King Fahd University of Petroleum & Minerals
Computer Engineering Dept

CSE 642 – Computer Systems Performance

Term 041

Dr. Ashraf S. Hasan Mahmoud

Rm 22-148-3

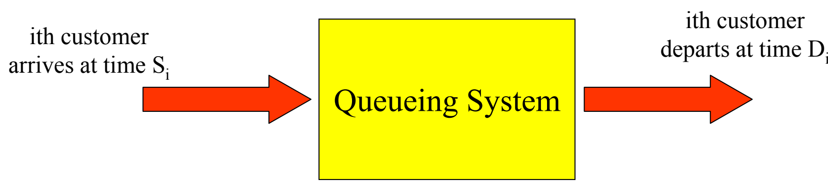
Ext. 1724

Email: ashraf@ccse.kfupm.edu.sa

Queuing Model

- **Consider the following system:**

$A(t)$ $\quad N(t) = A(t) – D(t)$ $\quad D(t)$



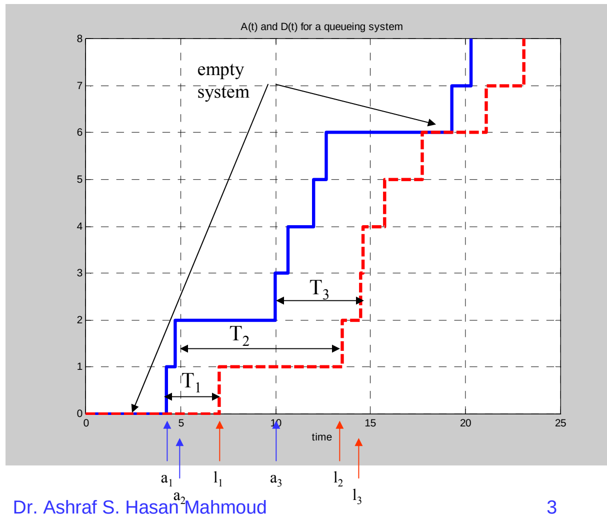
$T_i = D_i – A_i$

$W_i = T_i – S_i$
$= D_i – A_i – S_i$

A(t) – number of arrivals in (0, t]
D(t) – number of departures in (0, t]
N(t) – number of customers in system in (0,t]
T_i – duration of time spent in system for ith customer
W_i – duration of time spent waiting for service for ith customer

Example: Queueing System

- a_i and l_i arrival and departure instances
- $T_i = l_i - a_i$ – time spent in the system
- If A(t) = D(t) → system is empty
- The graph is shown for FCFS service



Little's Formula

- **Consider the time average of the number of customers in the system N(t) during (0,t],**

$$\langle N \rangle_t = \frac{1}{t}\int_0^t N(\tau)d\tau$$

i.e. average area under the curve for N(t)

$\langle N \rangle_t$ is also given by

$$\langle N \rangle_t = \frac{1}{t}\sum_{i=1}^{A(t)} T_i$$

Little's Formula – cont'd

- The average arrival rate $<\lambda>_t$ is given by

$$\langle\lambda\rangle_t = \frac{A(t)}{t}$$

- Combining the previous equations we get:

$$\langle N\rangle_t = \langle\lambda\rangle_t \frac{1}{A(t)}\sum_{i=1}^{A(t)} T_i$$

- Let the quantity $<T>_t$ be the average time a customer spends in the system, then

$$\langle T\rangle_t = \frac{1}{A(t)}\sum_{i=1}^{A(t)} T_i$$

Little's Formula – cont'd

- Combining the last two equations:

$$\langle N\rangle_t = \langle\lambda\rangle_t \langle T\rangle_t$$

- Which relates the time averages of the arrival rate, the number of customers in the system and the average time spent in the system
- Let t → ∞, then one can write:

$$E[N] = \lambda E[T]$$

Under what conditions will $<N>_t$ → E[N] for t → ∞?

Little's Formula – cont'd

- **Little's formula:**

  E[N] = λE[T]

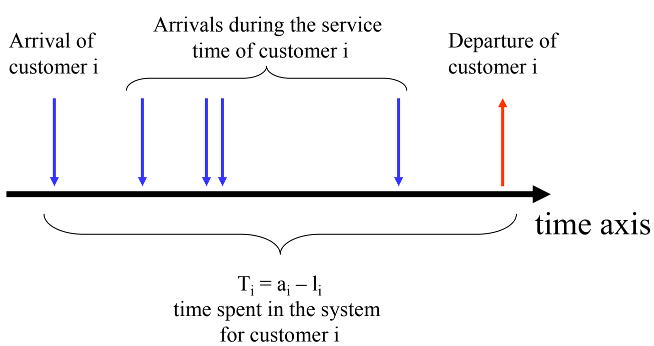
Holds for many service disciplines and for systems with arbitrary number of servers. It holds for many interpretations of the system as well

Note: $\sum_{i=1}^{A(t)} T_i = \sum_{i=1}^{A(t)} d_i - l_i = \sum_{i=1}^{A(t)} d_i - \sum_{i=1}^{A(t)} l_i$ **does not depend on the service order**

Intuitiveness of Little's Formula

- **Little's formula:**

  E[N] = λE[T]



- **Formula applies to many interpretations of "system"!**

Example 1:

- **<u>Problem</u>: Let Ns(t) be the number of customers being served at time t, and let τ denote the service time. If we designate the set of servers to be the "system" then Little's formula becomes:**

$$E[Ns] = \lambda E[\tau]$$

where E[Ns] is the average number of busy servers for a system in the steady state.

Example 1: cont'd

Note: for a single server Ns(t) can be either 0 or 1 ➔ E[Ns] represents the portion of time the server is busy. If p_0 = Prob[Ns(t) = 0], then we have

$$1 - p_0 = E[Ns] = \lambda E[\tau], \text{ Or}$$
$$p_0 = 1 - \lambda E[\tau]$$

The quantity $\lambda E[\tau]$ is defined as the utilization for a single server. Usually, it is given the symbol ρ

$$\rho = \lambda E[\tau]$$

For a c-servers system, we define the utilization (the fraction of busy servers) to be

$$\rho = \lambda E[\tau] / c$$

Poisson Process

- Refer to the Summation process example in the Random Processes package

- Def: Poisson process to be the point process for which the number of events (successes), X(t), in a t-second interval is given by the Poisson distribution

$$P_k(t) = P(X(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad k = 0,1,\ldots$$

where λ is the average rate of success per time unit

Poisson Process - Properties

- The random process X(t) is a Markov Process. For arbitrary times: $t_1<t_2<\ldots<t_k<t_{k+1}$

$$\text{Prob}[X(t_{k+1}) = x_{k+1}/X(t_k)=x_k, \ldots, X(t_1)=x_1]$$

$$= \text{Prob}[X(t_{k+1}) = x_{k+1}/X(t_k)=x_k]$$

- Independent increments
- Stationary increments

Poisson Process – Interarrival Time

- Let T be the random time between two consecutive events
- The distribution function is given by

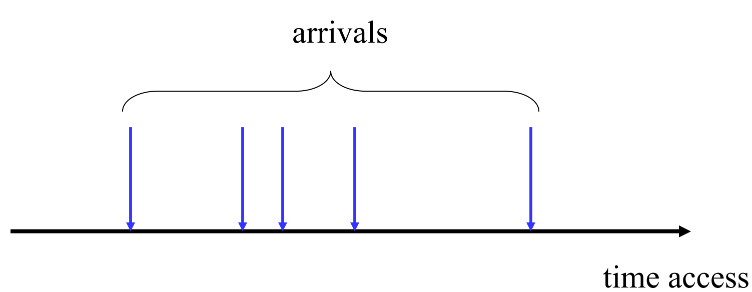
$F_T(t) = P(T \leq t)$
$= P(\text{at least one arrival in t seconds})$
$= 1 - P(\text{0 arrivals in t seconds})$
$= 1 - P_0(t)$
$= 1 - e^{-\lambda t}$

Therefore $f_T(t)$ is equal to $\lambda e^{-\lambda t}$ for $t \geq 0$

- Poisson Process ≡ interarrival times are independent and exponentially distributed

Uniformity Property

- Def – give a number of arrivals in an interval, the arrivals are uniformly distributed throughout the interval!



