MODULE 6: General Markovian Queueing Systems

LECTURE 20

Queues with Bulk Arrivals

General Markovian Queueing Systems

- We have considered so far Markovian queueing systems that can be modelled by a BDP.
  ♦ In a birth-death queueing model, the arrivals occur singly and the customers are served one at a time.
  ♦ These were analytically tractable easily and made our life easier.
- We will now consider queueing systems that can still be modelled by a CTMC (but by a non-birth-death process).
- The models are Markovian (Markovian property is retained) and allow transitions beyond nearest neighbours.
- The Chapman-Kolmogorov equations, backward and forward Kolmogorov equations, (global) balance equations are all still valid.
  ▶ The essence of the approach to solving these non-birth-death Markovian models remains the same (though the complexity might increase).
- From here onwards, we consider only queueing systems in equilibrium (and hence will not worry about the time-dependent behaviour).

Bulk Queues

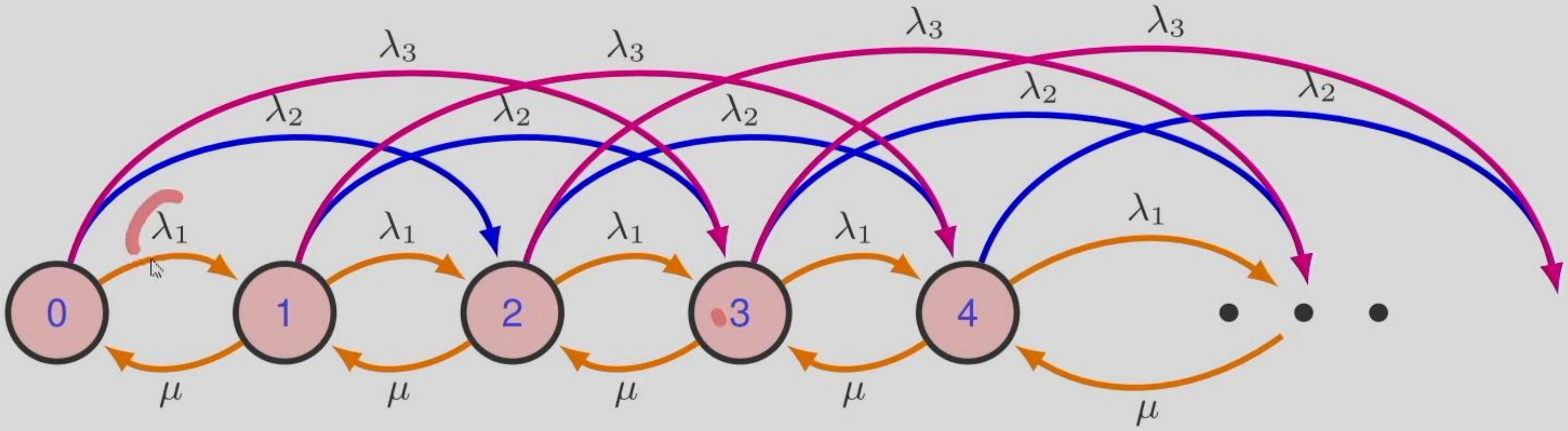
- There are many systems in which the arrivals can occur and services can be performed in groups (i.e., in bulk or in batches).
  - ▼ When a train or bus arrives (or leaves), people arrive (or leave) in groups.
  - ▼ People may go to a restaurant together and may be served in batches.
  - ▼ Lifts or elevators or boats handle passengers in batches.

  These type of systems are called as **bulk queues**.
- Bailey (1954) introduced the concept of bulk service in queues and Gaver (1959) introduced bulk arrival queues. The literature is now quite vast.
- We consider bulk queues that can be modelled by a non-birth-death but still a Markov process (i.e, CTMC).
- Specifically, we consider systems where either the arrivals are in batches or service is provided for batches of customers.

