Queueing Networks

Raj Jain
Washington University in Saint Louis
Jain@eecs.berkeley.edu or Jain@wustl.edu

A Mini-Course offered at UC Berkeley, Sept-Oct 2012

These slides and audio/video recordings are available on-line at:

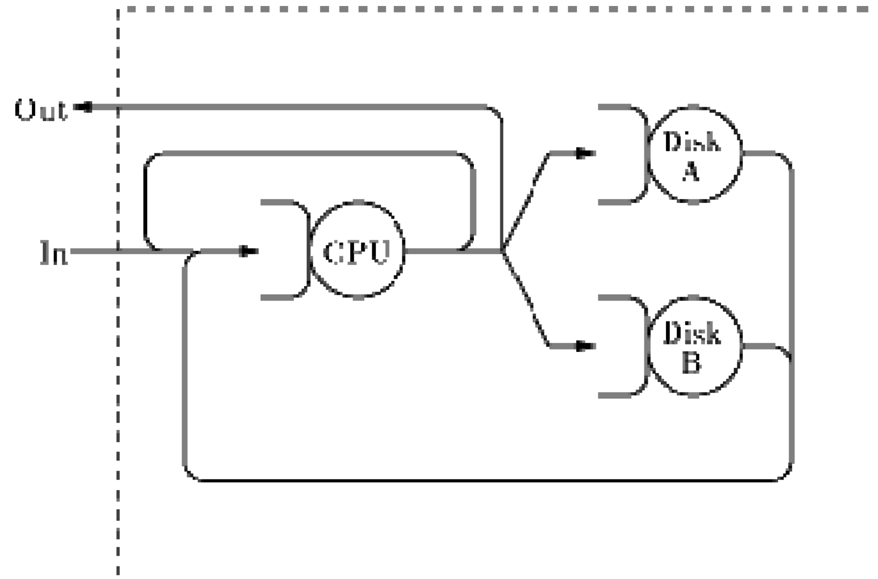
http://amplab.cs.berkeley.edu/courses/queue

and http://www.cse.wustl.edu/~jain/queue

1. Open and Closed Queueing Networks
2. Product Form Networks
3. Queueing Network Models of Computer Systems

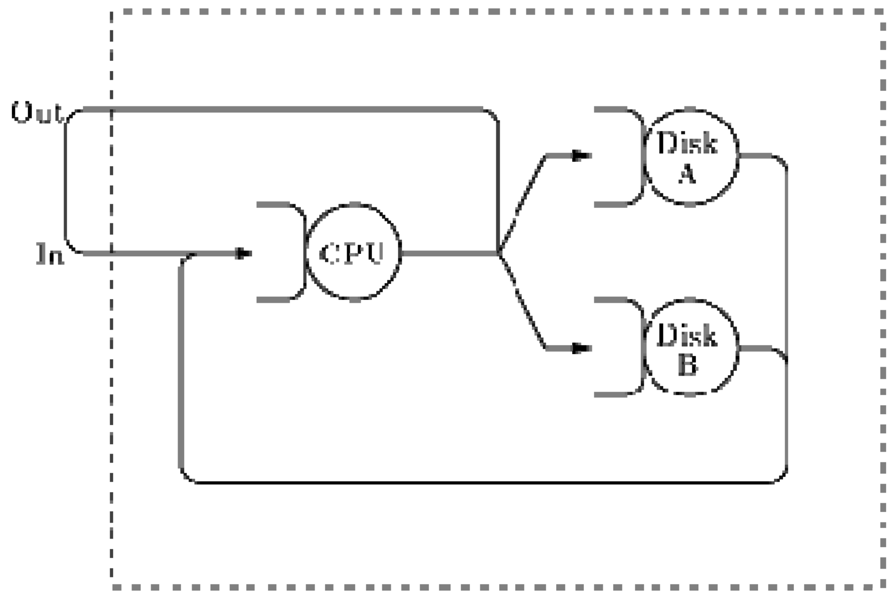
Open Queueing Networks

- **Queueing Network**: model in which jobs departing from one queue arrive at another queue (or possibly the same queue)
- **Open queueing network**: external arrivals and departures
  - Number of jobs in the system varies with time.
  - Throughput = arrival rate
  - Goal: To characterize the distribution of number of jobs in the system.



