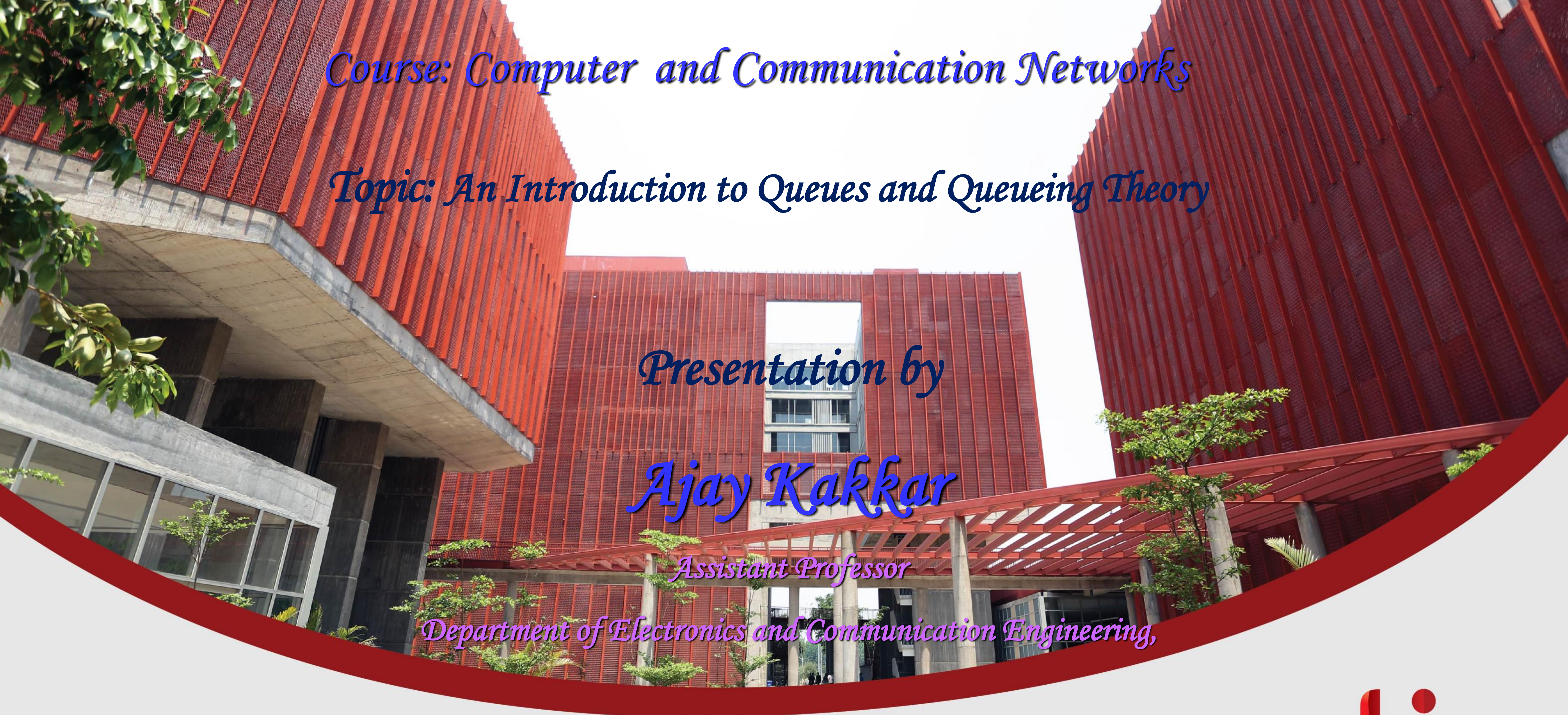


Course: Computer and Communication Networks

Topic: An Introduction to Queues and Queueing Theory

Presentation by

Ajay Kakkar

Assistant Professor

Department of Electronics and Communication Engineering,

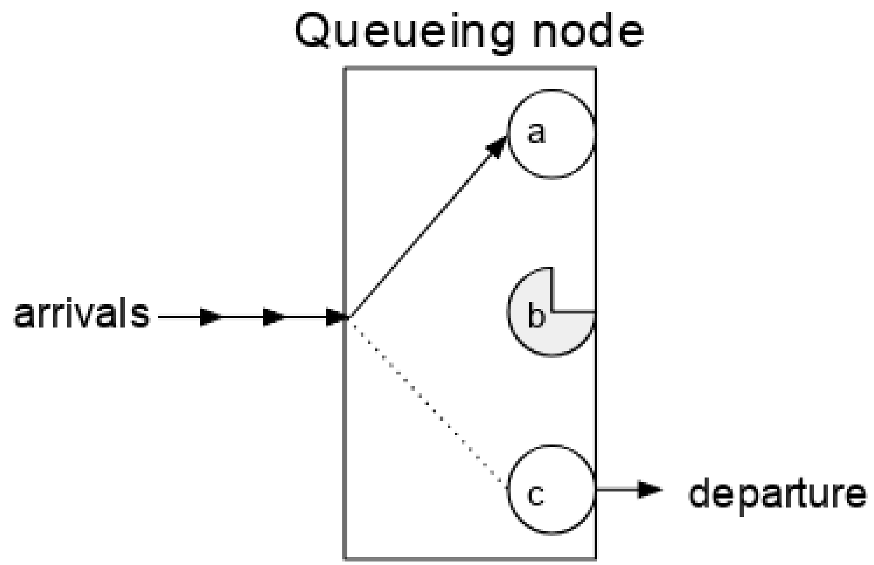
Thapar Institute of Engineering & Technology
(Deemed to be University)
Bhadson Road, Patiala, Punjab, Pin-147004
Contact No. : +91-175-2393201
Email : info@thapar.edu

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

An Introduction to Queues and Queueing Theory

Queue Model

1. **First Come First Served (FCFS)**
2. **Last Come First Served (LCFS)**
3. **Service in Random Order (SIRO)**
4. **Priority Service (PS)**



Queueing theory is the mathematical study of waiting lines, or queues.

A queueing model is constructed so that queue lengths and waiting time can be predicted.

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. Athanasios Papoulis, Probability Random Variables and Stochastic Processes, McGraw-Hill

An Introduction to Queues and Queueing Theory

1. First Come First Served

Process	Duration (sec)	Order	Arrival Time (sec)
P1	24	1	0
P2	3	2	0
P3	4	3	0

P1 waiting time is 0 second
P2 waiting time is 24second
P3 waiting time is 27second

Average waiting time is
= (0+24+27)/3=17 seconds

E.g. checkout counter at super market

Completion Time: Time at which process completes its execution.

Turn Around Time: Time Difference between completion time and arrival time. Turn Around Time = Completion Time – Arrival Time

Waiting Time(W.T): Time Difference between turn around time and burst time. Waiting Time = Turn Around Time – Burst Time