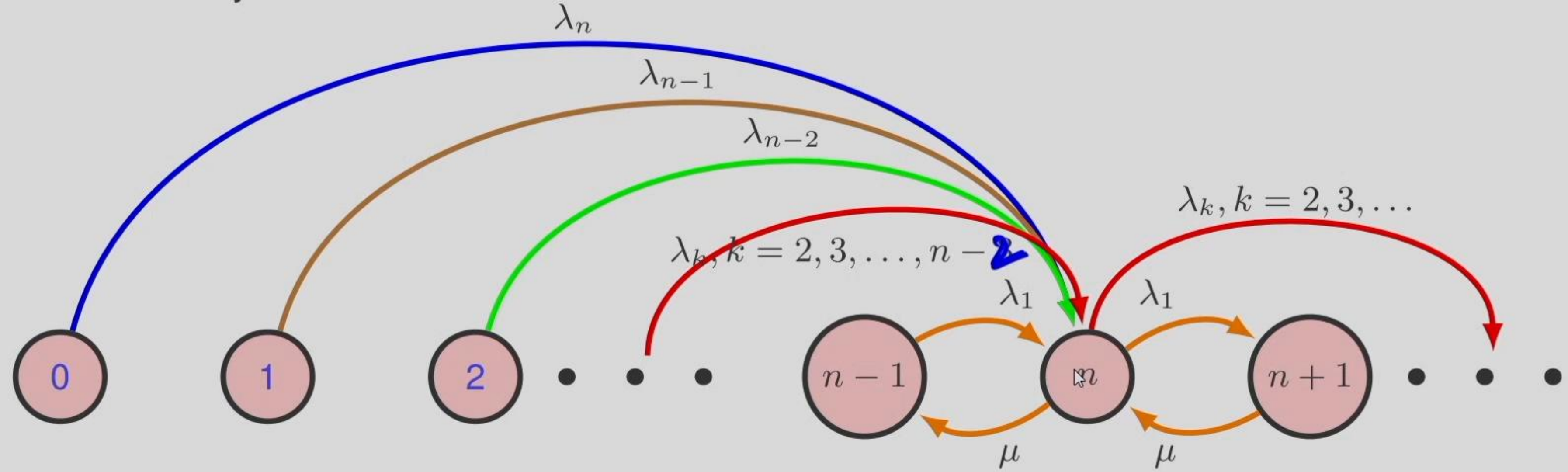
Bulk Input (or Arrival) Queues ($M^{[X]}/M/1$ Queues)

- Let us consider a system with all the underlying assumptions as that of an $M/M/1$ system. Further, assume, in addition to the assumption that the arrival stream follows a Poisson process, that the number of customers in any arriving module is a random variable X with PMF $\{c_n\}_{n\geq 1}$.
- The system described above is denoted by $M^{[X]}/M/1$.
  ♦ $c_1 = P\{X = 1\} = 1$ will give the usual $M/M/1$ system.
- The total arrival process is a compound Poisson process (with X denoting the batch size).
  - If λ_n is the arrival rate of a Poisson process of batches of size n, then clearly $c_n = \dfrac{\lambda_n}{\lambda}$, where λ is the composite arrival rate of all batches, i.e., $\lambda = \sum_{n=1}^{\infty} \lambda_n$.
  - The total arrival process then arises from the overlap of the set of Poisson processes with rates $\lambda_n, n = 1, 2, \ldots$.
- We, as usual, consider the number N of customers in the system in steady-state and the underlying process is still Markovian.

Sample Rate Diagram: The following diagram gives the transition rates for the simple case of $M^{[X]}/M/1$ queue, where X takes the value either 1 or 2 or 3.



- General Case: The transition rate diagram only with respect to a general state n, depicting the transitions only into and out of state n.



- For arbitrary batch size distribution, the balance equations are given by

$$\lambda p_0 = \mu p_1$$

$$(\lambda + \mu) p_n = \mu p_{n+1} + \lambda \sum_{k=1}^{n} p_{n-k} c_k, \qquad n \geq 1.$$

The last term in the second equation says that a total of n in the system can arise from the presence of $n - k$ in the system followed by an arrival of a batch size k.