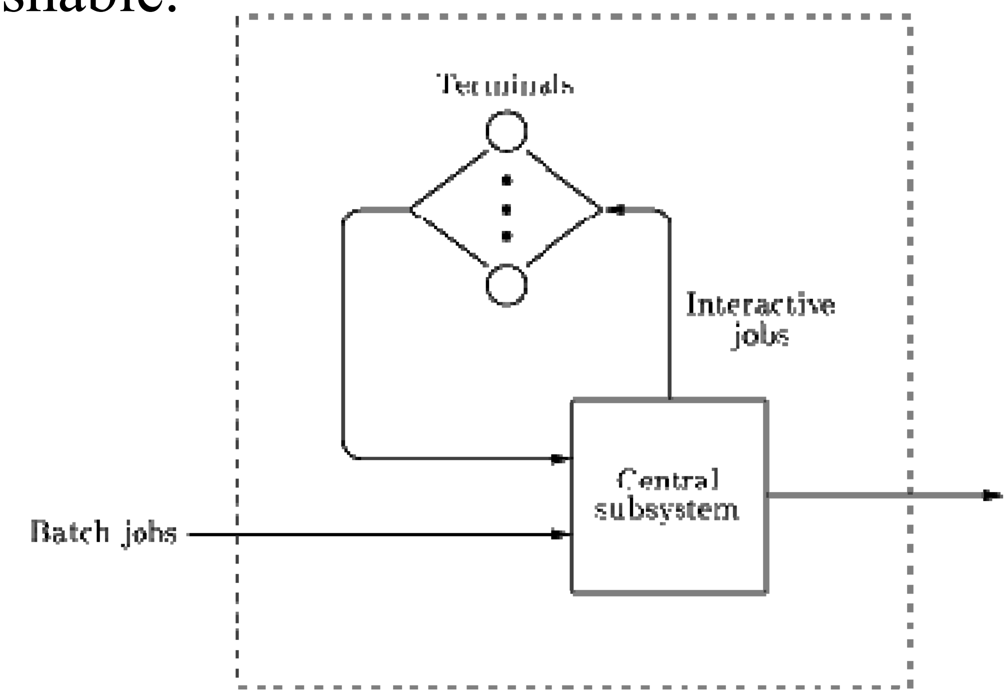
Closed Queueing Networks

- Closed queueing network: No external arrivals or departures
  - Total number of jobs in the system is constant
  - `OUT' is connected back to `IN.'
  - Throughput = flow of jobs in the OUT-to-IN link
  - Number of jobs is given, determine the throughput

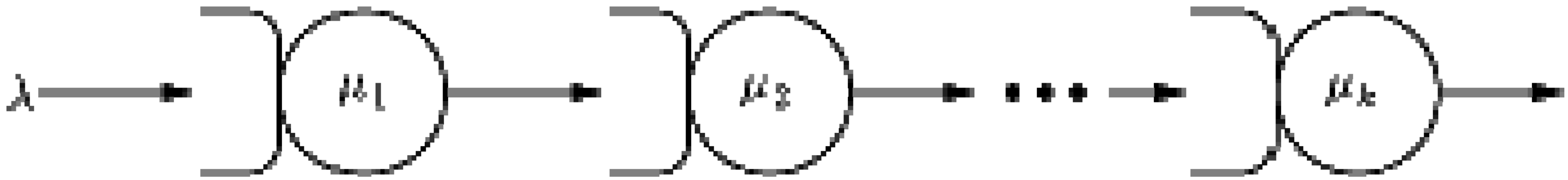


Mixed Queueing Networks

- **Mixed queueing networks**: Open for some workloads and closed for others ⇒ Two classes of jobs. **Class** = types of jobs. All jobs of a single class have the same service demands and transition probabilities. Within each class, the jobs are indistinguishable.



