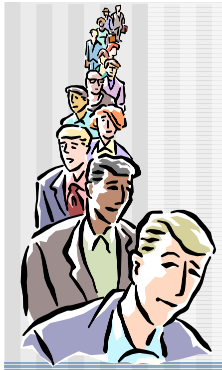
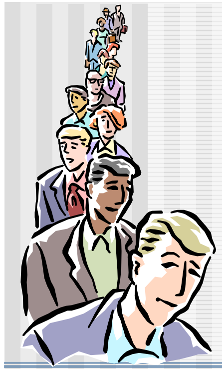


M/**G**/1/GD/∞/∞ System

- Pollaczek-Khinchin (PK) Equation

$$L_q = \frac{\lambda^2 \sigma_s^2 + \rho^2}{2(1-\rho)}$$

- Steady-state probabilities
  - $\pi_0 = 1 - \rho$
- Finding L, W_q, W



M/M/R/GD/**K**/**K** System

K machines → R repair people

- Draw the transition diagram
- Derive the steady-state probabilities:

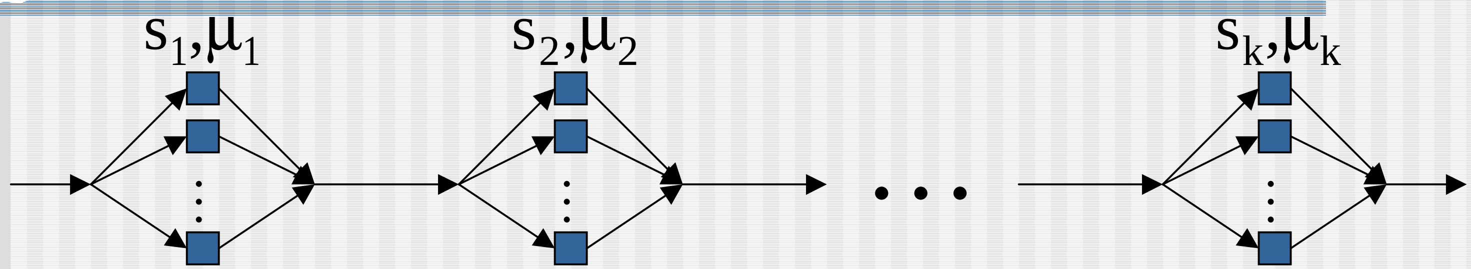
$$\pi_j = \begin{cases} \binom{K}{j} \rho^j \pi_0 & , \; j = 0,1,\ldots,R \\ \binom{K}{j} \dfrac{j!}{R!\,R^{j-R}} \rho^j \pi_0, & j = R+1, R+2, \ldots, K \end{cases}$$

- Find L, L_q, W, W_q

Example 1

- The Gotham Township Police Department has 5 patrol cars. A patrol car breaks down and requires service once every 30 days. The police department has two repair workers,each of whom takes an average of 3 days to repair a car. Breakdown times and repair times are exponential.
- E(# police cars in good condition)
- E(down time for car needing repairs)
- Fraction of time a particular repair worker is idle

Queues in Series



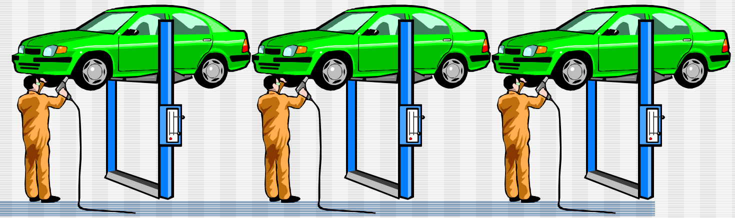
Jackson's Theorem: If

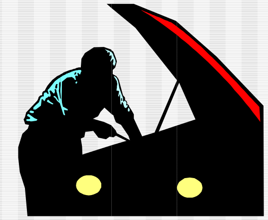
- Interarrival times are exponential, rate λ
- Service time at each server is exponential
- Each stage has infinite capacity waiting room

$\Rightarrow$ Interarrival times for each stage exponential, rate λ

$\Rightarrow$ If $\lambda < s_j \mu_j$, each stage is $M/M/s_j/GD/\infty/\infty$ system.

