of π_0.

$$\pi_k = \left(\frac{\lambda}{\mu}\right)^k \pi_0$$

2) Using $\lambda < \mu$ and the fact that all π_k's sum to 1, compute π_0 (as a function of λ and μ).

$$1 = \sum_{k=0}^{\infty} \pi_k = \pi_0 \sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k = \pi_0 \frac{1}{1-\lambda/\mu},$$

because the sum converges as $\lambda/\mu < 1$. Hence, we have

$$\pi_0 = 1 - \lambda/\mu.$$

3) Using the results above, compute the expected number of packets in the system at any given time. As you learnt in class, you should get $\frac{\lambda}{\mu-\lambda}$. You may find it useful that $\rho = \lambda/\mu < 1$.

Observe that $\pi_0 = 1 - \rho$. Then, the expected number of packets in the system is

$$\begin{aligned}\sum_{k=0}^{\infty} k\pi_k &= \pi_0 \sum_{k=0}^{\infty} k\rho^k = rho(1-\rho)\sum_{k=0}^{\infty} k\rho^{k-1} \\ &= \rho(1-\rho)\sum_{k=0}^{\infty} \frac{\partial}{\partial\rho}\rho^k = \rho(1-\rho)\frac{\partial}{\partial\rho}\left(\sum_{k=0}^{\infty}\rho^k\right) \\ &= \rho(1-\rho)\frac{\partial}{\partial\rho}\frac{1}{1-\rho} = \frac{\rho}{1-\rho} \\ &= \frac{\lambda}{\mu-\lambda}.\end{aligned}$$

Notice that the sum converges for $\rho < 1$, and this allowed to swap the derivative and the sum.

4)What is the expected time T_1 that a packet spends in the system (queue and service) if the arrival rate is λ and the departure rate is 3μ?

$$T_1 = \frac{1}{3\mu - \lambda}.$$

M/G/1 Queue

- Packet arrivals are memoryless with rate λ
- A packet/file/flow takes a random transmission time with CDF $F_T(t)$:= Prob (T ≤ t)

$$\rho = \frac{\lambda}{\mathbb{E}\{T\}} < \mathbf{1}$$

- Independent and identical across time
- Infinite buffer

$$\text{Mean Waiting Time} = \frac{\lambda\mathbb{E}(T^2)}{2\,(1-\rho)}$$

- Interesting Observations:
  - Waiting time decreases if you receive smaller packets w transmit time = T/2 at higher arrival rate of 2λ
  - Producing less variable packet sizes is beneficial!
    - Size-based queuing to differentiate between packets/files is beneficial

How much buffering?

RFC 3439 rule of thumb: average buffering equal to "typical" RTT (say 250 msec) times link capacity C

- e.g., C = 10 Gpbs link: 2.5 Gbit buffer

recent recommendation: with *N* flows, buffering equal to

$$\frac{\text{RTT} \cdot \text{C}}{\sqrt{\text{N}}}$$

ARP protocol: same LAN

ARP table: each IP node (host, router) on LAN has table

- IP/MAC address mappings for some LAN nodes:

  < IP address; MAC address; TTL>

- TTL (Time To Live): time after which address mapping will be forgotten (typically 20 min)

A wants to send datagram to B
- B's MAC address not in A's ARP table.

A broadcasts ARP query packet, containing B's IP address
- destination MAC address = FF-FF-FF-FF-FF-FF
- all nodes on LAN receive ARP query

B receives ARP packet, replies to A with its (B's) MAC address
- frame sent to A's MAC address (unicast)

A caches (saves) IP-to-MAC address pair in its ARP table until information becomes old (times out)
- soft state: information that times out (goes away) unless refreshed

ARP is "plug-and-play":
- nodes create their ARP tables *without intervention from net administrator*

Framing – Byte count

Frame begins with a count of the number of bytes in it
- Simple, but difficult to resynchronize after an error

Framing – Bit stuffing

Stuffing done at the bit level:
- Frame flag has six consecutive 1s (not shown)
- On transmit, after five 1s in the data, a 0 is added
- On receive, a 0 after five 1s is deleted

Analysis (ALOHA)

$x_i = \begin{cases} 1 & \text{if node } i \text{ gets its packet through} \\ 0 & \text{o.w.} \end{cases}$

