# Ôn tập NT531 - Đánh giá hiệu năng hệ thống mạng máy tính

## 1. Phân loại mô hình hàng đợi

### Queuing models: A/B/n

- A: số lượng process đến
- B: số lượng proccess phục vụ
- n: số lượng server

Các tiêu chuẩn notation: *tra cứu trong tài liệu*

Ví dụ:

- M/M/n: hệ thống thuần với Poisson arrival process với **n** server
- G1/G/1: hệ thống thông thường (General) với 1 server

### Queing models: A/B/n/K/NX

- tương tự như trên
- K: capacity của hệ thống
- N: số lượng khách hàng
- X: các quy tắc xử lý hàng đợi

> ***Note:*** nếu K = n, tức là hệ thống bị đầy/quá tải, thường được note là **A/B/n-Loss**

### Queuing disciplines (Quy tắc xử lý hàng đợi)

- **FIFO** - First In First Out: là một hàng đợi công bằng hoặc một hàng đợi có thứ tự, thường được gọi là FCFS - First Come First Serve
- **LIFO** - Last Come First Out: giống với Stack, được sử dụng trong hệ thống lưu trữ, tên gọi khác là LCFS - Last Come First Serve

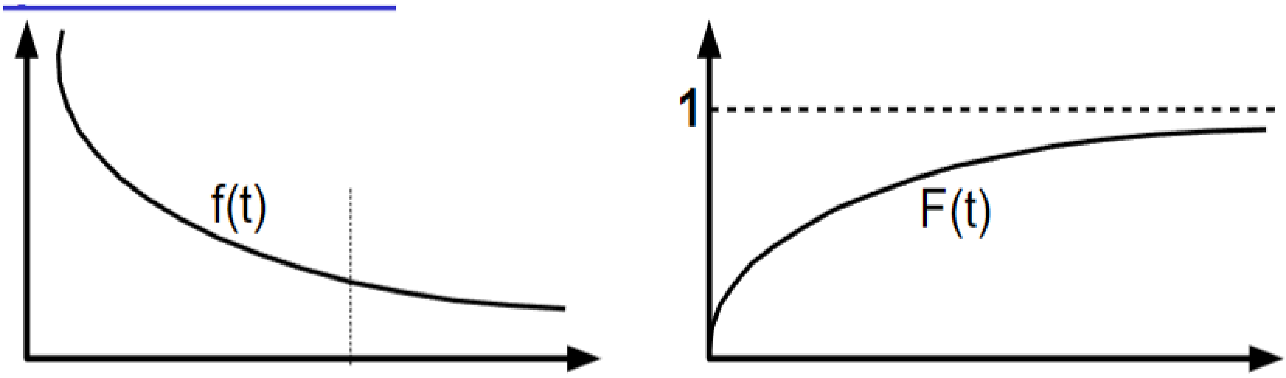
### Phân loại hàng đợi

- SIRO - Service In Random Order
- SJF - Shortest Job First (SJN - Shortest Job Next, SPF - Shortest Processing time First, etc.)
- RR - Round Robin
- PS - Processor Sharing
- NON-PREEMEPTIVE: HOL - Head-of-the-line

- PREEMPTIVE: Resume, Without Resampling, With Resampling, GD - General Discipline

# 2. Arrival and Service Process

## Exponential Model




- Probability density function - Hàm mật độ xác xuất (pdf): $f(t) = \lambda . e^{-\lambda t}$
- Cumulative function - Hàm phân phối tích luỹ : $F(t) = 1 - e^{-\lambda t}$ = Probability{inter-arrival time ≤ t}
- Mean value - Giá trị trung bình: $E[x] = 1/\lambda$
- Variance - Phương sai: $Var[x] = 1/\lambda^2$
- Conditional Probability - Xác xuất có điều kiện: $P = 1 - e^{-\lambda \times dt}$ với thời điểm tới kế tiếp trong khoảng $(t, t + dt)$
- Mô hình này còn được gọi là *memory less system* vì xác xuất của arrival mới phụ thuộc vào xác xuất của arrival trước nó
- Trường hợp sử dụng: Khi đề bài nói đến thời gian arrival giữa các lần (inter-arrival times)

Ví dụ: Let us assume that engineering personnel use an online terminal to make routine calculations. If the time each engineer spends in a session at the terminal has an exponential distribution with an average value of 36 minutes, find:

a. The probability that an engineer spends less than 30 minutes at the terminal
b. More than an hour
c. 90% of the session end in less than t minutes. What is the value of t?