Series Networks



- *k M/M/1* queues in series
- Each individual queue can be analyzed independently of other queues
- Arrival rate= λ. If μ_i is the service rate for i^{th} server:

$$\text{Utilization of } i^{th} \text{ server} \rho_i = \lambda / \mu_i$$

$$\text{Probability of } n_i \text{ jobs in the } i^{\text{th}} \text{ queue } = (1 - \rho_i)\rho_i^{n_i}$$

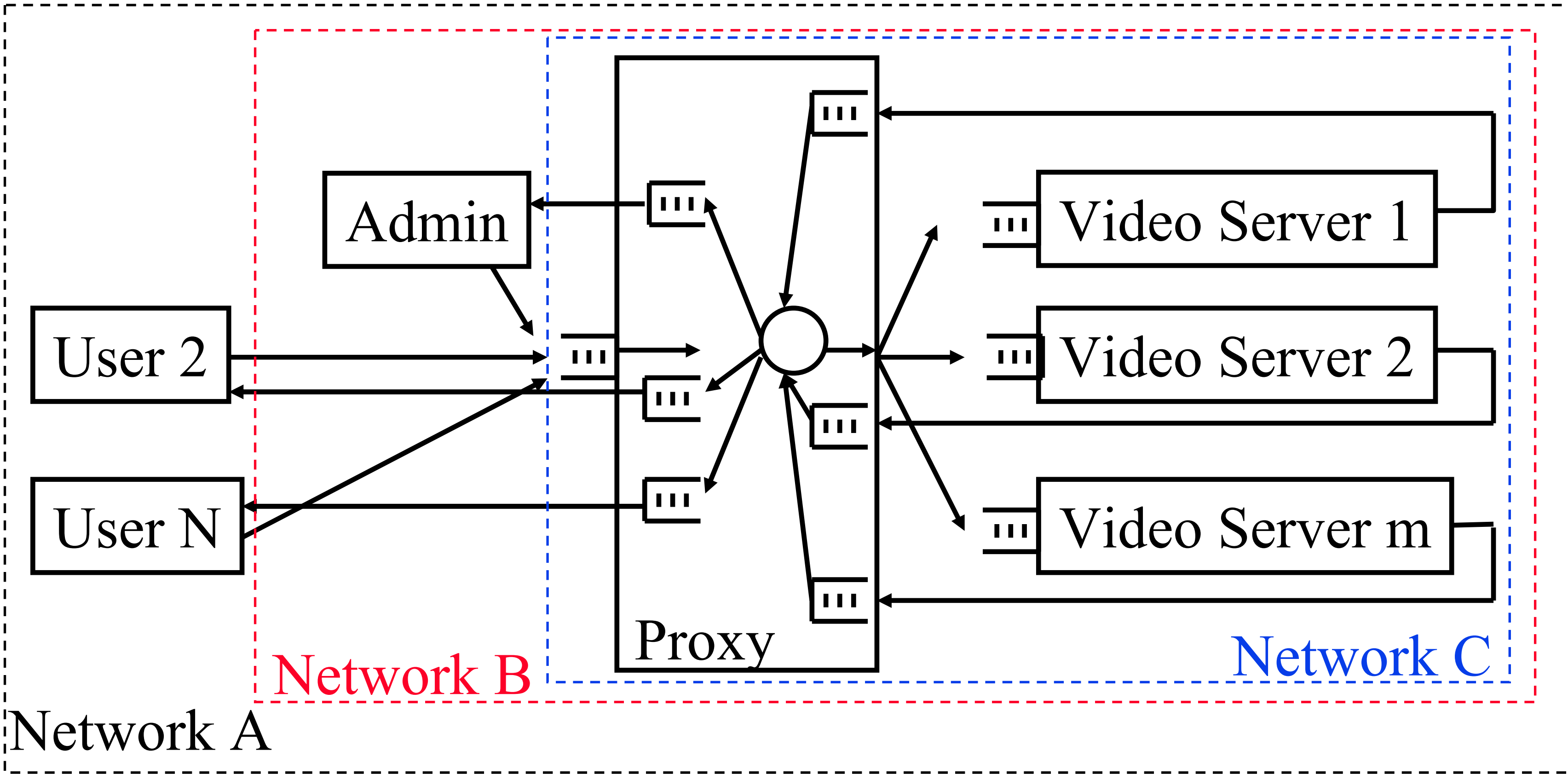
Series Networks (Cont)