[We assume N node contending for channel]

$\text{throughput} = \mathbb{E}\left(\sum_{i=1}^{N} x_i\right) = \sum_{i=1}^{N} \mathbb{E}(x_i)$

$\mathbb{E}(x_i) = \text{Prob}(\text{node } i \text{ gets its packet through})$

$= p(1-p)^{N-1}$

$\text{throughput} = Np(1-p)^{N-1}$

Question: what is the optimal choice of p?

set $\frac{d}{dp}\text{throughput} = 0 \implies p^* = 1/N$

$\text{throughput}(N) = \left(1-\frac{1}{N}\right)^{N-1}$

$N=2 \implies 1/2$

$N=3 \implies \left(\frac{2}{3}\right)^2 = \frac{4}{9}$

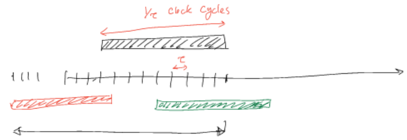
$\vdots$

As $N \to \infty$

$\lim_{N\to\infty}\left(1-\frac{1}{N}\right)^{N-1} = e^{-1} = .36$

$\left(1+\frac{a}{N}\right)^N \to e^a$

Analysis (Pure ALOHA)



Prob that station 1 gets a packet throug

$= p(1-p)^{(n-1)(2/\tau-1)} = S(p)$

avg # of tx per unit of time $= n\,S(p)\cdot\frac{1}{\tau}$

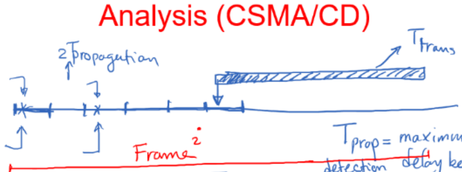
This is maximized when $p^* = \left[(n-1)(2/\tau-1)+1\right]^{-1}$

Plug this in and look at asymptotics

throughput $= \frac{1}{2}\left(1-\frac{1}{M}\right)^M$ for $M=(n-1)(2/\tau-1)$

$\to \frac{1}{2}e^{-1}$

Analysis (CSMA/CD)



T_{prop} = maximum prop & detection delay between two node

$\text{efficiency} \overset{?}{=} \frac{T_{Trans}}{T_{Trans} + 2(e-1)\,T_{propagation}}$

Let's look at time t:

$M(t)$ = # of full frames up to time t:

$\sum_{i=1}^{M(t)} T^i_{frame} \le t \le \sum_{i=1}^{M(t)+1} T^i_{frame}$

So all is left is to compute

$\mathbb{E}(\#\text{ of failures in each frame})$.

the first success happens at n^{th} transmission

with prob $(1-1/e)^{n-1}\,1/e$

$\implies \mathbb{E}(\#\text{ of failures}) = \sum_{n=0}^{\infty}(n-1)(1-1/e)^{n-1}\left(\frac{1}{e}\right)$

$= e-1$

Plugging this in we get

$\text{Effeciency} = \frac{T_{Tran}}{T_{Trans} + 2(e-1)\,T_{prop}}$

By making $T_{tran} \gg T_{prop}$

Can be as close to 100% as we want.

Network Layer

$\frac{1}{M(t)}\sum_{i=1}^{M(t)} \underbrace{T^i_{frame}}_{x_i} \le \frac{t}{M(t)} \le \frac{1}{M(t)+1}\cdot\sum_{i=1}^{M(t)+1} T^i_{frame}\cdot\frac{M(t)+1}{M(t)}$

x_i's iid $\implies \frac{1}{n}\sum_{i=1}^{n} x_i \to \mathbb{E}(x_i)$

$\overset{t\to\infty}{\implies} \mathbb{E}(T_{frame}) \le \frac{t}{M(t)} \le \mathbb{E}(T_{frame})\cdot\left(\frac{1+M(t)}{M(t)}\right)$

$\frac{1+M(t)}{M(t)} \to 1$

On the other hand $\text{effeciency} = \frac{M(t)\,T_{trans}}{t}$

$= \frac{T_{Trans}}{\mathbb{E}(T_{frame})} = \frac{Trans}{T_{Trans} + 2\,\mathbb{E}(\#\text{ of failed attempts})\,T_{prop}}$

Network Layer