Tóm tắt:

- $\lambda = \frac{1}{36}$ (1 phút có 1/36 session trên terminal)

a. Xác xuất người kỹ sư dành ít hơn 30 phút sử dụng terminal

$P_{x<30} = 1 - e^{-\frac{1}{36} \times 30} = 0.565$

b. Nhiều hơn 1 giờ => lấy phần bù

$P_{x>60} = 1 - P_{x\leq 60} = e^{-\frac{1}{36}\times 60} = 0.189$

c. Xác xuất 90% session kết thúc ít hơn t phút => tìm t

$P_{x<t} = 1 - e^{-\frac{1}{36}\times t} = 0.9 \Rightarrow t = 82,893$

# Poison Model

- Mean arrival rate - Tỉ lệ trung bình: $\lambda$
- Observation period - Chu kỳ quan sát: $T$
- Poisson distributed - Phân phối Poisson: $P_x(T) = \frac{(\lambda\times T)^x}{x!} \times e^{-\lambda\times T}$
- Mean value - Giá trị trung bình: $E[x] = \lambda \times T$
- Variance - Phương sai: $Var[x] = \lambda \times T$
- Trường hợp sử dụng: Khi đề bài cho khoảng thời gian cố định.

Ví dụ: On average, 1 call arrives every 5 seconds. Assume that inter-arrival time distribution is exponential. During a period of 10 seconds, what is the probability that:

a. No call arrives
b. One call arrives
c. Two calls arrive
d. More two calls arrive

Tóm tắt:

- $\lambda = \frac{1}{5}$ (1 giây có 1/5 cuộc gọi)
- T = 10 giây

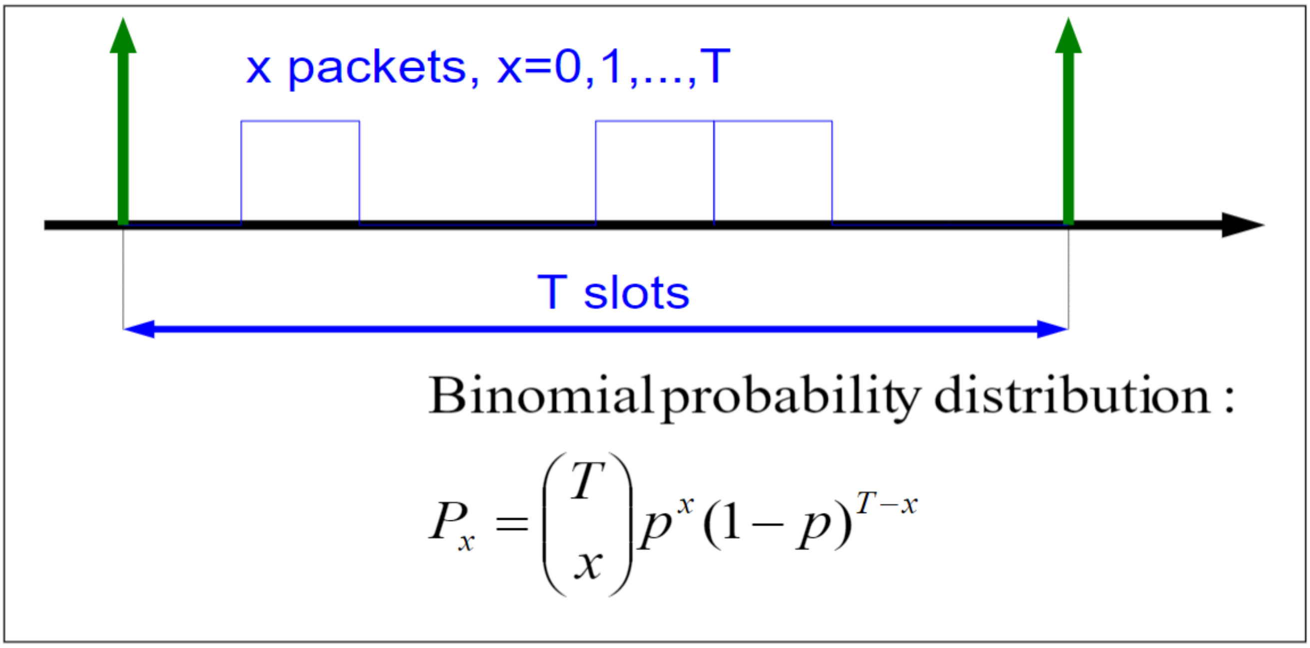
a. Không cuộc gọi => x = 0 => $P_0 = 0.135$