Uniformity Property – cont'd

- **Proof:**

Suppose that we are given than one arrival occurs in the interval [0,t],
Let Y be the arrival time of the single customer → 0 < y < t
Let X(y) be the number of events up to time y → X(t) – X(y) is the increment in the interval (y, t]

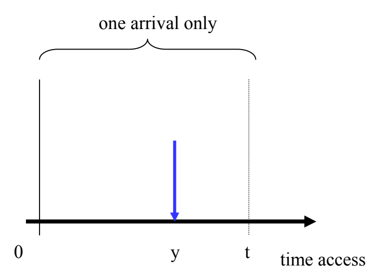
$$P(Y \le y) = P(X(y) = 1 / X(t) = 1]$$

$$= \frac{P(X(y) = 1 \text{ and } X(t) - X(y) = 0]}{P(X(t) = 1)}$$

$$= \frac{P(X(y) = 1)\, P(X(t) - X(y) = 0)}{P(X(t) = 1)}$$

$$= \frac{\lambda y e^{-\lambda y}\, e^{-\lambda(t-y)}}{\lambda t e^{-\lambda t}}$$

$$= y / t$$

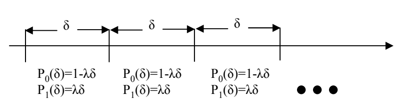


Kolmogorov Forward Differential Equations

- **Consider the incremental time interval δ, so small that λδ << 1 for all λ**
- **Using the Poisson density function and knowing that $e^{-\lambda\delta} \approx 1 - \lambda\delta + O(\delta)$ – where O(δ) are higher order terms of δ (i.e. lim O(δ)/ δ = 0 as δ → 0)**
- **One can write:**

$$P_0(\delta) = 1 - \lambda\delta + O(\delta)$$
$$P_1(\delta) = \lambda\delta + O(\delta)$$
$$P_i(\delta) = O(\delta) \quad \text{for } i \ge 2$$

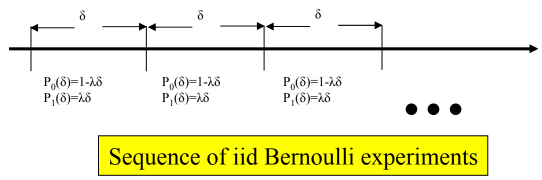
This means, we choose δ small such that the likelihood of more than one arrival during δ is close to zero



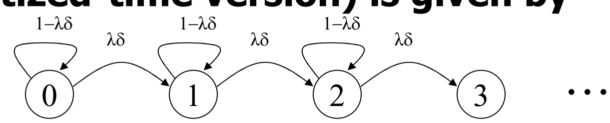
Sequence of iid Bernoulli experiments

Kolmogorov Forward Differential Equations – cont'd

- **This means, we choose δ small such that the likelihood of more than one arrival during δ is close to zero**



- **The corresponding state diagram (for the discretized-time version) is given by**



Kolmogorov Forward Differential Equations – cont'd

- **Let us study the evolution of $P_n(t)$ with respect to time, t**
  - **Remember $P_n(t)$ is the probability of n arrivals in an interval t**
- **Consider the change in $P_n(t)$ in the incremental interval (t, t+ δ)**

Kolmogorov Forward Differential Equations – cont'd

- **Case n = 0**

$P_0(t+\delta)$ = P(no arrivals in (0,t+ δ))
= P(no arrivals in (0,t)) P(no arrivals in (t,t+ δ))
= $P_0(t)(1 - \lambda\,\delta)$

- **Case n > 0**

$P_n(t+\delta)$ = P(n arrivals in (0,t+ δ))
= P(n arrivals in (0,t)) P(no arrivals in (t,t+ δ))
+ P(n-1 arrivals in (0,t)) P(1 arrival in (t,t+ δ))
= $P_n(t)(1 - \lambda\,\delta) + P_{n-1}(t)(\lambda\,\delta)$

- **The above equations can be written as**

$$[P_0(t+\delta) - P_0(t)]/\,\delta = -\,\lambda\, P_0(t), \text{ and}$$

$$[P_n(t+\delta) - P_n(t)]/\,\delta = -\,\lambda\, P_n(t) + \lambda\, P_{n-1}(t), \quad n> 0$$

Kolmogorov Forward Differential Equations – cont'd

- **Take the limit as δ → 0, the previous equations can be written as:**

$$dP_0(t)/dt = -\,\lambda\, P_0(t), \text{ and}$$

$$dP_n(t)/dt = -\,\lambda\, P_n(t) + \lambda\, P_{n-1}(t), \quad n> 0$$

- **Verify that $P_k(t)$ given by**

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad k = 0,1,\ldots$$

is a solution for the Kolmogorov Forward differential equations

Kolmogorov Forward Differential Equations – cont'd

- **Another form for the Kolmogorov D.E. is as follows:**

$$\frac{d\tilde{P}(t)}{dt} = \Lambda \tilde{P}(t)$$

where $\tilde{P}(t) = [P_0(t) \quad P_1(t) \quad P_2(t) \quad \ldots.]^T$

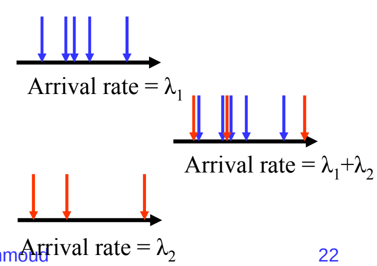
$$\Lambda = \begin{bmatrix} -\lambda & 0 & 0 & 0 & \ldots \\ \lambda & -\lambda & 0 & 0 & \ldots \\ 0 & \lambda & -\lambda & 0 & \ldots \\ 0 & 0 & \lambda & -\lambda & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Λ is the infinitesimal generator matrix

Note sum of columns = zero

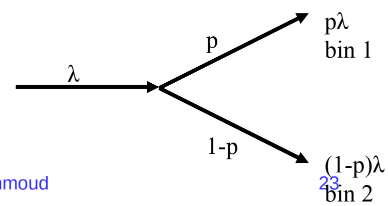
Adding Poisson Processes

- **Sum of two INDEPENDENT Poisson processes**
- **Consider an incremental interval δ**
  - **The probability of an arrival from either source is $\lambda_1\delta(1-\lambda_2\delta)+(1-\lambda_1\delta)\lambda_2\delta \approx (\lambda_1+\lambda_2)\delta$**
  - **The probability of arrivals from both source is $\lambda_1\delta\,\lambda_2\delta = \lambda_1\lambda_2\delta^2 \approx 0$**
- **Therefore, the sum is a Poisson process with rate $(\lambda_1+\lambda_2)$**



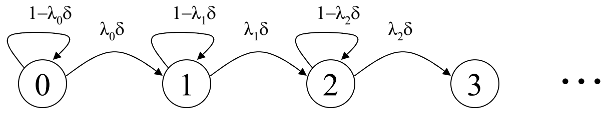
Splitting Poisson Processes

- Splitting of a Poisson processes
- Consider an incremental interval δ
  - The probability of an arrival to bin 1: λδp
  - The probability of an arrival to bin 1: λδ(1-p)
  - Since subsequence arrivals to either bins are independent and identically distributed
- Therefore, the arrivals processes to bin 1 and 2 Poisson with rate pλ and (1-p)λ, respectively



Pure Birth Processes

- Poisson process is a member of a wider class of "pure birth processes"
- In general the probability of an arrival in an interval δ can be function of the number in the system, $\lambda_n\delta$
- The corresponding state diagram will be



Pure Birth Processes – cont'd

- **In the same manner, you can show that the corresponding Kolmogorov D.E are given by**

$dP_0(t)/dt = -\lambda_0 P_0(t)$, and

$dP_n(t)/dt = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t)$, n> 0

Subject to the condition $\sum_{n=0}^{\infty} P_n(t) = 1$

Pure Birth Processes – cont'd

- **Putting the Kolmogorov D.E.s in a matrix form:**

$$\frac{d\tilde{P}(t)}{dt} = \Lambda \tilde{P}(t)$$

Necessary and sufficient condition for stability is $\sum 1/\lambda_n = \infty$

where $\tilde{P}(t) = [P_0(t) \quad P_1(t) \quad P_2(t) \quad \ldots]^T$

$$\Lambda = \begin{bmatrix} -\lambda_0 & 0 & 0 & 0 & \cdots \\ \lambda_0 & -\lambda_1 & 0 & 0 & \cdots \\ 0 & \lambda_1 & -\lambda_2 & 0 & \cdots \\ 0 & 0 & \lambda_2 & -\lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Λ is the infinitesimal generator matrix

Note sum of columns = zero

Example: Yule-Furry Process

- For Yule-Furry process, $\lambda_n = n\,\lambda$ – linear rate with system population
- The evolution equations are then given by

$$dP_n(t)/dt = -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t); \quad n \geq k$$

- For the initial condition $P_k(0)=1$ for some $k > 0$, show that

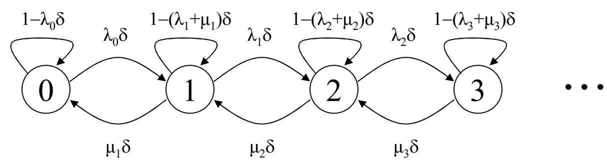
$$P_n(t) = \binom{n-1}{k-1} e^{-n\lambda t}\left(1-e^{-\lambda t}\right)^{n-k} \qquad n \geq k, t \geq 0$$

is a solution

Poisson Arrivals See Time Averages (PASTA)

