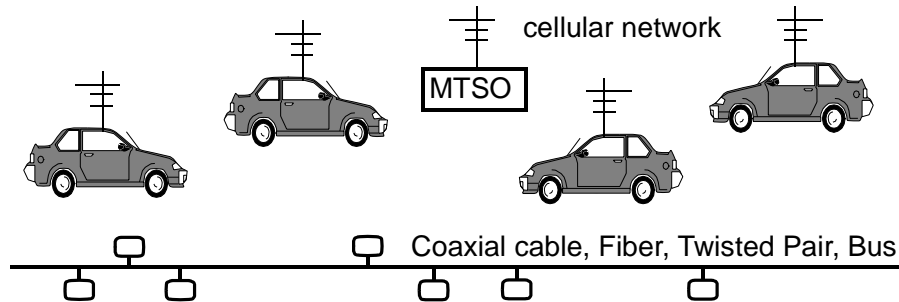




## Multiaccess Problem

How to let distributed users (efficiently) share a single broadcast channel?  
⇒ How to form a queue for distributed users?

- The protocols we used to solve this multiaccess problem are called multiaccess protocols. They are the lower sublayer of Data Link Control layer in the OSI model.
- The queueing theory studies properties of waiting queues. The mathematical formula of queueing theory can be used to evaluate the efficiency of different queueing system designs. In our applications, the efficiency of various multiaccess protocols.

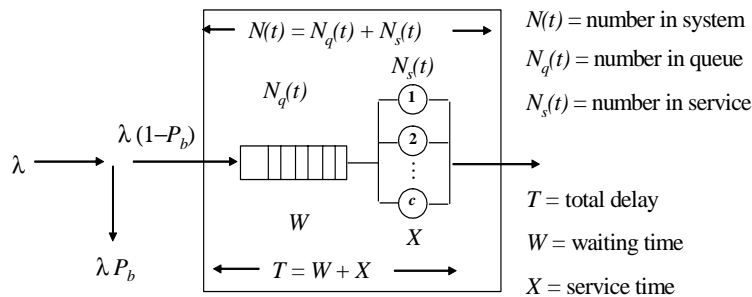


chow

Queueing—11/7/01—Page 1-



## Queueing Theory (See Appendix A1-3, Ch 5.5.1)



Parameter of interest to queueing analysis:

$\lambda$  average (avg, or mean) arrival rate (requests/sec) (packets/sec)

$\mu$  avg service rate (requests/sec) (packets/sec)

$\rho = \lambda/\mu$  utilization or traffic density, the ratio of system load to system capacity

$N$  avg no. of requests in the system, including those in buffers and in servers.

$T_w$  avg waiting time for a request at the buffer

$T_s$  avg service time for a request

$T$  avg delay in the system =  $T_s + T_w$ ; its inverse is the avg system throughput.

$P_B$  probability of request lost (due to buffer full situation)

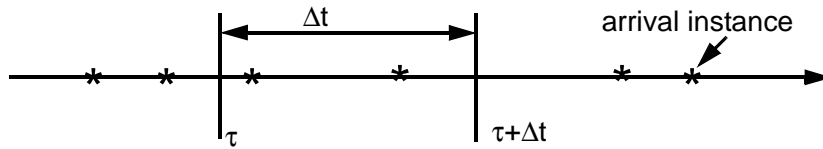
chow

Queueing—11/7/01—Page 2-



## Poisson Process

The random process used most frequently to model the arrival pattern.



The statistics of the Poisson process can be observed at any starting time:

1. Prob[instance occurrence in  $(\tau, \tau+\Delta t)$ ] =  $\lambda\Delta t + o(\Delta t)$ ;  $\lambda$  is mean arrival rate;
2. Prob[no instance occurrence in  $(\tau, \tau+\Delta t)$ ] =  $1 - \lambda\Delta t + o(\Delta t)$ ;
3. Arrivals are memoryless An arrival in one time interval of length  $\Delta t$  is independent of arrivals in previous or future intervals.

