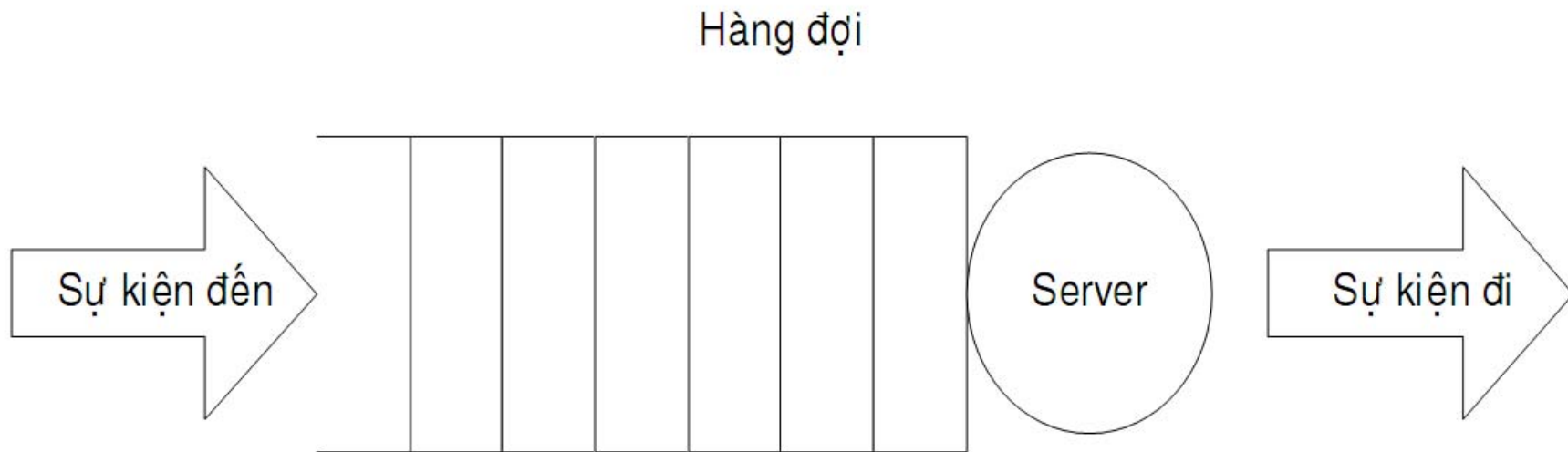


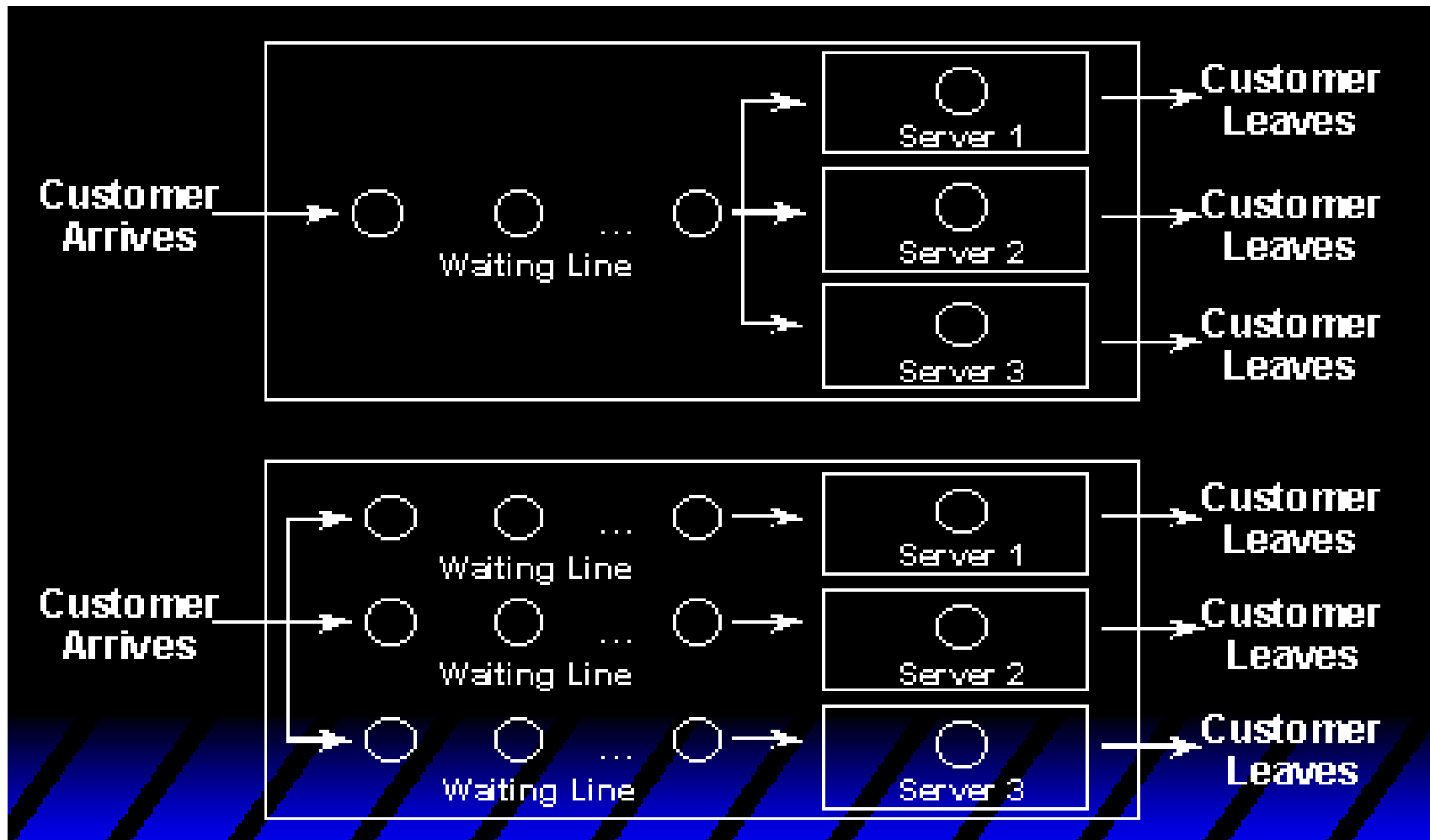
Lý thuyết xếp hàng và ứng dụng

- **Trong các hệ thống dịch vụ, chủ thể phục vụ (server) lần lượt phục vụ các đối tượng sử dụng dịch vụ. Số lượng chủ thể có thể nhiều hơn 1**
- **Ví dụ:**
 - Các hệ thống điện thoại: khi số lượng lớn khách hàng quay số để kết nối đến một trong những đường ra hữu hạn của tổng đài.
 - Trong mạng máy tính: khi mà gói tin được chuyển từ nguồn tới đích và đi qua một số lượng các nút trung gian. Hệ thống hàng đợi xuất hiện tại mỗi nút ở quá trình lưu tạm thông tin tại bộ đệm.

