Advanced Network Technologies

Queueing Theory

Dr. Wei Bao | Lecturer
School of Computer Science

THE UNIVERSITY OF SYDNEY

Drag picture to placeholder or clic

Outline

- Markov Chain
- Queueing System and Little's Theorem
- M/M/1 Queue foundations
- M/M/1 Queue

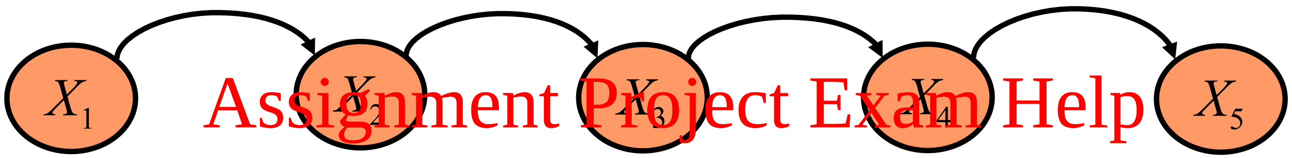
- Markov Chain

Markov Chain

- A stochastic process
  - $X_1, X_2, X_3, X_4 \ldots$
  - $\{X_n, n = 1, 2, \ldots\}$
  - X_n takes on a fini of possible values.
  - $X_n \in \{1,2, \ldots, S\}$
  - i: ith state
  - Markov Property: The state of the system at time $n+1$ depends only on the state of the system at time n

$$\Pr\left[X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_2 = x_2, X_1 = x_1\right] = \Pr\left[X_{n+1} = x_{n+1} \mid X_n = x_n\right]$$

Markov Chain



- *Stationary Assumption:* Transition probab ependent of time (n)

$$\Pr\left[X_{n+1} = b \mid X_n = a\right] = p_{ab}$$

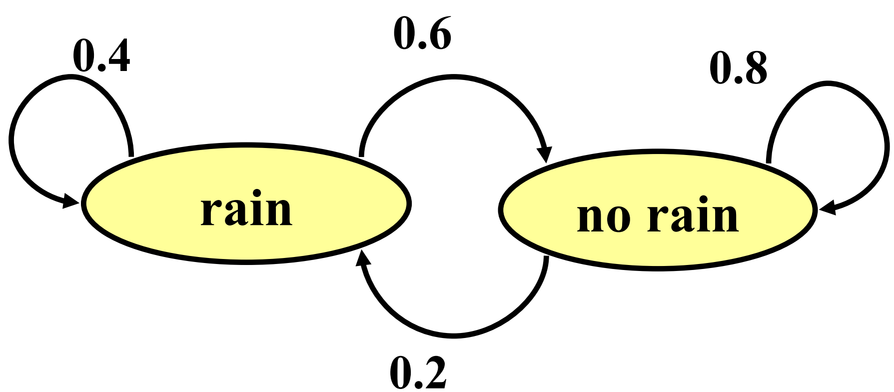
An example

Weather:

- **raining today** → **40% rain tomorrow**
  → **60% no rain tomorrow**

- **not raining toda** → **orrow**
  → **80% rrow**

Stochastic FSM:



An example

Weather:

- **raining today** → **40% rain tomorrow**

  → **60% no rain tomorrow**

- **not raining toda** → **orrow**

  → **80% rrow**

Matrix:

$$P = \begin{pmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{pmatrix}$$

- Stochastic matrix:
  Rows sum up to 1

Transition matrix

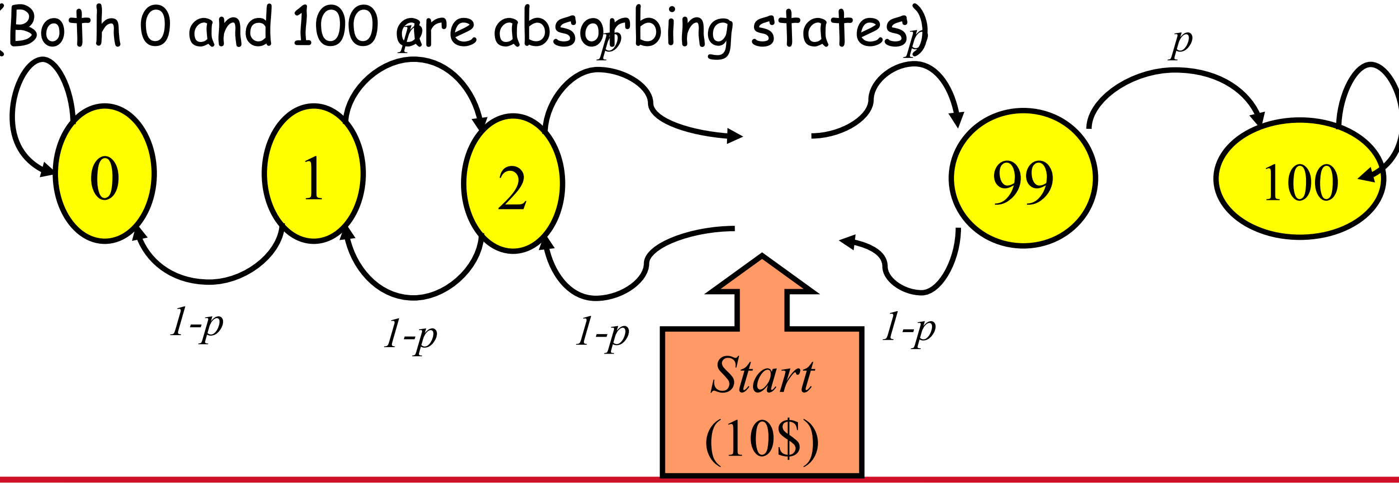
$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1S} \\ p_{21} & p_{22} & \cdots & p_{2S} \\ \vdots & \vdots & & \vdots \\ p_{S1} & p_{S2} & \cdots & p_{SS} \end{pmatrix}$$

THE UNIVERSITY OF SYDNEY

Gambler's Example

- Gambler starts with \$10
- At each play we have one of the following:
  - Gambler wins \$1 with probability p
  - Gambler looses
  - Game ends wh r gains a fortune of \$100

(Both 0 and 100 are absorbing states)



Gambler's Example

- **transient state**

if, given that we start in state *i*, there is a non-zero probability that we will never return to *i*

- **recurrent state**

Non-transient

- **absorbing** state

impossible to leave this state.

Coke vs. Pepsi Example

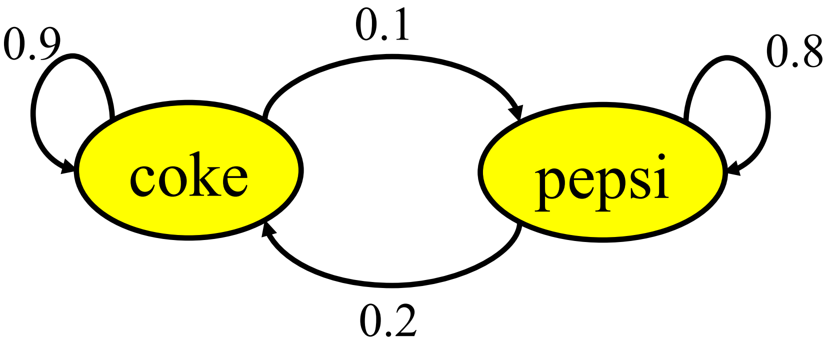
- Given that a person's last cola purchase was Coke, there is a 90% chance that his next cola purchase will also be Coke.

