Module 7:Application of stochastic processes in queueing theory

Lecture 29:Application of stochastic processes in Queueing Theory

The Lecture Contains:

Introdution

Queueing systems represent an example of a much broader class of interesting dynamic systems, which we will denote as systems of flows. Few examples of queueing systems are:

1. Waiting in traffic light.
2. Sequencing of different jobs in different machines.
3. Withdrawing cash at the teller counter.
4. Buying a movie ticket at the cinema hall ticket counter.
5. Holding the telephone as it rings.
6. Flow of computer programmes through a time sharing computer system.

In a flow system we say that some commodity is flowing or is moving or is being transferred through a finite capacity channel in order that it goes from one point to another point. Generally flows are
(i) ***steady flow*** and
(ii) ***unsteady flow***.

Previous Next

For a queueing system we generally should define or know the following

- Inter arrival time(s) and its distributions, i.e., $A_i(t) = P_i[\text{time between arrivals} \leq t_i]$, which being important is assumed as *i.i.d*.
- The service time(s) (time to serve customer) and this has the probability distribution represented as $B_j(x) = P_j[\text{service time} \leq x_1]$.
- Storage capacity, K_j , and in general problems we consider K_j as ∞.
- Processing policy, e.g., First In First Out (FIFO), Last In First Out (LIFO), Random Order Processing, Priority Based Processing, etc.

We will first consider the G/G/m queueing system, where
G = Inter arrival time distribution is arbitrary
G = Service distribution is arbitrary
m = number of servers
Remember here the processing policy is also arbitrary

Previous Next

General system notations

a: Arrival distribution

- M : Markovian/Poission ***arrival*** distribution
- D : Constant/Deterministic time of ***arrival***
- E_k : Erlang/Gamma distribution of time of ***arrival*** (sum of *n* number of exponential distribution is Gamma distribution)
- GI: General distribution for interval ***arrival*** time

b: Departure/Service time distribution

- M: Markovian/Poission ***departure/service*** distribution
- D: Constant/Deterministic ***departure/service*** time
- E_k: Erlang/Gamma distribution of time of ***departure/service*** (sum of *n* number of exponential distribution is Gamma distribution)
- G: General distribution for ***departure/service*** time

c:Number of parallel servers, n_s , (in general the number can be 1, 2,..., ∞)

d: Queue discipline

- FIFO/FCFS: First in first out or first come first server
- LIFO/LCFS: Last in first out or last come first server
- SIRO: Service in random order
- GD: General discipline (e.g., with priorities, weightages, etc.)

e: The maximum number of customers allowed in the system (remember it is the number being served by server system plus the number standing in the queue system)

f : The size of the calling source, i.e., from where the customers are coming and joining the queue.

So for example (M/M/3):(FIFO/25/500) would imply that we have

(i) M: Markovian/Poission ***arrival*** distribution
(ii) M: Markovian/Poission ***departure/service*** distribution
(iii) Number of servers is 3
(iv) We have the discipline of the queue as first in first out
(v) Maximum number of customers allowed in the system is 25
(vi) The size of the outside source from where customers are joining the queue is finite in number and is 500

Example of Single servers

(M/M/1):(GD/∞/∞)

(M/M/1):(FCFS/∞/∞)

Let us denote the following

- C_n = denotes the nth customer in the system at time t
- $\alpha(t)$ = number of arrivals in (0, t) time interval
- $\delta(t)$ = number of departures in (0, t) time interval
- $N(t)$ = number of customer in the system at time t, i.e., $N(t) = \{\alpha(t) - \delta(t)\}$
- $U(t)$ = the remaining time required to empty the system of all customers present at time t
- $\lambda_t = \dfrac{\alpha(t)}{t}$, i.e., average arrival rate in the interval (0, t)
- γ_t =accumulated customer seconds upto time t
- $T_t = \dfrac{\gamma(t)}{t}$, i.e., system time per customer averaged over all customer in the queueing system in the interval (0, t)
- τ_n = arrival time for C_n
- $t_n = (\tau_n - \tau_{n-1})$ *inter arrival time* for C_n , and as per our basic assumption we have

