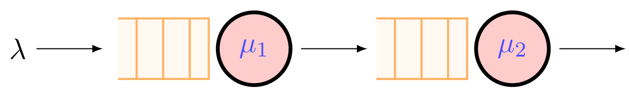
Introduction to Queueing Theory

Prof. N. Selvaraju
Department of Mathematics
Indian Institute of Technology Guwahati, India

Lecture - 29

Burke's Theorem, General Setup, Tandem Networks

Hi and hello, everyone; what we have seen are Queueing networks, and in which just recall from the previous lecture what we have been doing. We considered a two-node queueing network, two-node, which is connected serially as given



And it is something like $M/M/1 \to \bullet/M/1$, first node was $M/M/1$ and from which the input went to the second node, and that node has a single server of with exponentially distributed service time with a single server. So, in this now, if this is the system as we call which is a network of queues or queueing network is a two-node network series which is connected in series one after the other.

- The system state can be represented as a two-dimensional CTMC with state space $S = \{(n_1, n_2) : n_1, n_2 = 0, 1, 2, \dots\}$.
- Denote the probability of n_1 customers in the first node and n_2 customers in the second node in steady state by p_{n_1,n_2}.
- Then, the steady state solution for this system exists under the condition that $\rho_1 = \lambda/\mu_1 < 1$ and $\rho_2 = \lambda/\mu_2 < 1$ and can be obtained from the balance equations given by

$$\begin{aligned}(\lambda + \mu_1 + \mu_2)p_{n_1,n_2} &= \lambda p_{n_1-1,n_2} + \mu_1 p_{n_1+1,n_2-1} + \mu_2 p_{n_1,n_2+1}, \quad n_1 \geq 1, n_2 \geq 1\\ (\lambda + \mu_1)p_{n_1,0} &= \lambda p_{n_1-1,0} + \mu_2 p_{n_1,1}, \quad n_1 \geq 1\\ (\lambda + \mu_2)p_{0,n_2} &= \mu_1 p_{1,n_2-1} + \mu_2 p_{0,n_2+1}, \quad n_2 \geq 1\\ \lambda p_{0,0} &= \mu_2 p_{0,1}\end{aligned}$$

It can be shown that

$$p_{n_1,n_2} = \rho_1^{n_1}\rho_2^{n_2}p_{0,0} \quad \text{and} \quad p_{0,0} = (1-\rho_1)(1-\rho_2).$$

Thus,

$$p_{n_1,n_2} = \left[(1-\rho_1)\rho_1^{n_1}\right]\left[(1-\rho_2)\rho_2^{n_2}\right],\ n_1, n_2 \geq 0.$$

which you can write as in a way as $p_{n_1}p_{n_2}$, this is what we are observing, and p_{n_1,n_2} means that the joint distribution of the number in the system in the whole system means both the networks put together is equal to the product of their individual marginal distributions that is what we observed.

And what does that mean? That means that the second node which happens here is also an $M/M/1$ queue that is independent of the first one. So, this dot can become M; that is how it becomes, that is what if it happens that way then that would result in this way. So, it is behaving as if that. Now, whether that is true? This can be proved if we can characterize the output process from the first node. If it turns out the output process if we can characterize it, then this result is actually can be proved that this is true; it is not just a mere observation with respect to this particular system that you are talking about, but it is in general, it has to be true, and why it is true is what you have to look at it. And Burke's theorem is what helps us to do this; this Burke did this work in 1956, and he gave this result for an output process of an $M/M/c$ queue, and that is what we are going to see next in this lecture.

So, what is Burke's theorem, which characterizes the output process of an $M/M/c$ system is the following?

Theorem. *[Burke's Theorem]*
In an $M/M/c$ queueing system in steady state, the interdeparture times are IID exponential random variables with parameter λ. In other words, the output process is Poisson with the same parameter as the input process.

So, let us see how this can be shown to be true because the whole queueing network, at least in our analysis of networks that whatever we have been considering, like this output process, plays a critical. So far, we did not bother too much on the output process, but here it is very critical. So, it needs to be understood how this is true. So, let us see how it can be proved.

