

Chapter 9 Analysis of Waiting Lines

9.1 Single-server queues (SINGLEQ)	80
9.2 Multiple-server queues (MULTIQ)	83
9.3 Queuing equations	85

In almost every organization, there are examples of processes which generate waiting lines called queues. The worksheets in this chapter automate the most practical models for analyzing the behavior of queues. SINGLEQ includes 4 alternative models for a single queue processed by a single server or service facility. All models in SINGLEQ assume that arrivals to the queue are generated by a Poisson process. The models differ in the probability distribution that generates service times, the queue capacity or waiting space available, and the size of the source population feeding the queue. MULTIQ analyzes problems in which customers form a single queue processed by 2 or more servers. MULTIQ assumes that arrivals are Poisson, service times are exponential, and there is no restriction on queue capacity or the size of the source population. All equations used in this chapter are summarized in Section 9.3.

9.1 Single server queues (SINGLEQ)

Figure 9-1 automates four single-server queuing models distinguished by the alternative assumptions listed at the top of the worksheet. Pay careful attention to assumptions to ensure that you apply the correct model. Operation of any model is easy: all you have to do is change the input parameters in rows 8-12. Inputs are not applicable in cells that display "xxxxx."

The basic queuing model is in column D, with Poisson arrivals, exponential service times, and both infinite waiting capacity and source population. Let's use this model to analyze an operations problem faced by the Cattleman's National Bank of Angleton, Texas. Cattleman's runs a branch office, with one teller, in the student union at Rosharon State University. Cattleman's has received many complaints about waiting times on Friday, the peak business day at Rosharon. Stop-watch studies show that customer arrivals on Friday average 36 per hour, Poisson-distributed, and the teller serves 40 customers per hour, exponentially-distributed.

Just how bad is waiting time? Enter these arrival and service rates in SINGLEQ and the results indicate that the mean time in the queue is 0.225, while mean time in the system (waiting time plus service time) is 0.250. These numbers are interpreted as fractions of the time unit used to measure the arrival and service rates. The rates are per hour, so the mean time in the queue is 0.225 of an hour or $0.225 \times 60 = 13.50$ minutes. Mean time in the system is then $0.250 \times 60 = 15.00$ minutes. (In some queuing problems, arrivals and service rates are per day, so you would multiply by 24 hours to obtain the solution in hours. In other problems, rates might be per 15-minute interval, so you would multiply by 15 to get the solution in minutes.) The results also show that the mean number in the queue is 8.1 customers and the mean number in the system is 9.0. The probability that the server is idle is only .10, so the probability that the server is busy is .90. A table showing the probability that exactly N customers are in the system is available in the lower half of the worksheet.

Cattleman's is considering replacing the teller with a banking machine to reduce Friday waiting times. An analysis shows that more than 90% of Rosharon Friday transactions are cash withdrawals, that is, cashing of personal checks and withdrawals from savings accounts. A stop-watch study at other Cattleman's machine locations reveals that service time for cash withdrawals is not exponential. Thus we cannot use the basic model in column D. But we can use the model in column G, a variation of the basic model which allows for any service-time distribution. Column G requires the mean and standard deviation of service times. The stop-watch study gives a mean of 90 seconds and a standard deviation of 8 seconds. The service rate for model input is $(60 \times 60)/90 = 40$ per hour, the same as the teller. The standard deviation is 8 seconds and must be converted for model input as follows: $8/(60 \times 60) = .002222$.

Entering these values in column G shows that the banking machine can cut the time in the queue to $0.113 \times 60 = 6.78$ minutes, about half the time for a human teller. Time in the system is $.138 \times 60 = 8.28$ minutes, another substantial cut. Numbers of customers in the queue and system are also reduced. How can this be if the banking machine has the same service rate as the human teller? The answer is that machine service times are much less variable than with a human teller. Queuing statistics always improve when you reduce the variability of service times.