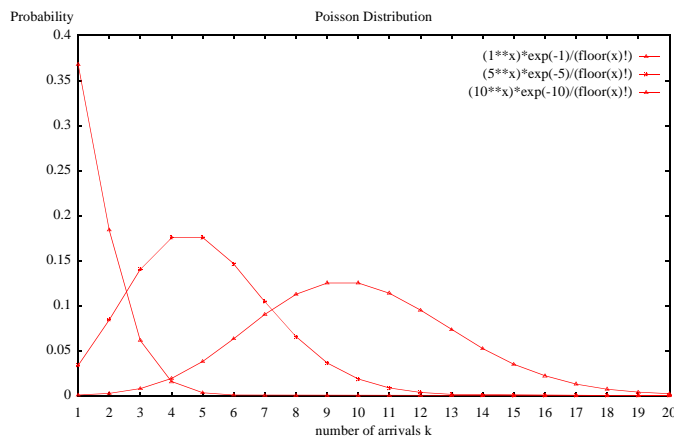
$o(\Delta t)$  implies that other terms are higher order in  $\Delta t$  and approaches 0 faster than  $\Delta t$ . Prob[2 or more arrivals in  $(\tau, \tau+\Delta t)$ ] =  $o(\Delta t)$ .

Note that the probability is independent of  $t$ .



## Poisson Distribution

Let  $N(\Delta t)$  = no. of arrivals in the interval  $(\tau, \tau+\Delta t)$  Prob[ $N(\Delta t)=k$ ] =  $(\lambda\Delta t)^k e^{-\lambda\Delta t}/k!$



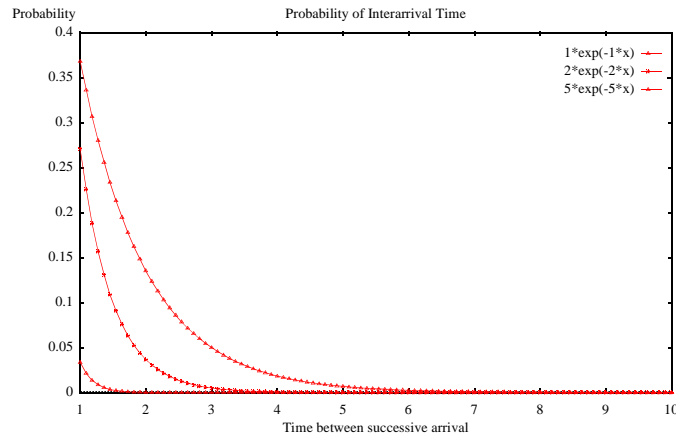
$$E[N(t)=k] = \sum_{k=0}^{\infty} k \frac{(\lambda t)^k}{k!} e^{-\lambda t} = \lambda t \quad \text{mean value}$$

$$\delta_k^2 = E[k \text{ } \text{and} \text{ } E[k]]^2 = E(k^2) \text{ } E^2(k) = \lambda t \quad \text{variance}$$



## Interarrival Time in Poisson Process

The interarrival time of the Poisson process is an exponential random variable with probability density function  $f(t) = \lambda e^{-\lambda t}$ , where  $t$  is the time between successive arrivals.



chow

Queueing—11/7/01—Page 5-



## Queueing Model Classification

*Arrival Process / Service Time / Servers / Max Occupancy*

Interarrival times  $\tau$

M = exponential

D = deterministic

G = general

Arrival Rate:

$$\lambda = 1 / E[\tau]$$

Service times  $X$

M = exponential

D = deterministic

G = general

Service Rate:

$$\mu = 1 / E[X]$$

1 server

$c$  servers

infinite

$K$  customers

unspecified if

unlimited

Multiplexer Models: M/M/1/K, M/M/1, M/G/1, M/D/1

Trunking Models: M/M/c/c, M/G/c/c

User Activity: M/M/ $\infty$ , M/G/ $\infty$

chow

Queueing—11/7/01—Page 6-

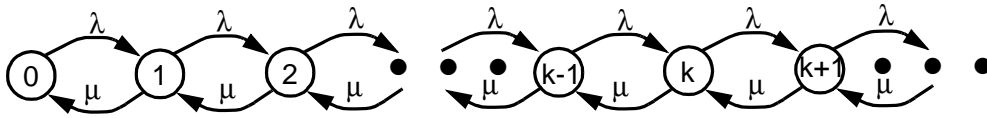


## M/M/1 Queue

Poisson arrival process, exponential service time, single server

$p_k = \text{Prob}[k \text{ customers in the system}]$

To analyze M/M/1 Queue, let us examine its system behavior described by the following state transition diagram. State number represent the number of customers in the system.