Proof:- Let $N(t)$ be the number in the system at time t and $t'_1, t'_2, \ldots, t'_n, t'_{n+1}, \ldots$ denote the successive departure instants so that $L = t'_{n+1} - t'_n$ is the n^{th} interdeparture interval.

Let $F_k(t) = P\{N(t'_n + t) = k, t'_{n+1} - t'_n > t\}, \quad t > 0, k = 0, 1, 2, \ldots$ be the joint probability that there are k in the system at time t after the last departure and that t is less than the interdeparture time.

Then, the CDF of L is

$$C(t) = P\{L \leq t\} = 1 - \sum_{k=0}^{\infty} F_k(t)$$

Since the input is a Poisson process, the probability that a departing customer leaves k in the system is equal to the probability that the number in the system is k and we therefore have

$$F_k(0) = p_k \quad \text{(of } M/M/c \text{ queue)}, \quad k = 0, 1, 2, \ldots.$$

Now, for an infinitesimal interval of length Δt,

$$\begin{aligned}
F_0(t + \Delta t) &= (1 - \lambda\Delta t)F_0(t) + o(\Delta t) \\
F_k(t + \Delta t) &= (1 - \lambda\Delta t)(1 - k\mu\Delta t)F_k(t) + \lambda\Delta t(1 - k\mu\Delta t)F_{k-1}(t) + o(\Delta t), \qquad 1 \leq k \leq c \\
F_k(t + \Delta t) &= (1 - \lambda\Delta t)(1 - c\mu\Delta t)F_k(t) + \lambda\Delta t(1 - c\mu\Delta t)F_{k-1}(t) + o(\Delta t), \qquad k \geq c
\end{aligned}$$

Moving $F_n(t)$ from the right-hand-side of each of the above equations to the LHS, dividing by Δt, and taking the limit as $\Delta t \to 0$, we obtain the differential-difference equations as

$$\begin{aligned} F_0'(t) &= -\lambda F_0(t) \\ F_k'(t) &= -(\lambda + k\mu)F_k(t) + \lambda F_{k-1}(t) \qquad 1 \le k \le c \\ F_k'(t) &= -(\lambda + c\mu)F_k(t) + \lambda F_{k-1}(t), \qquad k \ge c \end{aligned}$$

Using the boundary condition $F_k(0) = p_k$ and solving the above (in a similar manner as we did for the Poisson process), we obtain

$$F_k(t) = p_k e^{-\lambda t}, \qquad \text{where} \quad p_{k+1} = \begin{cases} \frac{\lambda}{(k+1)\mu} p_k, & 1 \le k \le c \\ \frac{\lambda}{c\mu} p_k, & k \ge c \end{cases}$$

Thus, we obtain the CDF of L as

$$C(t) = 1 - \sum_{k=0}^{\infty} p_k e^{-\lambda t} = 1 - e^{-\lambda t} \quad \text{and this implies that} \quad L \sim Exp(\lambda).$$

There are two more components to this the random variables $N(L)$ and L, where L is the duration, are independent, and the inter-departure times are also independent. So, that is what we will see in the next two steps. So, this is the first step that $L \sim Exp(\lambda)$.

Now, again consider

$$P\{N(t'_{n+1} + 0) = k, t \le t'_{n+1} - t'_n < t + \Delta t\} = F_{k+1}(t)\; P\{\text{one service completion in } (t, t + \Delta t)\}.$$

For $k+1 \le c$, $k+1$ servers are busy and rate of service is $(k+1)\mu$, while the rate of service is $c\mu$ when $k+1 > c$.

Thus, the right-hand-side expression for the above reduces to $p_{k+1}e^{-\lambda t}(k+1)\mu\Delta t + o(\Delta t)$ for $k+1 \le c$ and to $p_{k+1}e^{-\lambda t}c\mu\Delta t + o(\Delta t)$ for $k+1 > c$. Substituting the appropriate expressions for p_{k+1}, both of them reduces to $p_k\lambda e^{-\lambda t}\Delta t + o(\Delta t)$.

