Lecture 20
Chapter # 6 (contd.)
FUNDAMENTALS OF QUEUING MODELS

Transient State & Steady State

▪ When a queuing system has recently begun operation,
- the state of the system (number of customers in the system) will be greatly affected by the initial state and by the time that has since elapsed.

▪ The system is said to be in a *transient condition.*
▪ However, after sufficient time has elapsed, the state of the system becomes essentially independent of the initial state and the elapsed time (except under unusual circumstances).
▪ The system has now essentially reached *a steady-state condition*, where the probability distribution of the state of the system remains the same (the steady-state or *stationary* distribution) over time.
▪ Queuing theory has tended to focus largely on the steady-state condition.

You can actually divide the system operation into two parts or into two states first is the transient state when the operation has just begun where system may show some abrupt behavior or it may be effected by the number of customers may be the arrival of customers is with very short interval of time & number of servers are less & there are more customers in the queue then in the server. May be bottle neck will appear while providing services but with the passage of time may be the number of customers are reduced & the system start establishing stable state.

Once the stable condition achieved we actually reached into the stable state condition. & we are more concerned about the steady state condition of our queueing system.

Steady State Notations

The following notation assumes that the system is in a steady-state condition:

P_n = probability of exactly n customers in queueing system.

L = expected number of customers in queueing system $= \sum_{n=0}^{\infty} nP_n$.

L_q = expected queue length (excludes customers being served) $= \sum_{n=s} (n-s)P_n$.

$\mathcal{W}$ = waiting time in system (includes service time) for each individual customer.

$W = E(\mathcal{W})$.

$\mathcal{W}_q$ = waiting time in queue (excludes service time) for each individual customer.

$W_q = E(\mathcal{W}_q)$.

M / M / 1 QUEUE ANALYSIS

M/M/1 means that inter-arrival time and the service time are exponentially distributed with Markovian or Memoryless property & there's just a single server in the system. The analysis can be perform by means of

THE BIRTH-AND-DEATH PROCESS

We actually begun our analysis of a single queue system by assuming that all jobs arrives at the queue one at a time & this assumption specifically says that groups of jobs cannot arrive as a single batch all at the same time.

With this assumption we can describe the state of the system using single integer N let suppose which is the total number of jobs in the system & it includes both the jobs in the queue & jobs being serviced.

▪ Most elementary queuing models assume that the inputs (arriving customers) and outputs (leaving customers) of the queuing system occur according to the *birth-and-death process*.

▪ The state of the system at time t ($t \geq 0$), denoted by N(t),

- number of customers in the queuing system at time t.

▪ The birth-and-death process describes probabilistically how N(t) changes as t increases.

▪ Broadly speaking, it says that *individual* births and deaths occur *randomly*, where their mean occurrence rates depend only upon the current state of the system.

Since we are using Markovian distribution the birth & death of processes will be totally random.

What is birth & death process?

The arrival of a new job into a system is called Birth

& whenever a server completes a job it's called the Death of the process.

▪ The next transition in the state of the process is either

- $n \rightarrow n + 1$ (a single birth)
  or
- $n \rightarrow n - 1$ (a single death),
- depending on whether the former or latter random variable is smaller.

▪ Because of these assumptions, the birth-and-death process is a special type of *continuous time Markov chain*.

Also the basic assumption in this analysis suggests that the next state depends only upon the current state & it is also independent of how long the system has been in the current state. So it is the special type of the Markov chain.

▪ Queuing models that can be represented by a continuous time Markov chain are far more tractable analytically than any other.

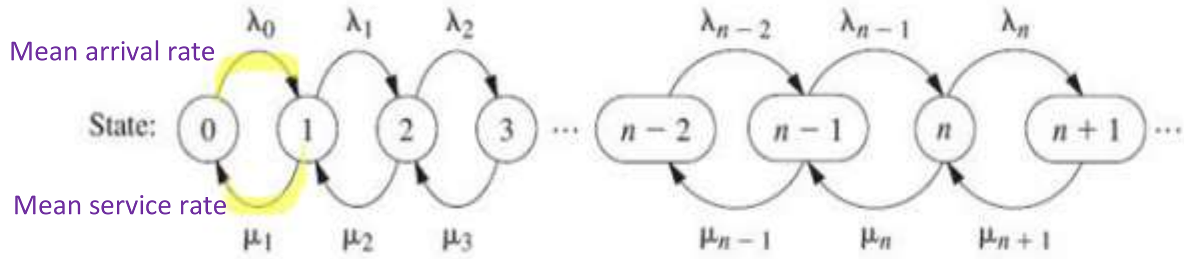

Fig 1: Rate diagram for the birth-and-death process

▪ The arrows in this diagram show the only possible *transitions* in the state of the system. Transition is only possible with the neighboring state.
▪ The entry for each arrow gives the mean rate for that transition.
In the steady state, the average rate at which the system enters at some state n must be equal to the average rate at which the system moves out of that particular state n & we can see that from the figure. It means the rate at which the system's state moves to the right across the line must be equal to the rate it moves back to the left.
If this for not true, then the number of jobs in the system would grow without any bound. & this steady state behavior of the birth and death process is often referred to as the conversation of flow.
▪ These results yield the following key principle:

Rate In = Rate Out Principle

▪ For any state of the system n (n = 0, 1, 2, ...),

- mean entering rate = mean leaving rate.

