Lecture-25: Queueing Networks

1 Migration Processes

Corollary 1.1. *Consider an $M/M/s$ queue with Poisson arrivals of rate λ and each server having independent and exponential service of rate μ. If $\lambda < s\mu$, then the output process in steady state is also Poisson with rate λ.*

Proof. Let X_t denote the number of entities in the system at time t. Since $M/M/s$ process is a birth and death process, it follows from the previous proposition that $X : \Omega \to \mathbb{Z}_+^{\mathbb{R}}$ is time reversible at stationarity. The stationarity of the forward process holds only if the queue is stable, i.e. $\lambda < s\mu$.

Going forward in time, the time instants at which X_t increases by unity are the arrival instants of a Poisson process. Hence, by time reversibility, the time points at which X_t increases by unity when we go backwards in time also constitutes a Poisson process. But these instants are exactly the departure instants of the forward process. Hence, the result holds when $\lambda < s\mu$. □

Lemma 1.2. *For an ergodic $M/M/1$ queue in steady state, the following are true.*
(a) The number of entities present in the system at time t is independent of the sequence of past departures.
(b) For FCFS service discipline, the sojourn time in the system (waiting in the queue plus the service time) by an entity is independent of the departure process prior to its departure.

Proof. Recall that the reversed process is an identical stochastic replica of the forward process.
(a) Since the arrival process is Poisson, the future arrivals are independent of the number of entities in the system at the current instant. Looking backwards in time, future arrivals are the past departures. Hence by time reversibility, the number of entities currently in the system are independent of the past departures.
(b) Consider the case when an entity arrives into the system at time T_1. The entity leaves at time $T_2 > T_1$. Since the service discipline is assumed first come first serve and the arrival is Poisson, it is seen that the sojourn time $T_2 - T_1$ is independent of the arrivals after T_1. Looking backwards in time, these are departures prior to the departure instant for the reversed process. Thus, from the time reversibility, we see that the waiting time $T_2 - T_1$ of an entity is independent of the departures prior to its departure.

□

2 Network of Queues

2.1 Tandem Queues

Time reversibility of $M/M/s$ queues can be used to study what is called as a tandem or sequential queueing system. For instance, consider a queueing system with two queues in sequence, with each queue having one dedicated server. Service time of server i is random independent and distributed exponentially with rate μ_i. Customers arrive to the queue 1 according to a Poisson process with rate λ. After being served by server 1, entities join queue 2 for its service. Assume there is infinite waiting room at both servers. Since the departure process of queue 1 is Poisson, as discussed previously, the arrival process to queue 2 is also Poisson with rate λ. Time reversibility concept can be used to give a much stronger result.

Theorem 2.1. *For the ergodic tandem queue in steady state, the following are true.*
(a) The limiting number of entities X_∞^1, X_∞^2 present at server 1 and server 2 respectively at stationarity, are independent, and for $s = 1$

$$P\left\{X_\infty^1 = n_1, X_\infty^2 = n_2)\right\} = {\rho_1}^{n_1}(1-\rho_1){\rho_2}^{n_2}(1-\rho_2).$$

(b) For FCFS discipline, the waiting time at server 1 is independent of the waiting time at server 2, at stationarity.

Proof. Proofs follow by looking at reversed processes.

(a) By part (a) of previous lemma, the number of entities at server 1 at any time, is independent of the past departures from server 1. However, the past departures are same as the arrival to server 2 until this time. The past departures at server 1, in conjunction with the service times at server 2, determine the number of entities at server 2 at this time. This implies the independence of the number of entities in both servers at any given time. The expression for the joint density follows from the independence of two queues and the expression for the invariant distribution of an $M/M/1$ queue.

(b) By part (b) of the previous lemma, the waiting time of an entity at server 1 is independent of the past departures from server 1 prior to its departure. The past departures at server 1, in conjunction with the service times at server 2, determine entity's waiting time at server 2. Hence the result follows.

□

2.2 Jackson Network

Consider a system of k queues each with a dedicated server with random independent service times distributed exponentially with rate μ_i for server $i \in [k]$. We assume an infinite waiting time at each of the k queues. To each queue, entities arrive from outside the system, according to a Poisson process with rate r_i. Once an entity is served at server i, the customer joins queue j with probability P_{ij}, such that $\sum_{j\in[k]} P_{ij} \leqslant 1$. The probability of the customer departing the system is $1 - \sum_{j\in[k]} P_{ij}$. If we denote λ_j as the total rate at which the entities join queue j, then λ_j can be obtained as a solution to

$$\lambda_j = r_j + \sum_{i\in[k]} \lambda_i P_{ij}, \quad j \in [k].$$

We denote the number of entities in the queue $i \in [k]$ at time $t \in \mathbb{R}$ by $X_i(t) \in \mathbb{Z}_+$. Denoting $X(t) \triangleq (X_1(t), X_2(t), \dots X_k(t))$, one can show that X evolves as a continuous-time Markov chain. We focus on this process at stationarity. That is, we analyze the k-queue system by a stationary continuous-time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{R}}$ with state space $\mathcal{X} \triangleq \mathbb{Z}_+^k$. Fix $n \in \mathbb{Z}_+^k$, then we are interested in computing the joint invariant distribution

$$\pi(n) \triangleq \lim_{t\to\infty} P\{X(t) = n\} = \lim_{t\to\infty} P\Big(\cap_{i=1}^k \{X_i(t) = n_i\}\Big).$$

From the tandem queue results, we expect the marginal queue occupancies to be independent random variables. From the joint distribution, we can write the marginal distribution for each queue $i \in [k]$ as

$$\pi_i(m_i) \triangleq \lim_{t\to\infty} P\{X_i(t) = m_i\} = \sum_{n\in\mathbb{Z}_+^k : n_i = m_i} \pi(n).$$

Since, X is a CTMC, only a single transition takes place in any infinitesimal time interval. Let $X(t) = n \in \mathbb{Z}_+^k$, then the possible transitions from this Markov process at time $t \in \mathbb{R}$ and the associated rates are the following. We denote the unit vector in the ith direction by e_i.