Birth And Death Processes

- **The corresponding state diagram is as shown:**



- **The Kolmogorov D.E are given by**

$dP_0(t)/dt = - \lambda_0 P_0(t) + \mu_1 P_1(t)$, and

$dP_n(t)/dt = - (\lambda_n+\mu_1)P_n(t) + \lambda_{n-1}P_{n-1}(t) + \mu_{n+1}P_{n+1}(t)$, $n> 0$

Subject to the condition $\sum_{n=0}^{\infty} P_n(t) = 1$

Birth And Death Processes – cont'd

- **Putting the Kolmogorov D.E.s in a matrix form:**

$$\frac{d\tilde{P}(t)}{dt} = M\tilde{P}(t)$$

+ve solution exists if $0 \le \lambda_n < \mu_n$

where $\tilde{P}(t) = [P_0(t) \quad P_1(t) \quad P_2(t) \quad \ldots]^T$

$$M = \begin{bmatrix} -\lambda_0 & \mu_1 & 0 & 0 & \cdots \\ \lambda_0 & -\lambda_1 - \mu_1 & \mu_2 & 0 & \cdots \\ 0 & \lambda_1 & -\lambda_2 - \mu_2 & \mu_3 & \cdots \\ 0 & 0 & \lambda_2 & -\lambda_3 - \mu_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Global Balance Equations

- Steady state solution ➔ $dP(t)/dt = 0$
- The resulting set of equations:

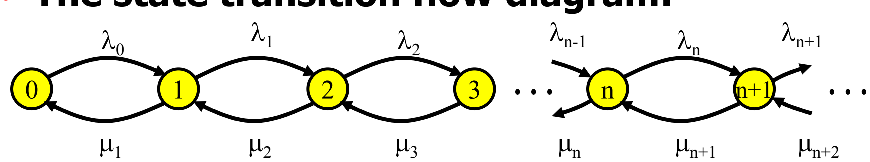
$$\lambda_0 P_0 = \mu_1 P_1, \text{ and}$$

$$(\lambda_n+\mu_n)P_n = \lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1}, \quad n > 0$$

In addition to the normalizing condition $\sum_{n=0}^{\infty} P_n = 1$

Global Balance Equations – cont'd

- The state transition flow diagram:



- We can show the solution for the global balance equation is given by

$$P_n = P_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i}$$

and

$$P_0 = \left[1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i}\right]^{-1}$$

⇐ **The basis for all queueing formula to come!!**

Queueing Models: M/M/1

- Making the substitutions: $\lambda_n = \lambda$ and $\mu_n = \mu$, and defining $\rho = \lambda/\mu$, one can write

$$P_n = (1-\rho)\rho^n \qquad n = 0,1,2,\ldots$$

or

$$P(z) = \frac{1-\rho}{1-z\rho}$$

- The mean and variance of number of customers in system, E[N] and Var[N] are given by

$$E[N] = \frac{\rho}{1-\rho} \qquad Var[N] = \frac{\rho}{(1-\rho)^2}$$

- The mean delay in the M/M/1 queue can be obtained through the application of Little's formula:

$$E[D] = E[N]/\lambda = \frac{1}{\mu-\lambda}$$

M/M/1- Delay Distribution

- The probability of n customers as a departing customer departs after spending t seconds in system is given by

$$P_n = \int_0^\infty \frac{(\lambda t)^n e^{-\lambda t}}{n!} d(t)dt \qquad n = 0,1,\ldots$$

or

$$P(z) = \sum_{n=0}^{\infty} p_n z^n = \sum_{n=0}^{\infty} z^n \int_0^\infty \frac{(\lambda t)^n}{n!} e^{-\lambda t} d(t)dt$$

$$P(z) = \int_0^\infty e^{-\lambda t(1-z)} d(t)dt = D(\lambda(1-z))$$

Note this probability is the same as the probability of n customers in system – Also equal to the probability of finding n customers in system by an arriving customer (refer to PASTA property)

d(t) is the PDF for the total delay time

Therefore, D(s) is given by

$$D(s) = \frac{\mu-\lambda}{s+\mu-\lambda}$$

i.e. the delay for M/M/1 queue is exponentially distributed with mean 1/(μ-λ),

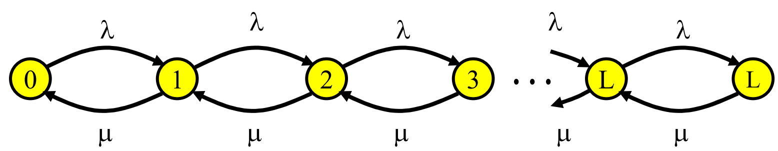
$$d(t) = (\mu-\lambda)e^{-(\mu-\lambda)t} \qquad t \ge 0$$

Queueing Models: M/M/1/L

- **Finite Capacity Case: $\lambda_j = \lambda$ for $j<L$**
  0 for $j \geq L$

  also $\mu_j = \mu$

- **The state-transition flow diagram of M/M/1/L queue is as shown below**



Queueing Models: M/M/1/L – cont'd

Steady-state pmf is given by

$$P_n = \begin{cases} \dfrac{(1-\rho)\rho^n}{1-\rho^{L+1}} & n \leq L \\ 0 & n > L \end{cases}$$

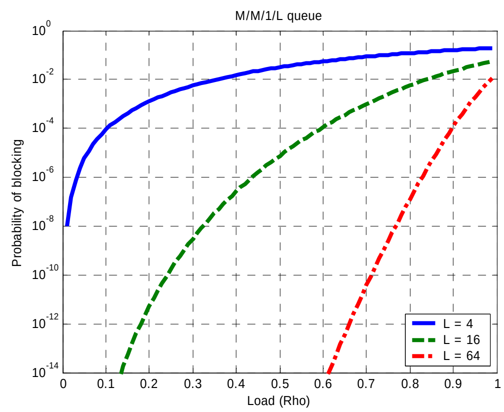
- **What is P(z) equal to?**

- **In particular, the blocking probability, P_L, is given by the relation above for n = L**

Queueing Models: M/M/1/L – cont'd

- **In particular, the blocking probability, P_L, is given by the relation above for n = L**

$$P_L = \frac{(1-\rho)\rho^L}{1-\rho^{L+1}}$$



Example: M/M/1/L – cont'd

- **<u>Problem</u>: A voice signal is digitized at a rate of 8000 bps. The average length of a voice message is 3 min. Messages are transmitted on a DS-1 line, which has the capacity of 1.344 Mbps. While waiting for transmission, the messages are stored in a buffer which has a capacity of 10^7 bit. Plot the blocking probability versus the voice message arrival rate.**

Example: M/M/1/L – cont'd

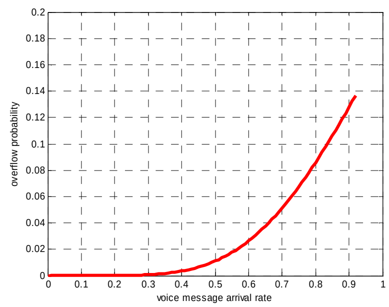
- **Solution:**

```
%
% Example 3.7 - voice multiplexing - page 91
clear all
LineWidth = 3;

DS1_Capacity     = 1.344e6; % bits/sec
BuffSizeBits     = 1e7; % different than textbook
BPSPerVoiceMsg   = 8000;% bps per voice msg
VoiceMsgDuration = 3*60; % second;
VoiceMsgSizeBits = VoiceMsgDuration * BitsPerVoiceMsg;
ServiceTime      = VoiceMsgSizeBits / DS1_Capacity;
% # of msgs buffer can fit
BufferSizeMsgs   = floor(BuffSizeBits/VoiceMsgSizeBit

Step  = 0.01;
Lamda = [0:Step:(1-Step)/ServiceTime];
Rho   = Lamda * ServiceTime;
PB    = (1-Rho).*Rho.^BufferSizeMsgs./(1-Rho.^(BufferSizeMsgs+1));
%
% Plot results
figure(1)
h = plot(Lamda, PB,'-r');
set(h, 'LineWidth', LineWidth);
xlabel('voice message arrival rate'); grid
ylabel('overflow probability');
axis([0 1 0 0.2]);
```

Note since voice message size is 1440000 bits, then buffer size can not be 10^6 bits as stated in the textbook. Here we use buffer size of 10^7 bits which means, buffer can accommodate 6 voice messages before it overflows.
Refer to example 3.7 page 91 in textbook



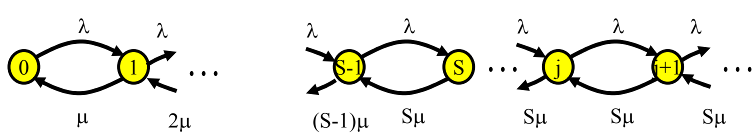
Queueing Models: M/M/S – Multiserver Systems

- **Assume S servers system, therefore:**

$$\mu_j = j\mu \text{ for } j \leq S$$
$$S\mu \text{ for } j > S$$

and $\lambda_j = \lambda$ **for all j**

- **The state-transition flow diagram of M/M/S queue is as shown below**



Erlang C model – Blocked calls are QUEUED

Queueing Models: M/M/S – Multiserver Systems – cont'd

- **Solving the balance equations, results in**

$$P_j = \begin{cases} \dfrac{P_0 \rho^j}{j!} & j \le S \\ \dfrac{P_0 \rho^j}{S! S^{j-S}} & j > S \end{cases}$$

P_0 is calculated as

$$P_0 = \left[\sum_{j=0}^{S-1} \frac{\rho^j}{j!} + \frac{S\rho^S}{S!(S-\rho)} \right]^{-1}$$

- **The traffic utilization, ρ = λ/ μ**
- **Note the condition for solution validity is ρ/S < 1**

i.e. in the S-server case, the traffic load ranges 0 to S.