- We will use the generating function approach to solve the system of equations. Define

$$C(z) = \sum_{n=1}^{\infty} c_n z_n \quad |z| \leq 1 \tag{1}$$

$$P(z) = \sum_{n=0}^{\infty} p_n z_n \quad |z| \leq 1 \tag{2}$$

as the generating functions of the batch-size probabilities $\{c_n\}$ and the steady-state probabilities $\{p_n\}$, respectively.

- $C(z)$ is known and hence is an input. We need to solve for p_n's.

- Multiplying each equation by z^n and summing over n, we have

$$\lambda \sum_{n=0}^{\infty} p_n z^n + \mu \sum_{n=1}^{\infty} p_n z^n = \frac{\mu}{z} \sum_{n=1}^{\infty} p_n z^n + \lambda \sum_{n=1}^{\infty} \sum_{k=1}^{n} p_{n-k} c_k z^n$$

- Now, observe that $\sum_{k=1}^{n} p_{n-k} c_k$ is the probability mass function for the sum of the steady-state system size and batch size. And, the PGF of this sum is the product of the respective PGFs.

$$\text{i.e.,} \quad \sum_{n=1}^{\infty} \sum_{k=1}^{n} p_{n-k} c_k z^n = \sum_{k=1}^{\infty} c_k z^k \sum_{n=k}^{\infty} p_{n-k} z^{n-k} = C(z)P(z).$$

- Substituting in the above,

$$\lambda P(z) + \mu[P(z) - p_0] = \frac{\mu}{z}[P(z) - p_0] + \lambda C(z) P(z)$$

$$\implies P(z) = \frac{\mu(1-z)p_0}{\mu(1-z) - \lambda z[1 - C(z)]}, \qquad |z| \le 1.$$

  - ▶ We need to find p_0 using $P(1) = 1$.

- Rewriting the $P(z)$ above,

$$P(z) = \frac{p_0}{1 - (\lambda/\mu)z\overline{C}(z)}, \qquad \text{where } \overline{C}(z) = \frac{1 - C(z)}{1 - z}.$$

■ Observe that $\overline{C}(z)$ is the generating function of the complementary batch-size probabilities $\overline{C}_n = P\{X > n\} = 1 - P\{X \leq n\}$ (Work it out, if you wish to understand this!).

- Then $1 = P(1) = \dfrac{p_0}{1 - (\lambda/\mu)\overline{C}(1)}$ and $\overline{C}(1)$ can be found from $\overline{C}(z)$, using L'Hopital rule (once).

We then get $\overline{C}(1) = \sum_{n=0}^{\infty} P(X > n) = E[X]$. Therefore,

$$p_0 = 1 - (\lambda/\mu)E[X] = 1 - \rho, \quad \left(\rho = \frac{\lambda E[X]}{\mu}\right)$$

and hence

$$P(z) = \frac{1 - \rho}{1 - (\lambda/\mu)z\overline{C}(z)}, \qquad |z| \leq 1.$$

We can then find p_n's from the $P(z)$ above.

♦ And, $\rho < 1$ is the necessary and sufficient condition for the steady state to exist.

Performance Measures

- We first find L from $P(z)$ as

$$L = P'(1) = (1-\rho)(\lambda/\mu)\frac{\overline{C}(1) + \overline{C'}(1)}{(1-(\lambda/\mu)\overline{C}(1))^2} = \frac{(\lambda/\mu)(E(X) + \overline{C'}(1))}{1-\rho}.$$

Now, $\overline{C'}(1)$ can be found from $\overline{C}(z)$, using L'Hopital rule (twice) to $\overline{C'}(z) = [1 - C(z) - (1-z)C'(z)]/(1-z)^2$ as $\overline{C'}(1) = C''(1)/2 = E[X(X-1)/2]$. Therefore,

$$L = \frac{(\lambda/\mu)(E[X] + E[X^2])}{2(1-\rho)} = \frac{\rho + (\lambda/\mu)E[X^2]}{2(1-\rho)} = \frac{\rho}{1-\rho}\left[\frac{E(X) + E(X^2)}{2E(X)}\right], \quad \left(\rho = \frac{\lambda E[X]}{\mu}\right)$$

- Note that the arrival rate of customers is $\lambda E(X)$ and hence the other performance measures using Little's law are given by

$$W = \frac{L}{\lambda E[X]}, \quad W_q = W - \frac{1}{\mu}, \text{ and } L_q = \lambda E[X] W_q = L - \rho.$$

06_MarkovianQ.pdf

10 / 50

Example (M/M/1 Queue)

If $c_1 = P\{X = 1\} = 1$ and $c_n = 0$ for $n > 1$, then we obtain the corresponding results for an $M/M/1$ queue.