At equilibrium state the following equations hold

$$\lambda p_0 = \mu p_1$$

$$(\lambda + \mu)p_k = \lambda p_{k-1} + \mu p_{k+1} \quad k \geq 1$$

Alternatively,  $\lambda p_k = \mu p_{k+1} \quad k \geq 1$

Solving  $p_k = \frac{\lambda}{\mu} p_{k-1} = \rho p_{k-1}$  where  $\rho = \frac{\lambda}{\mu}$  by definition

$$p_k = (1 - \rho)\rho^k \quad k \geq 0$$

Mean number of customers in the system  $N = \sum_{k=0}^{\infty} k p_k = \frac{\rho}{1 - \rho}$

chow

Queueing—11/7/01—Page 7-



## M/M/1 Queue & Little's Law

Little's Law  $N = \lambda T$  where  $T$  is the mean delay inside the system.

Mean delay for M/M/1 system

$$T = \frac{N}{\lambda} = \frac{\rho/\lambda}{1 - \rho} = \frac{1/\mu}{1 - \rho} = \frac{1}{\mu(1 - \rho)} = \frac{1}{\mu'(1 - \rho)}$$

where  $C =$  server service rate in # operations/sec (#bits/sec for transmission system)

$1/\mu' =$  mean # operations/customer (#bits/packet for TX system)

chow

Queueing—11/7/01—Page 8-



## Evolution of Queues

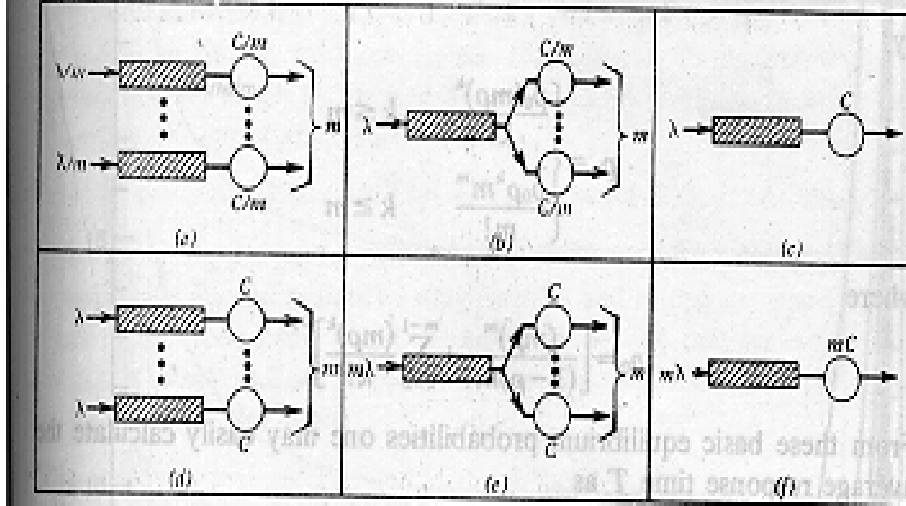


Figure 5.5 Evolution of queueing structures.

multiple queues → single queue multiple servers → single queue shared server

chow

Queueing—11/7/01—Page 9-



## Why we use statistical multiplexing?

Advanced Queueing analysis presents the following interesting result:

$$T(1, \lambda, C) \leq T(m, \lambda, C) \quad \text{while} \quad W(1, \lambda, C) \geq W(m, \lambda, C)$$

where  $(m, \lambda, C)$  is a system with  $m$  servers, total capacity  $C$ , arrival rate  $\lambda$ ,

$T$  is the total system delay and  $W$  is the waiting time in the queue.

Sharing a single high speed server increase “contention” delay in the queue but decrease overall delay due to much shorter service time.

