Lecture 6: Queueing Theory III

Lecturer: Süleyman Kerimov *Date: February 17, 2026*

Disclaimer: *These notes are primarily adapted from expositional texts, including work by Jyotiprasad Medhi, Karl Sigman, and János Sztrik. These notes are not meant to be complete or fully rigorous; some proofs are not given, incomplete, or only outlined, as they are discussed in class.*

Discussion 6.1. *Continuing our discussion from Lecture 5 on priorities (Discussion 5.4), now let's consider a non-preemptive case, where we have an M/M/1 queue with two types of customers, but now the arrival of a type 1 customer does not disrupt the service of a type 2 customer. After the server is done serving a type 2 customer, we start serving type 1 customers if there are any. Find the average number of customers and average waiting times of both types.*

6.1 Jackson Networks

Jackson's network model is defined as follows. Assume that customers from one node (queueing system) i proceed to an arbitrary node, and new customers may arrive to a node from outside (say customers arrive to node i according to a Poisson process with rate λ_i). Suppose that there are k nodes, where the ith node ($i = 1, \dots, k$) consists of c_i exponential servers with parameter μ_i (that is, each node contains a $M/M/c$ queueing system). Customers after receiving service at the ith node proceed to the jth node with probability p_{ij}.

Customers at node i depart from the system with probability

$$q_i = 1 - \sum_{j=1}^{k} p_{ij}.$$

Consider Jackson's general network model with k nodes. The arrivals can be categorized into two groups: the external arrivals (with rate λ_i) and internal arrivals (with rate $\sum_{j=1}^{k} p_{ji}\lambda_j$). Therefore, the effective arrival rate to node i (or the effective rate of flow through node i) is

$$\alpha_i = \lambda_i + \sum_{j=1}^{k} p_{ji}\alpha_j, \qquad i = 1, 2, \dots, k; \tag{6.1}$$

where these equations are also referred as traffic (flow balance, conservation, etc.) equations.

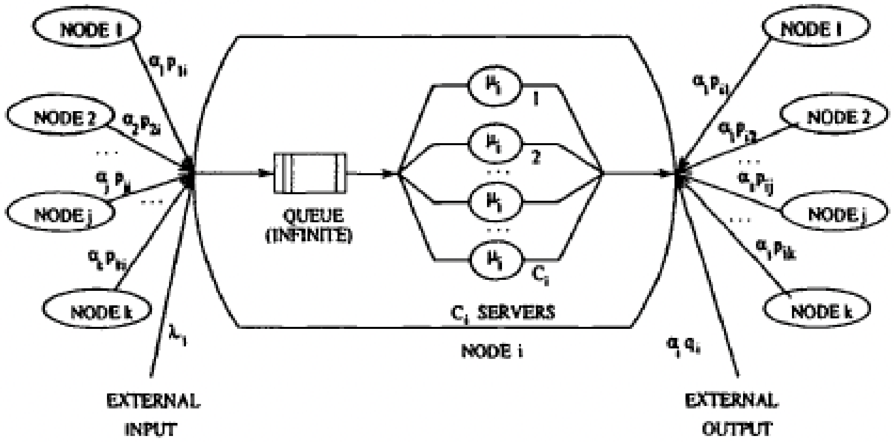

Figure 5.1 Node i in a Jackson Network.

Theorem 6.2 (Jackson's Theorem)**.** *Let $(n_1, n_2, \ldots, n_k)$ denote the state of the complete system in which there are n_i (in the queue and in service) at node i in a Jackson network of Markovian queues in equilibrium, and let $p(n_1, \ldots, n_k)$ be the probability that the system is in the state $(n_1, \ldots, n_k)$.*

Assume that

$$\rho_i = \frac{\alpha_i}{\mu_i} < 1, \qquad i = 1, 2, \ldots, k,$$

where $\{\alpha_i\}$ are given by the balance equations

$$\alpha_i = \lambda_i + \sum_{j=1}^{k} \alpha_j p_{ji}, \qquad i = 1, 2, \ldots, k. \tag{6.2}$$