Figure 9-1

	A	B	C	D	E	F	G
1	SINGLEQ.XLS		SINGLE-SERVER QUEUING MODELS				
2	ASSUMPTIONS:						
3	Arrival distribution			Poisson	Poisson	Poisson	Poisson
4	Service-time distribution			Exponential	Exponential	Exponential	ANY
5	Capacity			Infinite	FINITE	Infinite	Infinite
6	Source population			Infinite	Infinite	FINITE	Infinite
7	INPUT:						
8	Arrival rate			36	14	1	36
9	Service rate			40	20	8	40
10	Capacity (#in line + #in svc.)			xxxx	3	xxxx	xxxx
11	Source population (max. 20)			xxxx	xxxx	4	xxxx
12	Std. dev. of service time			xxxx	xxxx	xxxx	0.002222
13	QUEUE STATISTICS:						
14	Probability server idle (P0)			0.100	0.395	0.575	0.100
15	Probability server busy (1-P0)			0.900	0.605	0.425	0.900
16	Nbr. in queue (Lq)			8.100	0.464	0.172	4.082
17	Nbr. in system (Ls)			9.000	1.069	0.597	4.982
18	Time in queue (Wq)			0.225	0.038	0.050	0.113
19	Time in system (Ws)			0.250	0.088	0.175	0.138
20	Proportion of customers lost			#N/A	0.135	#N/A	#N/A
21	PROB. OF N CUSTOMERS IN SYSTEM:						
22	0			0.100	0.395	0.575	0.100
23	1			0.090	0.276	0.287	#N/A
24	2			0.081	0.193	0.108	#N/A
25	3			0.073	0.135	0.027	#N/A
26	4			0.066		0.003	#N/A
27	5			0.059			#N/A
28	6			0.053			#N/A
29	7			0.048			#N/A
30	8			0.043			#N/A
31	9			0.039			#N/A
32	10			0.035			#N/A
33	11			0.031			#N/A
34	12			0.028			#N/A
35	13			0.025			#N/A
36	14			0.023			#N/A
37	15			0.021			#N/A
38	16			0.019			#N/A
39	17			0.017			#N/A
40	18			0.015			#N/A
41	19			0.014			#N/A
42	20			0.012			#N/A

Unlike the basic model, in the any-service-time model it is impossible to compute the probability of any specific number of customers in the system, except when the system is empty. Otherwise, all queuing statistics in the basic model are available in the any-service-time model.

Another important variation of the basic model is shown in column E, where there is some maximum number of customers that can be in the system at any one time. The maximum includes customers waiting as well as in service. The finite-capacity model is useful in analyzing another Cattleman's location, a drive-in branch in the suburbs of Santa Fe. The branch has one teller window and space for 3 cars (1 in service and 2 in line). The branch is located on a busy street and when the system is full, customers are reluctant to wait in the street. Arrivals during peak business times are 14 per hour, with service at the rate of 20 per hour. How many customers are lost (forced to drive on because the waiting line is full)? Enter the arrival and service rates and a capacity of 3 in the model in column E and the proportion of customers lost is 0.135. What if Cattleman's expands the driveway to hold 4 cars (1 in service and 3 waiting)? The proportion lost falls to 0.087. Expansion to 5 cars reduces the proportion lost to 0.057.

The final single-server model in column F applies when the source population feeding the queue is limited. This model is widely used to analyze repair operations that support a specific population of equipment. For example, in the DeWalt Lubricants plant in Sugar Land, Texas, 4 machines fill 32-ounce cans of motor oil. The machines are the source population in the model and require frequent adjustment by a mechanic. Adjustment problems (arrivals) are Poisson at the rate of 1 per hour, and the mechanic can serve 8 per hour, exponentially-distributed. How much production time is lost per hour due to machines in service or waiting for adjustments? The answer is given by the time in the system of $0.175 \times 60 = 10.50$ minutes. DeWalt is considering adding a second bank of 4 filling machines identical to the present machines. If the single mechanic serves all 8 machines, what happens to lost production time? Assume that the arrival rate will double to 2 per hour. The source population is now 8. Enter these changes in SINGLEQ and the time in the system becomes $0.531 \times 60 = 31.86$ minutes.

9.2 Multiple-server queues (MULTIQ)

Many service facilities operate with a single waiting line that feeds several servers. Customers must wait on the first available server. The MULTIQ model in Figure 9-2 analyzes this type of system under the following assumptions: Poisson arrivals, exponential service times, and both infinite waiting capacity and source population. MULTIQ is a useful management tool at Cattleman's Bank, where all multiple-teller branches operate with a single waiting line.

Cattleman's management attempts to keep customer time in the queue to 1 minute or less. This goal is achieved by changing the number of teller windows that are open as arrival rates change during the business day. In Cattleman's large commercial branch in Thompsons, Texas, arrivals are Poisson and average 40 per hour during the 9 AM - 10 AM period on Monday mornings. The service rate per teller is 20 per hour. How many teller windows must be open at 9 AM on Mondays?

In Figure 9-2, the time in the queue with 3 tellers is $.022 \times 60 = 1.32$ minutes, so we must operate with 4 tellers. Enter this value and the result is time in the queue of $.004 \times 60 = .24$ (about 14 seconds). Now try a number of servers equal to 2 in MULTIQ. The queue statistics display ERR. Why? The arrival rate of 40 is equal to the total service rate of 2 tellers $\times 20 = 40$ per hour. All queuing systems blow up when arrival rates are equal to service rates. The reason is that servers can never catch up when arrivals vary above the mean. The same problem occurs in SINGLEQ if you set an arrival rate equal to a service rate.