Example (Constant Batch Size)

- If, for some $K > 1$, $c_K = P\{X = K\} = 1$ and $c_n = 0$ for $n \neq K$, then

$$L = \frac{\rho + K\rho}{2(1-\rho)} = \frac{K+1}{2}\frac{\rho}{1-\rho}, \qquad (\rho = \lambda K/\mu)$$

which is equal to the M/M/1 mean system size multiplied by $(K+1)/2$. Also,

$$L_q = L - \rho = \frac{2\rho^2 + (K-1)\rho}{2(1-\rho)}.$$

The inversion of $P(z)$ to get p_n's is reasonably easy for small values of K.

06_MarkovianQ.pdf *

Create account Login

Example (Geometric Batch Size)

Let $X \sim Geo(\alpha)$, i.e., $c_n = (1-\alpha)\alpha^{n-1}$, $n = 1, 2, \ldots, 0 < \alpha < 1$.

Then $\rho = \lambda E[X]/\mu = (\lambda/\mu)/(1-\alpha)$ and $C(z) = (1-\alpha)\sum_{n=1}^{\infty}\alpha^{n-1}z^n = \dfrac{z(1-\alpha)}{1-\alpha z}$.

$$\begin{aligned}
\implies P(z) &= \frac{(1-\rho)(1-z)}{1-z-(\lambda/\mu)z[1-C(z)]} = \frac{(1-\rho)(1-z)}{1-z-(\lambda/\mu)z[1-z\frac{(1-\alpha)}{(1-\alpha z)}]} \\
&= \frac{(1-\rho)(1-\alpha z)}{1-z[\alpha+(\lambda/\mu)]} \\
&= (1-\rho)\left(\frac{1}{1-z[\alpha+(\lambda/\mu)]} - \frac{\alpha z}{1-z[\alpha+(\lambda/\mu)]}\right).
\end{aligned}$$

Utilizing the formula for the sum of geometric series, we get

$$\begin{aligned}
p_n &= (1-\rho)\{[\alpha+(\lambda/\mu)]^n - \alpha[\alpha+(\lambda/\mu)]^{n-1}\} \\
&= (1-\rho)(\lambda/\mu)[\alpha+(\lambda/\mu)]^{n-1}, \qquad n > 0
\end{aligned}$$

202%

Example (A machine-line production system)

- Consider a multistage machine-line process that produces an assembly in quantity.
- After the first stage, many items are found to have one or more defects, which must be repaired before they enter the second stage and this job of making the adjustment is done by one worker.
- It is observed that the number of defects is 1 or 2 (most of the times).
- The interarrival times for units with one defect $\sim Exp(\lambda_1)$ and with two defects $\sim Exp(\lambda_2)$, where $\lambda_1 = 1/h$ and $\lambda_2 = 2/h$.
- The worker's service time distribution is $Exp(\mu)$, where $1/\mu = 10$ minutes.
- Subsequently, it was observed that the rate of defects have increased, and it was decided to put an additional worker who who will handle exclusively the 2-defect items, while the original worker will handle only 1-defect items.
- When to add the additional worker will be based on a cost analysis, especially based on L.

Example (contd...)