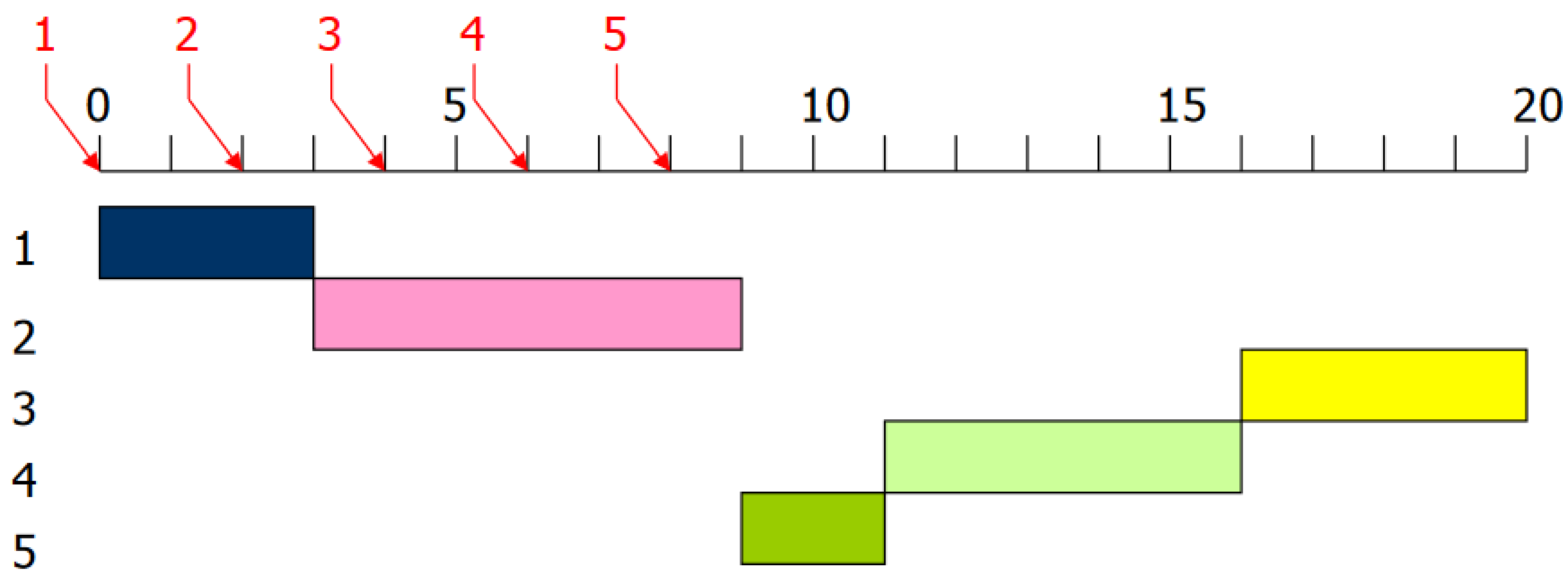
If arrival times as 0, turn around and completion times are same.

Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. Athanasios Papoulis, Probability Random Variables and Stochastic Processes, McGraw-Hill

An Introduction to Queues and Queueing Theory

2. Last Come First Served

- Execution of threads in reversed order of arrival at the ready list.
- Occupation of processor until end or voluntary yield.
- Remark: Rarely used in the pure form