b. 1 cuộc gọi => x = 1 => $P_1 = 0.271$

c. 2 cuộc gọi => x = 2 => $P_2 = 0.271$

d. >2 cuộc gọi => $P_{x>2} = 1 - \sum_{i=0}^{2} P_i = 0.323$

# Binominal Model




- Independence slots - Số slot độc lập: $T$
- Arrival in 1 slot - Xác xuất arrival trong 1 slot: $p = P$
- No arrival in 1 slot - Xác xuất không có arrival trong 1 slot: $1 - p = P$
- Binomial probability distribution - Phân phối Binominal: $P(x) = \binom{T}{x} p^x (1-p)^{T-x}$
- Trường hợp sử dụng: Khi đề bài nói tới số lượng arrival thành công trong một slot và có 2 trường hợp xảy ra (arrival - thành công, no arrival - không thành công)

Ví dụ: At time 0, there are 5 conversations in progress on a trunk group. Assume that the conversation times are **exponentially distributed** with mean 3 minutes. Calculate the probability that after 1 minute exactly 2 of those conversations are still in progress!

Tóm tắt:

- T = 5 (có 5 cuộc hội thoại trong 1 nhóm)
- $E[x] = 1/\lambda = 3$ (thời gian trung bình là 3 phút cho 1 cuộc hội thoại)
- t = 1, x = 2 (trong 1 phút có đúng 2 cuộc hội thoại đang diễn ra)

Áp dụng Exponentail Model:

- $P_{x \leq 1} = 1 - e^{-\lambda \times t} = 0.283$
- $P_{x > 1} = 1 - P_{x \leq 1} = 0.717$

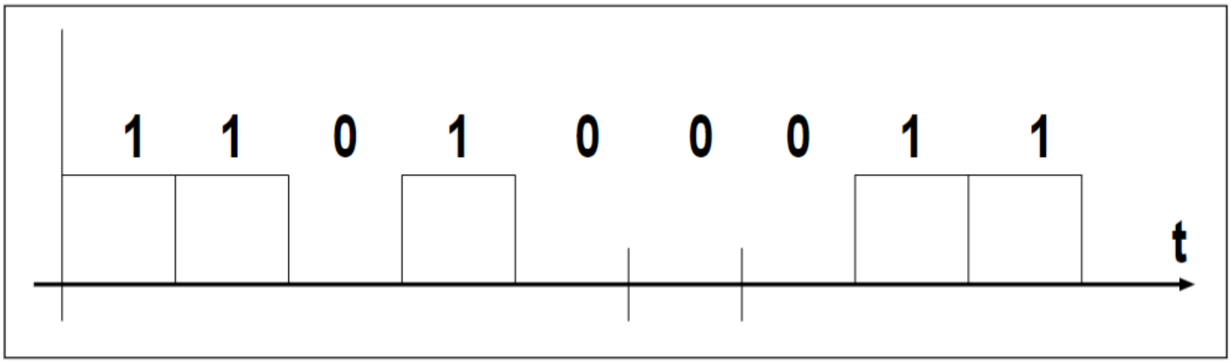
Chọn 2 cuộc hội thoại ngẫu nhiên, không quan tâm thứ tự từ 5 cuộc hội thoại: $C_5^2$

Xác xuất sau 1 phút có đúng 2 cuộc hội thoại: $P_1 = C_5^2 \times (P_{x>1})^2 \times (P_{x \leq 1})^{5-2} = 0.117$