  $P[t_n \leq t] = A_n(t)$
- x_n = service time for C_n and as per our assumption we have $P[x_n \leq x] = B_n(x)$
- w_n = waiting time in the queue for C_n
- $s_n = (x_n + w_n)$ total time spent by C_n, this is also known as the *system time*, i.e., the time spent in the system or queue by C_n

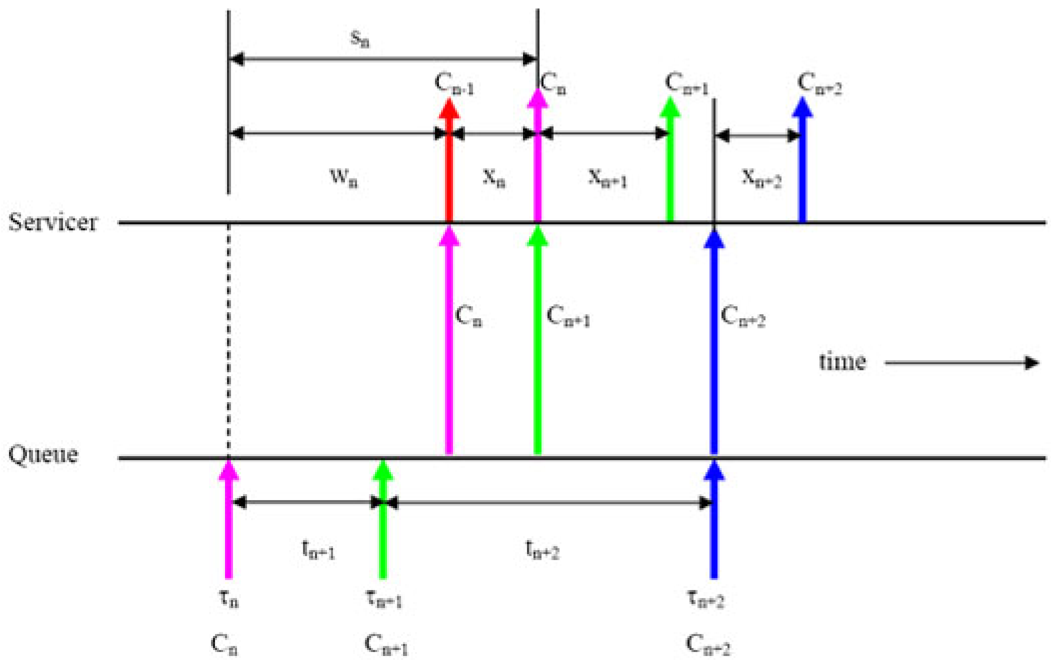

Figure 7.1: A simple schematic diagram to explain the different nomenclature in scheduling problems

We denote

- $(n-1)^{th}$ customer with red, i.e.,
- n^{th}customer with pink, i.e.,
- $(n+1)^{th}$customer with green, i.e.,
- $(n+2)^{th}$ customer with blue, i.e.,

Previous Next

What we can glean from the set of information given above

Thus for the general nth customer in the queue we have defined the following, which are: (i) his/her inter arrival time, (ii) his/her service time, (iii) his/her waiting time, (iv) his/her system time.
Thus the main data which we will have in front of us would be the sequence of $\{t_1, t_2, t_3, \ldots\}$, $\{x_1, x_2, x_3, \ldots\}$, $\{w_1, w_2, w_3, \ldots\}$ and $\{s_1, s_2, s_3, \ldots\}$.

What is of interest to us

- $E[t_n] = \tilde{t}_n$ and $\lim_{n\to\infty} \tilde{t}_n = \tilde{t} = \frac{1}{\lambda}$, where l is the average arrival time
- We denote, again for our convenience the following $E[t_n] = t_n$

Metric # 1 (average waiting time or cumulative waiting time, considering there is only one (1) machine only)

Minimize the average waiting time (cumulative waiting time), i.e., minimize: $\overline{w}_n$, which is $\overline{w}$.

Previous Next