If $p_i(n)$ denotes the probability that there are n customers in the system (in queue plus service) for the $M/M/c_i$ queue with input rate α_i and service rate μ_i for each of the c_i servers, i.e.,

$$p_i(n) = p_i(0)\,\frac{\left(\dfrac{\alpha_i}{\mu_i}\right)^n}{n!}, \qquad n = 0, 1, 2, \ldots, c_i, \tag{6.3}$$

$$= p_i(0)\,\frac{\left(\dfrac{\alpha_i}{\mu_i}\right)^n}{c_i!\; c_i^{\,n-c_i}}, \qquad n = c_i + 1, \ldots, \tag{6.4}$$

then we have

$$p(n_1, n_2, \ldots, n_k) = p_1(n_1)\, p_2(n_2) \cdots p_k(n_k). \tag{6.5}$$

Proof Sketch. Let $p_t(n_1, \ldots, n_k)$ be the probability that the complete system is in state $(n_1, \ldots, n_k)$ at time t.

Let

$$q_i = 1 - \sum_j p_{ij}, \qquad a_i(n) = \min\{n_i, c_i\} = \begin{cases} n_i, & \text{if } n < c_i, \\ c_i, & \text{if } n \geq c_i, \end{cases}$$

$$\delta_i = \min\{n_i, 1\} = \begin{cases} 1, & n_i \geq 1, \\ 0, & n_i = 0. \end{cases}$$

Our goal is to write the differential equations satisfied by p_t. Therefore, we will consider the state changes in an infinitesimal interval $(t, t+h)$ following the interval $(0,t)$. p_t. Consider the following four mutually exclusive ways to move from t to $t+h$:

(A) State at t is $(n_1, \ldots, n_k)$ and there are no arrivals or departures occur to or from any node externally. We get

$$\Pr(A) = p_t(n_1, \ldots, n_k) \left[1 - \left(\sum_i \lambda_i \right) h - \sum_i a_i(n_i) \mu_i h \right] + o(h). \tag{6.6}$$

(B) State at t is $(n_1, \ldots, n_i + 1, \ldots, n_k)$ and there is one service completion at i in $(t, t+h)$, and this completion departs from the system (with probability q_i). We get

$$\Pr(B) = \sum_{i=1}^{k} p_t(n_1, \ldots, n_i + 1, \ldots, n_k) \left[a_i(n_i + 1) \mu_i q_i \right] h + o(h). \tag{6.7}$$

(C) State at t is $(n_1, \ldots, n_i - 1, \ldots, n_k)$ and there is one arrival from the external source to node i in the interval $(t, t+h)$. We get

$$\Pr(C) = \sum_{i=1}^{k} p_t(n_1, \ldots, n_i - 1, \ldots, n_k) \left[\lambda_i h \delta_i \right] + o(h). \tag{6.8}$$

(D) State at t is $(n_1, \ldots, n_i + 1, \ldots, n_j - 1, \ldots, n_k)$: there is one service completion at node i in $(t, t+h)$, and the one whose service is completed moves to node j with probability p_{ij}. Thus,

$$\Pr(D) = \sum_i \sum_j p_t(n_1, \ldots, n_i + 1, \ldots, n_j - 1, \ldots, n_k) \left[a_i(n_i + 1) \mu_i h p_{ij} \right] + o(h). \tag{6.9}$$

Merging all cases, we have

$$p_{t+h}(n_1, \ldots, n_k) = \Pr(A) + \Pr(B) + \Pr(C) + \Pr(D). \tag{6.10}$$

(6.10) can be written as

$$p_{t+h}(n_1,\ldots,n_k) - \Pr(A) = \Pr(B) + \Pr(C) + \Pr(D). \tag{6.11}$$

$$p_{t+h}(n_1,\ldots,n_k) - p_t(n_1,\ldots,n_k)\left[1 - \left(\sum_i \lambda_i\right)h - \sum_i a_i(n_i)\mu_i h\right] - o(h) = \Pr(B) + \Pr(C) + \Pr(D). \tag{6.12}$$

Taking the limit as $h \to 0$ and solving for $p'(t) = 0$ gives the equations satisfied by steady-state probabilities:

$$\begin{aligned}
\left[\sum_i \lambda_i + \sum_i a_i(n_i)\mu_i\right] p(n_1,\ldots,n_k) &= \sum_i a_i(n_i+1)\mu_i q_i\, p(n_1,\ldots,n_i+1,\ldots,n_k) \\
&+ \sum_i \lambda_i \delta_i\, p(n_1,\ldots,n_i-1,\ldots,n_k) \\
&+ \sum_i \sum_j a_i(n_i+1)\mu_i p_{ij}\, p(n_1,\ldots,n_i+1,\ldots,n_j-1,\ldots,n_k).
\end{aligned}$$

Finally, one can show that (6.5) uniquely satisfies the equations above.