- Joint probability of queue lengths:

$$
\begin{aligned}
&P(n_1, n_2, n_3, \ldots, n_M) \\
&= (1-\rho_1)\rho_1^{n_1}(1-\rho_2)\rho_2^{n_2}(1-\rho_3)\rho_3^{n_3}\cdots(1-\rho_M)\rho_M^{n_M} \\
&= p_1(n_1)p_2(n_2)p_3(n_3)\cdots p_M(n_M)
\end{aligned}
$$

$\Rightarrow$ product form network

Quiz 32A



Identify open/closed/mixed networks:

A. Network A is ____________

B. Network B is ____________

C. Network C is ____________

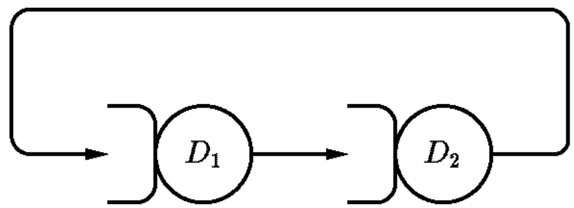
Product-Form Network

- Any queueing network in which:

$$P(n_1, n_2, \ldots, n_M) = \frac{1}{G(N)} \prod_{i=1}^{M} f_i(n_i)$$

- When $f_i(n_i)$ is some function of the number of jobs at the ith facility, $G(N)$ is a normalizing constant and is a function of the total number of jobs in the system.

Example 32.1

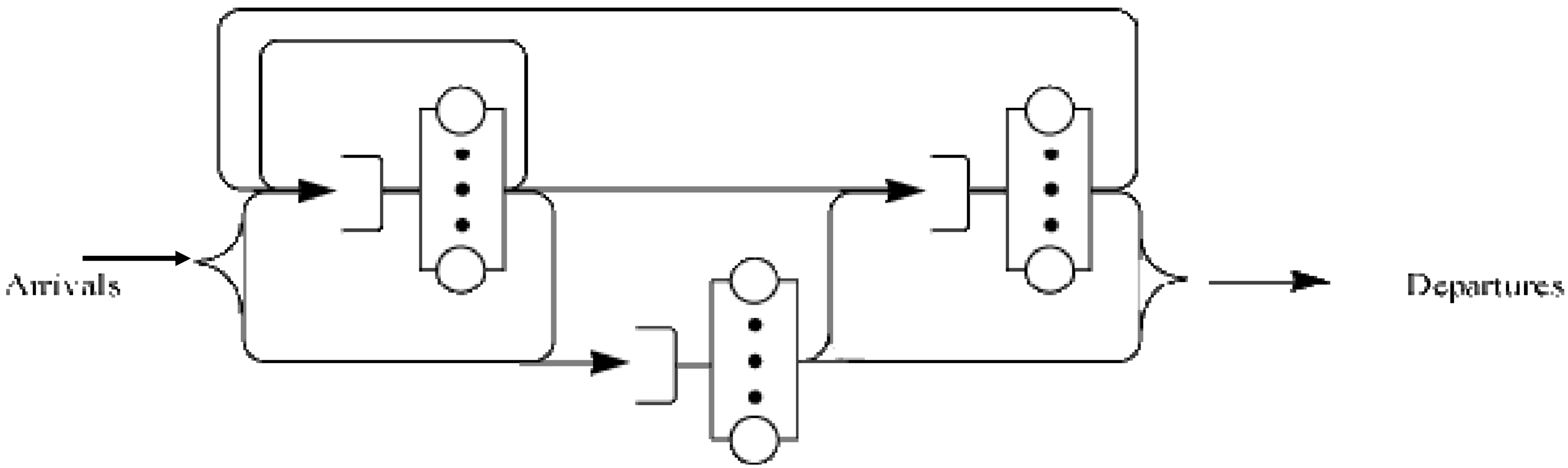


- Consider a closed system with two queues and N jobs circulating among the queues:
- Both servers have an exponentially distributed service time. The mean service times are 2 and 3, respectively. The probability of having n_1 jobs in the first queue and $n_2=N-n_1$ jobs in the second queue can be shown to be:

$$P(n_1, n_2) = \frac{1}{3^{N+1} - 2^{N+1}} \left(2^{n_1} \times 3^{n_2}\right)$$

- In this case, the normalizing constant $G(N)$ is $3^{N+1}-2^{N+1}$.
- The state probabilities are products of functions of the number of jobs in the queues. Thus, this is a ***product form network***.

General Open Network of Queues



- Product form networks are easier to analyze
- Jackson (1963) showed that any arbitrary open network of m-server queues with exponentially distributed service times has a product form

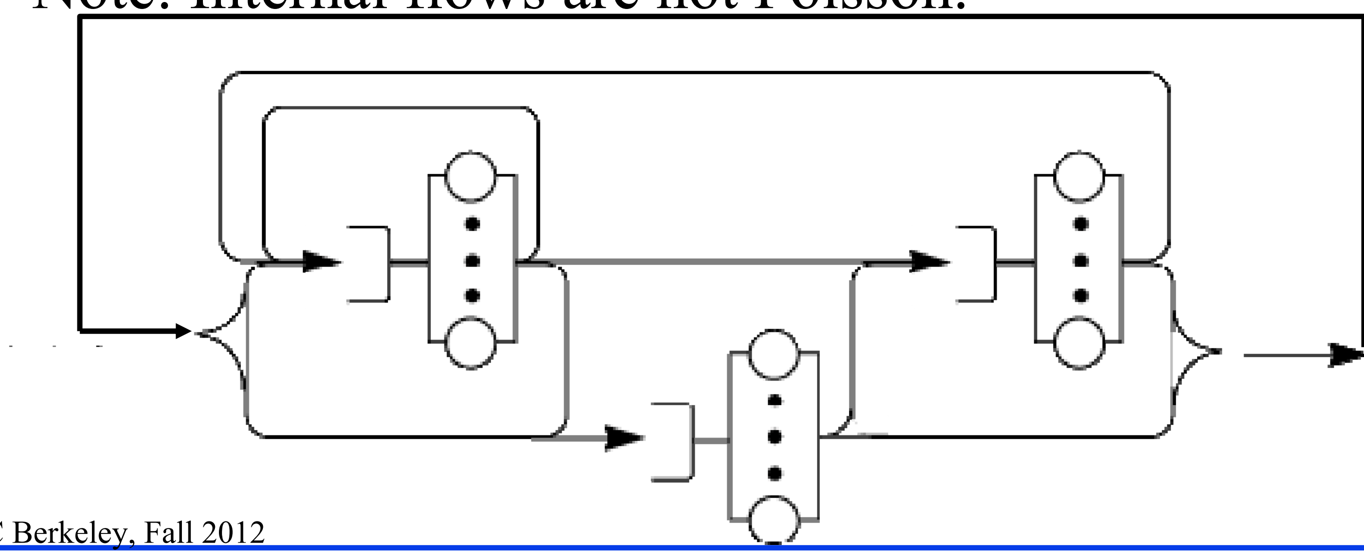
General Open Network of Queues (Cont)

- If all queues are single-server queues, the queue length distribution is:

