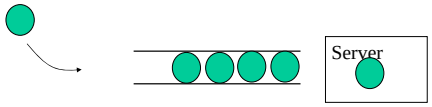
Queueing Theory

- A very basic introduction to queueing theory.
- Queueing theory is a very complicated topic, here we will deal with the fundamentals only



Queueing Theory

- An arrival requests, a server to process/serve them, and a queue for arrivals that arrive while the server is busy
- Examples ….
- Usually, defined as A/B/c/K
  - A is the interarrival time distribution
  - B is the service time distribution
  - c is the number of servers
  - K is the system capacity (queue + being served)

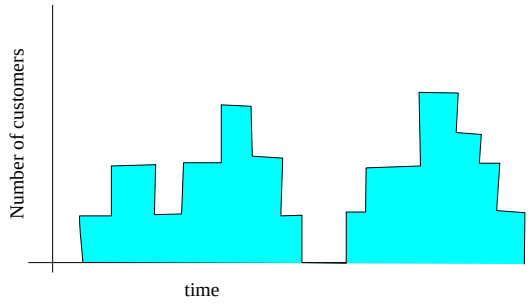
Queueing Theory

- For A and B, we usually use
  - M for exponential distribution
  - D for deterministic distribution
  - G for General distribution (GI)
- When the system capacity is infinite, we use A/B/c
- The simplest queue is M/M/1
- Examples …

Queueing Theory

- λ is the arrival rate (arrival per second)
- $1/\lambda$ is the interarrival time
- If the arrival is Poisson, the interarrival is exponential
- μ is the service rate, $1/\mu$ is the service time.
- The traffic intensity (occupancy) is defined as $\rho=\lambda/\mu$

Queueing Theory



Queueing Theory

$$\rho = \frac{\lambda}{\mu}$$

$$N = \frac{\rho}{1-\rho}$$

$$T_{total} = \frac{1}{\mu - \lambda}$$

$$N = \frac{\lambda}{T}$$