LESSON 11:
DELAY SYSTEMS

Objective:

The objective here is to learn Delay systems in teletraffic.

Introduction:

These telecommunication networks such as data networks places the call or message arrivals in a queue in the absence of resources and services them as and when resources become available. Servicing is not done until the resource becomes available. Delay systems are also called Lost Call Delayed (LCD) systems.

Delay systems are analyzed using queuing theory. Queuing theory is sometimes called waiting line theory. Queuing theory was used by early teletraffic researchers. But now-a-days theory is used for the analysis of a wide variety of applications outside telecommunications. We have following examples of delay line systems in telecommunications.

1. Message switching
2. Packet switching
3. Digit receiver access.
4. Automatic call distribution
5. Call processing

Delay Systems

The various elements of a queuing system are shown in Fig.11.1.

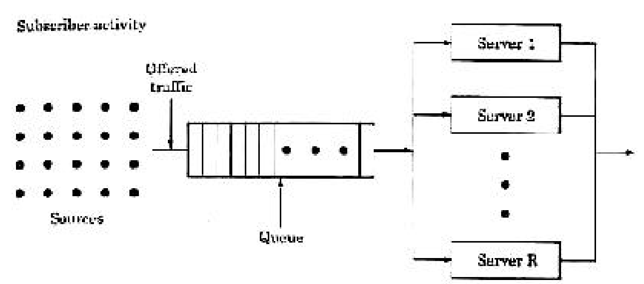

Fig.11.1Components or Elements of a queuing system.

There is a large population of sources, which generate traffic to the network. The queuing system provides a service facility that contains a number of identical servers. Each server is capable of providing the desired service to a request. When all the servers or links or trunks are busy, a request arriving at the network is placed in a queue until the server becomes available. In the analysis of queuing systems, we .

have to deal with a number of random variables. These variables are number of waiting requests, interarrival times between requests and time spent by a request in the system. Some of the important random variables associated with queuing system are given in Table 11.1 below.

S.No.	Random Variable	Notation
1	State of the system: It is also known as number of calls present in the system	K
2	Queuing length: It is defined as number of calls or requests waiting to be serviced	K_q
3	Number of calls in service	K_s
4	Mean wait time	t_q
5	Mean service time	t_k
6	Mean call arrival rate	R
7	Mean service rate	R'
8	Mean interarrival time	τ
9	Probability that there are K calls in the queuing system	P_K
10	Traffic Intensity = (R / R')	I_T
11	Server utilization = (I_T / K'_s)	U_s

Table 11.1

The number of requests present in the system or the state of the system is given by the sum of the requests in the queue and these are serviced. No request can be pending in the queue unless all the servers are busy.

The number of requests present in the queues or the state of the system = Number of requests in the queue + Number of requests being serviced

or $K = K_q + K_s$ …………Eq.1

The mean time a call spends in the system is the sum of the mean wait time t_q and mean service or holding time t_h

Mean time a call spends

= Mean wait time + Mean holding time

$= t_q + t_h$ ………..Eq.2

A queued operation in a delay system enables better utilization of servers or links or trunks than does a loss system. Queuing has the effect of smoothening out the traffic flow as far as servers are concerned. Peaks in call arrival process build up the queues length. If there is no loss of traffic then we need for infinite queuing capacity because there is no statistical limit on the number of call arrivals occurring in a short period of time. In a practical queuing system, only finite queue capacities are possible. Therefore, there is a blocking probability in delay systems, however, it may be small. Here, we assume that a delay system has infinite queue capacity in an operational sense. There is a necessary condition for its stable operation. This condition is given as

Mean call arrival rate < 1 or

Offered traffic < 1 Mean call service rate Number of servers

If above condition is not satisfied then queue length would become infinite sooner or later. Consequently, the system would never be able to clear the traffic offered to it. A queuing system is characterized by a set of 6 parameters. D.G Kendell has used concise 6-parameter notation to represent different types of queuing systems. He has used following 6-parameter notation A / B / c / K/ m/ Z. The parameter specifications are as follows:

A = call arrival process specification

B = Service time distribution

c = Number of servers

K = Queue Capacity

m = Number of sources (input population)

X = Service discipline

The value of the parameters K and m may be either a finite number or an infinite number. Queue discipline is the rule used for choosing the next customer to he serviced from the queue. Following queue disciplines are commonly used:

1. First – come – first – served (FCFS) selection
2. Random selection
3. Priority based selection.

The parameters K, m and Z may be omitted from queues specification, in which case they assume some default values.