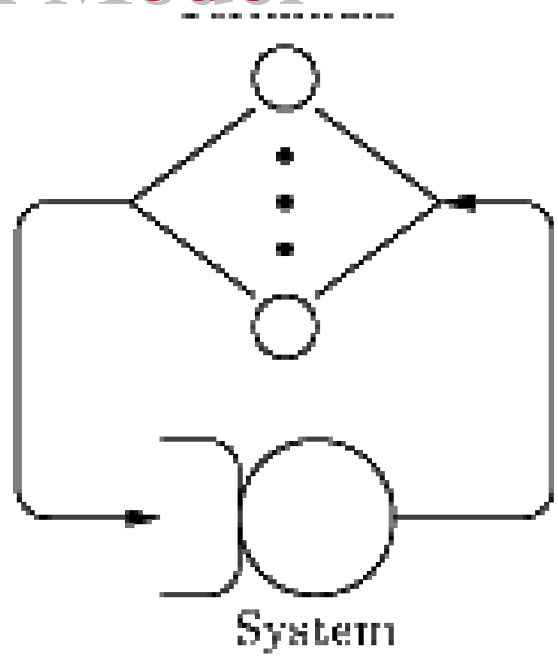
$$
\begin{aligned}
&P(n_1, n_2, n_3, \ldots, n_M) \\
&= (1-\rho_1)\rho_1^{n_1}(1-\rho_2)\rho_2^{n_2}(1-\rho_3)\rho_3^{n_3}\cdots(1-\rho_M)\rho_M^{n_M} \\
&= p_1(n_1)p_2(n_2)p_3(n_3)\cdots p_M(n_M)
\end{aligned}
$$

Closed Product-Form Networks

- Gordon and Newell (1967) showed that any arbitrary closed networks of m-server queues with exponentially distributed service times also have a product form solution.
- Baskett, Chandy, Muntz, and Palacios (1975) and then Denning and Buzen (1978) showed that product form solutions exist for an even broader class of networks.
- Note: Internal flows are not Poisson.



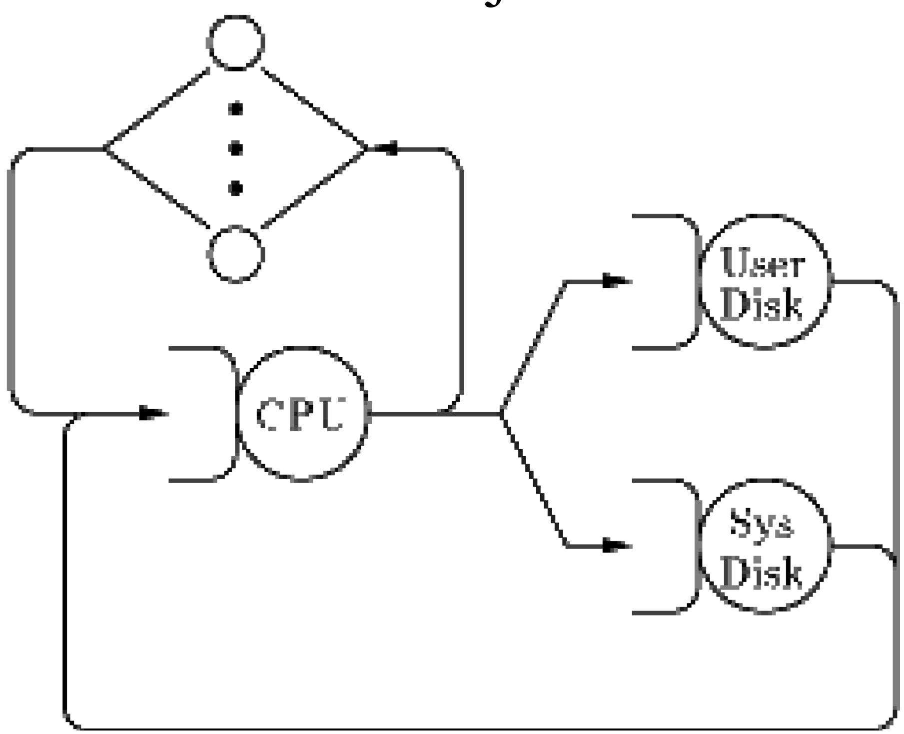
Machine Repairman Model



- Originally for machine repair shops
- A number of working machines with a repair facility with one or more servers (repairmen).
- Whenever a machine breaks down, it is put in the queue for repair and serviced as soon as a repairman is available
- Scherr (1967) used this model to represent a timesharing system with *n* terminals.
- Users sitting at the terminals generate requests (jobs) that are serviced by the system which serves as a repairman.
- After a job is done, it waits at the user-terminal for a random ``think-time'' interval before cycling again.

