

# Fundamentals of Queueing Networks

A *queueing network model* (QNM) of a computer system is a collection of service stations connected via directed paths along which the customers of the system move. The *stations* represent various system resources, and the *customer* represent jobs, processes, or other active entities. The *customers* move from one station to another, queueing up at each for service. The service requirement of a customer at a station is a random variable that is described by a probability distribution..

# Closed and Open Queueing Networks

- A queueing network in which there is no restriction on the number of customers is called an *open* or *infinite population* model. In such models, the customers initially arrive from an external source and eventually leave the system.
- In models where the number of customers is fixed (i.e., no arrivals from, or departures to, the external world), the model is called a *closed* or model with the number of circulating customers known as the *population*.

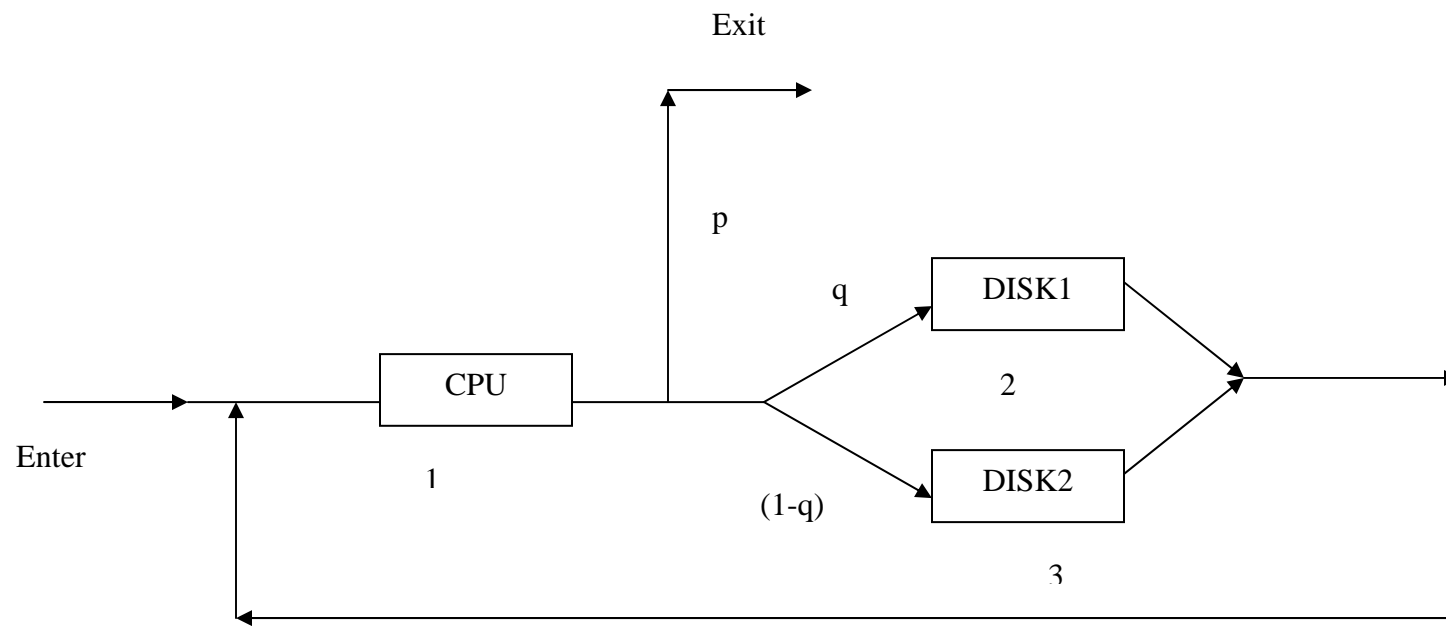


Figure 1

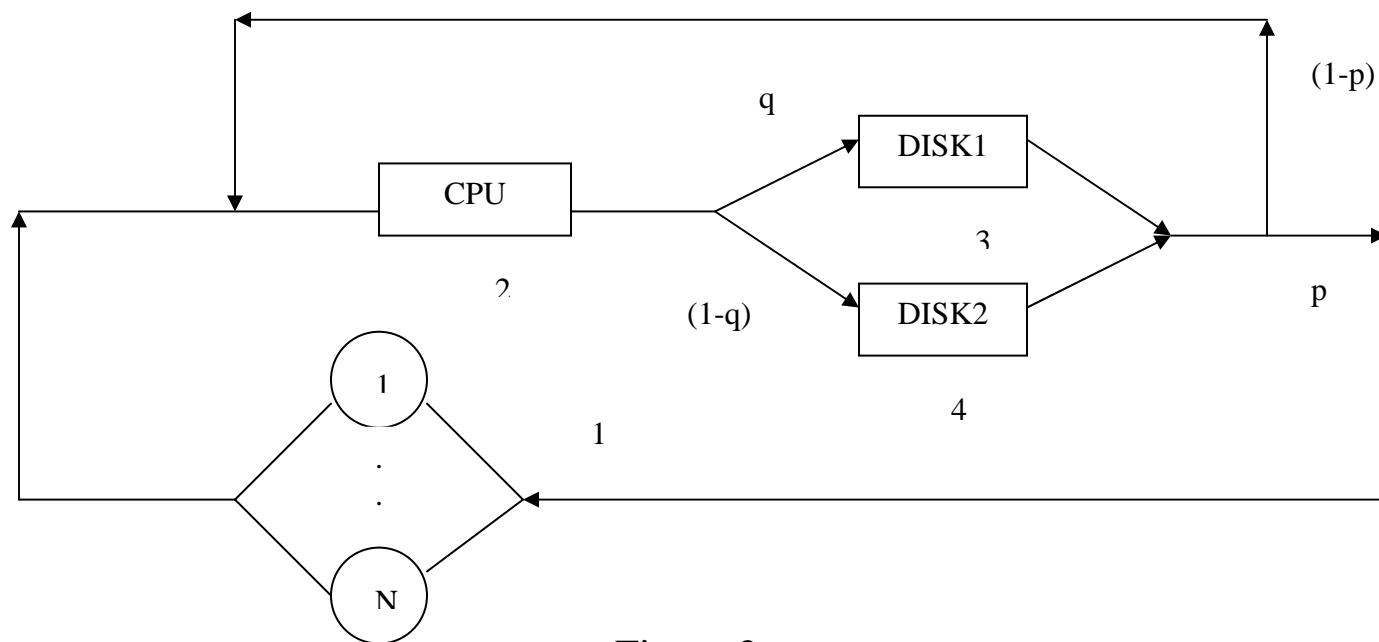


Figure 2

# Performance Parameters

- A simple QNM requires the following *inputs*:
  - (1) Number of stations, henceforth denoted as  $M$ ,
  - (2) Service-time distribution and scheduling discipline at each station,
  - (3) Routing probabilities of customers among stations,
  - (4) Population (for closed models) or interarrival-time distribution (for open models)
- The *output* parameters of interest are the various performance measures at each station, such as response time, queue length, throughput, and utilization.

# Performance Parameters List:

In the following list, we define certain basic input and output parameters for the stations of a simple queueing network and establish notations for them. The subscript  $i$  refers to the station to which the parameters apply. The network in question may be open or closed. In the latter case, it is customary to include the population level  $N$  as an argument of the output parameters.

1. *Average service time  $s_i$* : Average time spent in serving a customer at station  $i$ . We shall often speak of the (average) service rate, denoted  $\mu_i$ . This is defined as simply  $1/s_i$ .
2. *External arrival rate  $\lambda_i$* : Average rate at which customers arrive to station  $i$ , from the *external world*. This applies to open networks only.

# Performance Parameters List (continued):

3. *Routing probability  $q_{ij}$* : Fraction of departures from station  $i$  headed to station  $j$  next.
4. *Throughput  $\lambda_i$  or  $\lambda_i(N)$* : Average number of service completions per unit time at station  $i$ .
5. *Average response time  $R_i$  or  $R_i(N)$* : Average time a customer spends at station  $i$ , either waiting to be served or receiving service.
6. *Average waiting time  $W_i$  or  $W_i(N)$* : Average time a customer spends at station  $i$  waiting to be served.