## Bernouli Model

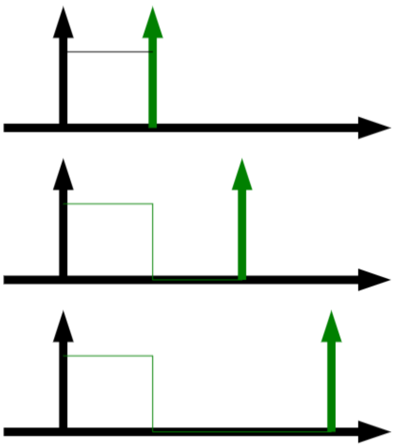
p=P{arrival in slot}

1-p=P{no arrival in slot}




### Geometric Model

p=P{arrival in slot}
1-p=P{no arrival in slot}



$$t = 1 \rightarrow p_1 = p$$

$$t = 2 \rightarrow p_2 = (1-p)p$$

$$t = 3 \rightarrow p_3 = (1-p)^2 p$$

- - - - - -

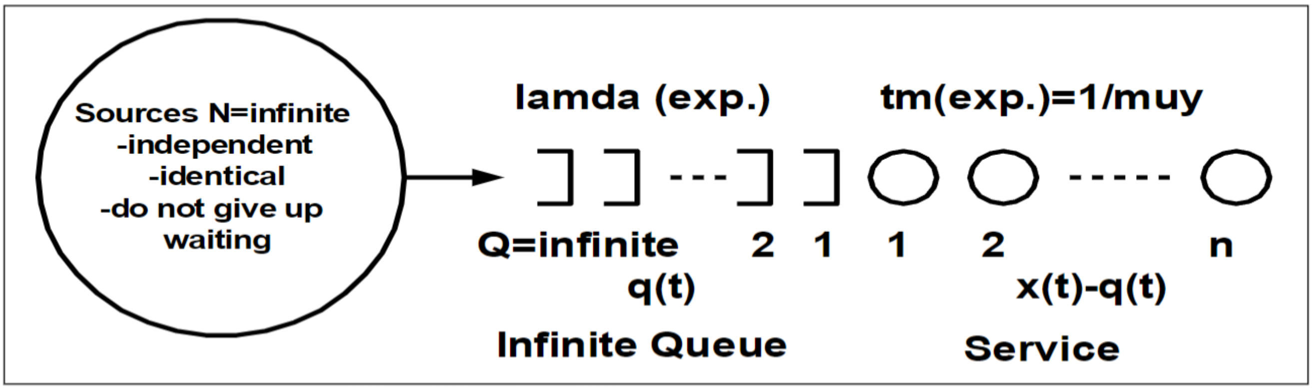
$$t = n \rightarrow p_n = (1-p)^n p$$

Pn=P{1 arrival in next n time slots}

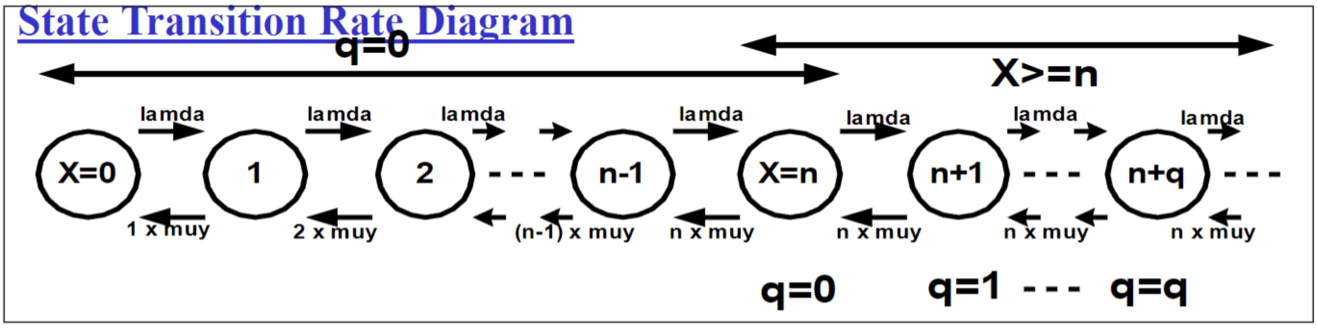
## 3. Erlang's Delay System (M/M/n - System)