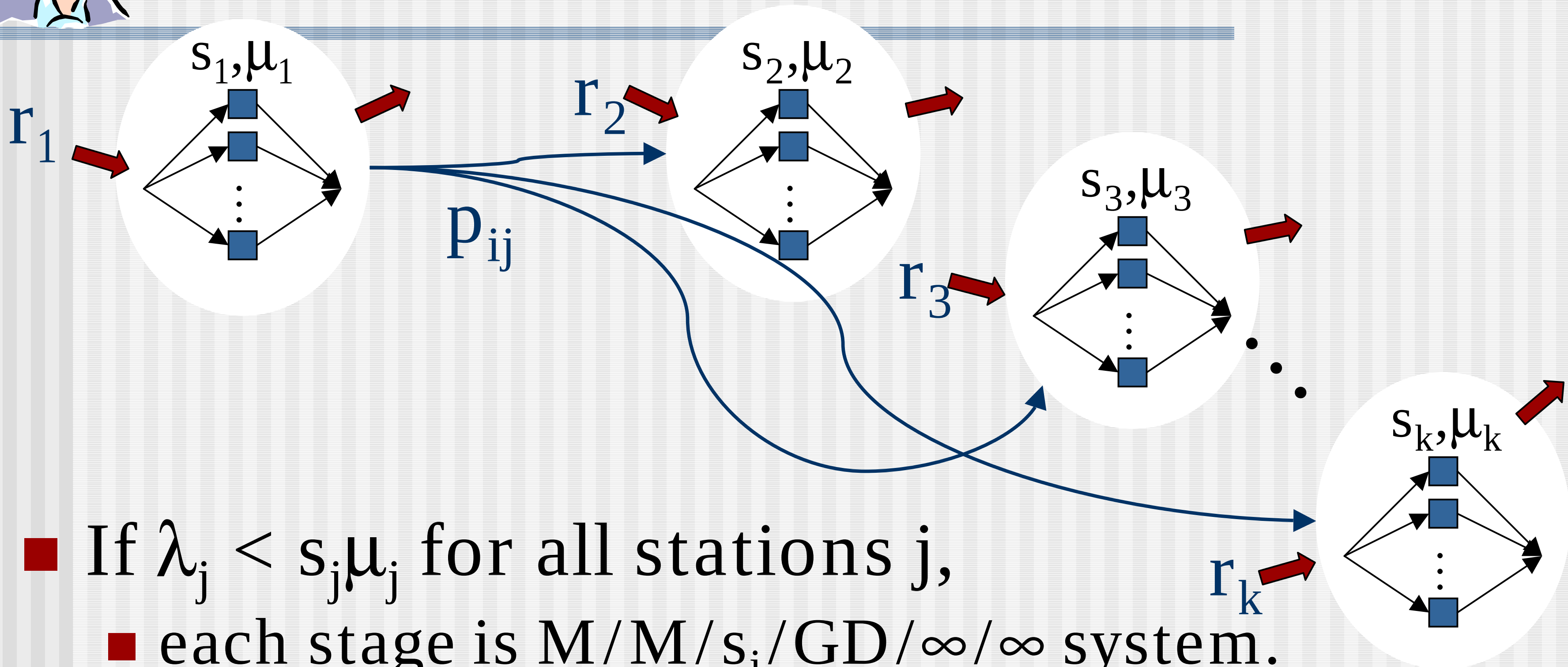
Example 2



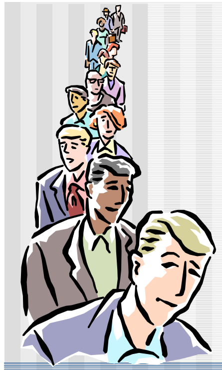


- The last two things that are done to a car before its manufacture is complete are installing the engine and putting on the tires. An average of 54 cars per hour arrive requiring these two tasks. One worker is available to install the engine and can service 60 cars per hour. After the engine is installed, the car goes to the tire station and waits for its tires to be attached. Three workers serve at the tire station. Each works on one car at a time and can put tires on a car in an average of 3 minutes. Both interarrival times and service times are exponential. Determine:
- E(queue length at each work station)
- Total E(time car spends waiting for service)

Open Queuing Networks

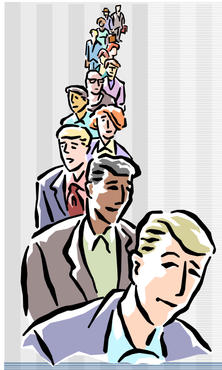


- If $\lambda_j < s_j\mu_j$ for all stations j,
  - each stage is $M/M/s_j/GD/\infty/\infty$ system.
- $L = L_1 + L_2 + \ldots + L_k$
- $\lambda = r_1 + r_2 + \ldots + r_k$, $W = L/\lambda$



Example 3

- TWO servers: An average of 8 customers arrive from outside at server 1, and an average of 17 customers per hour arrive from outside at server 2. Interarrival times are exponential. Server 1 can serve at an exponential rate of 20 customers per hour, and server 2 can serve at an exponential rate of 30 customers per hour. After completing service at server 1, half of the customers leave the system, and half go to server 2. After completing service at server 2, ¾ of the customers complete serve, and ¼ return to server 1.