Thus,

$$P\{N(t'_{n+1} + 0) = k, t \le t'_{n+1} - t'_n < t + \Delta t\} = p_k \lambda e^{-\lambda t}\Delta t + o(\Delta t),$$

proving that $N(t'_{n+1} + 0)$ and $t'_{n+1} - t'_n$ are independent. i.e., $N(L)$ and L are independent.

So, effectively what this means is that this L which is the inter-departure time duration or the time for the inter-departure times L and $N(L)$, the number of in the system in that intervals or in that point, at the time t that is how we are defining it they are independent.

Now, we will use this to show the independence of the inter-departure times because that is what we have to see; we have shown that it is exponential, but we have to show that they are independent as well, then only it becomes exactly the same as the Poisson process the input process.

Let Λ represent the set of lengths of an arbitrary number of interdeparture intervals subsequent to the interval of length L.

So, you consider an initial interval of length L, then consider the remaining ones, which is basically a sequence; it could be any number of departures could have happened in during that period, or you take suppose 5 departure lengths or 6 departure lengths and so on. So, that is what Λ represents.

The Markov property implies that

$$P(\Lambda|N(L)) = P(\Lambda|N(L), L).$$

Since $N(L)$ and L are independent, we have

$$P(N(L), L) = P(N(L))\, P(L).$$

The joint probability function of the initial interval length, the state at the end of the interval, and the set of subsequent interval lengths may be written as

$$\begin{aligned} P(L, N(L), \Lambda) &= P(\Lambda|N(L), L)P(N(L), L) = P(\Lambda|N(L))P(N(L))P(L) = P(\Lambda, N(L))P(L) \\ \implies \quad P(L, \Lambda) &= \sum_{N(L)=0}^{\infty} P(\Lambda, N(L))P(L) = P(L)P(\Lambda), \end{aligned}$$

thus proving the mutual independence of all the intervals. Let Λ represent the set of lengths of an arbitrary number of interdeparture intervals subsequent to the interval of length L. The Markov property implies that

$$P(\Lambda|N(L)) = P(\Lambda|N(L), L).$$

Since $N(L)$ and L are independent, we have

$$P(N(L), L) = P(N(L))\, P(L).$$

The joint probability function of the initial interval length, the state at the end of the interval, and the set of subsequent interval lengths may be written as

$$\begin{aligned} P(L, N(L), \Lambda) &= P(\Lambda|N(L), L)P(N(L), L) = P(\Lambda|N(L))P(N(L))P(L) = P(\Lambda, N(L))P(L) \\ \implies \quad P(L, \Lambda) &= \sum_{N(L)=0}^{\infty} P(\Lambda, N(L))P(L) = P(L)P(\Lambda), \end{aligned}$$

thus proving the mutual independence of all the intervals.

This is what Burke's theorem basically; what we need is the main idea of what is the output process in an $M/M/c$ system. So, the example that we have considered there is a two-node series network, which is basically the $M/M/1$ system we have considered for simplicity though it is also true for $M/M/1$ system. So, the output process, remember here we are not doing anything here like all arrivals go through the system once, and then they leave they do not come back, or no such thing is happening in the normal $M/M/1$ system, $M/M/c$ system. The output process is also a Poisson process with the same rate as the input process; this is a very useful, powerful result in the context of our queueing network.

So, that is the reason why in the two-node network, the result came out to be the product form of the distribution itself, not just the product form solution; it is a distribution itself that came out to be the product form. So, now, we consider we move on to a general queueing network after having seen certain ideas about an open network, closed network, mixed network, and what are the additional specification that we need to give to describe a particular queueing network in a general setting. We move on to the study of queueing networks in the context, but we restrict our study not just to any general network but with certain conditions.

- The system is a network of queues consisting of a group of k nodes.
- Each node represents a service facility with c_i servers at node $i, i = 1, 2, 3, \ldots, k$ and are assumed to have infinite buffers.
- In general, customers enter the system at any node, traverse from node to node and depart from any node (all customers need not have the same path).