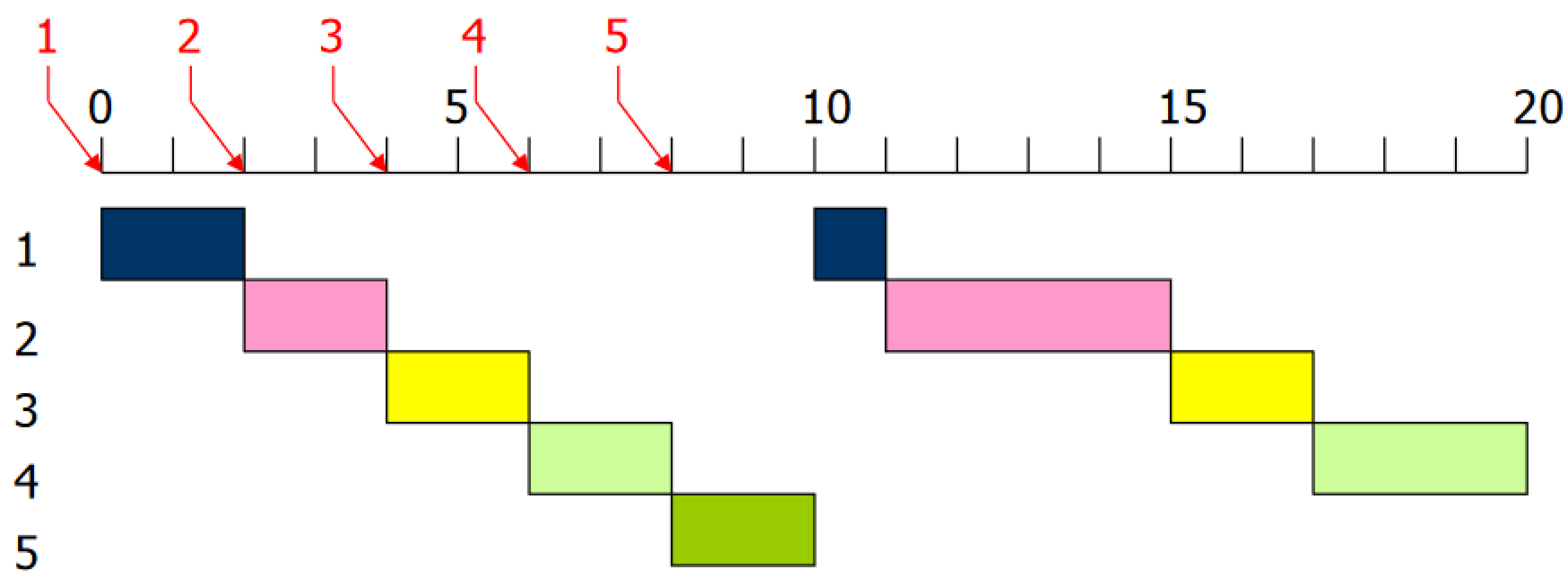


Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. Athanasios Papoulis, Probability Random Variables and Stochastic Processes, McGraw-Hill

An Introduction to Queues and Queueing Theory

Last Come First Served - Preemptive Resume

- Newly arriving thread at ready list preempts the currently running thread.
- Preempted thread is appended to ready list.
- In case of no further arrivals, the ready list is processed without preemption.
- Goal: Preference to short threads.
- A short thread has a good chance to finish before another thread arrives.
- A long thread is likely to be prempted several times.