From the fig; every state has been given an integer number and the number given to the states actually indicates the total number of jobs in the system as well.
& the transitions are allowed only between the adjacent state. It shows the one step behavior; only one customer can arrive at a time.
Moreover, it shows a balance flow as well.
▪ The equation expressing this principle is called the *balance equation* for state n.
▪ After constructing the balance equations for all the states in terms of the unknown P_n probabilities,

- we can solve this system of equations (plus an equation stating that the probabilities must sum to 1) to find these probabilities.

P_n = probability that the system is in state n.

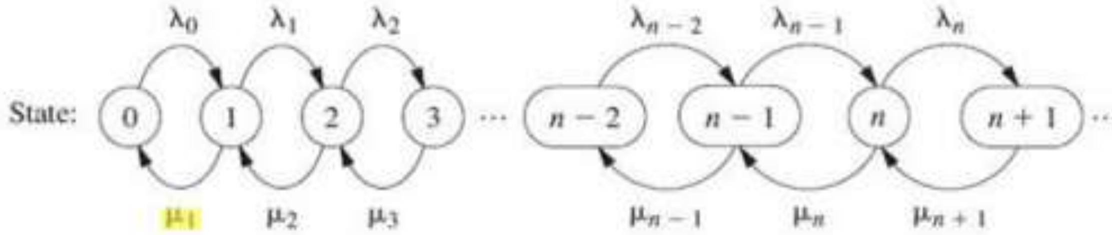

Fig 1: Rate diagram for the birth-and-death process

▪ To illustrate a balance equation, consider state 0. The process enters this state *only* from state 1.
▪ Given that the process is in state 1, the mean rate of entering state 0 is μ_1.
▪ From any *other* state, this mean rate is 0.
▪ Therefore, the overall mean rate at which the process leaves its current state to enter state 0 (the *mean entering rate*) is
the eq to enter state 0;

$$\mu_1 P_1 + 0(1 - P_1) = \mu_1 P_1$$

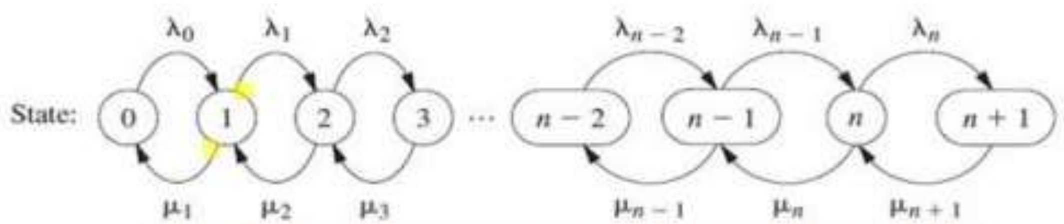

Fig 1: Rate diagram for the birth-and-death process

• By the same reasoning, the *mean leaving rate* must be $\lambda_0 P_0$, so the balance equation for state 0 is cause only transition of state 0 to state 1

$$\mu_1 P_1 = \lambda_0 P_0$$

rate of leaving = rate of entering

• For every other state there are two possible transitions both into and out of the state.
• Therefore, each side of the balance equations for these states represents the *sum* of the mean rates for the two transitions involved.
• Otherwise, the reasoning is just the same as for state 0.

<u>Writing Flow Equations</u>

Mean arrival rate and the mean leaving rate are equal.

$\lambda_0=\lambda_1=\lambda_2=\ldots=\lambda_n=\lambda$

$\mu_0=\mu_1=\mu_2=\mu_n=\mu$

$\lambda/\mu=\rho$

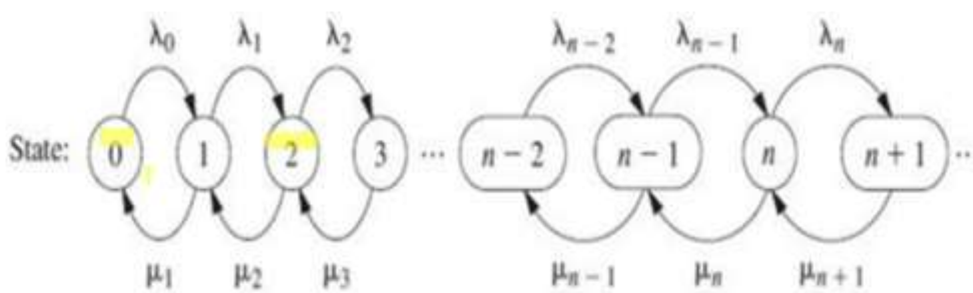

Fig 1: Rate diagram for the birth-and-death process

State 0:

$\lambda P_0 = \mu P_1$

$\Rightarrow P_1 = \rho P_0$

State 1:

$\lambda P_0 + \mu P_2 = \lambda P_1 + \mu P_1$

$\Rightarrow \mu P_2 = (\lambda + \mu)\frac{\lambda}{\mu} P_0 - \lambda P_0$

$= [\frac{\lambda}{\mu} + 1 - 1]\ \lambda P_0$

$\therefore \mathbf{P_2 = \rho^2 P_0}$

<u>Writing Flow Equations</u>

So, from these two equations, we can generalize the relation:

$$P_n = \rho^n P_0$$

$\because \sum_{n=0}^{\infty} P_n = 1$

$\Rightarrow P_0 + \rho P_0 + \rho^2 P_0 + \ldots\ldots. + \infty = 1$

$\Rightarrow \boldsymbol{P_0}(\frac{1}{1-\boldsymbol{\rho}}) = 1$

$\therefore P_0 = 1 - \rho$

Q1: Explain what does $\rho = 1 - P_0$ indicate?

ρ is the probability that server is busy and $1 - P_0$ suggests the probability of at least one customer.

Generalizing, we will get:

$$\therefore \mathbf{P_n = \rho^n (1 - \rho)}$$