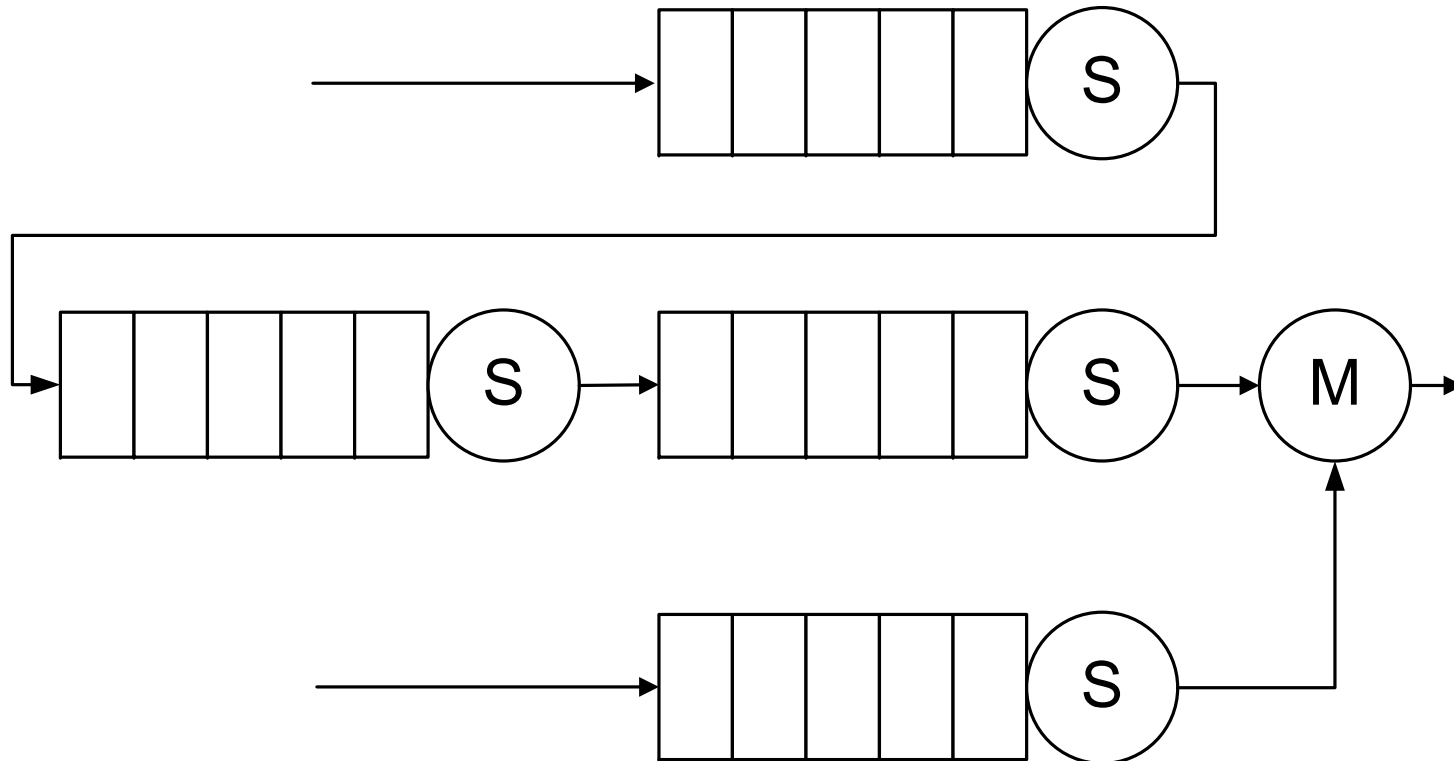
- **Mạng viễn thông**
- **Kiểm soát lưu lượng giao thông**
- **Đánh giá hiệu năng hệ thống máy tính**
- **Y tế và chăm sóc sức khỏe**
- **Không lưu, bán vé**
- **Dây truyền sản xuất**