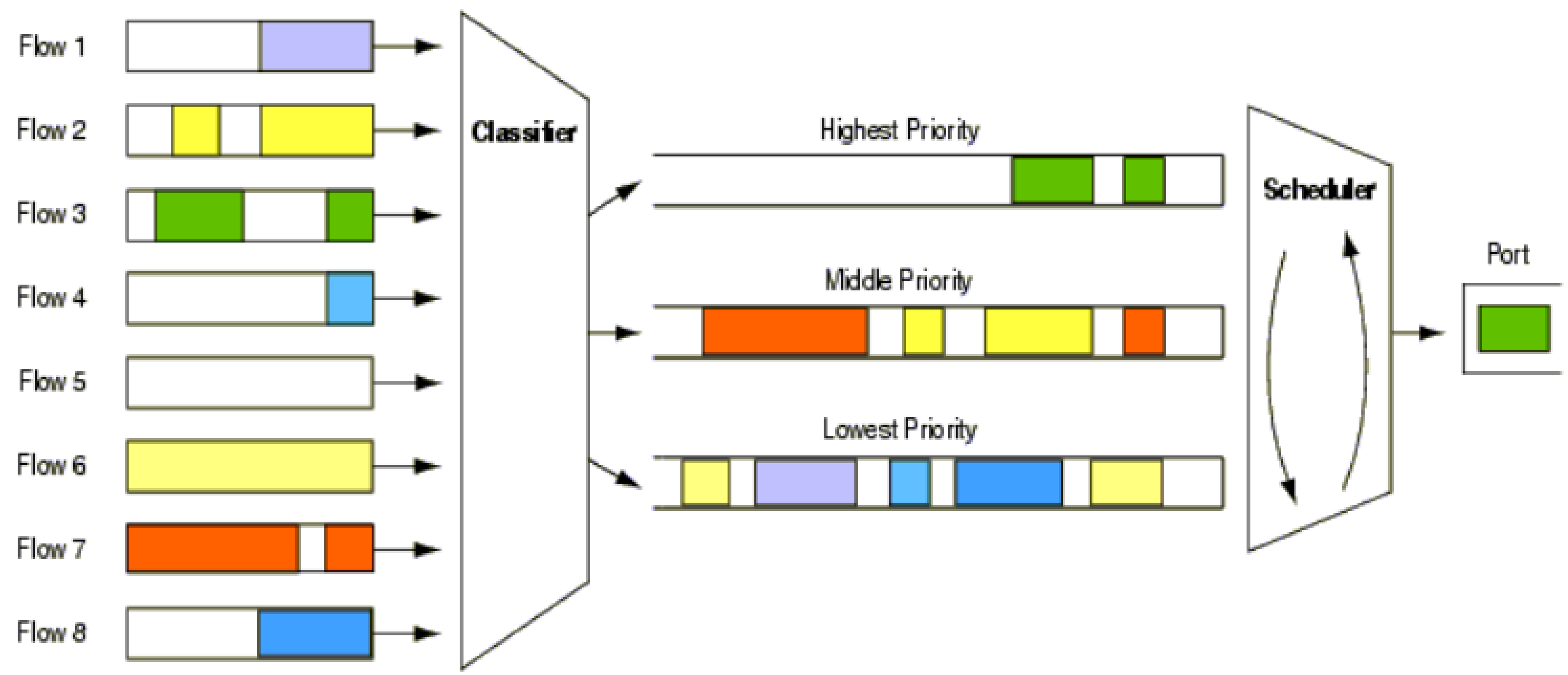


Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. Athanasios Papoulis, Probability Random Variables and Stochastic Processes, McGraw-Hill

An Introduction to Queues and Queueing Theory

3. **Service in Random Order (SIRO)**
If the next customer to enter service is randomly chosen from those customers waiting for the service it is called SIRO