Central Server Model

- Introduced by Buzen (1973)
- The CPU is the ``central server'' that schedules visits to other devices
- After service at the I/O devices the jobs return to the CPU



Types of Service Centers

Three kinds of devices

1. **Fixed-capacity service centers**: Service time does not depend upon the number of jobs in the device

For example, the CPU in a system may be modeled as a fixed-capacity service center.

2. **Delay centers or infinite server**: No queueing. Jobs spend the same amount of time in the device regardless of the number of jobs in it. A group of dedicated terminals is usually modeled as a delay center.

3. **Load-dependent service centers**: Service rates may depend upon the load or the number of jobs in the device., e.g., *M/M/m* queue (with $m \geq 2$)

A group of parallel links between two nodes in a computer network is another example

Quiz 32B

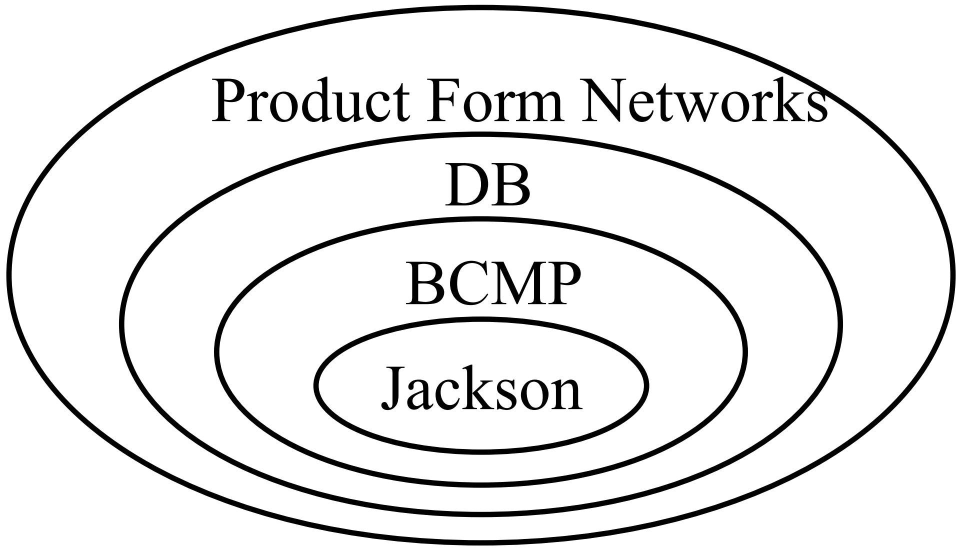
- The probability function for jobs in a system with m queues is:

$$P(n_1, n_2, , n_m) = \frac{g(n_1)g(n_2)g(n_{m-1})}{g(N - n_m)}$$

Is this a product form network? ________

- Identify the type of server:
  A. Multi-core CPU: ____________
  B. Single-core CPU (No dynamic frequency scaling): ____________
  C. Single-core CPU (with dynamic frequency scaling): ____________
  D. Hard disk drives: ____________
  E. Solid state drives: ____________
  F. Multiple users each handling one window: ____________
  G. A user handling multiple windows: ____________

Summary



- Open, Closed, and Mixed queueing networks
- Product form networks: Any network in which the system state probability is a product of device state probabilities
- Jackson: Network of M/M/m queues.
  BCMP: More general conditions
  Denning and Buzen: Even more general conditions
- Service centers: Fixed capacity, delay centers, load dependent

Homework 32

- Select a system in which jobs go to another queue after finishing service at one queue. Draw the queueing network model.

A. Show the possible paths the jobs follow.

B. Number of classes of jobs in the system?

C. Is the system open/closed/mixed?

D. Are the transition probabilities fixed for each class and at each server exit?