# Performance Parameters List (continued):

7. *Average queue length  $Q_i$  or  $Q_i(N)$* : Average number of customers at station  $i$ , *including those being served*.
8. *Average waiting line length  $L_i$  or  $L_i(N)$* : Average number of customers at station  $i$ , *excluding those being served*.
9. Utilization  $U_i$  or  $U_i(N)$ : Fraction of time that station  $i$  is busy.  
(As such, this definition applies only when  $s_i$  is independent of  $n$ ).
10. Queue-length distribution  $P_i(n)$  or  $P_i(n|N)$ : Probability of finding  $n$  customers at station  $i$ .



# Operational Definitions of Performance Measures

Suppose that the system of interest is station  $i$  belonging to a closed network with population  $N$ . Let us define the following parameters over the observation period  $T$ :

- $A_i(n)$  Number of arriving customers who find  $n$  customers at station  $i$ .
- $D_i(n)$  Number of departing customers who leave behind  $n-1$  customers at station  $i$  (i.e., the number of departures while the station is in state  $n$ ).
- $T_i(n)$  Duration of time for which there are  $n$  customers at station  $i$ .

# Operational Definitions of Performance Measures (continued)

- Note that  $A_i(N) = 0$  and  $D_i(0) = 0$ , but  $T_i(n)$  may be nonzero for all values of  $n = 1..N$ . Let  $B_i$  denote the total busy period. Obviously,

$$B_i = T - T_i(0) = \sum_{n=1}^N T_i(n)$$

- Let  $A_i$  and  $D_i$  denote, respectively, the total number of arrivals and departures during time  $T$ . Then

$$A_i = \sum_{n=0}^{N-1} A_i(n) \quad \text{and} \quad D_i = \sum_{n=1}^N D_i(n)$$

# Operational Definitions of Performance Measures (continued)

- Let  $X_i(N)$  denote the overall arrival rate to station  $i$ . Then

$$X_i(N) = A_i / T$$

- Let  $Y_i(n|N)$  denote the arrival rate to station  $i$  when  $n$  customers are present there. This is often known as restricted arrival rate and is given by

$$Y_i(n|N) = A_i(n) / T_i(n), \quad 0 \leq n \leq N-1$$

Note that  $Y_i(N|N) = 0$ .

# Operational Definitions of Performance Measures (continued)

- The departure rate or throughput of station  $i$ , denoted  $\lambda_i(N)$  is given by

$$\lambda_i(N) = D_i / T$$

- The average service time of station  $i$  when  $n$  customers are present there, denoted  $s_i(n)$ , is given by

$$s_i(n|N) = T_i(n) / D_i(n), \quad 1 \leq n \leq N$$

In analogy with  $Y_i(n|N)$ , we can think of service rate  $\mu_i(n)$  as the *restricted departure rate*.

# Operational Definitions of Performance Measures (continued)

- If the services are homogeneous, we denote  $s_i(n)$  and  $\mu_i(n)$  as simply  $s_i$  and  $\mu_i$ , respectively.
- If the arrivals are homogeneous,  $Y_i(n|N)$  is independent of  $n$  for  $n < N$  and denoted as  $Y_i(N)$ .
- Flow balance means that  $A_i = D_i$ , which implies that  $X_i(N) = \lambda_i(N)$ .