- If a person's last cola purchase was Pepsi, there is an 80% chance … chase will also be Pepsi.

stat… n diagram:

transition matrix:

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$



Coke vs. Pepsi Example

Given that a person is currently a Pepsi purchaser, what is the probability that he will purchase Coke two purchases from now?

Pr[Pepsi→?→Coke

Pr[Pepsi→Coke→C oke] =

0.2 * 0.9 + 0. = 0.34

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix}$$

Pepsi → ?   ? → Coke

Coke vs. Pepsi Example

Given that a person is currently a Coke purchaser, what is the probability that he will purchase Pepsi **three** purchases from now?

$$P^3 = \begin{bmatrix}0.9 & 0.1\\0.2 & 0.8\end{bmatrix}\begin{bmatrix} & \\ 0.34 & 0.\end{bmatrix}\begin{bmatrix}.781 & 0.219\\.438 & 0.562\end{bmatrix}$$

Coke vs. Pepsi Example

- Assume each person makes one cola purchase per week
- Suppose 60% of all people now drink Coke, and 40% drink Pepsi
- What fraction of people will be drinking Coke three weeks from now?

$$P = \begin{bmatrix} 0.[illegible] & [illegible] \\ 0.2 & 0.8 \end{bmatrix} \qquad \begin{bmatrix} [illegible] & 0.219 \\ [illegible] & 0.562 \end{bmatrix}$$

$$\Pr[X_3 = \text{Coke}] = 0.6 * 0.781 + 0.4 * 0.438 = 0.6438$$

Q_i - the distribution in week i

$Q_0 = (0.6, 0.4)$ - initial distribution

$Q_3 = Q_0 * P^3 = (0.6438, 0.3562)$

Coke vs. Pepsi Example

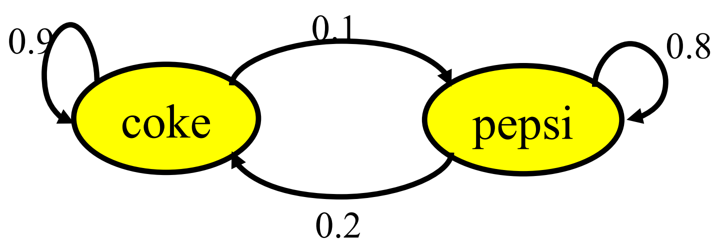
Simulation:

2/3

$\begin{bmatrix} 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 \end{bmatrix}$

stationar

$\Pr[X_i = \text{Coke}]$



week - i

THE UNIVERSITY OF SYDNEY

Steady State and Stationary distribution

$$\lim_{n \to \infty} P(X_n = i) = \pi_i$$

$n \to \infty$

$$\boldsymbol{\pi} = \boldsymbol{\pi} \cdot P$$

Steady State and Stationary distribution

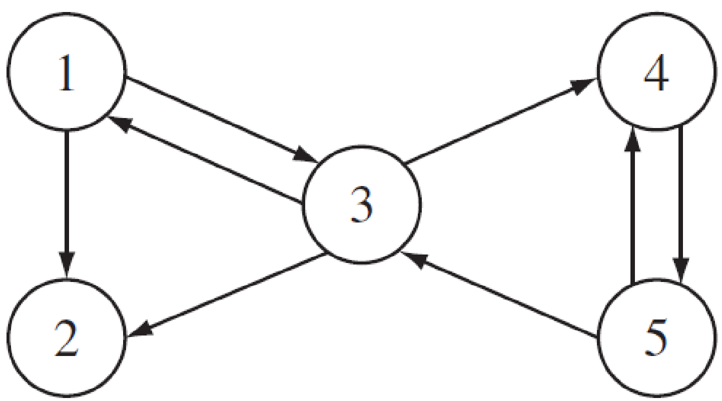
$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0 & \end{bmatrix} \qquad \boldsymbol{\pi} = \begin{bmatrix} \tfrac{2}{3} & \tfrac{1}{3} \end{bmatrix}$$

$$P^{10} = \begin{bmatrix} 0.6761 & 0.3239 \\ 0.6478 & 0.3522 \end{bmatrix} \qquad \begin{bmatrix} 0.6667 & 0.3333 \\ 0.6667 & 0.3333 \end{bmatrix}$$

$$\begin{bmatrix} \tfrac{2}{3} & \tfrac{1}{3} \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} \tfrac{2}{3} & \tfrac{1}{3} \end{bmatrix}$$

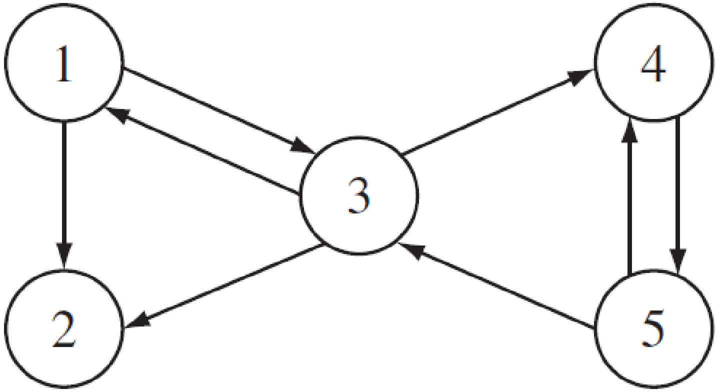
Steady State and Stationary distribution

PageRank: A Web surfer browses pages in a five-page Web universe shown in figure. The surfer selects the next page to view by selecting with equal probability from the pages pointed to by the current page. If a page has no outgoing link (e.g., page 2) … y of the pages in the universe with … e probability that the surfer views page i.



Steady State and Stationary distribution

Transition matrix P



Steady State and Stationary distribution

Stationary Distribution:
Solve the following equations:

$$\sum_{i=1} \pi_i$$

π = (0.12195, 0.18293, 0.25610, 0.12195, 0.317072)

Search engineer. page rank: 5, 3, 2, 1, 4

› Queueing System and Little’s Theorem

Queueing System

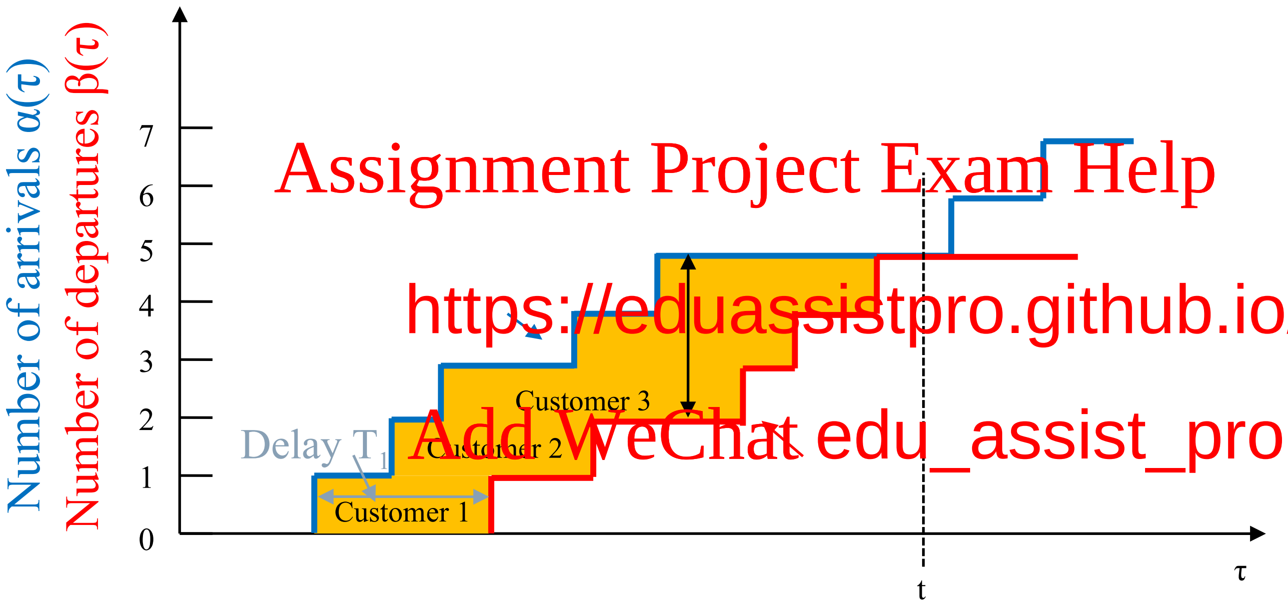
- Customers = Data packets
- Service Time = Packet Transmission Ti packet length and transmission speed)
- Queueing delay = time spent in buffer before transmission
- Average number of customers in systems
– Typical number of customers either waiting in queue or undergoing service
- Average delay per customer
– Typical time a customer spends waiting in queue + service time