- We will find the average number of units in the system, L, under the assumption of only two possible batch sizes, from the given data.

$$\lambda = \lambda_1 + \lambda_2 = 3, \quad \mu = 6, \quad c_1 = \frac{\lambda_1}{\lambda} = \frac{1}{3}, \quad c_2 = \frac{\lambda_2}{\lambda} = \frac{2}{3}$$

$$E[X] = \frac{1}{3} + 2 \times \frac{2}{3} = \frac{5}{3}, \qquad E[X^2] = \frac{1}{3} + 2^2 \times \frac{2}{3} = 3$$

$$\rho = \lambda \frac{E[X]}{\mu} = \frac{5}{6}, \qquad L = \frac{\rho + (\lambda/\mu)E[X^2]}{2(1-\rho)} = \frac{\frac{5}{6} + \frac{3}{2}}{2(1 - \frac{5}{6})} = 7.$$

- Extra: If needed, we can get p_n's from $P(z) = \dfrac{1}{6 - 3z - 2z^2}$ as

$$p_n = (0.116)(0.880)^n + (0.050)(-0.379)^n, n \geq 0.$$

Example (contd...)

Cost Analysis:

- C_1 = The cost per unit time per waiting repair.
- C_2 = The cost of a worker per unit time.

- When there is only one worker, the system is an $M^{[X]}/M/1$ system with the assumption of only two possible batch sizes (1 and 2) and L is computed as above.
  Then the expected cost per unit time of the single-server system is $C = C_1 L + C_2$
- If a second worker is added, then there is an additional C_2 cost, and we now have two queues:
  - The singlet line is $M/M/1$ and $L_1 = \dfrac{\frac{\lambda_1}{\mu}}{1 - \frac{\lambda_1}{\mu}}$
  - The doublet line is $M^{[X]}/M/1$ with fixed batch size of 2 and $L_2 = \dfrac{3\frac{\lambda_2}{\mu}}{1 - 2\frac{\lambda_2}{\mu}}$.
  - Hence, the average number of units in the system is

$$L = L_1 + L_2 = \frac{\frac{\lambda_1}{\mu}}{1 - \frac{\lambda_1}{\mu}} + \frac{3\frac{\lambda_2}{\mu}}{1 - 2\frac{\lambda_2}{\mu}}.$$

06_MarkovianQ.pdf *

Create account Login

Example (contd...)

- Therefore, with the second worker, the new expected cost is

$$C^* = C_1\left(\frac{\frac{\lambda_1}{\mu}}{1-\frac{\lambda_1}{\mu}} + \frac{3\frac{\lambda_2}{\mu}}{1-2\frac{\lambda_2}{\mu}}\right) + 2C_2.$$

- Decision is based on the comparative magnitude of C and C^*. An additional worker is added whenever $C^* < C$ or

$$C_1\left(\frac{\frac{\lambda_1}{\mu}}{1-\frac{\lambda_1}{\mu}} + \frac{3\frac{\lambda_2}{\mu}}{1-2\frac{\lambda_2}{\mu}}\right) + 2C_2 < C_1 L + C_2$$

$$\Rightarrow C_1\left(\frac{\frac{\lambda_1}{\mu}}{1-\frac{\lambda_1}{\mu}} + \frac{3\frac{\lambda_2}{\mu}}{1-2\frac{\lambda_2}{\mu}}\right) + C_2 < C_1 L = C_1\left(\frac{\rho + (\lambda/\mu)E[X^2]}{2(1-\rho)}\right) = C_1\left(\frac{\frac{\lambda_1}{\mu} + 3\frac{\lambda_2}{\mu}}{1-\frac{\lambda_1}{\mu}-2\frac{\lambda_2}{\mu}}\right)$$

06_MarkovianQ.pdf *

Create account Login

Example (contd...)

- That is

$$C_2 < C_1 \left(\frac{\frac{\lambda_1}{\mu} + 3\frac{\lambda_2}{\mu}}{1 - \frac{\lambda_1}{\mu} - 2\frac{\lambda_2}{\mu}} - \frac{\frac{\lambda_1}{\mu}}{1 - \frac{\lambda_1}{\mu}} - \frac{3\frac{\lambda_2}{\mu}}{1 - 2\frac{\lambda_2}{\mu}} \right)$$

- Using the values of the parameters, we arrive at a decision criterion of

$$C_2 < \frac{19C_1}{5}.$$