# Some Interesting Distributions of the Number of Customers in the System

- *Distribution seen by an arriving customer*, the probability that an arriver finds  $n$  customers in the system, given by

$$PA_i(n|N) = A_i(n) / A_i \quad \text{for } 0 \leq n < N$$

- *Distribution seen by a departing customer*, the probability that a departer leaves behind  $n$  customers in the system, given by

- 

$$PD_i(n|N) = D_i(n+1) / D_i \quad \text{for } 0 \leq n < N$$

# Some Interesting Distributions of the Number of Customers in the System (continued)

- *Distribution seen by a random observer*, a fraction of time the system contains  $n$  customers, given by

- 

$$P_i(n|N) = T_i(n) / T \quad \text{for } 0 \leq n \leq N$$

- As a simple example

$$U_i(N) = B_i / T = (T - T_i(0)) / T = 1 - P_i(0|N)$$

# Flow Balance Law

Let  $\lambda_i$  denote the throughput of station  $i$  in a flow-balanced network with  $M$  stations. Then for any station  $i$ , the total arrival rate should also be  $\lambda_i$ . This arrival rate can be expressed in terms of throughputs of other stations by accounting for all the flows into station  $i$ . This gives

$$\lambda_i = \Lambda_i + \sum_{j=1}^M \lambda_j q_{ij}, \quad 1 \leq i \leq M$$

Where  $\Lambda_i$  is the external arrival rate to station  $i$ , and  $q_{ij}$  is the probability that a customer exiting station  $j$  goes to station  $i$  directly.



## Flow Balance Law (continued)

- If the network is open, the previous system of equations will be non-homogeneous, and will have a unique solution. This means that given the external arrival rates, we can determine the throughputs of all stations exactly.
- If the network is closed, the previous system of equations will become homogeneous and will not have a unique solution. It will in fact include the set of exactly  $M-1$  linearly independent equations, which can be solved to determine the ratios of throughputs.

## Example:

Compute the (relative) throughputs for each station in Figures 1 and 2.

## Solution:

Let  $\Lambda$  denote the external arrival rate, and  $\lambda_i$  the throughput of station  $i$ . Then for Figure 1, we can write the following equations

$$\begin{aligned}\lambda_1 &= \Lambda + \lambda_2 + \lambda_3 \\ \lambda_2 &= (1-p) \lambda_1 \times q \\ \lambda_3 &= (1-p) \lambda_1 \times (1-q)\end{aligned}$$

## Solution (continued)

Suppose that we define visit ratios relative to the external source, i.e.,  $v_i = \lambda_i / \Lambda$ . Then

$$v_1 = 1/p \quad v_2 = q(1-p)/p \quad \text{and} \quad v_3 = (1-q)(1-p)/p$$

Note that we can determine exact throughputs here. For example, if  $p = 0.1$ ,  $q = 0.8$ , and  $\Lambda = 10/\text{sec}$ , then we have:

$$\lambda_1 = 100/\text{sec}, \lambda_2 = 72/\text{sec}, \text{ and } \lambda_3 = 18/\text{sec}.$$

## Solution (continued)

For the model of Figure 2, we have

$$\lambda_2 = \lambda_1 + (1-p) (\lambda_3 + \lambda_4), \quad \lambda_3 = q \lambda_2, \quad \lambda_4 = (1-q) \lambda_2$$

It is easy to see that these equations are linearly independent, but we cannot write any more independent equations. Thus, exact throughputs cannot be determined, but visit ratios can. For example, if we let  $v_2 = 1$ ,  $p = 0.2$ , and  $q = 0.6$ , then  $v_3 = 0.2$ ,  $v_1 = 0.6$ , and  $v_4 = 0.4$ .

## Some Fundamental Results:

$$\lambda_i(N) = \frac{D_i}{T} = \frac{1}{T} \sum_{n=1}^N D_i(n) = \sum_{n=1}^N \frac{D_i(n)}{T_i(n)} \frac{T_i(n)}{T} = \sum_{n=1}^N \frac{P_i(n|N)}{s_i(n)}$$

Put in standard form, the throughput law then states

$$\lambda_i(N) = \sum_{n=1}^N \mu_i(n) P_i(n|N)$$

If the services are homogeneous, i.e.,  $\mu_i(n) = \mu_i$  for all  $n > 0$ , the above equation can be written as

$$\lambda_i(N) = \mu_i [1 - P_i(0|N)],$$

or,

$$U_i(N) = \lambda_i(N) s_i$$

## Some Fundamental Results: (continued)

A more direct way of obtaining the *utilization law* is as follows:

$$U_i(N) = \frac{B_i}{T} = \frac{B_i}{D_i} \times \frac{D_i}{T} = s_i \lambda_i(N)$$

Like the throughput law, we have the following *arrival law*:

$$X_i(N) = \sum_{n=0}^{N-1} Y_i(n | N) P_i(n | N)$$

The proof follows trivially from the definition:

$$X_i(N) = \frac{A_i}{T} = \sum_{n=0}^{N-1} \frac{A_i(n) T_i(n)}{T_i(n) T} = \sum_{n=0}^{N-1} Y_i(n | N) P_i(n | N)$$