$= \frac{N}{T} = \frac{Number}{Time}$

- W: average waiting time in queue
- X: average service time
- T: average time spent in system ($T = W + X$)
- N_Q = average number of customers in queue
- ρ = utilization = average number of customers in service
- N = average number of customer in system ($N = N_Q + \rho$)
- Want to show later: $N = \lambda T$ (Little's theorem)
- λ Average arrival rate

THE UNIVERSITY OF SYDNEY

Little’s Theorem



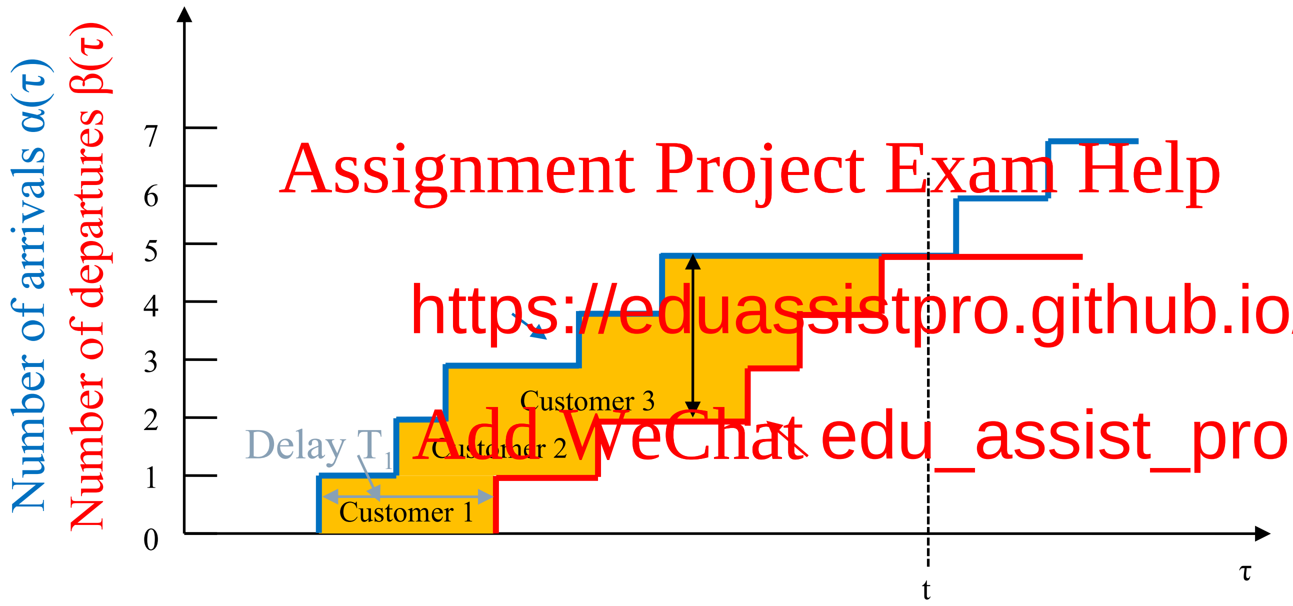
$\alpha(t)$ = Number of customers who arrived in the interval $[0, t]$

$\beta(t)$ = Number of customers who departed in the interval $[0, t]$

$N(t)$ = Number of customers in the system at time t, $N(t) = \alpha(t) - \beta(t)$