- What fraction of the time is server 1 idle?
- Find E(# customers at each server)
- Find E(time a customer spends in the system)
- What if server 2 could serve only 20 customers per hour?



M/G/**s**/GD/**s**/∞ System

- L, L_q, W, W_q are of limited interest
- Queue never forms so $L_q = W_q = 0$
- $W = W_s = 1/\mu$
- Fraction of arrivals turned away = $\lambda\pi_s$
- $L = L_s = \lambda(1-\pi_s)/\mu$
- **Erlang's Loss Formula**:

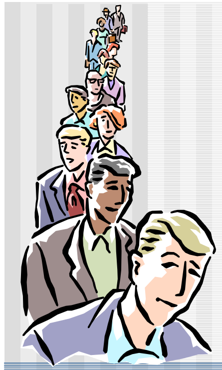
$$\pi_s \text{ depends only on } \lambda/\mu$$

Example 4

- An average of 20 ambulance calls per hour are received by Newport Hospital. An ambulance requires an average of 20 minutes to pick up a patient and take the patient to the hospital. The ambulance is then available to pick up another patient. How many ambulances should the hospital have to ensure that there is at most a 1% probability of not being able to respond immediately to an ambulance call? Assume that interarrival times are exponentially distributed.

Queuing Disciplines

- FCFS = First Come First Serve
- SIRO = Service In Random Order
- LCFS = Last Come First Serve
- $E(W_{FCFS}) = E(W_{SIRO}) = E(W_{LCFS})$
- $Var(W_{FCFS}) < Var(W_{SIRO}) < Var(W_{LCFS})$



Priority Queues

- W_{qk} = E(type k customer's waiting time)
- W_k = E(type k cust time in system)
- L_{qk} = E(# type k customers in line)
- L_k = E(# type k customers in system)

$M_i/G_i/1/NPRP/\infty/\infty$ System

$$\rho_i = \frac{\lambda_i}{\mu_i}, a_0 = 0, a_k = \sum_{i=1}^{k} \rho_i \qquad \text{If} \quad \sum_{i=1}^{n} \rho_i < 1$$