Figure 9-2

	A	B	C	D	E
1	MULTI.Q.XLS		MULTIPLE-SERVER QUEUING		
2	ASSUMPTIONS:				
3	Arrival distribution				Poisson
4	Service-time distribution				Exponent.
5	Waiting capacity				Infinite
6	Source population				Infinite
7	INPUT:				
8	Arrival rate				40
9	Individual service rate (each server)				20
10	Nbr. of servers (max. 20)				3
11	QUEUE STATISTICS:				
12	Prob. all servers idle (P0)				0.111
13	Prob. at least 1 server busy (1-P0)				0.889
14	Prob. all servers busy (Pw)				0.444
15	Nbr. in queue (Lq)				0.889
16	Nbr. in system (Ls)				2.889
17	Time in queue (Wq)				0.022
18	Time in system (Ws)				0.072
19	PROB. OF N CUSTOMERS IN SYSTEM:				
20	0				0.111
21	1				0.222
22	2				0.222
23	3				0.148
24	4				0.099
25	5				0.066
26	6				0.044
27	7				0.029
28	8				0.020
29	9				0.013
30	10				0.009
31	11				0.006
32	12				0.004
33	13				0.003
34	14				0.002
35	15				0.001
36	16				0.001
37	17				0.001
38	18				0.000
39	19				0.000
40	20				0.000

9.3 Queuing equations

This section is a reference guide to equations used in SINGLEQ and MULTIQ. Notation is as follows:

- λ = mean arrival rate,
- μ = mean service rate,
- σ = standard deviation of service time,
- M = waiting capacity, and
- N = size of finite source population.

Single-server model with Poisson arrivals, exponential service times, infinite waiting capacity, and infinite source population:

$$L_q = \text{mean number in the queue} = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

$$L_s = \text{mean number in the system} = \frac{\lambda}{\mu - \lambda}$$

$$W_q = \text{mean time in the queue} = \frac{\lambda}{\mu(\mu - \lambda)}$$

$$W_s = \text{mean time in the system} = \frac{1}{\mu - \lambda}$$

$$P_w = \text{probability the service facility is busy (utilization factor)} = \frac{\lambda}{\mu}$$

Single-server model with Poisson arrivals, exponential service times, finite waiting capacity, and infinite source population:

$$P_0 = \text{probability of zero in the system} = \frac{1 - \left(\frac{\lambda}{\mu}\right)}{1 - \left(\frac{\lambda}{\mu}\right)^{M+1}}$$

$$P_w = \text{probability the service facility is busy (utilization factor)} = 1 - P_0$$

$$P_M = \text{proportion of customers lost because system is full} = \left(\frac{\lambda}{\mu}\right)^M P_0$$

$$L_s = \text{mean number in the system} = \frac{P_w - M\left(\frac{\lambda}{\mu}\right)P_M}{1 - \left(\frac{\lambda}{\mu}\right)}$$

$$L_q = \text{mean number in the queue} = L_s - \frac{\lambda(1 - P_M)}{\mu}$$

$$W_s = \text{mean time in the system} = \frac{L_s}{\lambda(1 - P_M)}$$

$$W_q = \text{mean time in the queue} = W_s - \frac{1}{\mu}$$

Single-server model with Poisson arrivals, exponential service times, finite waiting capacity, and finite source population:

$$L_q = \text{mean number in the queue} = N - \left(\frac{\lambda + \mu}{\lambda}\right)(1 - P_0)$$

$$L_s = \text{mean number in the system} = L_q + (1 - P_0)$$

$$W_q = \text{mean time in the queue} = \frac{L_q}{(N - L)\lambda}$$

$$W_s = \text{mean time in the system} = W_q + \frac{1}{\mu}$$

$$P_0 = \text{probability all channels are simultaneously busy (utilization factor)} = \frac{1}{\sum_{n=0}^N \frac{N!}{(N - n)!} \left(\frac{\lambda}{\mu}\right)^n}$$

Single-server model with Poisson arrivals, any service time distribution, infinite waiting capacity, and infinite source population:

$$L_q = \text{mean number in the queue} = \frac{\lambda^2 \sigma^2 + \left(\frac{\lambda}{\mu}\right)^2}{2\left(1 - \frac{\lambda}{\mu}\right)}$$

$$L_s = \text{mean number in the system} = L_q + \frac{\lambda}{\mu}$$

$$W_q = \text{mean time in the queue} = \frac{L_q}{\lambda}$$

$$W_s = \text{mean time in the system} = W_q + \frac{1}{\mu}$$

$$P_w = \text{probability the service facility is busy (utilization factor)} = \frac{\lambda}{\mu}$$