(i) An external arrival takes place in queue i, with rate

$$Q(n, n + e_i) = r_i.$$

(ii) If $n_i > 0$, a service completes and an entity departs from queue i and joins queue j, with rate

$$Q(n, n - e_i + e_j) = \mu_i P_{ij}.$$

(iii) If $n_i > 0$, a service completes and an entity departs from queue i exiting the system, with rate

$$Q(n, n - e_i) = \mu_i (1 - \sum_{j\in[k]} P_{ij}).$$

Theorem 2.2. *For each queue $i \in [k]$, we define load $\rho_i \triangleq \lambda_i/\mu_i$ such that $\rho_i < 1$, then the reversed stochastic process is a network process of the same type as the original. That is, the reversed process $\hat{X}$ is a CTMC with the generator matrix Q^* given by the following.*

(a) The system has external Poisson arrivals to queue i at rate $\lambda_i(1 - \sum_{j\in[k]} P_{ij})$.

(b) The service times at queue i is i.i.d. *exponential with rate μ_i.*

(c) A departure from queue j goes to queue i with probability $\bar{P}_{ji}$ defined as $\bar{P}_{ji} \triangleq \lambda_i P_{ij}/\lambda_j$.

(d) *Each queue evolves as an independent $M/M/1$ queue with load ρ_i. That is, the joint invariant distribution $\pi \in \mathcal{M}(\mathbb{Z}_+^k)$ is defined as $\pi(n) \triangleq \prod_{i=1}^k \pi_i(n_i) = \prod_{i=1}^k \rho_i^{n_i}(1-\rho_i)$ for each $n \in \mathbb{Z}_+^k$.*

Proof. We observe the following for the reversed process.
(a) The arrival rate to any queue $i \in [k]$ is $Q^*(n, n+e_i) = \lambda_i(1 - \sum_{j\in[k]} P_{ij})$.
(b) If $n_j > 0$, then the rate of joining a queue i after a service completion from queue j is

$$Q^*(n, n-e_j+e_i) = \mu_j \bar{P}_{ji} = \frac{\mu_j}{\lambda_j}\lambda_i P_{ij}.$$

(c) If $n_j > 0$, then the rate of customer departing the system after service completion from server j is

$$Q^*(n, n-e_j) = \mu_j(1 - \sum_{i\in[k]} \bar{P}_{ji}) = \frac{\mu_j}{\lambda_j}(\lambda_j - \sum_{i\in[k]} \lambda_i P_{ij}) = \frac{\mu_j}{\lambda_j} r_j.$$

It suffices to check that the detailed balanced conditions hold with the candidate generator matrix Q^* for the reversed process, the candidate invariant distribution π, and any state $n \in \mathbb{Z}_+^k$.

1. We first focus at the detailed balance equations associated with the external arrivals

$$\pi(n)Q(n, n+e_i) = \pi(n+e_i)Q^*(n+e_i, n).$$

Since $Q(n, n+e_i) = r_i, Q^*(n+e_i, n) = \frac{r_i}{\rho_i}$ and the candidate invariant has a product form, we have $\pi_i(n_i+1) = \rho_i \pi_i(n_i)$. That the detailed balance equations hold for these transitions.

2. Second for $n_i > 0$, we look at the detailed balanced equations corresponding the transitions where an entity joins queue j after getting service from queue i,

$$\pi(n)Q(n, n-e_i+e_j) = \pi(n-e_i+e_j)Q^*(n-e_i+e_j, n).$$

From the product form of the candidate invariant distribution π and the definition of Q and candidate construction Q^*, we get that

$$\pi_i(n_i)\pi_j(n_j)\mu_i P_{ij} = \pi_i(n_i-1)\pi_j(n_j+1)\mu_j \bar{P}_{ji}, \quad n_i \in \mathbb{N}.$$

From the candidate invariant distribution π and definition of $\bar{P}$, the detailed balance equations continue to hold for these transitions as well.

3. Finally for $n_i > 0$, we look at the detailed balanced equations corresponding the exits from the system from queue i,

$$\pi(n)Q(n, n-e_i) = \pi(n-e_i)Q^*(n-e_i, n).$$

From the product form for invariant distribution π, we have for each $n_i \in \mathbb{N}$

$$\pi_i(n_i)\mu_i(1 - \sum_{j\in[k]} P_{ij}) = \pi_i(n_i-1)\lambda_i(1 - \sum_{j\in[k]} P_{ij}).$$

From the candidate invariant distribution π, the detailed balance equations continue to hold for these transitions as well.

□

Corollary 2.3. *The departure process for each server $i \in [k]$ is an independent homogeneous Poisson process with rate $\lambda_i(1 - \sum_{j\in[k]} P_{ij})$.*

Proof. We have already shown that in the reversed process, entities arrive to queue i from outside the system according to independent Poisson processes having rates $\lambda_i(1 - \sum_{j\in[k]} P_{ij})$ for $i \in [k]$. Since an arrival from outside corresponds to a departure out of the system from queue i in the forward process, the result follows. □