For K and m, the default values are infinity.

The default queue descriptive is FCFS. The parameter c is a non – zero positive finite number. The parameters A and B may be assumed as anyone of the values given in Table 11.2, For example, the queue specification M / D/ 4 means a queues system with Poisson call arrival. Deterministic service time distribution, four servers, infinite queue capacity, infinite number of sources and FCFS queue discipline.

Values for A and B	Meaning	Remarks
GI	Call arrival process with general independent distribution for interarrival time	For A only
G	General service Time Distribution (No assumption considered)	For B only
E_R	Erlang – k interarrival or service time distribution	
M	Poisson arrival and exponential Service time distribution	M stands for Markov
D	Deterministic interarrival or service time distribution	It is a constant time distribution
H_K	Hyper exponential interarrival or service time distribution (with No. of stage k)	

Table :11.2. Different Values of Queues Parameter A and B

We can note that Kendell's notation can also be used to represent a loss system where the parameter K = 0.

Here, in analysis of delay system, we assume that infinite sources exist, infinite queuing capabilities exist, and the queue is serviced on FCFS basis. We also assume a Poisson call arrival process and service times are exponentially distributed or constant. These service time distributions represent the most random and the most deterministic service times. If a system that operates with some other distribution of service time then it has a performance somewhere between those produced by these two distributions. M/ M/ N_L and M/D/N_L queuing systems are used to model our telecommunication systems. Analysis of telecommunication systems is done to determine the probability distribution of waiting times and associated mean values. Often, the probability that waiting time exceeds a specified limit is of interest.

In particular, the probability that the waiting time is greater than zero represents the call congestion. Hence, it is of immediate interest.

The analysis of delay systems is started with a B-D process for call arrival and service rates, we have

R_k = R. for k = 0, 1, 2, 3

r_k = k r for k = 1, 2, 3, …… ,

N_L, r_k. = N_L r for K > N_L

The stability condition for the system is given by

$$\frac{R}{N_L r} < 1 \quad \text{or} \quad \frac{A}{N_L} < 1 \quad \ldots\ldots\ldots\ldots\text{Eq.3}$$

Many of the results obtained for the LCC model apply for LCD systems as well. The main difference between the LCC system and the M/ M/ N_L de1ay system is that the state of the M /M/ R system varies from 0 to ¥ whereas the loss system from 0 to N_L.

Any number of calls may enter a delay system and be serviced or be in the wait-stage. But no call can enter a loss system once all the servers are busy. Accordingly, we modify the following equation,

$$P_0 + P_1 + P_2 + \ldots + P_{NL} = 1$$

as,

$$P_0 + P_1 + P_2 + \ldots + P_{NL} + P_{NL+1} = 1 \quad \ldots\ldots\ldots\ldots\text{Eq.4}$$

Therefore from Eq.4 we get,

$$P_0 = 1 - \sum_{K=1}^{\infty} P_k \quad \ldots\ldots\ldots\ldots\text{Eq.5}$$

The loss system and delay system behave identically as long as the system is within the state N_L. Both the systems behave differently for states equal to or greater than state N_L. Substituting the birth and death rates of the M/ M/ N_L system in Eq.6 given below for k = N_L, and rearranging the terms, we have,

Relationship between Erlang C formula (or Erlang Second formula) and Erlang B formula (or Erlang first- Relation between Erlang formula):