Then

$$W_{qk} = \frac{\sum_{k=1}^{n} \lambda_k E(S_k^2)}{2(1-a_{k-1})(1-a_k)}, \qquad L_{qk} = \lambda_k W_{qk}$$

$$W_k = W_{qk} + \frac{1}{\mu_k}, \qquad L_k = \lambda_k W_k$$

Priority Queue Example 1

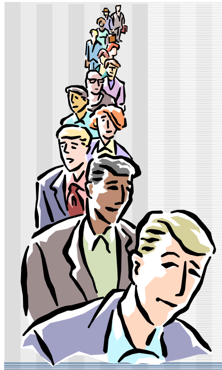
- A copying facility gives shorter jobs priority over long jobs. Interarrival times for each type of job are exponential, and an average of 12 short jobs and 6 long jobs arrive each hour. Let type 1 job = short job and type 2 job = long job. Then we are given that E(S1)=2 min, Var(S1)=2, E(S2)=4 min, Var(S2)=2. Determine the average length of time each type of job spends in the copying facility.

Priority Q E.g. 2

- Gotham Township has 5 police cars. The police dept receives two types of calls: emergency (type 1) and nonemergency (type 2) calls. Interarrival times are exponentially distributed, with an average of 10 emergency and 20 nonemergency calls each hour. Each call type has exponential service s, with a mean of 8 minutes (assume that, on the average, 6 of the 8 minutes is travel time from the police station to the call and back to the station). Emergency calls are given priority over nonemergency calls. On average, how much time will elapse between the placement of a nonemergency call and the arrival of a police car?

Customer-Dep Waiting Costs

- Assign priority such that
  - $c_1\mu_1 \geq c_2\mu_2 \geq \ldots \geq c_n\mu_n$
- Shortest Processing Time (SPT) Discipline:
  - Let $c_i = 1$
  - $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$
  - $1/\mu_1 \leq 1/\mu_2 \leq \ldots \leq 1/\mu_n$



$M_i/G_i/s/NPRP/\infty/\infty$ System

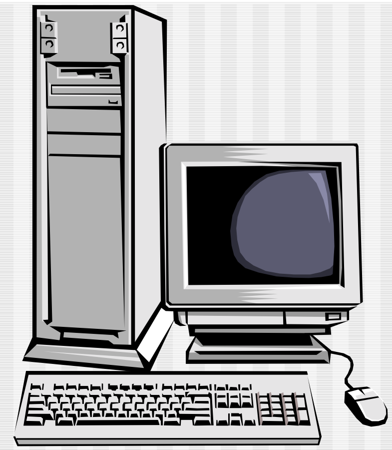
$$\rho_i = \frac{\lambda_i}{s\mu}, a_0 = 0, a_k = \sum_{i=1}^{k} \rho_i \qquad \text{If} \qquad \rho = \sum_{i=1}^{n} \rho_i < 1$$

Then

$$W_{qk} = \frac{P(j \geq s)}{s\mu(1-a_{k-1})(1-a_k)}, \qquad L_{qk} = \lambda_k W_{qk}$$

$$W_k = W_{qk} + \frac{1}{\mu_k}, \qquad L_k = \lambda_k W_k$$

Priority Queue E.g. 3



- Consider a computer system to which two types of computer jobs are submitted. The mean time to run each type of job is $1/\mu$. The interarrival times for each type of job are exponential, with an average of λi type i jobs arriving each hour. Consider the following three situations:
  - Type 1 jobs have priority over type 2 jobs, preempt
  - Type 1 jobs have priority over type 2 jobs, no preempt
  - All jobs are serviced on a FCFS basis.
- Under which system are type 1 jobs best off? Worst off? What about type 2 jobs?

$M_i/G_i/1/PRP/\infty/\infty$ System

Preemptive resume vs. preemptive repeat

$$W_k = \frac{1/\mu}{(1-a_{k-1})(1-a_k)}, \qquad a_0 = 0, a_k = \sum_{i=1}^{k} \frac{\lambda_i}{\mu}$$