Multiple-server model with Poisson arrivals, exponential service times, infinite waiting capacity, and infinite source population:

$$P_w = \text{probability all channels are simultaneously busy (utilization factor)} = \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k \frac{k\mu}{k\mu - \lambda} P_0$$

$$L_s = \text{mean number in the system} = \frac{\lambda\mu \left(\frac{\lambda}{\mu}\right)^k}{(k-1)!(k\mu - \lambda)^2} P_0 + \frac{\lambda}{\mu}$$

$$L_q = \text{mean number in the queue} = L_s - \frac{\lambda}{\mu}$$

$$W_s = \text{mean time in the system} = \frac{L_s}{\lambda}$$

$$W_q = \text{mean time in the queue} = \frac{L_q}{\lambda}$$

Exercises

- 9-1** Trucks arrive at Incaviglia Distribution to be unloaded in a pattern which can be characterized by the Poisson distribution. The average rate of arrivals is 30 per hour, and the service times are normally distributed with a mean service time of 1.5 minutes and a standard deviation of .2 minutes. The drivers make \$12 each hour and do not unload the trucks. What is the average idle time expense incurred by the trucking company each time a driver visits Incaviglia?
- 9-2** Turk Farrell works at McClane Insurance Company processing claims. The claims arrive at his desk according to a Poisson distribution with a mean of 10 claims per day. Turk processes the claims according to a uniform distribution with a mean of 11 claims per day and a standard deviation of 2 hours per claim during each 8-hour day. What is the average time elapsed before a newly arrived claim is processed?
- 9-3** Ramon Garcia plans to open a Dairy Queen restaurant near Astros Field. The restaurant will have a drive-in window, although Ramon is uncertain about the number of spaces to allow for cars. During peak business hours (before and after ball games), Ramon estimates that cars will arrive at the drive-in window at the rate of 20 per hour, Poisson distributed. He estimates that the average service time per car will be 2.5 minutes, exponentially distributed. Ramon assumes that any customer who arrives when the line is full will be lost. Ramon wants to keep the proportion of lost customers to less than 5 percent. How many spaces should the drive-in line have?
- 9-4** Motorists arrive at C. J. Nitkowski's Auto Analytical Center for state emissions inspections at a mean rate of 10 cars per hour, Poisson distributed. C. J. is able to perform 12 inspections per hour on average and his inspection time is exponentially distributed. He has room to accommodate 6 cars waiting in line for inspections. What percentage of the time will the Center work on inspections? What percentage of potential customers will go elsewhere for inspections because his waiting area is full?
- 9-5** Castilla Lumber Company has a logging operation near Texarkana, Arkansas. Cut logs are skidded from the cut site to a logging road where a crew loads log trucks. Vinny Castilla has sufficient log trucks to ensure that there is always at least one empty truck available for loading. Load skidders arrive at the truck loading area at an average rate of 3 per hour in a manner that follows a Poisson distribution. A crew of 5 can take care of 4 skidders an hour with service times that are exponentially distributed. Vinny wants a less costly way to load, since it costs \$60 an hour to operate a skidder, loaded or empty. There is a hydraulic off-loader on the market which can service 7 skidders an hour, with normally distributed service times, with a standard deviation of 2 minutes per skidder. It will cost twice as much as the manual crew, each of whose members get \$8.50 an hour. A second alternative for Vinny is to add another crew who could service as

many skidders as the original crew. However, this option will also call for the addition of a supervisor at \$14 an hour. Which of these alternatives is the less costly for Vinny?

9-6 One or more Astros players are assigned to the Shed to sign autographs on selected dates during the off-season. On these dates, fans arrive in a pattern which is well-described by a Poisson distribution at an average rate of 500 an hour. A single player can sign autographs at a mean rate of 60 an hour where the service time is exponentially distributed. Drayton McClane thinks that fans should not have to wait longer than 30 minutes to get an autograph. How many players should be assigned to work in the Shed on each autograph date?

9-7 The Astros Branch Ticket Office in Sugar Land has two agents. Customers arrive at the rate of 30 per hour, Poisson distributed. Each clerk can service 20 customers per hour, with exponentially distributed service times. Johnny Weekly, the Branch manager, is willing to increase the number of agents until he has decreased the mean length of the waiting line by half. How many agents does Johnny need to meet this goal?