Erlang C formula in terms of Erlang B formula. If we term the value of P_{NL} in Erlang B formula or Erlang first formula as P'_{NL} and P_{NL} in Erlang C formula or Erlang second formula as P''_{NL} then P''_{NL} is related to P'_{NL} as

$$P_{k-1}R_{k-1} + P_{k+1}r_{k+1} - (R_k + r_k)P_k = 0$$

for $k \geq 0$ Eq.6

$$N_L\, r\, P_{N_{L+1}} = (R + N_L r)\; P_{N_L} - R\, P_{N_{L-1}} \quad \text{...... Eq.7}$$

Using Eqn. $P_j = \dfrac{A^j P_0}{\underline{\angle j}}$ in Eqn.7

we get

$$N_L\, r\, P_{N_{L+1}} = (R + N_L r)\frac{A^{N_L}}{\underline{\angle N_L}}P_0 - R\frac{A^{N_L - 1}}{\underline{\angle N_L + 1}}P_0$$

where the value of P_0 is governed by Eqn. 5 Simplifying above Eqn., we get

$$N_L\, r\, P_{N_{L+1}} = \frac{RA^{N_L}}{\underline{\angle N_L}}P_0$$

or

$$P_{N_{L+1}} = \frac{RA^{N_L}P_0}{N_L\;\underline{\angle N_L}\; r} = \frac{R}{rN_L}\frac{A^{N_L}}{\underline{\angle N_L}}P_0$$

$$= \frac{A}{N_L}\frac{A^{N_L}}{\underline{\angle N_L}}P_0 = \frac{A}{N_L}P_{N_L} \quad \text{..... Eq.8}$$

Similarly, for $K = N_L + 1$, we get

$$P_{N_{L+2}} = \left(\frac{A}{N_L}\right)^2 P_{N_L} \quad \text{...... Eq.9}$$

Generalisation of above for $k > N_L$, we get

$$P_k = \left(\frac{A}{N_L}\right)^{k - N_L} P_{N_L} \quad \text{...... Eq.10}$$

Considering the inequality $\dfrac{A}{N_L} < 1$, we conclude that

$$\frac{A}{N_L} + \left(\frac{A}{N_L}\right)^2 + \left(\frac{A}{N_L}\right)^3 + \ldots = \frac{\left(\frac{A}{N_L}\right)}{1 - \left(\frac{A}{N_L}\right)} = \frac{A}{N_L - A} \quad \text{...... Eq.11}$$

From Eqn. $P_0 + P_1 + P_2 + \cdots + P_{N_L} - P_{N_{L+1}} = 1$
and Eqn. **11** we get

$$\sum_{k=0}^{N_L}\frac{A^k}{\underline{\angle k}}P_0 + \frac{A}{N_L - A}\frac{A^{N_L}}{\underline{\angle N_L}}P_0 = 1$$

Therefore,

$$P_0 = \frac{1}{\sum_{k=0}^{N_L}\frac{A^k}{\underline{\angle k}} + \frac{A^{N_L}}{\underline{\angle N_L}}\left(\frac{A}{N_L - A}\right)} \quad \text{..... Eq.12}$$

$$P_{N_L} = \frac{A^{N_L}}{\underline{\angle N_L}}P_0$$

$$= \frac{A^{N_L}}{\underline{\angle N_L}}\frac{1}{\sum_{k=0}^{N_L}\frac{A^k}{\underline{\angle k}} + \frac{A^{N_L}}{\underline{\angle N_L}}\left(\frac{A}{N_L - A}\right)} \quad \text{..... Eq.13}$$

Now the probability of a message being delayed is nothing. But the probability of finding the system in state N_L or above. This probability is given by

$$P(\text{delay} > 0) = \sum_{k=N_L}^{\infty} P_K \quad \text{......... Eq.14}$$

Therefore, we have

$$P(\text{delay} > 0) = \left[1 + \frac{A}{N_L - A}\right]P_{N_L} = \left(\frac{N_L}{N_L - A}\right)P_{N_L} \quad \text{...... Eq.15}$$

where P_{N_L} is obtained by substituting Eqn. 12 in Eqn. given below]

$$P_{N_L} = \frac{\frac{A^{N_L}}{\underline{\angle N_L}}}{1 + A + \frac{A^2}{\underline{\angle 2}} + \ldots + \frac{A^{N_L}}{\underline{\angle N_L}}}$$