Queueing Models: M/M/S – Multiserver Systems – cont'd

- **The probability of queueing is equal to the probability of finding all S servers busy, therefore,**

$$P_c(S,\rho) = \sum_{j=S}^{\infty} P_j = P_0 \frac{\rho^S}{S!} \frac{S}{(S-\rho)}$$

- **The mean number of customers in queue, E[Nq], is given by**

$$\overline{Q} = E[Nq] = \sum_{j=0}^{\infty} j P_{j+S} = P_0 \frac{\rho^S}{S!} \frac{S\rho}{(S-\rho)^2}$$

- **Therefore, the relation between average number of customers in queue and probability of queueing is given by**

$$\overline{Q} = \frac{P_c \rho}{(S-\rho)}$$

Exercise: M/M/S/∞

- **Show that the waiting time distribution is given by**

$$F_W(x) = 1 - \frac{p_c S}{S-\rho} e^{-\mu(S-\rho)x} \qquad x > 0$$

Refer to slides of "Queueing Models" for COE 541 for proof.

Example: M/M/S/∞

- **Problem: a 160 kb/s line is used for data transmission. Two options are provided**

a) Implement a 16-channel TDM scheme where every channel provides 10 kb/s.

b) Use the overall trunk as one *fat* data transmission pipe.

Assume data frames arrive at a Poisson rate λ and are exponentially distributed in length with average of 2000 bits per frame.

Which scheme provides less delay?

Example: M/M/S/∞ - cont'd

- **Solution:**

  a) S = 16 servers – Model M/M/S

  R_c = 10 kb/s → $E[\tau] = 1/\mu = 2000/10 = 200$ msec

  $\rho = \lambda/\mu = \lambda E[\tau] = 200\lambda$

  $E[T] = E[W] + E[\tau] = E[Nq]/\lambda + E[\tau]$

  $= P_c (1/\mu)/(S-\rho) + E[\tau]$

  b) S = 1 server – Model M/M/1

  R_c = 160 kb/s → $E[\tau] = 1/\mu = 2000/160 = 1.25$ msec

  $\rho = \lambda/\mu = \lambda E[\tau] = 1.25\lambda$

  $E[T] = E[W] + E[\tau] = E[Nq]/\lambda + E[\tau]$

  $= 1/(\mu - \lambda)$

Example: M/M/S/∞ - cont'd

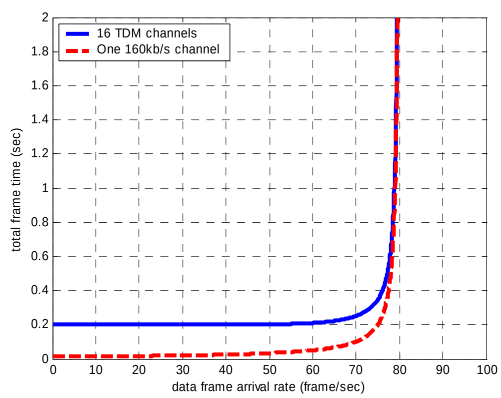
- **Solution:**

  For option (a)
  - minimum service time is equal to 200 msec

  For option (b)
  - minimum service time is equal to 1.25 msec

Option (b) provides better (less) system

Note: The x-axis in the textbook graph is not correct (Example 3.8 page 94). Verify?



Example: M/M/S/∞ - cont'd

```
%
% Example 3.8 - voice multiplexing - page 94
clear all
LineWidth = 3;
Line_Capacity    = 160e3; % bits/sec
NoOfChannels     = 16; % No of TDM channels
RateTDMChannel   = Line_Capacity/NoOfChannels;% bps per channel
AvgFrameSizeBits = 2000; % bits
%
% option (a) - 16 TDM channels - M/M/S queue
ServiceTime_a = AvgFrameSizeBits / RateTDMChannel;
S             = NoOfChannels;
Step          = 0.05;
Lamda_a       = [Step:Step:S/ServiceTime_a - Step];
Rho_a         = Lamda_a * ServiceTime_a;
[P0 PS Pc]    = Get_M_M_S(S, Rho_a);
W_a           = Pc.*Rho_a./(S-Rho_a)./Lamda_a;
T_a           = W_a + ServiceTime_a;
%
% option (b) - 1 160 kb/s channel - M/M/1 queue
ServiceTime_b = AvgFrameSizeBits / Line_Capacity;
Step          = 0.05;
Lamda_b       = [Step:Step:1/ServiceTime_b-Step];
Rho_b         = Lamda_b * ServiceTime_b;
T_b           = 1./(1./ServiceTime_b - Lamda_b);
%
% Plot results
figure(1)
h = plot(Lamda_a, T_a,'-', Lamda_b, T_b,'--r');
set(h, 'LineWidth', LineWidth);
xlabel('data frame arrival rate (frame/sec)'); grid
ylabel('total frame time (sec)');
legend('16 TDM channels', 'One 160kb/s channel', 2);
axis([0 100 0 2]);
```

```
function [P0, PS, Pc]    = Get_M_M_S(S, Rho);
% compute P0, PS, and Pc for an M/M/S queue given S and Rho
P0 = zeros(size(Rho));
PS = zeros(size(Rho));
Pc = zeros(size(Rho));
temp = zeros(size(Rho));
for i=0:S-1
    temp = temp + Rho.^i./factorial(i);
end
temp = temp + S.*Rho.^S./(factorial(S).*(S - Rho));
P0 = 1./temp;
PS = P0 .* Rho.^S./factorial(S);
Pc = PS .* S./(S - Rho);
```

Code to generate key probabilities (P_0, P_S, P_c) for M/M/S system

Code to compare between options (a) and (b)

Queueing Models: M/M/S/L

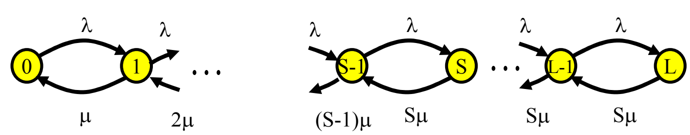
- **S server model with finite waiting room**
- **Assuming L ≥ S, we have**

$$\mu_j = j\mu \text{ for } j \le S$$
$$\quad\;\; S\mu \text{ for } j > S$$

and

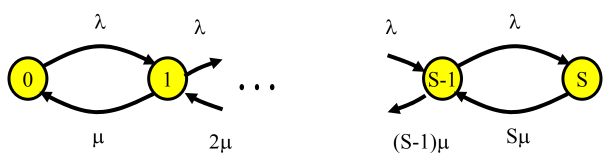
$$\lambda_j = \lambda \quad \text{for } j < L$$
$$\quad\;\; 0 \quad \text{for } j \ge L$$

- **The state transition flow diagram M/M/S/L queue**



Queueing Models: M/M/S/S

- Special case of M/M/S/L where L = S;
- The state transition flow diagram M/M/S/S queue



Erlang B model – Blocked calls are CLEARED

Queueing Models: M/M/S/S – cont'd

- Solving the balance equation yields:

$$P_n = \frac{P_0 \rho^n}{n!} \qquad n = 0,1,2,\ldots,S$$

and

$$P_0 = \left[\sum_{n=0}^{S} \frac{\rho^n}{n!}\right]^{-1}$$

- When an arrival finds all S servers busy, it is blocked or dropped (no waiting room) – Probability of blocking is given by

$$P_B(S,\rho) = \frac{\rho^S / S!}{\sum_{n=0}^{S} \frac{\rho^n}{n!}}$$

$$P_B(S,\rho) = \frac{\rho P_B(S-1,\rho)}{S + \rho P_B(S-1,\rho)}$$

where $P_B(0,\rho) = 1$

- **Insensitivity Property of Erlang-B formula: Blocking probability does NOT depend on the distribution of the service time, but rather its mean**

Example: M/M/S/S

- **Problem: constant length frames of 1000 bit each arrive an a multiplexer which has 16 output lines, each operating at a 50 kb/s rate. Suppose that frames arrive at an average rate of 1,440,000 frame per hour. There is no storage; thus if a frame is not served immediately it lost.**

  Calculate the blocking probability at the multiplexer.