## Some Fundamental Results: (continued)

Under flow balance, the arrival rate is the same as throughput, which means that the previous equation is another way of expressing the throughput. If the arrivals are homogeneous, i.e.,  $Y_i(n|N) = Y_i(N)$  for all  $n < N$ , the previous equation can be written as

$$X_i(N) = Y_i(N) (1 - P_i(N|N))$$

# Properties of Queue Length Distributions

In general, if arrivals (departures) occur singly, then in any system state  $n$ , at most one arrival (departure) can remain unmatched by a departure (arrival). This happens because the system cannot return to state  $n$  unless an opposite event occurs. It follows that

$$\forall n \quad |A_i(n) - D_i(n+1)| \leq 1$$

where  $D_i(n+1)$  can exceed  $A_i(n)$  only if we start with a nonempty system. Under flow balance, we cannot have an arriver who has not departed, or vice versa, and hence the above equation reduces to

$$A_i(n) = D_i(n+1)$$



# Properties of Queue Length Distributions (continued)

The earlier equation implies

$$A_i = D_i$$

or

$$PA_i(n|N) = PD_i(n|N)$$

# Properties of Queue Length Distributions (continued)

We can also get recursive equations for random observer's and arriver's distributions under these two assumptions. By definition,

$$P_i(n | N) = \frac{T_i(n)}{T} = \frac{T_i(n)}{D_i(n)} \frac{D_i(n)}{T} = \frac{T_i(n)}{D_i(n)} \frac{A_i(n-1)}{T_i(n-1)} \frac{T_i(n-1)}{T}$$

where we have used the identity  $D_i(n) = A_i(n-1)$ . Recognizing various terms, we get

$$P_i(n|N) = s_i(n) Y_i(n-1|N) P_i(n-1|N)$$

# Properties of Queue Length Distributions (continued)

Similarly, for the arriver's distribution, we get

$$PA_i(n | N) = \frac{A_i(n)}{A_i} = \frac{A_i(n)}{T_i(n)} \frac{T_i(n)}{D_i(n)} \frac{D_i(n)}{A_i} = Y_i(n | N) s_i(n) \frac{A_i(n-1)}{A_i}$$

which yields

$$PA_i(n|N) = s_i(n) Y_i(n|N) PA_i(n-1|N)$$

# Properties of Queue Length Distributions (continued)

We can show another interesting result popularly known as the *arrival theorem*. It states that

$$PA_i(n|N) = P_i(n|N-1)$$

The departure's distribution can be related of the random observer's distribution (without any assumptions) as follows:

$$PD_i(n | N) = \frac{D_i(n+1)}{T_i(n+1)} \frac{T_i(n+1)}{T} \frac{T}{D_i} = \frac{\mu_i(n+1)}{\lambda_i(N)} P_i(n+1 | N)$$

# Properties of Queue Length Distributions (continued)

Under flow balance and one-step behavior, we get

$$PA_i(n|N) = \mu_i(n+1) P_i(n+1|N)/\lambda_i(N)$$

Furthermore, using the arrival theorem, we get the so called *marginal local balance theorem*, which can be restated as

$$\mu_i(n) P_i(n|N) = \lambda_i(N) P_i(n-1|N-1)$$

# Properties of Queue Length Distributions (continued)

In practice, the arrivals are often homogeneous, and we can show some interesting results under this assumption. By definition,

$$PA_i(n | N) = \frac{A_i(n)}{A_i} = \frac{A_i(n)}{T_i(n)} \frac{T_i(n)}{T} \frac{T}{A_i} = \frac{Y_i(N)}{X_i(N)} P_i(n | N)$$

Therefore, we get

$$PA_i(n | N) = \frac{P_i(n | N)}{1 - P_i(N | N)}$$

That is, in a closed network under homogeneous arrivals, the arriver's distribution is a simple renormalization of the random observer's distribution

# Properties of Queue Length Distributions (continued)

For an open network, the previous equation means that the arriver's and random observer's distributions are identical. If the system shows one-step behavior and flow balance as well, we get

$$P_i(n | N) = \frac{\lambda_i(N)}{\mu_i(n)} \frac{P_i(n-1 | N)}{1 - P_i(N | N)} \triangleq \alpha(n) P_i(n-1 | N)$$

This recurrence relation, long with the requirement that all probabilities sum to 1, gives us the following expression

$$P_i(n | N) = \prod_{i=1}^n \alpha(i) \left[ 1 + \sum_k^N \prod_{j=1}^k \alpha(j) \right]^{-1}$$

# Properties of Queue Length Distributions (continued)

Finally, under homogeneous services, flow balance, and one-step behavior, we can write

$$P_i(n|N) = U_i(N) PA_i(n-1|N)$$

which means that

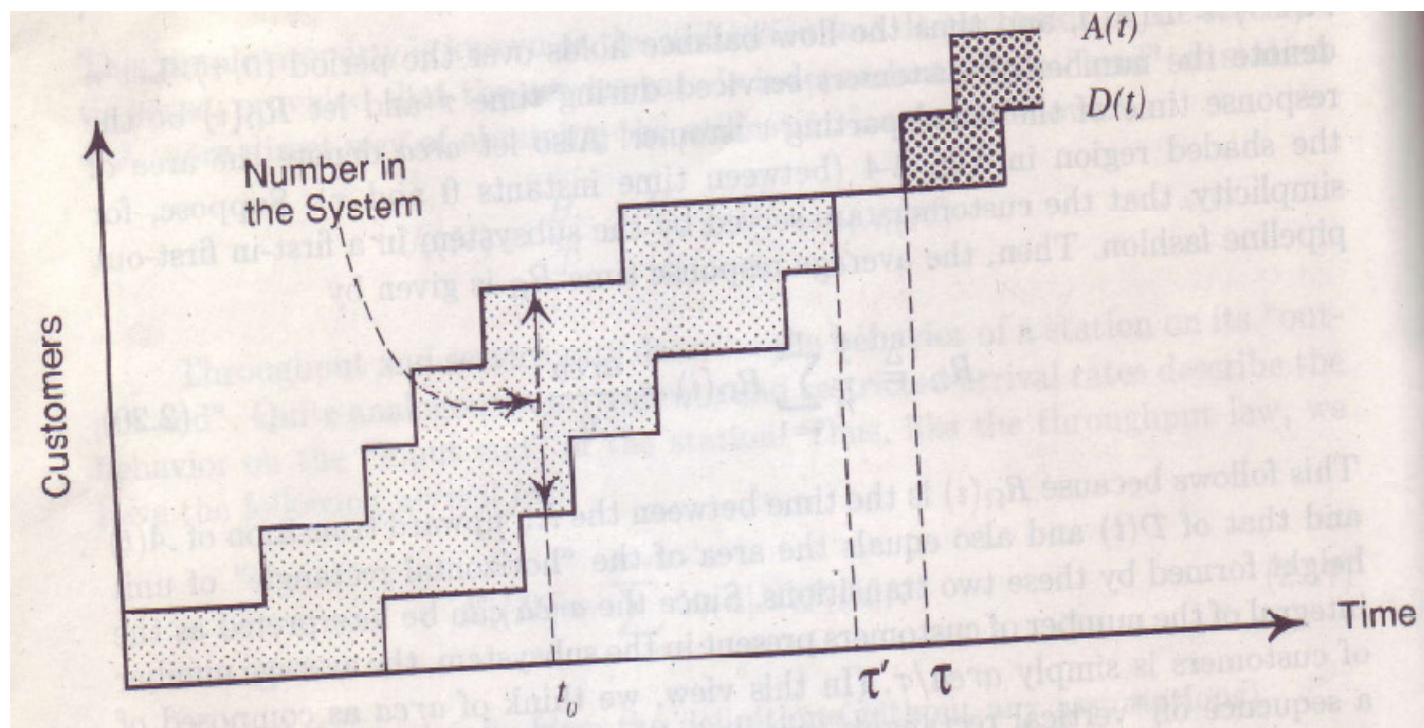
$$Q_i(N) \triangleq \sum_{n=1}^N nP_i(n|N) = U_i(N) \sum_{n=1}^N nPA_i(n-1|N)$$