  And in the process, there may be feedback, meaning that a customer may return to a node in which he has already received the service; it can happen in a production system. So, you pass through certain things like, you go to the next stage that is something is done then some defective happened then you have to start all over again. So, you will come back to this, it is possible that feedbacks are there, but if it is in some situation, feedback will not be there because if it is passing through a packet is passing through the network to outside, then it will not be coming back to itself generally. But it can be, but it is a different matter, but it may not be so that you may require that also.

- There may be feedback or may be only feed-forward, and need not visit all nodes too.
- The main characteristics of our networks are:
  1. Arrivals from the 'outside' to node i follow a Poisson process with rate γ_i.
  2. Service (holding) times at each channel at node i are IID $Exp(\mu_i)$ (may depend on the queue length).
  3. The probability that a customer who has completed service at node i will go next to node j (routing probability) is r_{ij}, $1 \leq i \leq k, 0 \leq j \leq k$ (independent of the state of the system) and r_{i0} indicates the probability that a customer will leave the system from node i.
- The networks having the above properties are called **Jackson networks** (Jackson, 1957, 1963).
  ⧫ They will have a product form solution in the true sense.
  It need not be that product form in terms of distribution functions, but the solution would be in product form; we already mentioned what that would mean we will come back to that again when we actually see it.
- Networks with $\gamma_i = 0, \ \forall i$ and $r_{i0} = 0, \ \forall i$ are referred to as 'closed' Jackson networks (the general case described above are 'open' Jackson networks).
- The finite-source queue (machine repairmen problem) is a closed network, with two nodes (one representing operating machines and the other the repair facility).
  ► Here, $i = 1$(represents the operating node), 2(represents failed machines); $j = 0, 1, 2; r_{12} = r_{21} = 1$ and all

other r_{ij}'s are zero.

If imperfect repairs happen, suppose in the same scenario meaning that you repair once and finally, you found out that there is still an imperfect repair one would call. It is the additional feature; repair means it is not completely repaired; one could have a scenario where there can be imperfect repairs, then it has to go to repair again, then there could be feedback to that. We are not considering that; it is generic one what we have seen. Even with the imperfect repair, it can still be considered as a closed queueing network, but it cannot be then part of this BDP business. So, and all other r_{ij}'s are 0 is what then you have here. So, this is a closed network; we have already seen one example of a closed network of this nature.

- We will start with open networks where

$$\gamma_i = \begin{cases} \lambda, & i = 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad r_{ij} = \begin{cases} 1, & j = i+1, 1 \leq i \leq k-1 \\ 1, & i = k, j = 0 \\ 0, & \text{otherwise} \end{cases}$$

So, it is just that a series network with k nodes what we have seen earlier 2 nodes now it is generalized to k node.

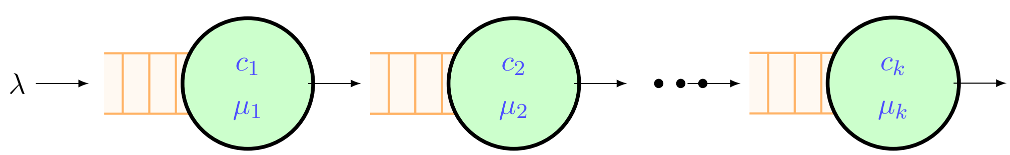
▶ These networks are called 'series' or 'tandem' queues where customers may enter from the outside only at node 1 and depart only from node k.

- We then generalize to a general open Jackson networks and then to closed Jackson networks.
- We consider only Markovian systems, wherein all the exogenous inputs are Poisson, the holding times are exponential, and the routing probabilities are known and state independent.

So, this is what the series.

- There are series of service stations through which each calling unit must progress prior to leaving the system.
  ■ Typical in manufacturing or assembly-line processes, clinical systems.