■

6.2 The M/G/1 Model

Now assume that we have a Poisson arrival process with rate λ, the service times are i.i.d. and follow a general distribution with mean $\mathbb{E}[S] = \frac{1}{\mu}$, and there is a single server. Again, for stability, we assume that $\rho = \frac{\lambda}{\mu} < 1$.

Let R be the residual service time and let P_k denote the probability that there are k customers in the system in the steady-state. By PASTA property, we have

$$\begin{aligned}
W_q &= \sum_{k=1}^{\infty} \left(E(R) + (k-1)E(S)\right) P_k \\
&= \sum_{k=1}^{\infty} E(R)P_k + \left(\sum_{k=1}^{\infty} (k-1)P_k\right) E(S) \\
&= E(R)\rho + L_q E(S).
\end{aligned}$$

where the equation follows since $1 - \rho = P_0$. By Little's Law, we get

$$W_q = \frac{\rho \mathbb{E}[R]}{1-\rho} \tag{6.13}$$

which is known as the Pollaczek-Khintchine mean value formula. So now we have to characterize $\mathbb{E}[R]$.

Proposition 6.3. *We have* $\mathbb{E}[R] = \frac{\mathbb{E}[S^2]}{2\mathbb{E}[S]}$.

There are various ways to prove this, but the most pedagogical one is via renewal reward theorem (see the next section). From Proposition 6.3, we have

$$E(R) = \frac{E(S^2)}{2E(S)} = \frac{\mathrm{Var}(S) + E[S]^2}{2E(S)} = \frac{1}{2}\left(C_S^2 + 1\right) E(S), \tag{2.33}$$

where C_S^2 is the squared coefficient of the service time S. From here, we get

Thus for the mean waiting time we have

$$W_q = \frac{\rho E(R)}{1-\rho} = \frac{\rho}{2(1-\rho)}\left(C_S^2 + 1\right) E(S).$$

And by Little's law, we get

$$L_q = \frac{\rho^2}{1-\rho} \frac{C_S^2 + 1}{2}.$$

Discussion 6.4. *Kingman's G/G/N formula:*

$$W \approx \frac{1}{\mu N} \cdot \frac{\rho^{\sqrt{2(N+1)}-1}}{1-\rho} \cdot \frac{C_A^2 + C_S^2}{2} + \frac{1}{\mu}.$$

Discussion 6.5. *Matching queues, and the relationship between regret and queue-lengths.*

6.3 The Renewal Reward Theorem

Definition 6.6. *A random point process $\psi = \{t_n\}$ for which the (non-negative) interarrival times $X_n = t_n - t_{n-1}$, $n \geq 1$, form an i.i.d. sequence is called a* renewal process.

Following the definition, t_n is called the nth *renewal epoch* and $F(x) := P(X \leq x)$, $x \geq 0$, denotes the common inter-arrival time distribution. $t_n = X_1 + \cdots + X_n$, and $N(t) = \max\{n : t_n \leq t\}$ is the counting process. The rate of the renewal process is denoted by $\lambda \triangleq 1/\mathbb{E}[X]$.