T_i = Time spent in the system by the i-th arriving customer

Little’s Theorem



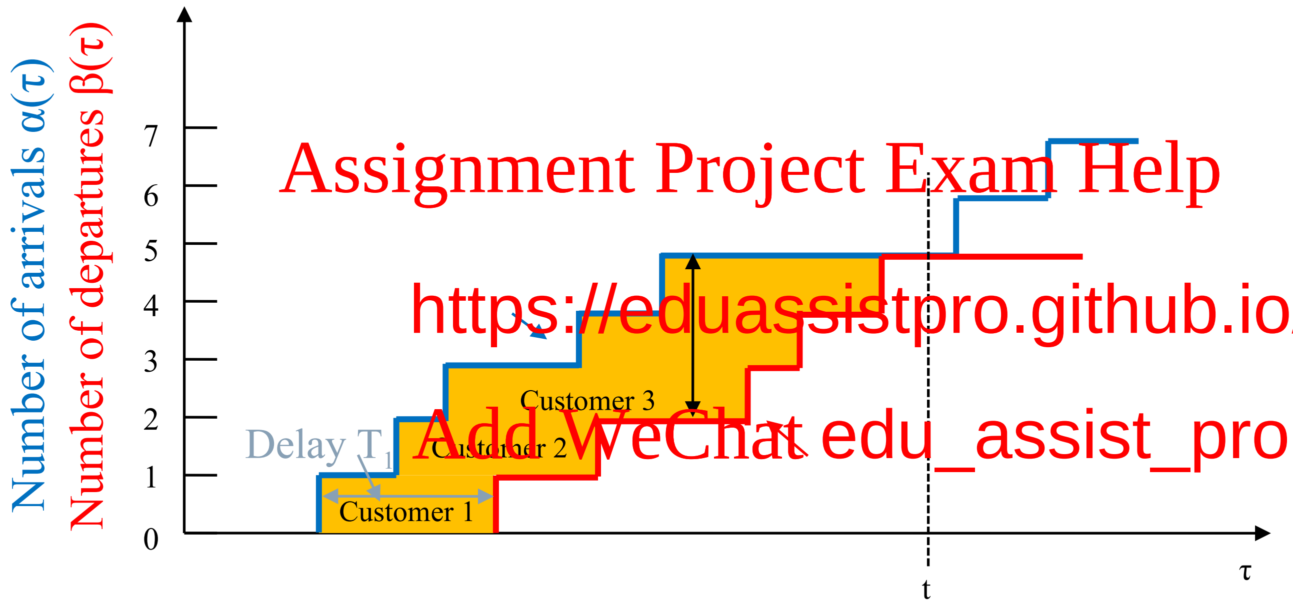
Average # of customers until t

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

Average # of customers in long-term

$$N = \lim_{t\to\infty}\frac{1}{t}\int_0^t N(\tau)d\tau$$

Little’s Theorem



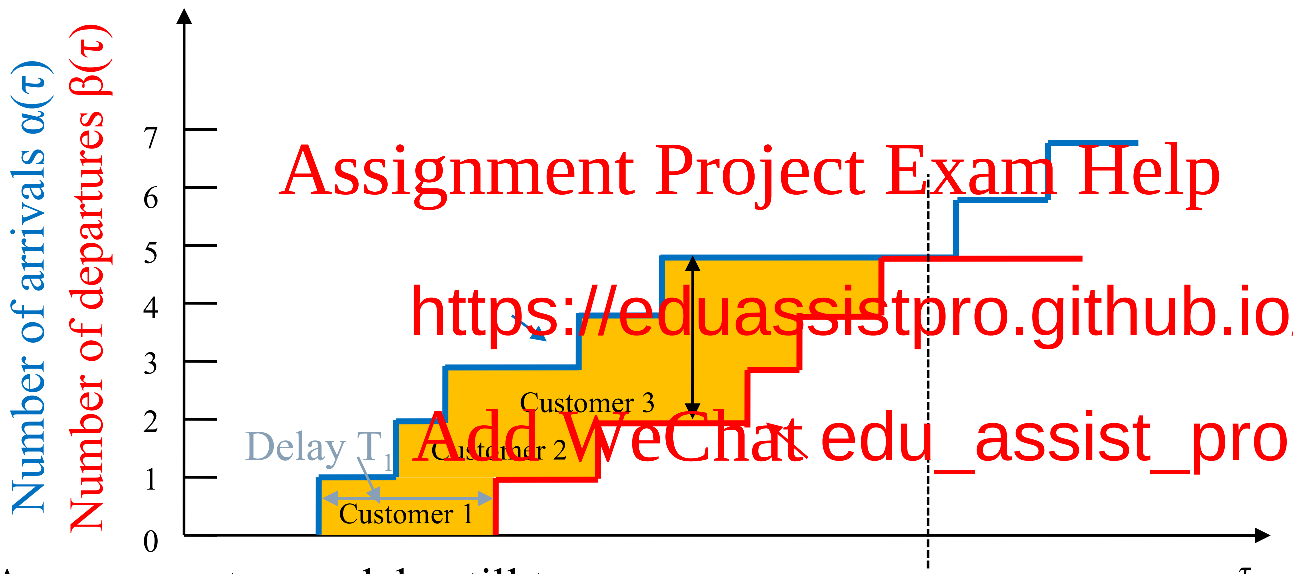
Average # arrival rate until t

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

Average # arrival rate in long-term

$$\lambda = \lim_{t \to \infty} \frac{\alpha(t)}{t}$$

Little's Theorem



Average customer delay till t

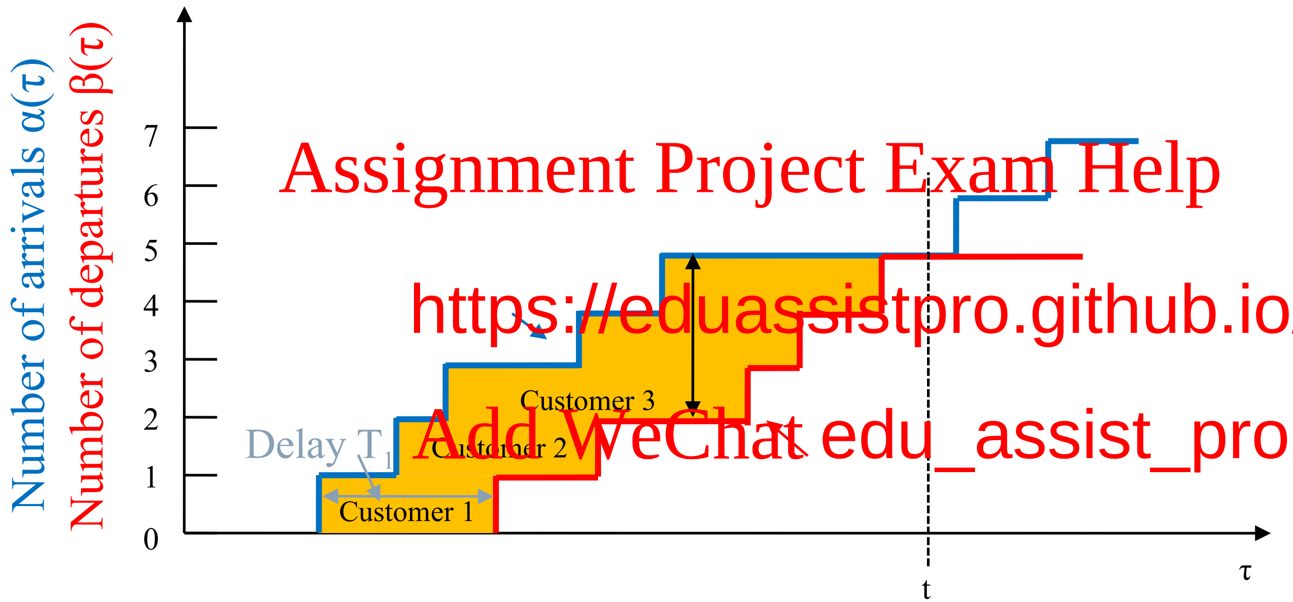
$$T_t = \frac{\sum_{i=1}^{\alpha(t)} T_i}{\alpha(t)}$$