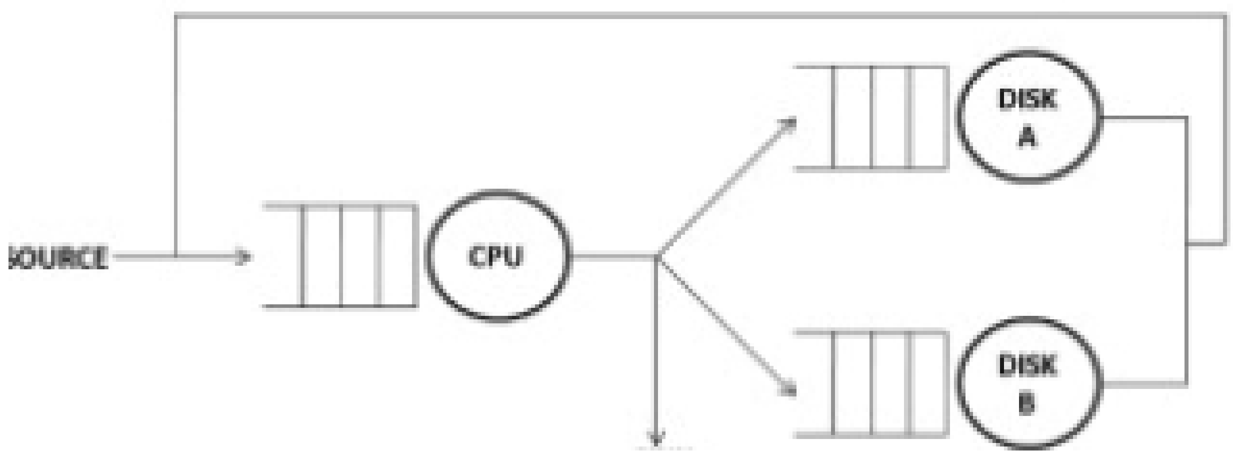
4. **Priority Service:**



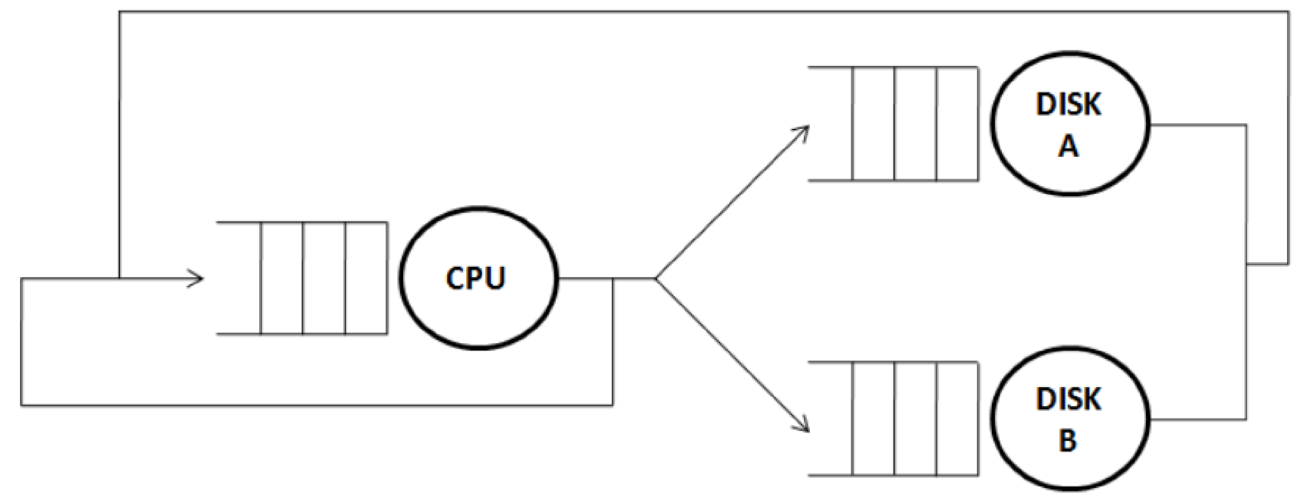
Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. Athanasios Papoulis, Probability Random Variables and Stochastic Processes, McGraw-Hill

An Introduction to Queues and Queueing Theory

An Open Queueing Network



A Closed Queueing Network



Reference: 1. D. Gross and C. Harris, Fundamentals of Queueing Theory, 3rd Edition, Wiley 2. Athanasios Papoulis, Probability Random Variables and Stochastic Processes, McGraw-Hill

Little’s Theorem

Little’s Theorem states:

$N = \lambda T$

Where N is the average number of customers in a queue, T is the average time a customer spends queuing and λ is the average rate of arrivals to the queue.

If a lot of people are in a queue (N is large) then they will have long delays (T is large);

if few people arrive in a queue (λ is small) then the average number of people in the queue is small (N is small).

Let us first make precise the definitions of N, λ and T and then make clear the assumptions on which the theorem rests.

N(τ) is the number of customers in the system at time τ.

α(τ) is the number of customers who arrived in the interval [0,τ].

β(τ) is the number of customers who have departed in the interval [0,τ].

ti is the time at which the ith customer arrived.

T(i) is the time spent queuing by the ith customer.

If Nt is the mean value of N(τ) taken over the interval [0,t] then it is clear that:

$$N_t = \frac{1}{t}\int_0^t N(\tau)d\tau$$

$$N = \lim_{t\to\infty} N_t$$

(Note that this limit is not guaranteed to exist — imagine, for example, a queue which keeps growing.) If the limit exists, N is the steady state time average of N(τ).