Example: M/M/S/S – cont'd

- **Solution:**

  frame arrival rate, λ = 1,440,000 frame/hour
  = 400 frame/sec
  frame service time, $1/\mu$ = 1000 / 50 kb/s
  = 0.02 sec
  Traffic intensity, ρ = λ / μ = 8
  Number of servers, S = 16 (verify $\rho/S < 1$)

  Using the iterative formula ➔

S	1	2	3	4	5	6	7	8
$P_B(S, \rho)$	0.8889	0.7805	0.6755	0.5746	0.4790	0.3898	0.3082	0.2356
S	9	10	11	12	13	14	15	16
$P_B(S, \rho)$	0.1731	0.1217	0.0813	0.0514	0.0307	0.0172	0.0091	0.0045

our answer

M/M/S/S – Infinite Servers Case

- **Special case of the M/M/S/S queue**
- **Let S → ∞, i.e. an arriving customer always has a server available**
- **The probability of system in state zero is given by**

$$P_0 = \left[\sum_{n=0}^{\infty} \frac{\rho^n}{n!}\right]^{-1} = e^{-\rho}$$

- **Therefore, the probability of system in state n ≥ 0 is computed as**

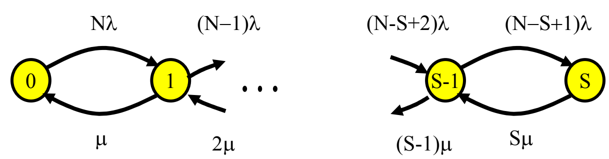
$$P_n = \frac{\rho^n}{n!} e^{-\rho}$$

Which is the Poisson distribution!!

Finite Source Queueing – Engset Distribution

- **Assume a finite population of N – each generate a message with rate λ (or with probability λδ in the interval (t, t+δ)). The next message is not transmitted till the prior one is served. Assume no storage case, i.e. if a source generates a message when no server is available, the message is lost and the source returns to idle state immediately.**

- **The state transition flow diagram is as shown:**



Finite Source Queueing – Engset Distribution – cont'd

- **You can show that the pmf is given by**

$$P_n = P_0 \binom{N}{n} \left(\frac{\lambda}{\mu}\right)^n \qquad n = 0,1,\ldots,S$$

and

$$P_0 = \left[\sum_{n=0}^{S} \binom{N}{n} \left(\frac{\lambda}{\mu}\right)^n \right]^{-1}$$

- **Remember P_S is the probability of blocking**
- **There is no blocking for N ≤ S – Why?**

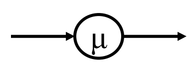
Method Of Stages – Erlang Distribution (E_r)

- **Single state server:**

$$f_T(t) = \mu e^{-\mu t} \qquad t \ge 0$$

$$F_T(s) = \frac{\mu}{s+\mu}$$

$$E[T] = 1/\mu \qquad Var[T] = 1/\mu^2 \qquad C_b = 1$$



- **r-Stage server:**

$$f_T(t) = \frac{r\mu(r\mu t)^{r-1} e^{-r\mu t}}{(r-1)!} \qquad t \ge 0$$

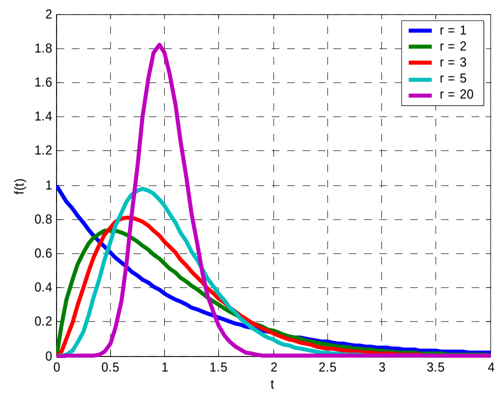
$$F_T(s) = \left(\frac{r\mu}{s+r\mu}\right)^r$$

$$E[T] = 1/\mu \qquad Var[T] = 1/(r\mu^2) \qquad C_b = 1/\sqrt{r}$$

r-stages

Method Of Stages – Erlang Distribution – cont'd

```
%
% Erlang distribution
LineWidth = 3;
Rs   = [1 2 3 5 20];
Mue = 1;
t    = [0:Step:4];
f    = zeros(length(Rs), length(
 
for i=1:length(Rs)
    r = Rs(i);
    f(i,:) = r*Mue*(r*Mue*t).^(
   1).*exp(-r*Mue*t)./ ...
             factorial(r-1);
end
 
figure(1);
h = plot(t, f); grid
set(h,'LineWidth', LineWidth);
xlabel('t');
ylabel('f(t)');
LegendStr = ['legend('];
for i=1:length(Rs)-1;
    LegendStr = [LegendStr '''r
   num2str(Rs(i)) ''','];
end
LegendStr = [LegendStr '''r = '
   num2str(Rs(length(Rs))) ''');'];
eval(LegendStr);
```



Erlang Distribution - Observations

- **Let r → ∞, the distribution of T approaches a constant (deterministic) value of 1/ μ**

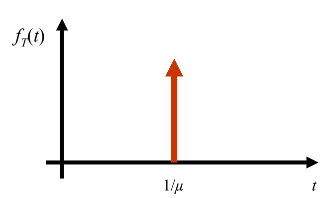
$$\lim_{r\to\infty} F_T(s) = \lim_{r\to\infty}\left(\frac{1}{1+s/r\mu}\right)^r = e^{-s/\mu}$$

or

$$f_T(t) = \delta(t - 1/\mu)$$

where

$$E[T] = 1/\mu \qquad Var[T] = 0$$



The Queue M/E$_r$/1

- Service time ~ r-stages Erlangian distribution
- System state:
  - Number of customers in system
  - Number of stages remaining in the service
- Define j = number of stages left in total system (i.e. for all customers)
- If system contains k customers
  - (k – 1) waiting
  - One is in service – let him be in the ith stage
- Therefore, j is given by

  j = (k - 1)r + (r – i + 1), or
  = rk – i + 1;

The Queue M/E$_r$/1 – cont'd

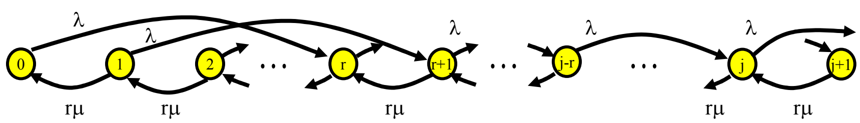
- Define P_j = Prob of j stages in system
- Define p_k = Prob of k customers in system
- P_j and p_k are related as follows:

$$p_k = \sum_{j=(k-1)r+1}^{kr} P_j \qquad k = 1,2,\ldots$$

Note: $p_0 = P_0$!!

The Queue M/E$_r$/1 – cont'd

- **The state-transition-rate diagram for number of stages is as shown**
- **Every arrival brings along r new stages to be completed!**
- **Note that state 0, 1, …, r-1 – are special boundary states!! – WHY?**



The Queue M/E$_r$/1 – Forward Equations

- **Forward equations in equilibrium,**

$$\lambda P_0 = r\mu P_1, \text{ and}$$

$$(\lambda + r\mu)P_j = \lambda P_{j-r} + r\mu P_{j+1}, \quad j=1,2,\ldots$$

Note: $P_j = 0$ for $j < 0$

- **Define P(z) to be** $P(z) = \sum_{j=0}^{\infty} P_j z^j$

- **Therefore,**

$$\sum_{j=1}^{\infty} (\lambda + r\mu) P_j z^j = \sum_{j=1}^{\infty} \lambda P_{j-r} z^j + \sum_{j=1}^{\infty} r\mu P_{j+1} z^j$$

$$(\lambda + r\mu)\left[\sum_{j=0}^{\infty} P_j z^j - P_0\right] = \lambda z^r \sum_{j=1}^{\infty} \lambda P_{j-r} z^{j-r} + \frac{r\mu}{z} \sum_{j=1}^{\infty} P_{j+1} z^{j+1}$$

$$(\lambda + r\mu)[P(z) - P_0] = \lambda z^r P(z) + \frac{r\mu}{z}[P(z) - P_0 - P_1 z]$$

The Queue M/E$_r$/1 – Forward Equations

- After simplifying, P(z) can be written as

$$P(z) = \frac{r\mu P_0 (1-z)}{r\mu + \lambda z^{r+1} - (\lambda + r\mu) z}$$

- P_0 can be found using the condition P(z = 1) = 1 ➔ $P_0 = 1 - \lambda/\mu$

- If we define $\rho = \lambda/\mu$, P(z) can be rewritten as

$$P(z) = \frac{r\mu (1-\rho)(1-z)}{r\mu + \lambda z^{r+1} - (\lambda + r\mu) z}$$

Example: M/E$_r$/1

- **Problem**: show that M/M/1 queue is a special case of M/E$_r$/1 where r = 1

- **Solution**: Using r = 1, P(z) reduces to

$$P(z) = \frac{\mu(1-\rho)(1-z)}{\mu + \lambda z^2 - (\lambda + \mu) z} = \frac{(1-\rho)(1-z)}{1 + \rho z^2 - (1+\rho) z} = \frac{(1-\rho)}{1-\rho z}$$

Which is the generating function for number of customers in an M/M/1 queue

The probability of k customers in system, p_k is given by

$$p_k = (1-\rho)\rho^k$$

M/E$_r$/1 Queue Solution

- **Problem**: How to invert P(z) in general for r > 1.
- **Solution**: P(z) in general is given by $P(z)=\frac{r\mu(1-\rho)(1-z)}{r\mu+\lambda z^{r+1}-(\lambda+r\mu)z}$

The denominator is a polynomial of degree r+1 ➔ It r+1 roots
It is clear that z = 1 is one of the roots
We must identify the remaining r roots

Let the denominator be $D(z)=(1-z)\left[r\mu-\lambda\left(z+z^2+\cdots+z^r\right)\right]$
Let the r zeros be denoted by $\{z_1, z_2, \ldots, z_r\}$

$$D(z)=r\mu(1-z)(1-z/z_1)(1-z/z_2)\cdots(1-z/z_r)$$

Then P(z) can be written as

$$P(z)=\frac{1-\rho}{(1-z/z_1)(1-z/z_2)\cdots(1-z/z_r)}$$

Finding the ZEROS of D(z) is the most challenging task!!