Average customer delay in long-term

$$T = \lim_{t \to \infty} \frac{\sum_{i=1}^{\alpha(t)} T_i}{\alpha(t)}$$

THE UNIVERSITY OF SYDNEY

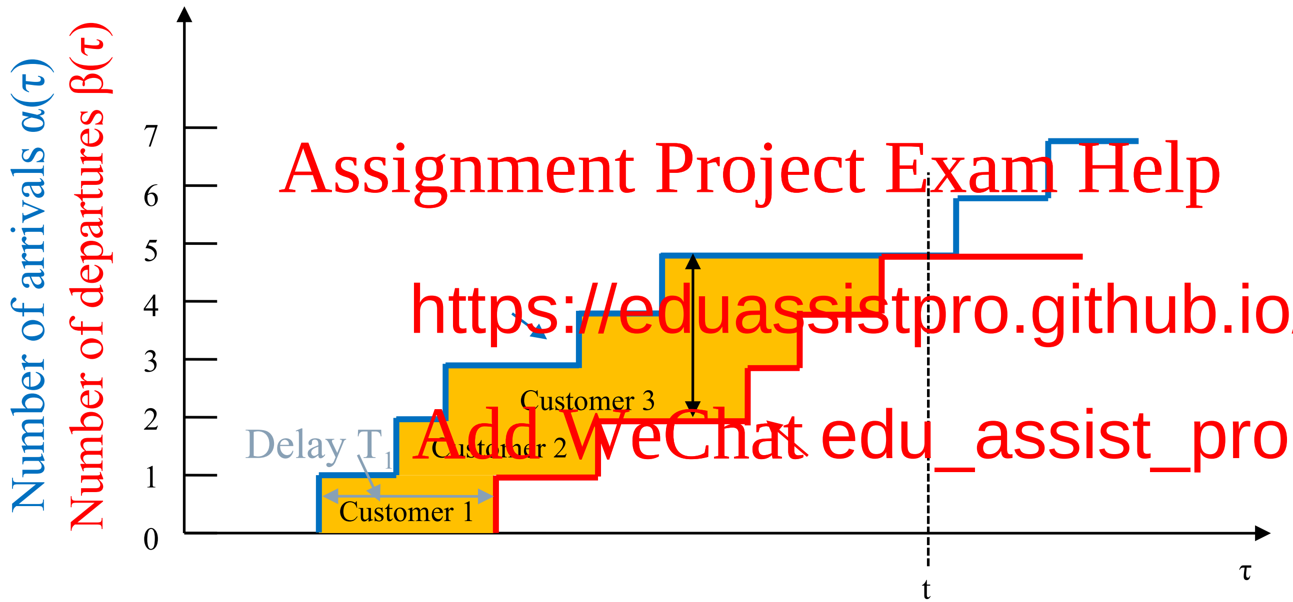
Little’s Theorem



Shaded area when the queue is empty: two ways to compute

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

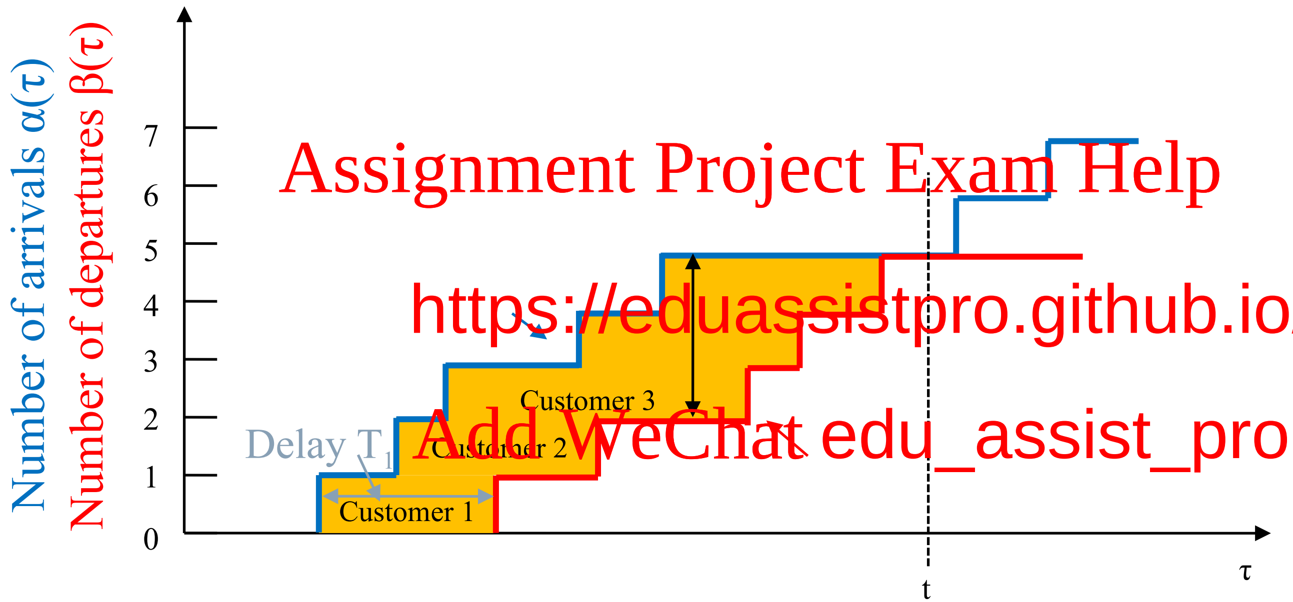
Little’s Theorem



Shaded area when the queue is empty: two ways to compute

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

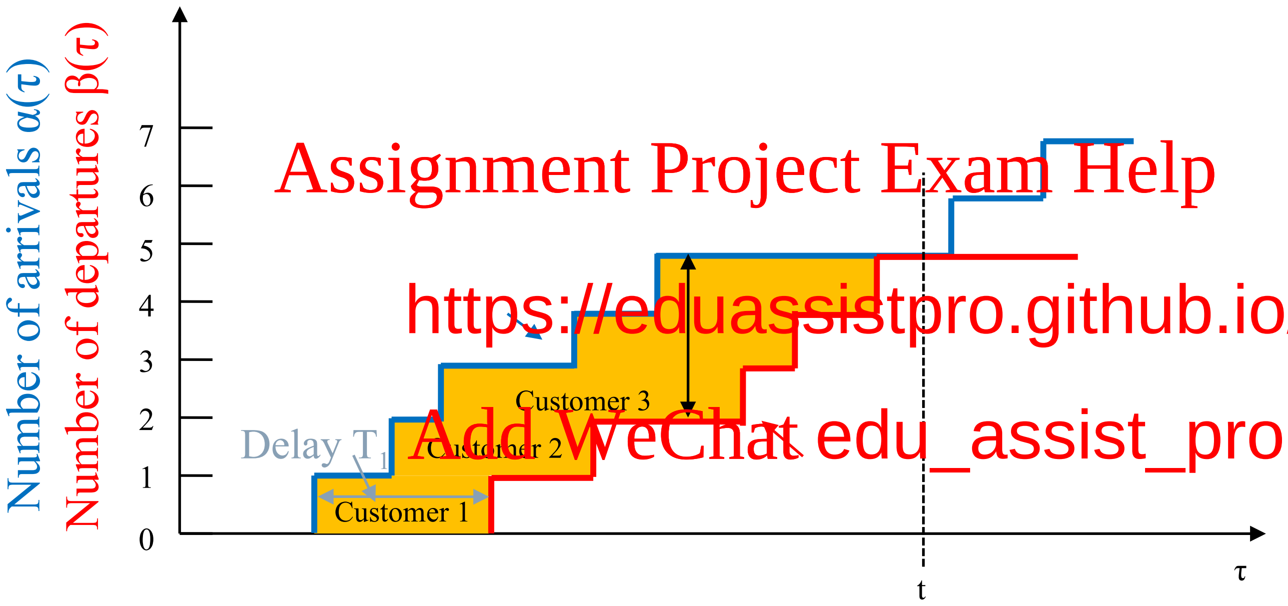
Little's Theorem



Shaded area when the queue is empty: two ways to compute

$$\underbrace{\frac{1}{t}\int_0^t N(\tau)d\tau}_{N_t} = \underbrace{\frac{\alpha(t)}{t}}_{\lambda_t}\ \underbrace{\frac{\sum_{i=1}^{\alpha(t)} T_i}{\alpha(t)}}_{T_t}$$

Little’s Theorem



Shaded area when the queue is empty: two ways to compute

$$N_t = \lambda_t T_t$$

$$N = \lambda T$$

Note that the above Little’s Theorem is valid for any service disciplines (e.g., first-in-first-out, last-in-first-out), interarrival and service time distributions.

λ

R Number

$$\lambda = \frac{N}{T} \qquad Rate = \frac{Number}{Time}$$

- $N = \lambda T$
- $N_Q = \lambda W$
- ρ = proportion of time that the server is busy = λX
- $T = W + X$
- $N = N_Q + \rho$

› M/M/1 Queue foundations

Exponential Distribution

- Exponential Distribution
- • The cumulative distribution function $F(x)$ and probability
- density function $f(x)$ are:
- $F(x) = 1 - e^{-\lambda x}$ $f(x) = \lambda e^{-\lambda x}$

The mean is equal to its standard deviation: $E[X] = \sigma_X = 1/\lambda$

Memoryless Property

- $P(X > s + t \mid X > t) = P(X > s)$ for all $s, t \geq 0$
- The only continuous distribution with this property
- Practice Q2 in Tutorial Week 4

Other Properties of Exponential Distribution

- Let $X_1, \ldots, X_n$ be i.i.d. exponential r.v.s with mean $1/\lambda$,
- then $X_1 + X_2 + \ldots + X_n$ (Practice Q1 in Tutorial Week 4)

- gamma distributio
- Suppose X_1 and X_2 are … h means
- $1/\lambda_1$ and $1/\lambda_2$, respecti…

$$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Counting Process

- A stochastic process $\{N(t), t \geq 0\}$ is a counting process if $N(t)$ represents the total # of events that have occurred up to time t.
- 1. $N(t) \geq 0$ and $N(t)$ is integer valued.
- 2. If $s < t$, then $N(s) \leq N(t)$
- 3. For $s < t$, $N(t) - N(s)$ in (s, t)
- • Examples:
- – # of people who have entered a particular t
- – # of packets sent by a mobile phone
- • A counting process is said to be independent increment if # of events which occur in disjoint time intervals are independent.
- • A counting process is said to be stationary increment if the distribution of # of events which occur in any interval of time depends only on the length of the time interval.