We can next define the average arrival rate over the time period [0,t].

$$\lambda_t = \frac{\alpha(t)}{t}$$

and, again, we assume that the following limit exists:

$$\lambda = \lim_{t\to\infty} \lambda_t$$

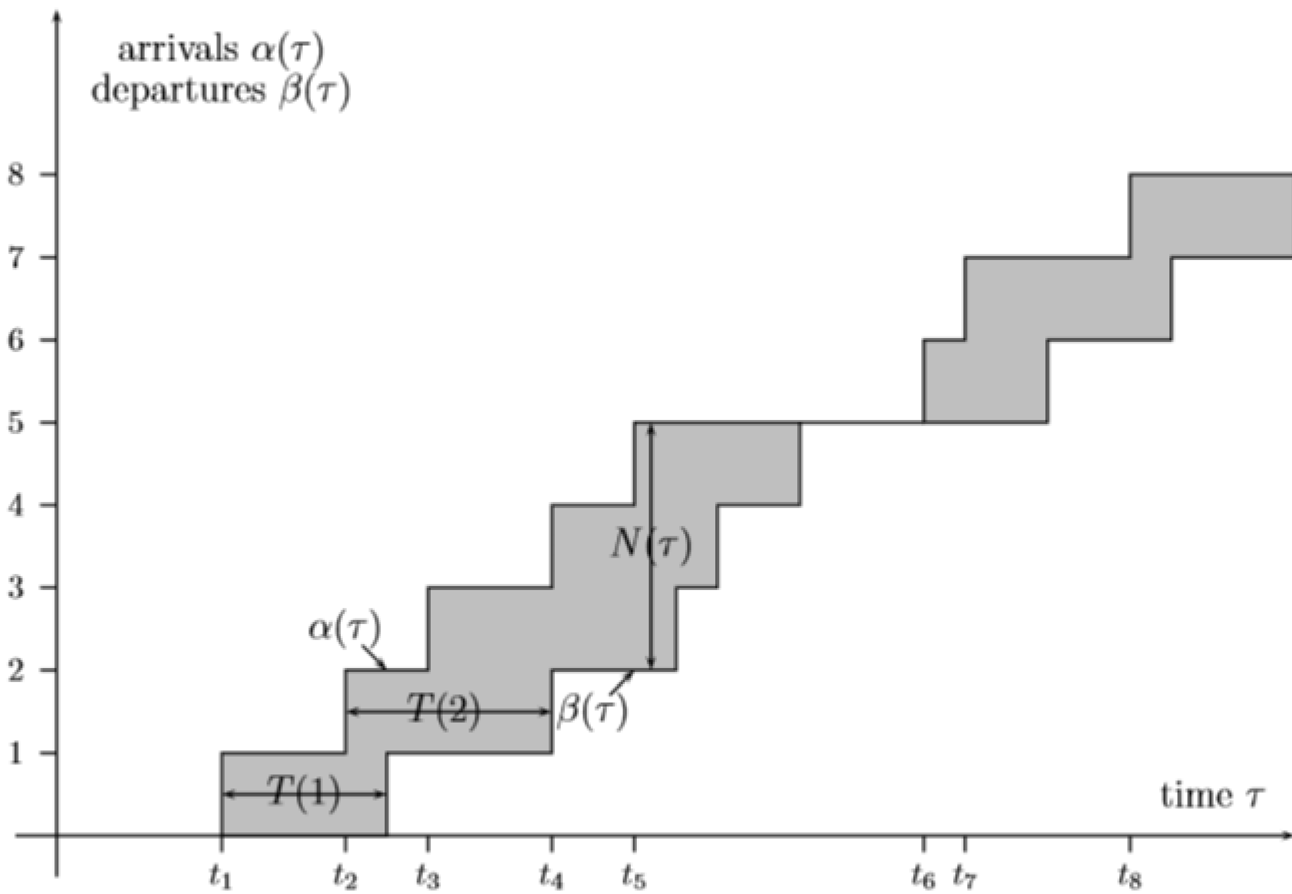
Finally, the average delay experienced by those customers who enter the system at times in [0,t] is given by:

$$T_t = \sum_{i=1}^{\alpha(t)} \frac{T(i)}{\alpha(t)}$$

And, for a third time, we assume that the following limit exists:

$$T = \lim_{t\to\infty} T_t$$

Little's Theorem Proof assuming FIFO



$$N(\tau) = \alpha(\tau) - \beta(\tau)$$

It is clear that if we choose a time t when the system again becomes empty then we can calculate the area of the shaded area A(t):