Eqn. 15 in known as Erlang second formula or Erlang Delay formula or Erlang *C* formula

$$P''_{NL} = \frac{P'(N_L - A)}{N_L - A(1 - P'_{NL})} \quad \text{.........Eq.16}$$

Substituting Eq.16 in Eq.15 we get

$$P(\text{delay} > 0) = \left(\frac{N_L}{N_L - A}\right) P_{N_L} = \left(\frac{N_L}{N_L - A}\right) P''_{N_L}$$

$$= \left(\frac{N_L}{N_L - A}\right) \frac{P'_{N_L}(N_L - A)}{N_L - A(1 - P'_{N_L})}$$

or $$P(\text{delay} > 0) = \frac{N_L P'_{N_L}}{(N_L - A)(1 - P'_{N_L})} \qquad \text{-----Eq.17}$$

where, P'_{N_L} = the blocking probability given by Erlang B formula.

For $M/M/N_L$ queue, it can be shown that the waiting time distribution is exponential. It is given by

$$P(\text{delay} > t) = P(\text{delay} > 0)e^{\left[-(N_L - A)\frac{t}{t_h}\right]} \qquad \text{-----Eq.18}$$

By integrating above equation over time, the average waiting time can be determined as

$$t_q = \frac{P(\text{delay} > 0)t_h}{N_L - A} \qquad \text{-----Eq.19}$$

This equation is applied to all call arrivals to system, some of which enter the queue and others which do not. It is given that a message is already put in the queue.

Therefore, the average waiting time can be expressed as

$$t'_q = \frac{t_h}{N_L - A} \qquad \text{-------Eq.20}$$

Example 11.1: A public call office (PCO) is installed in a busy part of a city. This telephone booth has been used by 150 persons daily. The average holding time for a call is 1.5 minute. The public has suggested that they required another PCO in the same locality as the wait times are unduly long. Analyse the situation using M /M /l queue and find if the suggestion deserves serious consideration.

Solution:

Average call arrival

R = 0.104 Per minute

Offered load,

$A = R\, t_h = 0.104 \times 1.5 = 0.1560$

For $M/M/N_L$ Delay system, we have given number of servers or links or trunks N_L equal to one. So it is single server delay system.

We know that

$$P_{N_L} = \frac{A^{N_L}}{\underline{\angle N_L}} \frac{1}{\sum_{k=0}^{N_L} \frac{A^k}{\underline{\angle k}} + \frac{A^{N_L}}{\underline{\angle N_L}} \frac{A}{N_L - A}}$$

For single server case, above Eqn. can be written as

$$P_{N_L} = \frac{A^1}{\underline{\angle 1}} \frac{1}{\sum_{k=0}^{1} \frac{A^k}{\underline{\angle k}} + \frac{A^1}{\underline{\angle 1}} \frac{A}{1 - A}}$$

$$= A \frac{1}{\sum_{k=0}^{1} \frac{A^k}{\underline{\angle k}} + \frac{A^2}{(1 - A)}} = \frac{A}{\frac{A^0}{\underline{\angle 0}} + \frac{A^1}{\underline{\angle 1}} + \frac{A^2}{(1 - A)}}$$

$$= \frac{A}{1 + A + \frac{A^2}{(1 - A)}} = \frac{A}{\frac{1 - A^2 + A^2}{(1 - A)}} = A(1 - A)$$

or $$P_{N_L} = A(1 - A) \qquad ...(1)$$

From Eqn. given below

$$P(\text{delay} > 0) = \left[1 + \frac{A}{N_L - A}\right] P_{N_L} = \left(\frac{N_L}{N_L - A}\right) P_{N_L}$$

$$= \left(\frac{N_L}{N_L - A}\right) A(1 - A) = \left(\frac{1}{1 - 0.156}\right) 0.156(1 - 0.156)$$

$$= 0.156$$

There is really no reason for the public to complain. But the problem statement implies that the traffic is spread over throughout the day which may not be true. The busy hour traffic is important.