Poisson Process

- The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process having rate $\lambda > 0$, if
- 1. $N(0) = 0$
- 2. The process has independent increments (i.e., # of events which occur in disjoint time intervals ar

- – for $0 < t_1 < t_2 < t_3 < t_4$,

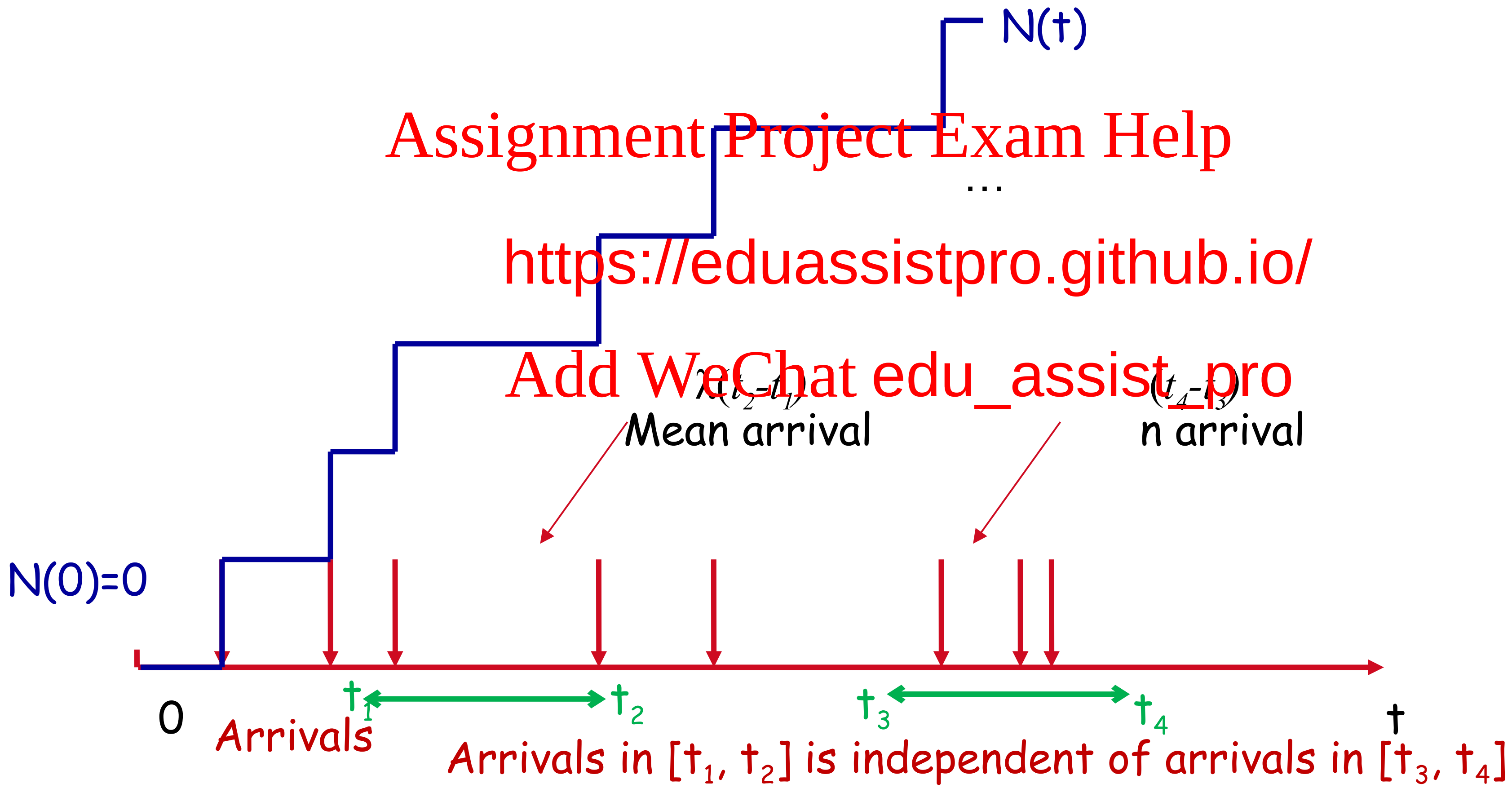
- – $P\{\, N(t_4) - N(t_3) = n \mid N(t_2) - N(t_1) = j \,\} = \ldots = n \,\}$

- 3. Number of events in any interval of length t is Poisson distributed with mean λt. That is, for all $s, t \geq 0$

$$E(N(t+s) - N(s)) = \lambda t$$

$$P(N(t+s) - N(s) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Poisson Process

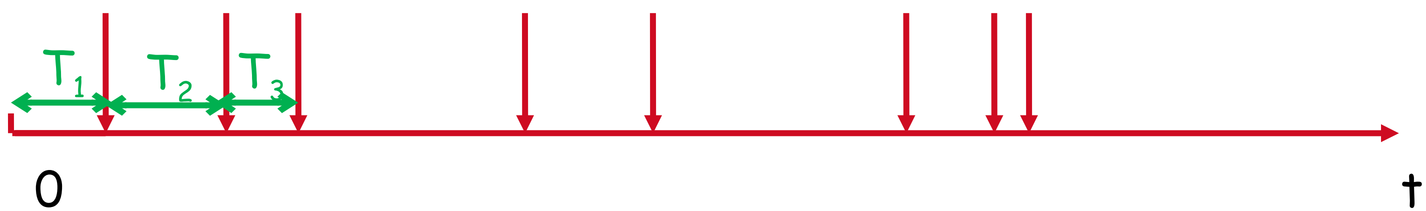


Poisson Process: Inter arrival time distribution

Exponential distribution with parameter λ

(mean 1/ λ)

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$



Number of arrivals in a short period of time

Number of arrival events in a very short period

$P\{N(t+h) - N(t) = 1\} = \lambda h + o(h)$; and

$P\{N(t+h) - N(t) \geq$

$o()$. Small o notation. The functio to be $o(h)$ if

$$\lim_{h \to 0} \frac{f(h)}{h} = 0$$

Summery

Poisson process:

Independent increments

# of arrivals: Poisson distributed

# of arrivals in a s rival, probability λh

Inter-arrival time stribution

› M/M/1 Queue

THE UNIVERSITY OF SYDNEY

Queues: Kendall's notation

- Notations Used in Queueing Systems
- *X*/*Y*/*Z*
- – *X* refers to the distribution of the interarrival times
- – *Y* refers to the distribution of service times
- – *Z* refers to the number o
- Common distributions:
- – *M* = Memoryless = exponential distribution
- – *D* = Deterministic arrivals or fixed-length service
- – *G* = General distribution of interarrival times or service times
- *M*/*M*/*1* refers to a single-server queuing model with exponential interarrival times (i.e., Poisson arrivals) and exponential service times.
- In all cases, successive interarrival times and service times are assumed to be statistically independent of each other.