M/E$_r$/1 Queue Solution – cont'd

- **Solution**:

We can perform partial fraction expansion on P(z) to obtain:

$$P(z)=(1-\rho)\sum_{i=1}^{r}\frac{A_i}{(1-z/z_i)}$$

where

$$A_i=\prod_{\substack{n=1\\ n\neq i}}^{r}\frac{1}{(1-z_i/z_n)}$$

Therefore, P(z) can be inverted as

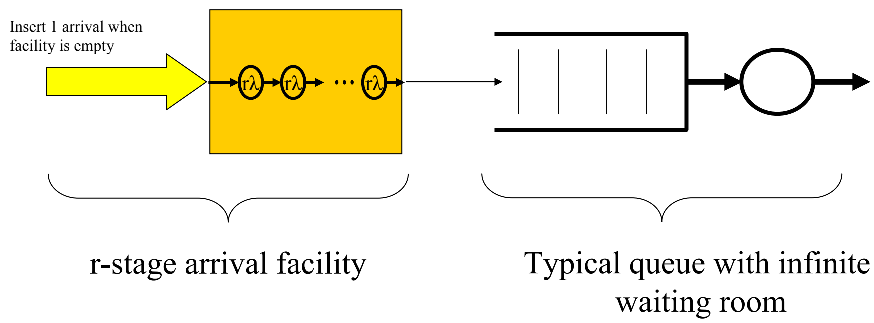
$$P_j=(1-\rho)\sum_{i=1}^{r}A_i(z_i)^{-j} \qquad j=1,2,\cdots,r$$

Note

- The distribution of the number of stages in the system is a weighted sum of geometric distributions.
- The above is NOT the distribution of customers in the system yet!!

The Queue $E_r/M/1$

- **Imagine the following configuration**



The Queue $E_r/M/1$ – cont'd

- **Interarrival time ~ r-stages Erlangian distribution**
- **Service time ~ exponential with rate μ**
- **System state:**
  - **Number of customers already in system**
  - **Number of arrival stages of customer to arrive**
- **Define j = number of arrival stages in system**
- **If system contains k customers**
  - **Arriving customer is in the i^{th} stage ($1 \le i \le r$) – i.e. he finished i-1 stages**
  - **k customer fully arrived – each brought r-stages of arrival**
- **Therefore, j is given by**

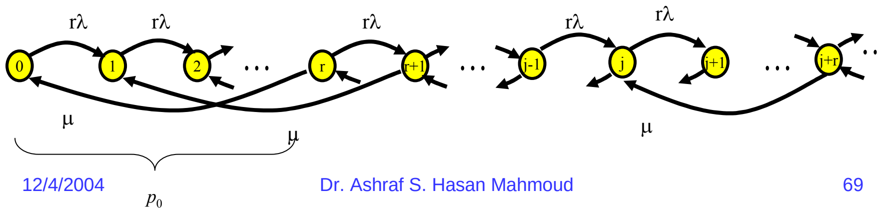
$$j = r k + i - 1$$

The Queue $E_r/M/1$ – cont'd

- **The state-transition-rate diagram for number of stages is as shown**
- **Define P_j = Prob of j arrival stages in system**
- **Define p_k = Prob of k customers in system**
- **P_j and p_k are related as follows:**

$$p_k = \sum_{j=kr}^{(k+1)r-1} P_j \qquad k = 0,1,2,\ldots$$

- **Every departure removes r stages of arrival from system!**



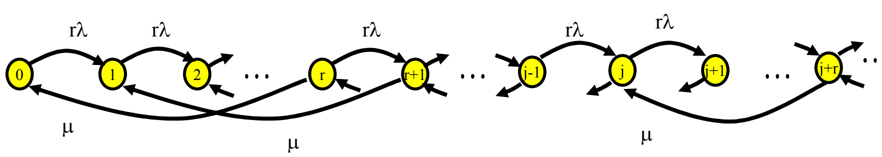
The Queue $E_r/M/1$ – Forward Equations

- **Forward equations in equilibrium,**

$$r\lambda P_0 = \mu P_r, \text{ and}$$

$$r\lambda P_j = r\lambda P_{j-1} + \mu P_{j+r}, \quad j=1,2,\ldots,r-1$$

$$(r\lambda+\mu)P_j = r\lambda P_{j-1} + \mu P_{j+r}, \quad j=r,r+1,\ldots$$



The Queue E_r/M/1 – Solution

- **Define P(z) to be** $$P(z)=\sum_{j=0}^{\infty}P_j z^j$$
- **Therefore,** $$\sum_{j=1}^{\infty}(\mu+r\lambda)P_j z^j-\sum_{j=1}^{r-1}\mu P_j z^j=\sum_{j=1}^{\infty}r\lambda P_{j-1}z^j+\sum_{j=1}^{\infty}\mu P_{j+r}z^j$$
$$(\mu+r\lambda)[P(z)-P_0]-\sum_{j=1}^{r-1}\mu P_j z^j=r\lambda zP(z)+\frac{\mu}{z^r}\left[P(z)-\sum_{j=0}^{r}P_j z^j\right]$$
- **Finally,** $$P(z)=\frac{(1-z^r)\sum_{j=0}^{r-1}P_j z^j}{r\rho z^{r+1}-(1+r\rho)z^r+1}$$

where $\rho=\lambda/\mu$

The Queue E_r/M/1 – Solution – cont'd

- **Consider the denominator of P(z), D(z)**

$$D(z)=r\rho z^{r+1}-(1+r\rho)z^r+1$$

- **D(z) has r+1 roots**
  - **z = 1 is one root**
  - **It can be shown that r-1 roots are within the unit circle – i.e. |z| < 1 (Rouche's Theorem)**
  - **Remaining zero, z_0, lies outside the unit circle, $|z_0| > 1$**

The Queue $E_r/M/1$ – Solution – cont'd

- **Consider the numerator of P(z), N(z)**

$$N(z) = (1 - z^r)\sum_{j=0}^{r-1} P_j z^j$$

- **N(z) has 2r-1 roots**
  - **r roots at z = 1**

- **Since P(z) is analytic on |z| < 1 ➔ P(z) is bounded for all |z| < 1 (i.e. no singularities inside the unit circle)**
  - **The remaining r roots of N(z) (contributed by the summation term) are inside the unit circle and cancel those r roots of D(z)**

The Queue $E_r/M/1$ – Solution – cont'd

- **Therefore, one can write**

$$\frac{D(z)}{(1-z)(1-z/z_0)} = K\sum_{j=0}^{r-1} P_j z^j$$

$$\frac{r\rho z^{r+1} - (1+r\rho)z^r + 1}{(1-z)(1-z/z_0)} = K\sum_{j=0}^{r-1} P_j z^j$$

- **This means, P(z) can be written as**

$$P(z) = \frac{(1-z^r)}{K(1-z)(1-z/z_0)}$$

$$= \frac{(1-z^r)(1-1/z_0)}{r(1-z)(1-z/z_0)}$$

since P(1) = 1 ➔ K = r/(1-1/z_0)

The Queue E_r/M/1 – Solution – cont'd

- We are now in a position to solve for the final pmf – performing the partial fraction expansion on P(z), yields

$$P(z) = \left(1 - z^r\right)\left[\frac{1/r}{1-z} + \frac{-1/(rz_0)}{1 - z/z_0}\right]$$

- If we let $\left[\frac{1/r}{1-z} + \frac{-1/(rz_0)}{1 - z/z_0}\right] \overset{Z^{-1}}{\Longleftrightarrow} f_j$
- Then $P_j = f_j - f_{j-r}$

Recall $f_j = 0$ for $j < 0$

- Clearly, $f_j = \begin{cases} \frac{1}{r}\left(1 - z_0^{-j-1}\right) & j \geq 0 \\ 0 & j < 0 \end{cases}$