Theorem 6.7 (Elementary renewal theorem)**.** *For a renewal process,*

$$\lim_{t\to\infty} \frac{N(t)}{t} = \lambda \quad a.s.$$

and

$$\lim_{t\to\infty} \frac{\mathbb{E}[N(t)]}{t} = \lambda.$$

Proof. We start with the first statement. Note that $t_n = X_1 + \cdots + X_n$, $n \geq 1$. Consider the time

$$t_{N(t)} \leq t < t_{N(t)+1}, \tag{6.14}$$

Then we have

$$t_{N(t)} = \sum_{j=1}^{N(t)} X_j, \qquad t_{N(t)+1} = \sum_{j=1}^{N(t)+1} X_j,$$

which can be written as

$$\frac{1}{N(t)} \sum_{j=1}^{N(t)} X_j \;\leq\; \frac{t}{N(t)} \;<\; \frac{1}{N(t)} \sum_{j=1}^{N(t)+1} X_j.$$

By the strong law of large numbers, the left and right hand sides converge to $E(X)$ as $t \to \infty$ almost surely. One can prove the second part via Wald's identity (which you will prove in your homework with some provided hints).

■

Now let $R(t) = \sum_{j=1}^{N(t)} R_j$ be the total amount of reward collected by time t, where $N(t)$ is the counting process for the renewal process. We want to calculate our long-run reward rate

$$\lim_{t\to\infty} \frac{R(t)}{t}.$$

Theorem 6.8 (Renewal reward theorem)**.** *For a positive recurrent renewal process in which a reward R_j is earned during cycle length X_j and such that $\{(X_j, R_j) : j \geq 1\}$ is i.i.d. with $\mathbb{E}[|R_j|] < \infty$, the long run rate at which rewards are earned is given by*

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[R]}{\mathbb{E}[X]} = \lambda \mathbb{E}[R] \quad a.s., \tag{6.15}$$

where (X, R) denotes a typical "cycle" (X_j, R_j); $\lambda = \{\mathbb{E}[X]\}^{-1}$ is the arrival rate for the renewal process.

Moreover,

$$\lim_{t\to\infty} \frac{\mathbb{E}[R(t)]}{t} = \frac{\mathbb{E}[R]}{\mathbb{E}[X]}. \tag{6.16}$$

Discussion 6.9. *Now let's prove Proposition 6.3 via the renewal reward theorem. Consider a renewal point process $\{t_n : n \geq 1\}$ with i.i.d. interarrival times $X_n = t_n - t_{n-1}$, $n \geq 1$. Define*

$$A(t) = t_{N(t)+1} - t, \qquad t \geq 0. \tag{6.17}$$

$A(t)$ is called the excess at time t, *or* remaining lifetime. *If $t_{n-1} \leq t < t_n$, then*

$$A(t) = t_n - t \leq X_n.$$

Note that if $\{t_n\}$ *is a Poisson process at rate* λ*, then by the memoryless property we have* $A(t) \sim \exp(\lambda)$*,* $t \geq 0$*. But for a general renewal process (as in residual service time in* $M/G/1$ *queue), we need to be smarter. We want to show that*

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t A(s)\,ds = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]} \quad a.s..$$

Note that we can view the i.i.d. X_j *as cycle lengths (service times), and* $r(t) = A(t)$ *as the generated reward rate at time* t*. Let* R_1 *be the generated reward in the first cycle. Then we have*

$$R_1 = \int_0^{X_1} A(s)\,ds = \int_0^{X_1} (X_1 - s)\,ds = \frac{X_1^2}{2}.$$

Since $\{(X_j, R_j)\}$*'s are i.i.d., by the renewal reward theorem, we almost surely have*

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t A(s)\,ds = \frac{\mathbb{E}[R]}{\mathbb{E}[X]} = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]}$$

Discussion 6.10 (Inspection paradox)**.** *Let* $S(t) = t_{N(t)+1} - t_{N(t)}$ *be the length of the interarrival time covering time* t*. If* $t_{j-1} \leq t < t_j$*, then we have* $S(t) = X_j$*. Define the reward rate as* $r(t) = S(t)$*. Then we get*

$$R_j = \int_{t_{j-1}}^{t_j} S(s)\,ds = \int_{t_{j-1}}^{t_j} X_j\,ds = X_j \int_{t_{j-1}}^{t_j} ds = X_j^2.$$

By the renewal reward theorem, we have almost surely that

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t S(s)\,ds = \frac{\mathbb{E}[R]}{\mathbb{E}[X]} = \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}$$

where the fact that $\frac{E[X^2]}{E[X]} \geq \mathbb{E}[X]$ *yields the inspection paradox.*