which along with Little's law gives the following interesting relationship

$$R_i(N) = s_i (1 + QA_i(N))$$

Where  $QA_i(N)$  is the average queue length seen by an arriving customer. Note that by the arrival theorem,  $QA_i(N) = Q_i(N-1)$ .





# Example:

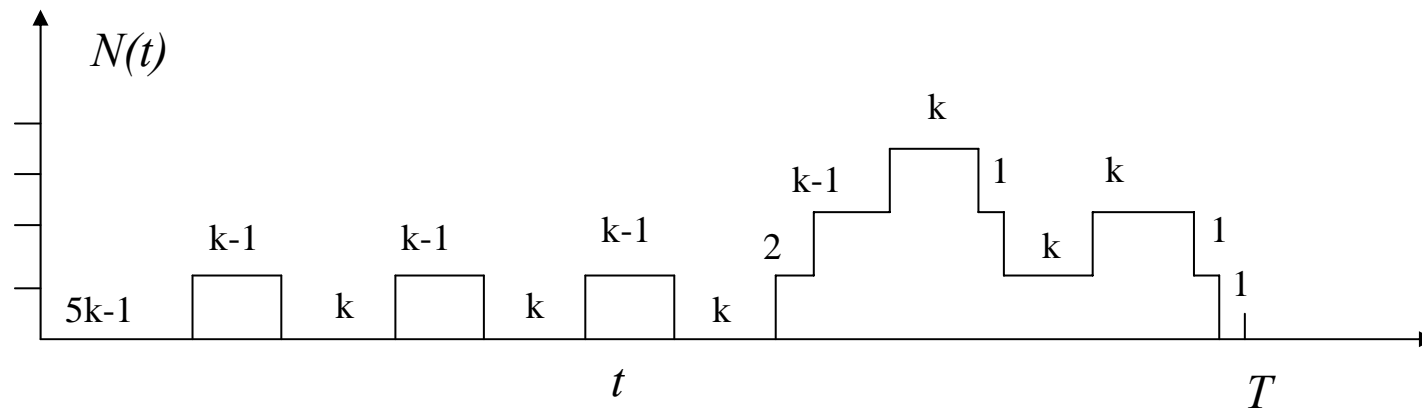


Figure 3: A sample behavior sequence.

## Example (continued):

$n$	$A_i(n)$	$D_i(n)$	$T_i(n)$	$P_i(n)$	$PA_i(n)$	$PD_i(n)$	$X_i(n)$	$s_i(n)$
0	4	-	$8k$	$8/15$	$4/7$	$4/7$	$1/(2k)$	-
1	2	4	$4k$	$4/15$	$2/7$	$2/7$	$1/(2k)$	$k$
2	1	2	$2k$	$2/15$	$1/7$	$1/7$	$1/(2k)$	$k$
3	-	1	$k$	$1/15$	-	-	-	$k$

Table 1: Performance parameters for the behavior sequence in Figure 3.

## Example (continued):

- Note that  $A_i = D_i = 7$  as expected from the balance property, which also means that the arrival rate  $X_i(N)$  and throughput  $\lambda_i(N)$  should be the same ( $= 7/15k$ ).
- Both arrivals and services are homogeneous since  $Y_i(n|N)$  and  $s_i(n)$  are independent of  $n$ .
- Note that the utilization law applies because  $U_i = 1 - P_i(0) = 7/15$  and  $s_i \lambda_i = k \times 7/(15k) = 7/15$ .
- Also  $PA_i(n)$  is simply a normalization of  $P_i(n)$ , and  $PD_i(n)$  as expected.

# Chains:

- A real computer system may serve many categories of jobs such that the level of service provided and/or the routing depends on the category. For example, the characteristics of batch jobs may be different from those of the interactive jobs. This aspect can be represented in a queueing network model by introducing the concept of *chains*.
- A chain forms a permanent categorization of jobs; a job belonging to one chain cannot switch to another chain.

# Classes:

- In addition to the permanent categorization of jobs, it is also desirable to allow customers to go through different phases of processing in the system. This necessarily requires the ability to switch from one phase to another, and thus needs a concept other than chains. We refer to these phases as *classes* and allow several classes within a chain.
- For example, the execution of both batch and interactive jobs would involve two distinct phases: program loading from disk into the main memory, and the actual execution. We can represent this by having two classes in each chain, such that all customers of a chain start out in class 1 to do program loading, and then switch to class 2 for execution. (In a closed model, class 2 jobs would switch back to class 1 after finishing execution).