Computer communication, at least in a more general setup, will come, but you have a specific sequence; it goes through only that sequence, mainly in manufacturing and clinical systems and hospitals, basically you would see here. So, it is basically what you have?



Now, from the knowledge that we have gained from our discussion of this two-node network that we have seen already. We see that each of these stations there what we said; analyze that two-node it actually you can treat each node separately as an independent entity and for the whole system, then you have the complete picture in hand that is what you have seen that that is also true here.

- The calling units arrive according to PP(λ), no restriction on the capacity of the waiting room between stations, and the service times of each server at station i are $Exp(\mu_i),\ i = 1, 2, \ldots, k$.

- From our knowledge gained from the consideration of the two-stage series network earlier, we see that each station can be analyzed separately as a single-stage queueing model.

- By Burke's theorem, all stations are independent $M/M/c_i/\infty$ models, as long as there are ample holding spaces in front of each node.

- Thus, the results of $M/M/c$ can be used on node individually and a complete analysis of this series or tandem queueing network is possible.

- The steady state probability $p_{n_1,n_2,\ldots,n_k}$ that there are n_i customers in the ith node is therefore given by

$$p_{n_1,n_2,\ldots,n_k} = p_{n_1}\, p_{n_2}\, \cdots\, p_{n_k}, \qquad n_i \geq 0, i = 1, 2, \ldots, k,$$

  where p_{n_i} is the probability that there are n_i customers in the system in an $M/M/c_i$ queue in steady state (which exists under the condition $\rho_i = \dfrac{\lambda}{c_i \mu_i} < 1$).

- The product-form result obtained above holds good in more general cases of Jackson networks too.

- It is possible to carry out analysis of a series network wherein only the last node can have capacity limitations (with blocked customers dropped from the system), using the ideas as above.

- Analysis of feedforward networks (i.e., networks in which customers are not allowed to revisit previously visited nodes) is quite similar to that of the series networks.

Let us take a simple example to see what the series network is.

Example. *[A Supermarket]*

- In a supermarket, as customers complete their shopping, they enter the lounge and wait if all the checkout counters are busy.

- Customers arrive according to a PP(40/h) and shopping times and checkout times are exponentially distributed with an average of 45 minutes and 4 minutes, respectively.
  ▶ *a) Minimum number of checkout counters needed?*
  ▶ *b) If we add one more counter to the minimum, what is the average waiting time in lounge? How many customers, on an average, will be in the lounge? How many people, on an average, will be in the entire supermarket?*

- This situation can be modelled by a two-node tandem queue.
  ▶ *First Stage:* $M/M/\infty$ *with* $\lambda = 40, \mu = \frac{4}{3}$.
  ▶ *Second Stage:* $M/M/c$ *with* $\lambda = 40, \mu = 15$.

- For steady-state convergence, we need $c\mu > \lambda \implies c > \frac{\lambda}{\mu} = 2.67 \implies c = 3$. That is, the minimum number of checkout counters needed is 3.

- *If we add one more to the minimum number of counters, then the new $c = 4$ and node-2 now is an $M/M/4$ system.*
- From the results of $M/M/c$, we have that $p_0 = 0.06$ and

$$W_q = 1.14 \textit{ minutes} \qquad \textit{and} \qquad L_q = \lambda W_q = 0.76.$$

  The total number of customers in the supermarket is

$$\begin{aligned} L &= L\,(\textit{of } M/M/\infty) + L\,(\textit{of } M/M/4) \\ &= \frac{\lambda}{\mu} + \lambda \left(W_q + \frac{\lambda}{\mu} \right) = \frac{40}{(4/3)} + 40 \left(0.019 + \frac{4}{60} \right) = 30 + 3.44 = 33.44 \end{aligned}$$

- Now, one can do similar calculations with the minimum number of checkout counters (i.e., 3) to see how much these numbers increases.

So, this is a series network. So, basically, what we have seen is that very simple network and how it can give you some insights about how the system operates if you have to model such kinds of things. So, this is about the series network; I mean, later on then, we will generalize to a more general network in the following lectures.
Thank you, bye.