M/M/1 Queue

- Arrival:
- Poisson arrival with rate λ
- Service:

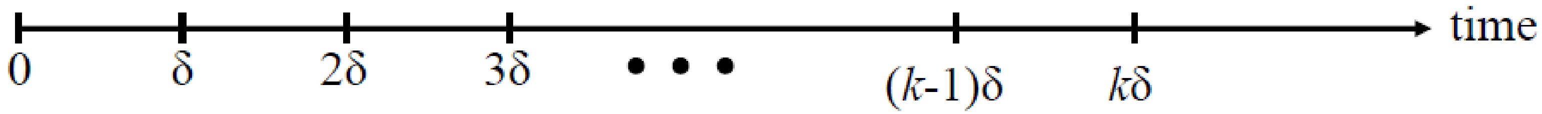
- Service time: ex

on with mean $1/\mu$

- μ: service rate,

- $\lambda < \mu$: Incoming rate < outgoing rate

Markov Chain Formulation



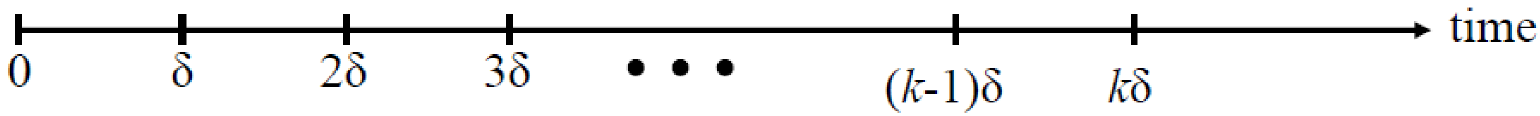
δ: a small value

N_k = Number of customers in t time $k\delta$

N_0 N_1 N_2... is a Markov Chain!

Q: How to compute the transition probability?

Markov Chain Formulation

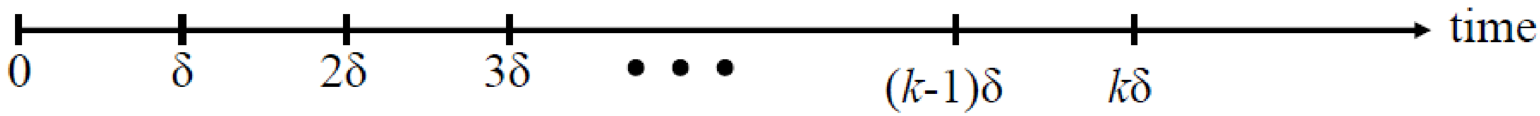


$P(0$ cust $\delta + o(\delta)$

$P(1$ customer arrive (δ)

$$P(\;2\;\text{customer}\;\;\text{arrives}) = o(\delta)$$

Markov Chain Formulation



$P(0 \quad \text{custome} \ldots (\delta))$

$$P(1 \quad \text{customer} \quad \text{leaves}) = \begin{cases} \ldots & i \quad 1 \\ 0 & i = 0 \end{cases}$$

$$P(\quad 2 \quad \text{customer} \quad \text{leaves}) = o(\delta)$$

No one in the system

Markov Chain Formulation

$$\textit{Aim to compute } P_{ij} = P\{N_{k+1} = j \mid N_k = i\}$$

$$\textit{For example, } P\{N_{k+1} = i \mid N_k = i\},\ i > 0$$

$P(0$ **custome** m **departs)**

$+P(1$ **customer arrives)** **mer departs)**

$+P(\text{other})$

$$\text{Result}: 1 - \lambda\delta - \mu\delta + o(\delta)$$

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

$$[\lambda\delta + o(\delta)][\mu\delta + o(\delta)] = o(\delta)$$

$$o(\delta)o(\delta) = o(\delta)$$

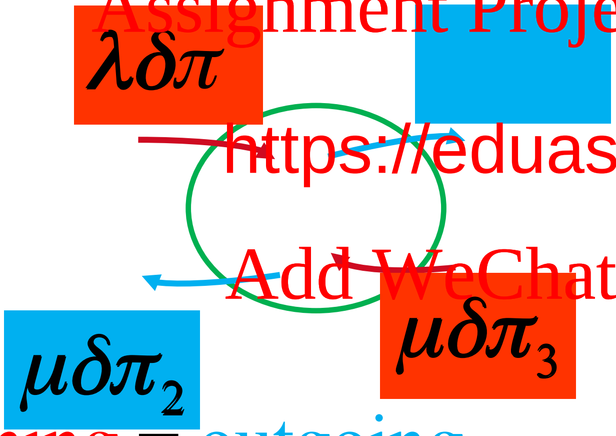
Result:

$\boldsymbol{\pi}_{\mathbf{i}}$ Stationary distribution of state *i*

The probability that there are *i* units in the system

Stationary Distribution Derivation

How to derive π_i *balance equation satisfied*

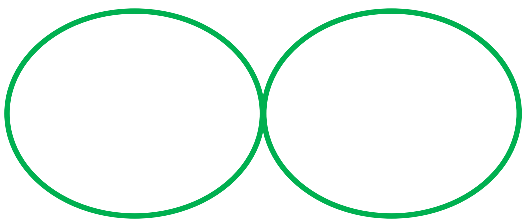


Incoming = outgoing

$$\lambda\delta\pi_2 + \mu\delta\pi_2 = \lambda\delta\pi_1 + \mu\delta\pi_3$$

Stationary Distribution Derivation

How to derive π_i



balance equation is performed at each state

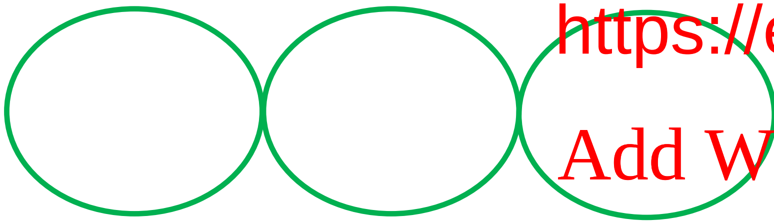
$$\lambda\delta\pi_0 = \mu\delta\pi_1$$

$$\lambda\delta\pi_1 + \mu\delta\pi_1 = \lambda\delta\pi_0 + \mu\delta\pi_2$$

$$\lambda\delta\pi_1 = \mu\delta\pi_2$$

Stationary Distribution Derivation

How to derive π_i



balance equation is performed at each state

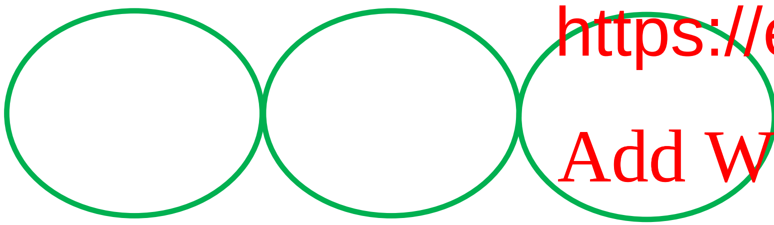
$$\lambda\delta\pi_0 = \mu\delta\pi_1$$

$$\lambda\delta\pi_1 = \mu\delta\pi_2$$

$$\lambda\delta\pi_2 + \mu\delta\pi_2 = \lambda\delta\pi_1 + \mu\delta\pi_3 \Rightarrow \lambda\delta\pi_2 = \mu\delta\pi_3$$

