Queuing System Analysis

- General description of a queuing system 
  \( A / B / m / K / M \)
- \( A \): Arrival process
  - \( M \) → Poisson (Markovian)
  - \( D \) → Deterministic
  - \( G \) → General distribution...
- \( B \): Service time process
- \( m \): Number of servers
- \( K \): System capacity
- \( M \): Customer population

For all systems to be analyzed, we define
- \( N \): Average number of customers in the system
- \( T \): Average customer delay in the system
- \( N_Q \): Average number of customers waiting in queue
- \( W \): Average customer waiting time in queue
- \( p_n \): Probability of \( n \) customers in the system

From these definitions, we get
\[
N = \sum_{n=0}^{\infty} np_n, \quad \text{and using Little’s Theorem,}
\]
\[
T = \frac{N}{\lambda} \quad \text{and} \quad W = \frac{N_Q}{\lambda}.
\]

Counting Process

A stochastic process \(\{A(t) | t \geq 0\}\) is a counting process if
- \(A(0) = 0\)
- \(A(t)\) assumes non-negative integer values
- \(s < t \Rightarrow A(s) \leq A(t)\)
- \(A(t) - A(s)\) is the "counting events" in \((s, t] \cap s \geq t\)

Poisson Process

A stochastic process \(\{A(t) | t \geq 0\}\) is a Poisson process with rate \(\lambda\) if
- \(A(t)\) is a counting process
- The number of arrivals that occur in disjoint time intervals are independent
- The distribution of number of arrivals in any time interval depends only on the length of the interval
- \(P(A(t) = n) = \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad \forall s, t, n\)
- \(P(A(\Delta t) = 1) = \lambda \Delta t + o(\Delta t), \quad \lambda > 0, \lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0\)
- \(P(A(\Delta t) \geq 2) = o(\Delta t)\)
\[
\Rightarrow P(A(\Delta t) = 0) = 1 - \lambda \Delta t + o(\Delta t)
\]
Properties of Poisson Process

- Expected value and variance of Poisson pdf:
  \( E[X] = \lambda, \sigma^2 = \lambda \)
- Distribution of interarrival times, \( T \):
  \( T \) is an exponentially distributed random variable

\( \lambda \) is a constant, is If equations hold, then we obtain the differential equation:

\[ \lambda = \alpha \lambda \]

\( \lambda_1, \lambda_2 \) are the respective, and rates are independent Poisson processes. Then, if two processes have rates \( \lambda_1, \lambda_2 \), respectively, the resulting process has rate \( \lambda_1 + \lambda_2 \).

\[ P(n, t) = \lambda^n e^{-\lambda t} / n! \]

Solving the differential equation, we obtain the Poisson process pdf:

\[ P(t) = \frac{e^{-\lambda t} \lambda^t}{t!} \]
Exponential Distribution

- We assume that the customer service times follow exponential distribution with parameter $\mu$, i.e., if $T_s$ is the service time of the $n^{th}$ customer, then $P(T_s \leq t) = 1 - e^{-\mu t}$, $t \geq 0$.
  Moreover, $T_s$ are mutually independent and independent of all interarrival times.

- Memoryless Property:
  \[
P(T_{s+1} \leq t \mid T_s > t_0) = \frac{P(T_{s+1} \leq t - t_0 \mid T_s > t_0)}{P(T_s > t_0)} = \frac{1 - e^{-\mu (t - t_0)}}{e^{-\mu t_0}} = e^{-\mu t_0} = P(T_{s+1} \leq t \mid T_s > t_0)
  \]