Z-transform Pairs:
$\delta(n) = 1$ $n = 0,1, \ldots$ $\leftrightarrow 1/(1-z)$
0 otherwise
$A\alpha^n \leftrightarrow A/(1-\alpha z)$
If $f_n \leftrightarrow P(z)$, then $f_{n-r} \leftrightarrow z^r P(z)$

The Queue E_r/M/1 – Solution – cont'd

- Therefore, P_j for $j \geq r$ is given by

$$P_j = \frac{1}{r} z_0^{r-j-1}\left(1 - z_0^{-r}\right) \quad j \geq r$$
$$= \rho(z_0 - 1) z_0^{r-j-1} \quad j \geq r$$

$\because D(z_0) = r\rho z_0^{r+1} - (1 + r\rho) z_0^r + 1 = 0$
$\Rightarrow r\rho(z_0 - 1) = 1 - z_0^{-r}$

- For $0 \leq j < r$, we observe $f_{j-r} = 0$ ➔ $P_j = f_j$ only.
- Hence, the over all pmf is given by

$$P_j = \begin{cases} \frac{1}{r}\left(1 - z_0^{-j-1}\right) & 0 \leq j < r \\ \rho(z_0 - 1) z_0^{r-j-1} & j \geq r \end{cases}$$

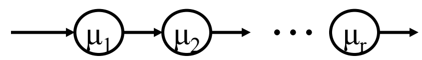
- It can be shown that the pmf for the number of customers in the system is given by

$$p_k = \begin{cases} 1 - \rho & k = 0 \\ \rho\left(z_0^r - 1\right) z_0^{-rk} & k > 0 \end{cases}$$

Generalization of the Erlangian Distribution – First attempt

- **The previous Erlangian distribution is limited in the sense that $C_b = 1/\sqrt{r} \le 1$**

- **Consider a series of r-stages; each with parameter μ_i**



$$f_T(t) = f_{T_1}(t) \otimes f_{T_2}(t) \otimes \cdots \otimes f_{T_r}(t)$$

$$T = T_1 + T_2 + \ldots + T_r$$

$$F_T(s) = \left(\frac{\mu_1}{s+\mu_1}\right)\left(\frac{\mu_2}{s+\mu_2}\right)\cdots\left(\frac{\mu_r}{s+\mu_r}\right)$$

$$E[T] = \sum_{i=1}^{r} 1/\mu_i \quad Var[T] = \sum_{i=1}^{r} 1/\mu_i^2 \quad C_b^{\,2} = \left(\sum_{i=1}^{r} 1/\mu_i^2\right)\Big/\left(\sum_{i=1}^{r} 1/\mu_i\right)^2$$

But C_b is again always less than 1 for any choice of $1/\mu_i$

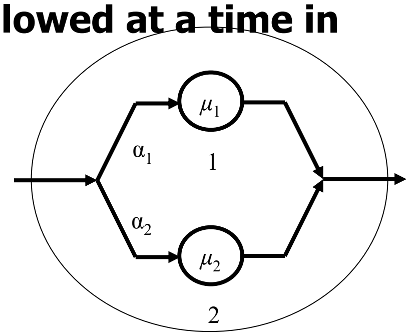
Generalization of the Erlangian Distribution – Second attempt

- **Consider the 2-stage parallel server**

- **Only one customer is allowed at a time in the service facility**

$$f_T(t) = \alpha_1\mu_1 e^{-\mu_1 t} + \alpha_2\mu_2 e^{-\mu_2 t}$$

$$F_T(s) = \alpha_1 \frac{\mu_1}{s+\mu_1} + \alpha_2 \frac{\mu_2}{s+\mu_2}$$

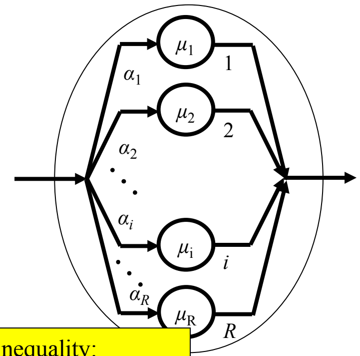


Service Facility

$$\alpha_1 + \alpha_2 = 1$$

Generalization of the Erlangian Distribution – Second attempt – cont'd

- Consider the R-stage parallel server
- $f_T(t)$ ~ hyperexponential distribution (denoted by H_R)



$$\sum_{i=1}^{R} \alpha_i = 1$$

$$f_T(t) = \sum_{i=1}^{R} \alpha_i \mu_i e^{-\mu_i t} \qquad t \geq 0$$

$$F_T(s) = \sum_{i=1}^{R} \alpha_i \frac{\mu_i}{s + \mu_i}$$

$$E[T] = \sum_{i=1}^{R} \frac{\alpha_i}{\mu_i} \qquad E[T^2] = 2\sum_{i=1}^{R} \frac{\alpha_i}{\mu_i^2}$$

$$C_b^2 = \frac{2\sum_{i=1}^{R} \frac{\alpha_i}{\mu_i^2}}{\left(\sum_{i=1}^{R} \frac{\alpha_i}{\mu_i}\right)^2} - 1 \geq 1$$

Cauchy-Schwartz inequality:

$$\left(\sum_i a_i b_i\right)^2 \leq \left(\sum_i a_i^2\right)\left(\sum_i b_i^2\right) \qquad a_i = \sqrt{\alpha_i} \qquad b_i = \frac{\sqrt{\alpha_i}}{\mu_i}$$

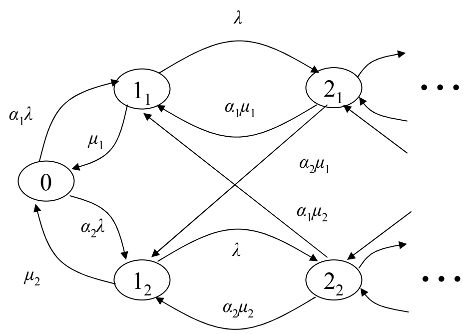
$$\left(\sum_{i=1}^{R} \frac{\alpha_i}{\mu_i}\right)^2 \leq \left(\sum_{i=1}^{R} \alpha_i\right)\left(\sum_{i=1}^{R} \frac{\alpha_i}{\mu_i^2}\right) \leq \left(\sum_{i=1}^{R} \frac{\alpha_i}{\mu_i^2}\right)$$

M/H$_R$/1, H$_R$/M/1, H$_{Ra}$/H$_{Rb}$/1 Queues

- Analysis by method of stages exists
  - Take into account the hperexponential service (or arrival) facility by merely specifying which stage within service (or arrival) facility the customer currently occupies. PLUS
  - Number of customers in system
- The above forms a Markov chain which may be analyzed as we did before

Example: M/H_2/1

- **State-transition-rate diagram is as shown**
- **k_i – implies system contains k customers and the customer in service is in service stage i**
  - **REMEMBER: only ONE customer can be in service facility**



Example: M/E_2/2/2

- **<u>Problem</u>: Consider an M/E_2/2/2 – a system with two servers, each with 2 identical stages. There is no storage room, and packets arriving to system while serving two packets are lost.**

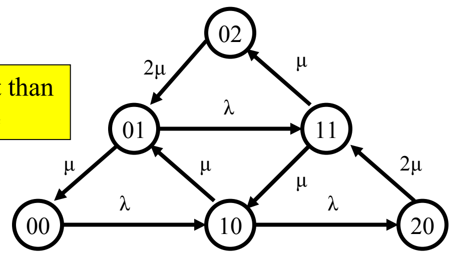
  Assume packets arrive with rate λ, while the service rate in a stage is given by μ.

  Compute the blocking probability for this system?

Example: M/E_2/2/2 – cont'd

- **Solution:**
  Let the state for such system be (i, j) – where i and j are, respectively, the number of packets in the first and seconds stages.
  ➔ possible states: (0,0), (0,1), (0,2), (1,0), (2,0), (1,1)
  The state transition flow diagram is as shown:

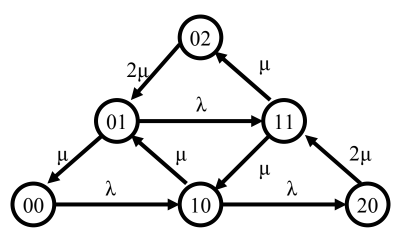
Note the state variable is different than that we used for the M/E_r/1 queue



Example: M/E_2/2/2 – cont'd

- **Solution: cont'd**
  We then proceed with writing the equilibrium equations:

$$\lambda P_{00} = \mu P_{01},$$
$$(\lambda + \mu) P_{01} = 2 \mu P_{02} + \mu P_{01},$$
$$2 \mu P_{02} = \mu P_{11},$$
$$(\lambda + \mu) P_{10} = \lambda P_{00} + \mu P_{11},$$
$$2 \mu P_{20} = \lambda P_{10},$$
$$2 \mu P_{11} = \lambda P_{01} + 2 \mu P_{20}$$



Note that Blocking probability, P_B is given by
$P_B = P_{11} + P_{02} + P_{20}$

Solving, the above equations: you can show that

$$P_B = \frac{2(\lambda/\mu)^2}{1 + 2(\lambda/\mu) + 2(\lambda/\mu)^2}$$

__Note:__
The blocking probability DOES not depend on the service time distribution, but rather on the mean service time – This is referred to as the insensitivity property of Erlang-B formula!!