Stationary Distribution Derivation

How to derive π_i



balance equation is performed at each state

$$\lambda\delta\pi_0 = \mu\delta\pi_1$$

$$\lambda\delta\pi_1 = \mu\delta\pi_2$$

$$\lambda\delta\pi_2 = \mu\delta\pi_3$$

$$\Rightarrow \quad \lambda\delta\pi_i = \mu\delta\pi_{i+1} \quad \text{For any i}$$

Stationary Distribution Derivation

How to derive π_i

balance equation is performed at each state

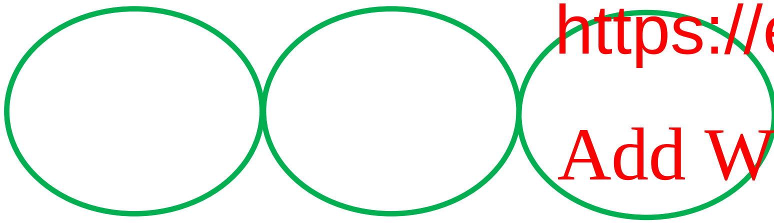
$$\pi_1 = \frac{\lambda}{\mu}\pi_0 \qquad \pi_2 = \left(\frac{\lambda}{\mu}\right)^2 \pi_0 \qquad \cdots \qquad \pi_i = \left(\frac{\lambda}{\mu}\right)^i \pi_0$$

$$\sum_{i=0}^{\infty} \pi_i = 1$$

Sum of geometric sequence

Stationary Distribution Derivation

How to derive π_i



balance equation is performed at each state

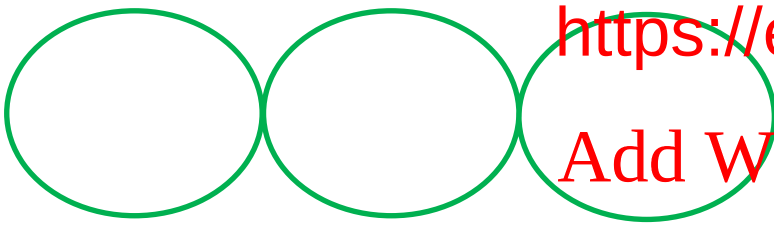
$\pi_1 = \rho\pi_0$ $\quad \pi_2 = (\rho)^2\pi_0$ $\quad \cdots \quad$ $\pi_i = (\rho)^i\pi_0$ $\quad \rho = \dfrac{\lambda}{\mu} < 1$

$$\sum_{i=0}^{\infty} \pi_i = 1$$

Sum of geometric sequence

Stationary Distribution Derivation

How to derive π_i



balance equation is performed at each state

$$\lim_{N\to\infty}\frac{\pi_0(1-\rho^N)}{1-\rho}=\frac{\pi_0}{1-\rho}=1$$

$$\pi_0 = 1-\rho$$

$$\pi_i = (1-\rho)\rho^i$$

Sum of geometric sequence

Average number of users in the system

$$E(N) = \sum_{n=0}^{\infty} n(1-\rho)\rho^n$$

$$= \rho(1-\rho)\sum_{n=0}^{\infty}$$

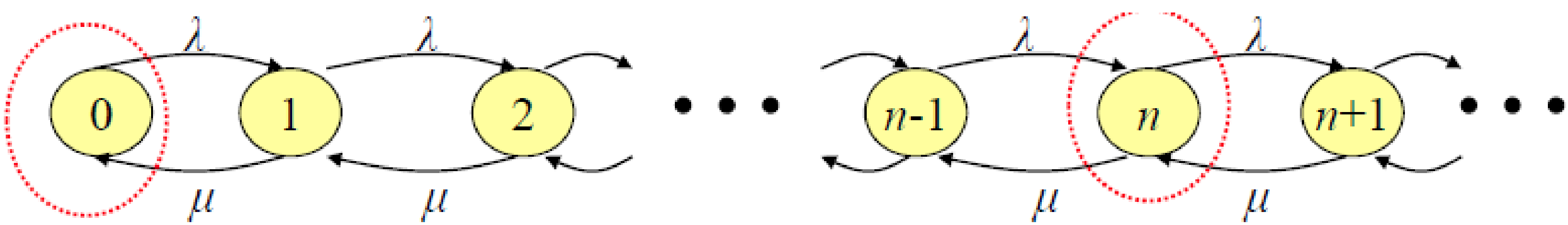
$$= \rho(1-\rho)\frac{\partial\left[\sum_{n=0}^{\infty}\rho^n\right]}{\partial\rho}$$

$$= \rho(1-\rho)\frac{\partial\left[\frac{\rho}{1-\rho}\right]}{\partial\rho} = \frac{\rho}{1-\rho}$$

Average waiting time

Average waiting time Little's Theorem

$$E(T) = \frac{\rule{4cm}{0.4pt}}{\lambda} \quad \rule{3.5cm}{0.4pt}$$


0
1
2
n-1
n
n+1
λ
μ

Stationary Distribution Derivation

balance equation is pe each state

$$\lambda \pi_0 = \mu \pi_1$$

$$\lambda \pi_1 + \mu \pi_1 = \lambda \pi_0 + \mu \pi_2$$

$$\cdots$$

Stationary Distribution Derivation

balance equation is pe *each state*

Following the same step, derive the same result

Stationary Distribution Derivation

Queueing delay goes to infinity when arrival rate approaches service rate!

M/M/m Queue

- Arrival:
- Poisson arrival with rate λ
- Service:

- Service time for ntial distribution with mean $1/\mu$

- service rate is $i\ \mu$, if there are $i<m$ users in the system
- service rate is $m\mu$, if there are $i>=m$ users in the system

M/M/m Queue

$$\lambda \pi_{i-1} = i\mu\pi_i \qquad i$$

$$\lambda \pi_{i-1} = m\mu\pi_i \qquad i > m$$

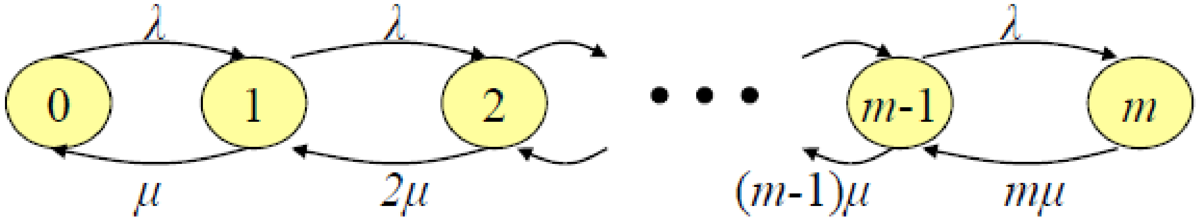
$$\pi_n = \begin{cases} \pi_0 \dfrac{(m\rho)^n}{n!} & n \le m \\ \pi_0 \dfrac{m^m \rho^n}{m!} & n > m \end{cases}$$

$$\rho = \frac{\lambda}{m\mu} < 1$$

Then, π_0 can be solved

M/M/∞ Queue

M/M/m/m Queue



M/M/m/n Queue

Arrivals will dropped if there are n users in the system.

Buffer size is n-m

How do you derive its stationary distribution?

Practice

- Analyze M/M/ ∞, M/M/m/n queues
  - Draw the state transition diagrams
  - Derive their stationary distributions
  - For M/M/m/n queue, c ncoming user is dropped. Calculate the probabili all users are served in the servers or there are no users at all.)