M/M/1 Queue

- Poisson arrivals, exponential service time distribution, one server, infinite queue capacity
- In M/M/1 systems, the future number of customers in the system depends on the past only through the current number of customers
- Therefore, we can use discrete state, continuous time Markov chain analysis
- Let us define the system state $S_t$ at time $t$ as the number of customers in the system

\[P(t) = \sum_{i=0}^{\infty} P(S_t = i) = 1 \]

\[P(S_{t+\delta} = j | S_t = i) = \sum_{j=0}^{\infty} \delta P(S_{t+\delta} = j \mid S_t = i) = \sum_{j=0}^{\infty} \delta P(S_{t+\delta} = j \mid S_t = i) = \delta \sum_{j=0}^{\infty} \lambda \delta + o(\delta)
\]

Exponential Distribution

- The memoryless property of the exponential distribution has two practical meanings
  - Poisson arrivals:
    The time until the next arrival is independent of the time of the last arrival
  - Exponential service time:
    The remaining service time for a customer being served is independent of the elapsed service time

M/M/1 Queue

- \[P(t) = \sum_{i=0}^{\infty} P(S_t = i), \sum_{i=0}^{\infty} P(t) = 1 \]

- \[P(S_{t+\delta} = j | S_t = i) = \sum_{j=0}^{\infty} \delta P(S_{t+\delta} = j \mid S_t = i) = \sum_{j=0}^{\infty} \delta P(S_{t+\delta} = j \mid S_t = i) = \delta \sum_{j=0}^{\infty} \lambda \delta + o(\delta)
\]

- Let the customer arrive with rate $\lambda$ and depart with rate $\mu$.
  1) \[P(S_{t+\delta} = n+1 | S_t = n) = \lambda \delta + o(\delta), \quad n = 0, 1, 2, \ldots \]
  2) \[P(S_{t+\delta} = n-1 | S_t = n) = \mu \delta + o(\delta), \quad n = 1, 2, 3, \ldots \]
  3) \[P(S_{t+\delta} = 0 | S_t = 0) = o(\delta) \]

\[P(t + \delta) = P(t)(1 - \lambda \delta + o(\delta)) + P(t+\delta)(\mu \delta + o(\delta)) + o(\delta)
\]

\[P(t + \delta) = P(t)(1 - \lambda \delta + o(\delta)) + P(t+\delta)(\mu \delta + o(\delta)) + P(t+\delta)(\lambda \delta + o(\delta))
\]

\[P(t + \delta) = P(t)(1 - \lambda \delta + o(\delta)) + P(t+\delta)(\mu \delta + o(\delta)) + o(\delta), \quad k > 0
\]
### M/M/1 Queue

Let us calculate the performance measures for this system:

\[ N = \sum p_i = \sum \frac{1}{(1-\rho)} \mu^i = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda} \]

\[ T = \frac{N}{\lambda} = \frac{1}{\mu - \lambda} \]

\[ N_0 = N - \rho = \frac{\rho}{1-\rho} \]

\[ W = \frac{N_0}{\lambda} = \frac{\rho}{\lambda(1-\rho)} \]

- Note that we assumed that arrival rate is smaller than the service rate. If this condition is not met, M/M/1 system does not reach steady state.

### M/M/1 Queue

\[ P_i = \frac{1}{i!} \left( \frac{\lambda}{\mu} \right)^i P_1 \]

\[ \left( \frac{\lambda}{\mu} \right)^i P_1 = \lambda^i P_1 + \mu P_i \]

\[ \Rightarrow \left( \frac{\lambda}{\mu} \right)^i P_1 = \lambda P_1 + \mu P_i \]

Plug into the general equation:

\[ P_i = \rho^i P_1, \quad \rho = \frac{\lambda}{\mu} \]

\[ \sum_i P_i = \sum_i \rho^i P_1 = \rho \sum_i \rho^i = 1 \]

If \( \lambda < \mu, P_i = \frac{1}{1-\rho} \]

\[ P_i = \rho^i (1-\rho), \quad k \geq 0 \]

- In a Markov chain, the sum of outgoing flow rates through a closed contour is equal to the sum of incoming flow rates.
- With the right selection of the contours, one can significantly reduce the steady state probability calculations.