# Characterization of Classes and Chains

We would like to answer the following questions concerning a network with multiple categories of customers:

1. How do we check if the network is *consistent*, i.e., connected and structurally sound?
2. How do identify the chains, and the classes contained within those chains?
3. How do we check if the network is *well-formed*, i.e., has a well-defined long-term behavior?

# Characterization of Classes and Chains (continued)

- To answer the above questions, we use a graph model, known as the *reachability graph (RG)*.
- The nodes of a RG are classes, and an arc from node  $c$  to node  $d$  means that a class  $c$  customer can switch to class  $d$  directly (with a nonzero probability).
- Let  $S_1, \dots, S_k$  be the strongly connected components (SCC) of RG. All classes in a SCC must show similar long-term behavior, because any class in a SCC can transform to any other within the same SCC.



# Classification of SCC's:

Each SCC can be classified into one of the following four categories:

**Closed** if in the long run it will have a finite number of customers permanently locked in it.

**Open** if it will have both arrivals from and departures to the external world (either directly or via the classes of other SCC's).

**Transient** if in the long run it will have no customers in it.

**Unstable** if in the long run it receives customers from the external world or classes of other SCC's, but does not shed any.

# Characterization of well-formedness:

- A consistent network is well-formed if and only if it does not have any unstable or transient SCC's.
- In a well-formed network, closed SCC's can be identified as closed chains, and open SCC's can be further grouped into open chains.

# Classification algorithm:

We start with the reachability graph, of the given model, and proceed as follows:

1. Construct a reduced reachability graph, say  $G_r$ , by replacing the SCC's of  $G$  by individual nodes. In  $G_r$ , an arc from node  $i$  to  $j$  indicates that some class in the SCC  $S_i$  can switch to some class in the SCC  $S_j$  with nonzero probability.
2. If class  $i$  customers arrive from the external world, regard this as switch from a ``dummy class'' to class  $i$ . Similarly, regard class  $i$  departures to the external world as a switch from class  $i$  to yet another dummy class. Add nodes and arcs to  $G_r$  corresponding to these dummy classes. Let the new graph be denoted as  $G'_r$ .

3. Label all nodes of  $G_r'$  as  $O$ ,  $C$ ,  $T$  or  $U$  (for open, closed, transient and unstable) recursively as follows:
- a) Label all “dummy nodes” as  $O$ , and all isolated nodes as  $C$ .
  - b) If a node has an incoming arc but at least one outgoing arc, label it as  $T$ .
  - c) If a node has at least one incoming arc and at least one outgoing arc, label it as follows: if all its source nodes are labeled as  $T$ , label it as  $T$ , else label it as  $O$ .
  - d) If a node has no outgoing arc but at least one incoming arc, label it as follows: if all its source nodes carry the label  $T$ , label it as  $C$ , else label it as  $U$ .

4. Remove from  $G_r'$ , all dummy nodes and nodes labeled as T or U. (The corresponding arcs are also removed.) Call the resulting graph  $G_r''$ .
5. All nodes marked as C in  $G_r''$  must be isolated and represent closed chains (with SCC population as the chain population). Of the remaining nodes, each connected component represents an open chain.