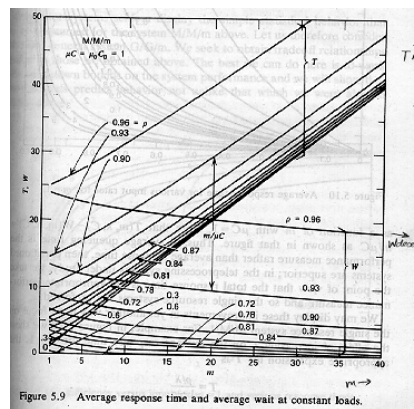


Figure 5.9 Average response time and average wait at constant loads.

chow

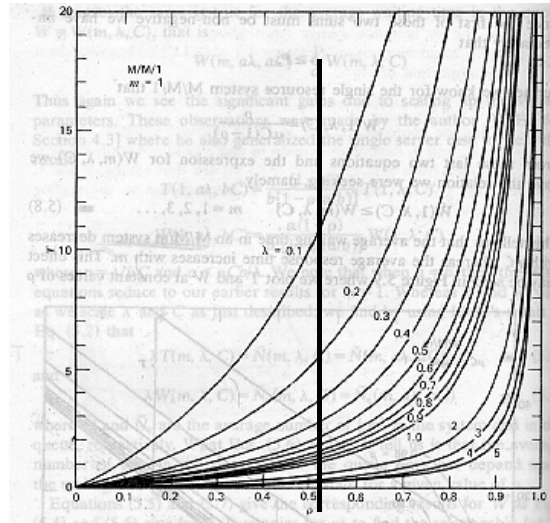
Queueing—11/7/01—Page 10-



## Scale factor of M/M/1 Queueing System

$T = \frac{\rho/\lambda}{1 - \rho}$  if  $\lambda \uparrow$  and  $C \uparrow$  so that  $\rho$  remains constant, then  $T \downarrow$ .

This benefit is in addition to economy of scale.



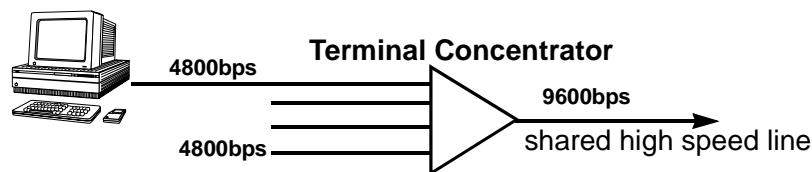
chow

Queueing—11/7/01—Page 11-



## Application of Queueing Theory

Case 1. Terminal Concentrators:



mean packet length  $1/\mu' = 1000$  bits/packet

input lines traffic are Poisson process with mean arrival rate  $\lambda_i = 2$  packets/sec.

Q1: What is the mean delay experienced by a packet from the time the last bit arrives at the concentrator until the moment that bit is retransmitted on the output line?

Use  $T = \frac{1}{\mu' C - \lambda}$ , where  $\lambda = 4 \times \lambda_i = 8$ ,  $\mu' C = 9.6$  packets/sec.

$\rightarrow T = 1/(9.6 - 8) = 0.625$  sec.

Q2: What is the mean number of packets in the concentrator, including the one in service?

A: Use the little's law  $N = \lambda T = 8 \times 0.625 = 5!$

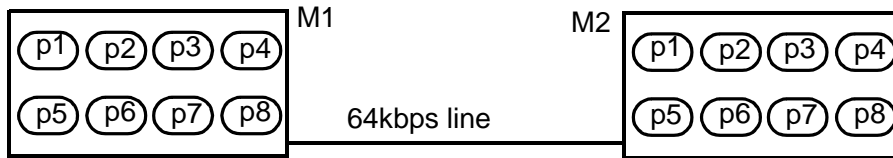
chow

Queueing—11/7/01—Page 12-



## Application of Queueing Theory

Dedicated vs. Shared Channels:



Eight parallel sessions using this 64kbps line. Each session generates Poisson traffic with  $\lambda_i=2$  packets/sec. Packet lengths are exponentially distributed with a mean of 2000 bits.

There are two design choices:

- Each session is given a dedicated 8kbps channel (via FDM or TDM).
- Packets of all sessions compete for a single 64kbp shared channel.

Which one gives a better response time?

A: a) For 8kbps channel,  $\lambda=2$  packets/sec,  $\mu'=1/2000$  packets/bit,  $C=8000$  bits/sec,  $\mu'C=4$  packets/sec,  $T=1/(\mu'C-\lambda)=1/(4-2)=0.5$  sec.

- For 64kbps shared channel,  $\lambda=8 \times 2=16$  packets/sec,  $\mu'=1/2000$  packets/bit,  $C=64000$  bits/sec,  $\mu'C=32$  packets/sec,  $T=1/(\mu'C-\lambda)=1/(32-16)=0.0667$  sec.

Reason?