$$A(t) = \int_0^t N(\tau)d\tau$$

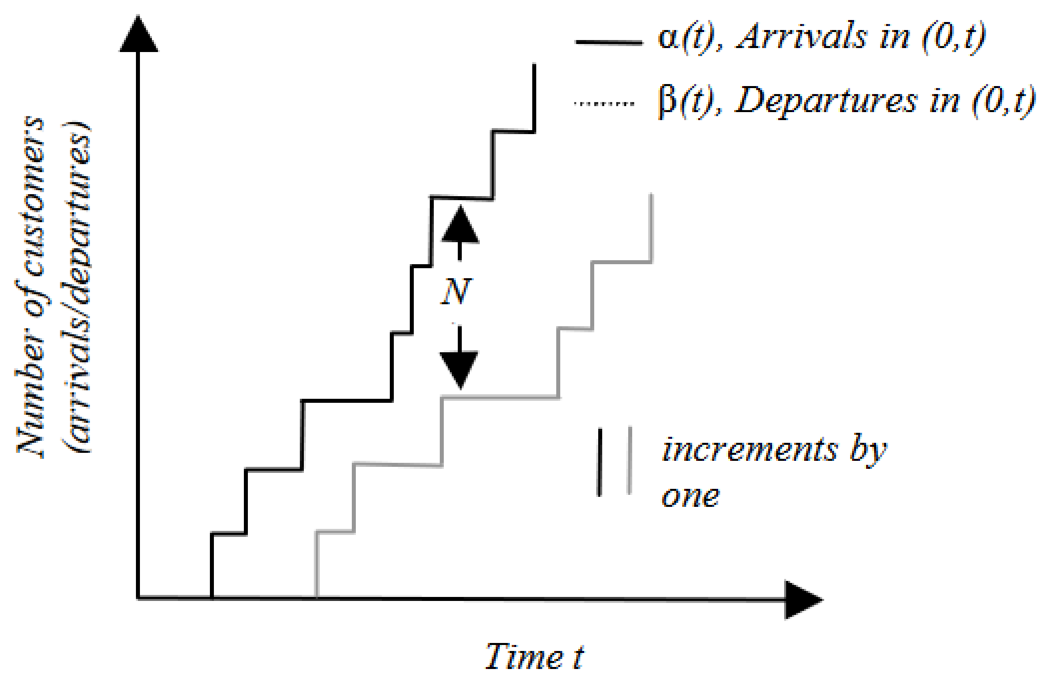
However, equally, we can consider the shaded area to be composed of horizontal strips of height 1 and width T(i) (for the ith customer). In this case, we have:

$$A(t) = \sum_{i=1}^{\alpha(t)} T(i)$$

$$\frac{1}{t}\int_0^t N(\tau)d\tau = \frac{1}{t}\sum_{i=1}^{\alpha(t)} T(i) = \frac{\alpha(t)}{t}\frac{\sum_{i=1}^{\alpha(t)} T(i)}{\alpha(t)}$$

$$N_t = \lambda_t T_t$$

Graphical verification of Little's Theorem



Graphical Illustration/Verification of Little's Result

Consider the time interval $(0,t)$ where t is large, i.e. $t\rightarrow\infty$

$$Area(t) = \text{area between } \alpha(t) \text{ and } \beta(t) \text{ at time } t = \int_0^t [\alpha(t) - \beta(t)]dt$$

$$\text{Average Time } W \text{ spent in system} = \lim_{t\rightarrow\infty} \frac{Area(t)}{\alpha(t)}$$

$$\text{Average Number } N \text{ in system} = \lim_{t\rightarrow\infty} \frac{Area(t)}{t} = \lim_{t\rightarrow\infty} \frac{\alpha(t)}{t} \frac{Area(t)}{\alpha(t)}$$

Since, $\lambda = \lim_{t\rightarrow\infty} \dfrac{\alpha(t)}{t}$

Therefore, $N = \lambda W$