**Example: A computer system model with  
six categories of jobs**

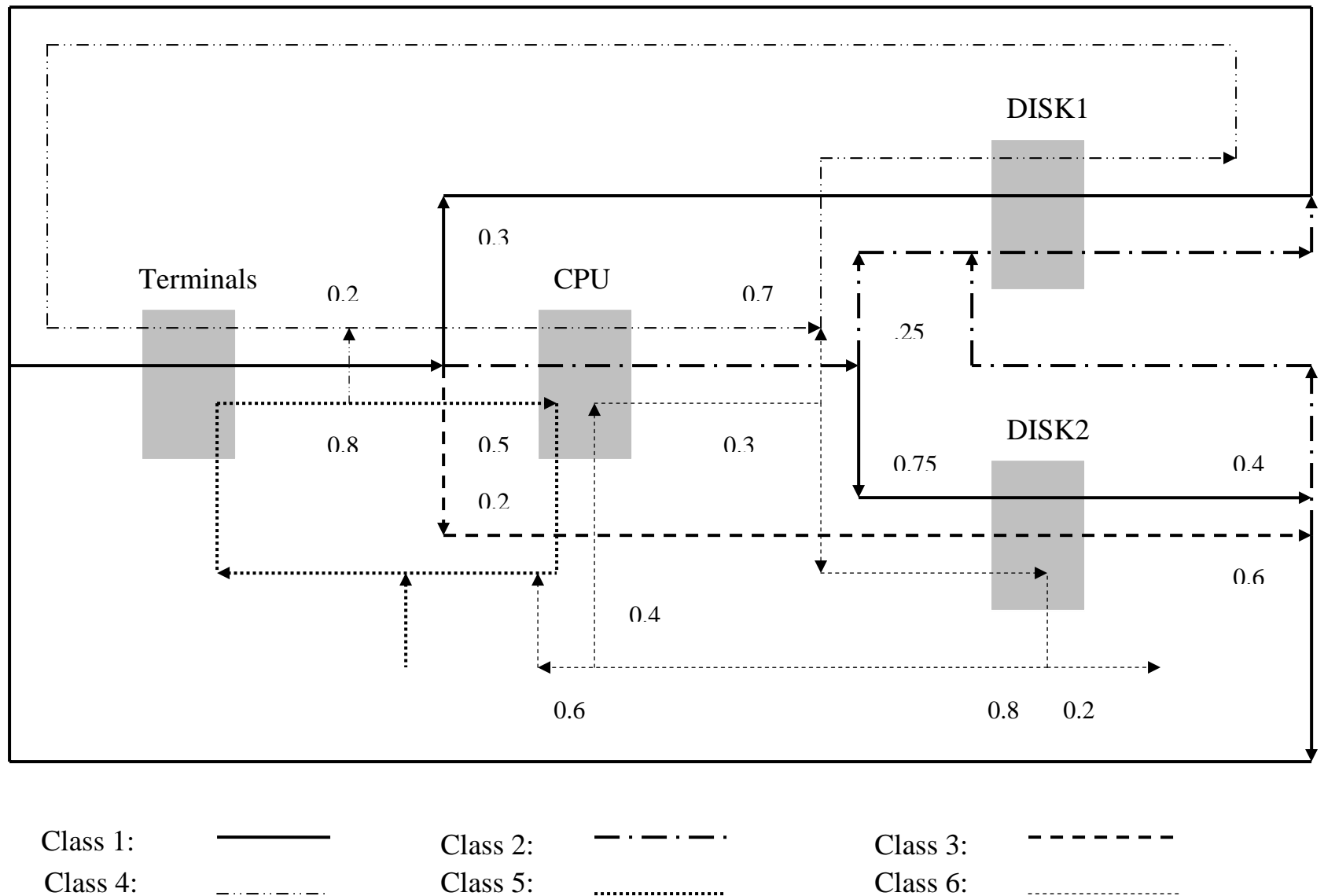


Figure 4: A computer system model with six categories of jobs.

## Example (continued):

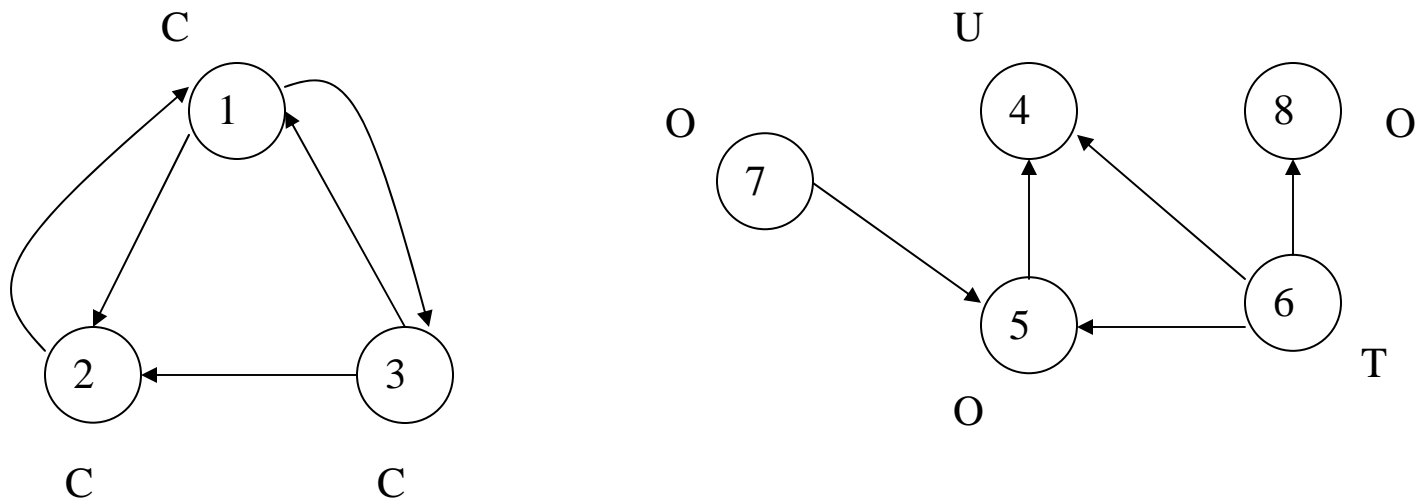


Figure 5: Reachability graph for the model of Figure 4.



# Characteristics of Multiple-Class Networks

- As indicated before, all classes in a well-formed network can be grouped into open and closed chains.
- If all chains in the model are closed, we call the mode closed, if all are open, we call the model open, and if some chains are closed and some open, we call the model mixed.
- For generality, we consider a mixed model with  $K$  chains, such that chains  $1 \dots C$  are closed, and the rest are open.
- We shall denote the population of chain  $i$  as  $N_i$ , and the population vector as  $\mathbf{N} = (N_1, \dots, N_K)$  where by convention,  $N_i = \infty$  for an open chain.

# Characteristics of Multiple-Class Networks

Under flow balance, the forced-flow law yields the following equation:

$$v_{irc} = \Lambda_{irc} + \sum_{j=1}^M \sum_{d=1}^{nc(r)} v_{jrd} q_{jdicr}$$

where  $v_{irc}$  denote the visit ratio for class  $c$  of chain  $r$  at station  $i$ ,  $q_{jdicr}$  denote the routing probability for a chain  $r$  job going to station  $i$  as a member of class  $c$  after being serviced at station  $j$  as a member of class  $d$ , and  $nc(r)$  denote the number of classes in chain  $r$ .

## **Example:**

Compute all visit ratios for the model of Figure 4 (after the removal of classes 4 and 6).

## Solution:

We shall denote the chain containing classes 1-3 as chain 1, and the one containing class 5 as chain 2. The class subscript is redundant for chain 2, and we shall omit it. Suppose (arbitrarily) that  $v_{111} = v_{12} = 1$ . Obviously, for chain 2,  $v_{22} = 0.8 \times v_{12} = 0.8$ . Chain 1 contains classes 1-3, which satisfy the following equations:

$$v_{212} = 0.50 \times v_{111} = 0.50$$

$$v_{311} = 0.30 \times v_{111} = 0.30$$

$$v_{411} = 0.75 \times v_{212} = 0.375$$

$$v_{413} = 0.20 \times v_{111} = 0.20$$

$$v_{312} = 0.25 \times v_{212} + v_{411} + 0.40 \times v_{413} = 0.125 + 0.375 + 0.08 = 0.580$$

# Operational Analysis of Multiple-Chain Networks

For simplicity in stating the results, we only consider multiple-chain networks with one class per chain. Similar results can be obtained for the more general case as well.