### Giả thiết đặt ra

- Số lượng server có hạn $(n < \infty)$
- Số lượng vị trí trong hàng đợi vô hạn $(Q = \infty)$
- Thời gian giữa các lần đến và thời gian được đáp ứng được phân phối Exponential
- Các nguồn giống hệt nhau và độc lập với nhau
- Người dùng không rời hàng đợi, và tất cả yêu cầu đều được đáp ứng
- Thuận theo nguyên tắc FIFO - First In First Out




- Phân phối thời gian giữa các lần đến là phân phối Exponential với tỷ lệ đến trung bình, gọi là $\lambda$
- Phân phối xác suất của thời gian chiếm dụng cá nhân là phân phối Exponentail với giá trị trung bình là $t_m = \frac{1}{\mu}$ , với $\mu$ là tỉ lệ đáp ứng
- Số lần tới trong $t_m$ là $A = \lambda \times t_m$
- Hệ thống hoàn toàn sẵn để các yêu cầu có thể đến bất kỳ máy chủ nào, với số lượng người dùng trong hệ thống tại thời điểm $t = (0, \infty)$ là $x(t)$
- Số lượng người dùng đang chờ trong hệ thống tại thời điểm $t = (0, \infty)$ là $q(t)$
- Số lượng người dùng được phục vụ trong hệ thống tại thời điểm $t = (0, \infty)$ là $x(t) - q(t)$




- Erlang's Delay Formular, trong trường hợp bình thường:
  - $P_0 = \frac{1}{\sum_{i=0}^{n-1} \frac{A^i}{i!} + \frac{n}{n-A} \cdot \frac{A^n}{n!}}$
  - $P_x = \frac{\frac{A^x}{x!}}{\sum_{i=0}^{n-1} \frac{A^i}{i!} + \frac{n}{n-A} \times \frac{A^n}{n!}}$, khi q = 0, x = (1, n-1)
  - $P_x = \frac{(\frac{A}{n})^q \times \frac{A^n}{n!}}{\sum_{i=0}^{n} \frac{A^i}{i!} + \frac{n}{n-A} \times \frac{A^n}{n!}}$, khi $x \geq n$ và q = x - n = $(0, \infty)$
- Erlang's Delay Formular, trong trường hợp không có delay
  - $D_n(A) = \frac{\frac{n}{n-A} \cdot \frac{A^n}{n!}}{\sum_{i=0}^{n-1} \frac{A^i}{i!} + \frac{n}{n-A} \cdot \frac{A^n}{n!}}$
  - $P_{\text{no delay}} = 1 - D_n(A)$

Ví dụ: Một hệ thống hàng đợi M/M/8 có tốc độ đến của gói tin là λ=135 packets/hour (135 gói tin/1 giờ). Thời gian phục vụ/xử lý gói tin trung bình là tm=2 phút (minutes). Hãy tính toán:

- Xác suất gói tin đến hệ thống và được phục vụ ngay lập tức (no delay)?

- Xác suất tất cả các server đều bận và hàng đợi trống (empty)?

Tóm tắt:

- n = 8 (ghi chú trong M/M/8)
- $\lambda = 135$ (135 packets/hour)
- $t_m = \frac{2}{60}$ (tm=2 phút)
- $A = \lambda \times t_m = 4.5$

a. Không có độ trễ (no delay)

$$D_n(A) = \frac{\frac{n}{n-A} \cdot \frac{A^n}{n!}}{\sum_{i=0}^{n-1} \frac{A^i}{i!} + \frac{n}{n-A} \cdot \frac{A^n}{n!}} = 0.104$$

$$P_{\text{no delay}} = 1 - D_n(A) = 0.896$$

b. Tất cả server đều bận và hàng đợi trống:

$x = n$ (tất cả 8 server đều bận), $q = 0$ (hàng đợi trống)

$$P_0 = \frac{1}{\sum_{i=0}^{n-1} \frac{A^i}{i!} + \frac{n}{n-A} \times \frac{A^n}{n!}} = 0.011$$

$$P_8 = (\frac{A}{n})^q \times \frac{A^n}{n!} \times P_0 = 0.046$$