More Generalization – Series-Parallel Service

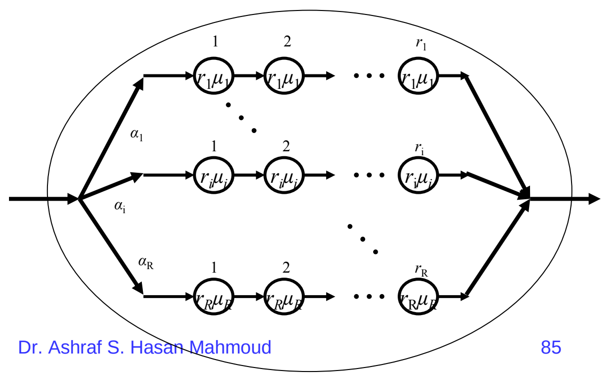
- **Note: $r_1, \ldots, r_i, \ldots, r_R$ are not necessarily equal**

$$f_T(t) = \sum_{i=1}^{R} \alpha_i \frac{r_i\mu_i (r_i\mu_i t)^{r_i-1} e^{-r_i\mu_i t}}{(r_i-1)!} \qquad t \geq 0$$

$$F_T(s) = \sum_{i=1}^{R} \alpha_i \left(\frac{r_i\mu_i}{s + r_i\mu_i} \right)^{r_i}$$

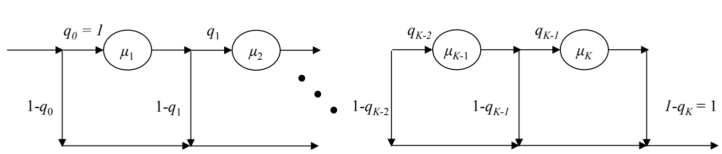
If the rates in stages within one branch are not equal, then

$$F_T(s) = \sum_{i=1}^{R} \alpha_i \prod_{i=1}^{r_i} \frac{\mu_{ij}}{s + \mu_{ij}}$$



More Generalization – Cox Network

- **Consider the network of stages shown – Cox Network**
- **Prob of going through exactly i stages:** $\prod_{j=0}^{i-1} q_j (1-q_i)$
- **Assume $q_0 = 1$, $q_K = 0$, then** $\sum_{i=1}^{K} \prod_{j=0}^{i-1} q_j (1-q_i) = 1$



Characterization of Cox Network – con't

- **The Laplace transform of the service time if i stages are used:**

$$M_{T/i}(s)=\prod_{j=1}^{i}\frac{\mu_i}{s+\mu_i}$$

- **The Laplace transform for the service time in K-stages network:**

$$M(s)=q_0(1-q_1)\frac{\mu_1}{s+\mu_1}+q_0q_1(1-q_2)\frac{\mu_1}{s+\mu_1}\frac{\mu_2}{s+\mu_2}$$
$$+q_0q_1q_2(1-q_3)\frac{\mu_1}{s+\mu_1}\frac{\mu_2}{s+\mu_2}\frac{\mu_3}{s+\mu_3}+\cdots+q_0q_1q_2\cdots q_{K-1}(1-q_K)\frac{\mu_1}{s+\mu_1}\frac{\mu_2}{s+\mu_2}\cdots\frac{\mu_K}{s+\mu_K}$$

$$M(s)=\sum_{i=1}^{K}\prod_{j=0}^{i-1}q_j(1-q_i)\prod_{k=1}^{i}\frac{\mu_k}{s+\mu_k}$$

Characterization of Cox Network – con't

- **M(s) given by**

$$M(s)=\sum_{i=1}^{K}\prod_{j=0}^{i-1}q_j(1-q_i)\prod_{k=1}^{i}\frac{\mu_k}{s+\mu_k}$$

is known as the Coxian distribution

- **You can show (refer to textbook), the mean is given by**

$$E[T]=\sum_{i=1}^{K}\prod_{j=0}^{i-1}q_j(1-q_i)\sum_{k=1}^{i}\frac{1}{\mu_k}$$
$$=\sum_{i=1}^{K}\frac{\prod_{j=0}^{i-1}q_j}{\mu_i}$$

Characterization of Cox Network – con't

- Note that for $q_i = 1$, and $\mu_i = \mu$ for all i, then the expression for M(s) reduces to

$$M(s) = \left(\frac{\mu}{s+\mu}\right)^K$$

which the K-stage Erlang-distribution previously discussed on slide 56

- The expected delay in this case is given by

$$E[T] = \frac{K}{\mu}$$

Example: M/G/N/N

- Consider a queueing system where
  - Arrivals are Poisson with rate λ
  - N servers and no waiting room
  - Each server is a Coxian server with K stages

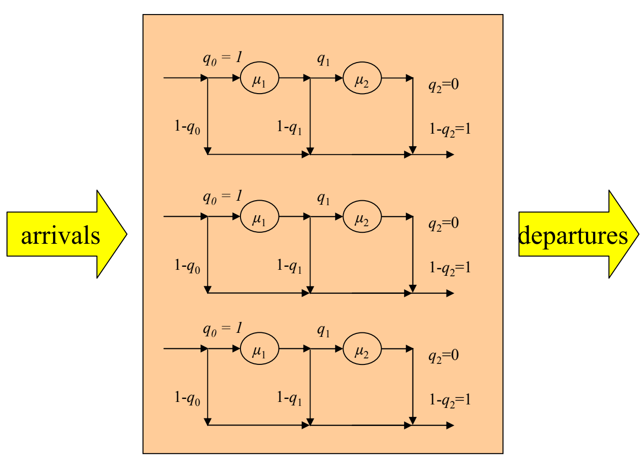
- Objective: compute blocking probability? And show that it depends only on the mean service rate and the mean arrival rate (i.e. no dependence on the probability distribution of the service time – the *insensitivity* property of the Erlang-B formula)

Example: M/G/N/N – cont'd

- **System state: K-dimensional vector**
  - **i.e. state = (k_1, k_2, …, k_K) – where k_i; i=1,2, …, K is the number of customers in stage I**
  - **Obviously, sum of k_is should be less or equal to N. Note it is equal to N if all servers are busy – remember too that only one customer can be in any server!!**

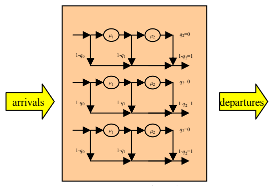
Example: M/G/N/N – cont'd

- **Consider a case where N = 3 and K = 2.**

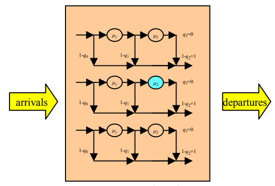


Example: M/G/N/N – cont'd

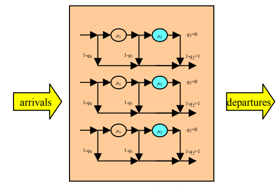
- **System States: examples**



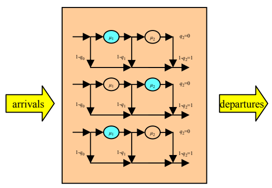
State = (0,0)



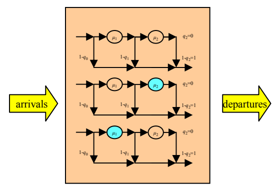
State = (0,1)



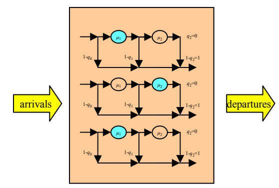
State = (0,3) – system full



State = (2,1) – system full



State = (1,1)



State = (2,1) – system full

Example: M/G/N/N – cont'd

- **Exercise: For the K= 2, N = 3 case explained before**
  - **A) draw the state transition diagram**
  - **B) show that the state equilibrium equations (3.76 and 3.77) are satisfied**
  - **C) Derive the detailed balance equation 3.78**
- **The exercise is worth 10% points bonus in the final exam**
- **Deliver a soft copy in power point of the detailed solution**
- **Deadline: January 3rd, 2005**

Example: M/G/N/N – Blocking Probability

- **Blocking probability is equal to the probability of system being in states where the sum of k_is is equal to N. i.e.**

$$P_B = \Pr ob\left(\sum_{i=1}^{K} k_i = N\right)$$

- **The textbook shows that the blocking probability is given by**

$$P_B = \frac{P(0)}{N!}\left[\lambda \sum_{i=1}^{K} \frac{\prod_{j=0}^{i-1} q_j}{\mu_i}\right]^N$$

where P(0) is a constant term found through the normalization equation

- **Refer for textbook for derivation details.**