**Notations:** We assume the network has  $M$  stations and  $K$  chains. The state of a station in such a network can be described by a vector, called the occupancy vector. We denote it as  $\mathbf{n}_i = (n_{i1}, \dots, n_{iK})$ , whose components  $n_{ir}$  give the number of chain  $r$  customers at station  $i$ . When there is no confusion, we shall shorten this notation to  $\mathbf{n}$  with chain  $r$  component denoted as  $n_r$ . Obviously,  $\mathbf{n}$  can range from  $\mathbf{0}$  to  $\mathbf{N}$ . We also need the notation  $\mathbf{e}_r$  which is a vector with  $K$  elements, of which the  $r$ th element is 1 and all others are zero. Suppose that we observe station  $i$  for  $T$  time units.

# Operational Analysis of Multiple-Chain Networks (continued)

Then we can introduce the following three measures for each chain  $r \in [1..K]$ :

- $A_{ir}(\mathbf{n})$ : Number of arriving chain  $r$  customers who find  $\mathbf{n}$  customers at station  $i$ .
- $D_{ir}(\mathbf{n})$ : Number of chain  $r$  departures while the station is in state  $\mathbf{n}$  (i.e., the number of departing customers that leave behind  $\mathbf{n} - \mathbf{e}_r$  customers).
- $T_{ir}(\mathbf{n})$ : Duration of time for which there are  $\mathbf{n}$  customers at station  $i$ .

# Operational Analysis of Multiple-Chain Networks (continued)

Let

$$A_{ir} = \sum_{n_1=0}^{N_1} \dots \sum_{n_{r-1}=0}^{N_{r-1}} \sum_{n_r=0}^{N_r-1} \sum_{n_{r+1}=0}^{N_{r+1}} \dots \sum_{n_K=0}^{N_K} A_{ir}(\mathbf{n})$$

We shall abbreviate summation like this by specifying the range of  $\mathbf{0}$  to  $N - e_r$  for vector  $\mathbf{n}$ . Thus, we define

$$A_{ir} = \sum_{\mathbf{n}=\mathbf{0}}^{N-e_r} A_{ir}(\mathbf{n})$$

$$D_{ir} = \sum_{\mathbf{n}=e_r}^N D_{ir}(\mathbf{n})$$

# Operational Analysis of Multiple-Chain Networks (continued)

Let  $X_{ir}(N)$  denote the total arrival rate of chain  $r$  customers to station  $i$ . Then

$$X_{ir}(N) = A_{ir} / T$$

Let  $Y_{ir}(\mathbf{n}|N)$  denote the restricted arrival rate of chain  $r$  to station  $i$  when  $n$  customers are present there. Then

$$Y_{ir}(\mathbf{n} | N) = \frac{A_{ir}(\mathbf{n})}{T_i(\mathbf{n})} \quad , \quad \mathbf{0} \leq \mathbf{n} \leq N - e_r$$

The throughput of chain  $r$  at station  $i$ , denoted  $\lambda_{ir}(N)$  is given by

$$\lambda_{ir}(N) = D_{ir} / T$$



# Operational Analysis of Multiple-Chain Networks (continued)

Let the average service time of chain  $r$  customers at station  $I$  when  $n$  customers are present there, be denoted as  $s_{ir}^*(n)$ .

Obviously,

$$s_{ir}^*(n) = \frac{T_i(n)}{D_{ir}(n)}, \quad e_r \leq n \leq N$$

It is also convenient to define service rate as simply the inverse of service time, i.e.,

$$\mu_{ir}^*(n) = 1 / s_{ir}^*(n)$$

# Operational Analysis of Multiple-Chain Networks (continued)

The three basic distributions for the multiple-chain case can be defined as follows:

- Distribution seen by chain  $r$  arrivers

$$PA_{ir}(\mathbf{n}|N) = A_{ir}(\mathbf{n}) / A_{ir}, \quad \mathbf{0} \leq \mathbf{n} \leq N - e_r$$

- Distribution seen by chain  $r$  departers

$$PA_{ir}(\mathbf{n}|N) = D_{ir}(\mathbf{n} + e_r) / D_{ir}, \quad \mathbf{0} \leq \mathbf{n} \leq N - e_r$$

- Distribution seen by a random observer

$$P_i(\mathbf{n}|N) = T_i(\mathbf{n}) / T, \quad \mathbf{0} \leq \mathbf{n} \leq N$$

# Operational Analysis of Multiple-Chain Networks (continued)

- The throughput law holds, since

$$\lambda_{ir}(N) = \frac{D_{ir}}{T} = \frac{1}{T} \sum_{n=e_r}^N D_{ir}(\mathbf{n}) = \sum_{n=e_r}^N \frac{D_{ir}(\mathbf{n})}{T_i(\mathbf{n})} \frac{T_i(\mathbf{n})}{T} = \sum_{n=e_r}^N \mu_{ir}^*(\mathbf{n}) P_i(\mathbf{n} | N)$$

- Similarly, we can prove the arrival law

$$X_{ir}(N) = \frac{A_{ir}}{T} = \frac{1}{T} \sum_{n=0}^{N-e_r} A_{ir}(\mathbf{n}) = \sum_{n=0}^{N-e_r} \frac{A_{ir}(\mathbf{n})}{T_i(\mathbf{n})} \frac{T_i(\mathbf{n})}{T} = \sum_{n=0}^{N-e_r} Y_{ir}(\mathbf{n} | N) P_i(\mathbf{n} | N)$$