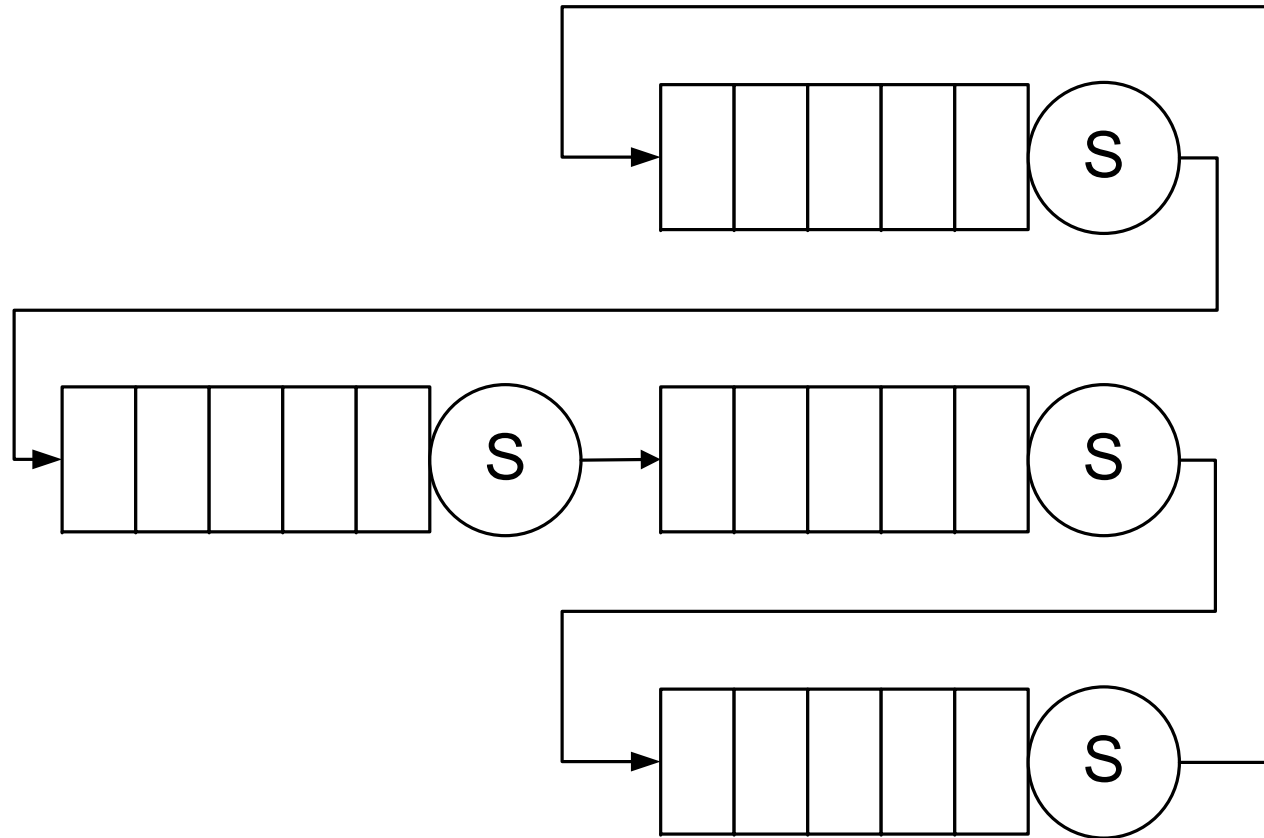
Tổng quan



Mạng hàng đợi mở



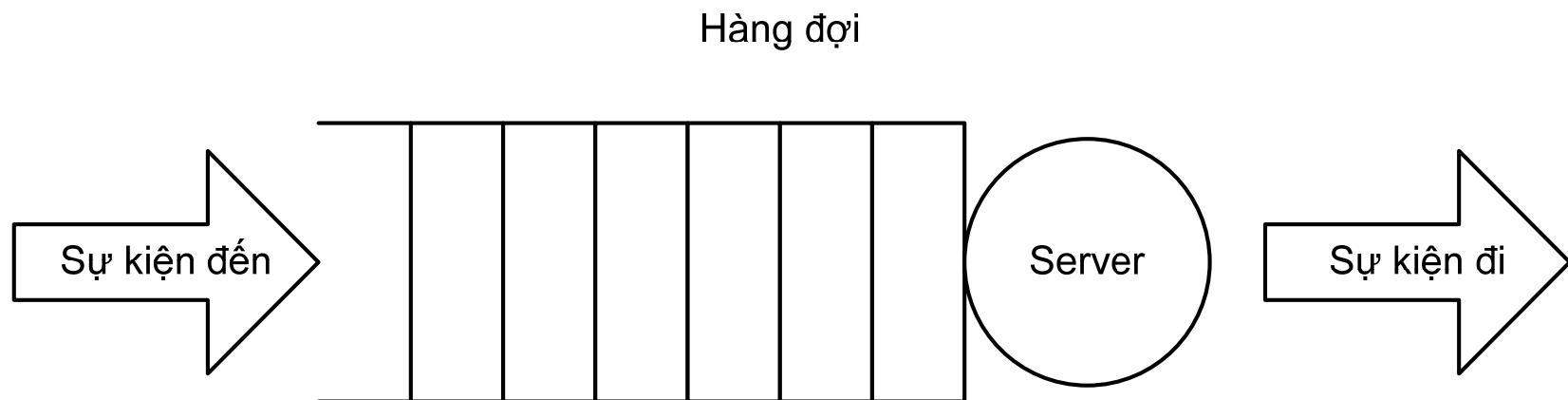
Mạng hàng đợi đóng



- **Có thể mô hình hóa mạng viễn thông như một tập hợp các hàng đợi**
 - Mỗi nút gồm một số giao tiếp mỗi giao tiếp gắn với một hoặc một số hàng đợi
 - Cấu trúc dữ liệu theo kiểu FIFO
- **Lý thuyết xếp hàng sẽ giúp phân tích các tham số:**
 - Chiều dài trung bình của hàng đợi
 - Thời gian đợi trung bình
 - Xác suất một hàng đợi có chiều dài nào đó
 - Xác suất mất nối

- **Hệ thống có bao nhiêu server? Tốc độ phục vụ của các server này ?**
- **Có bao nhiêu vị trí đợi trong hàng đợi?**
- **Có bất kỳ quy tắc nội bộ đặc biệt nào không (yêu cầu dịch vụ, mức độ ưu tiên...)?**
- **Miêu tả của tiến trình đến (phân bố khoảng thời gian đến)**

- Quy tắc phục vụ (FCFS, LCFS, RANDOM)
- Thời gian rỗi (phân bố thời gian rỗi)
- Mức độ ưu tiên



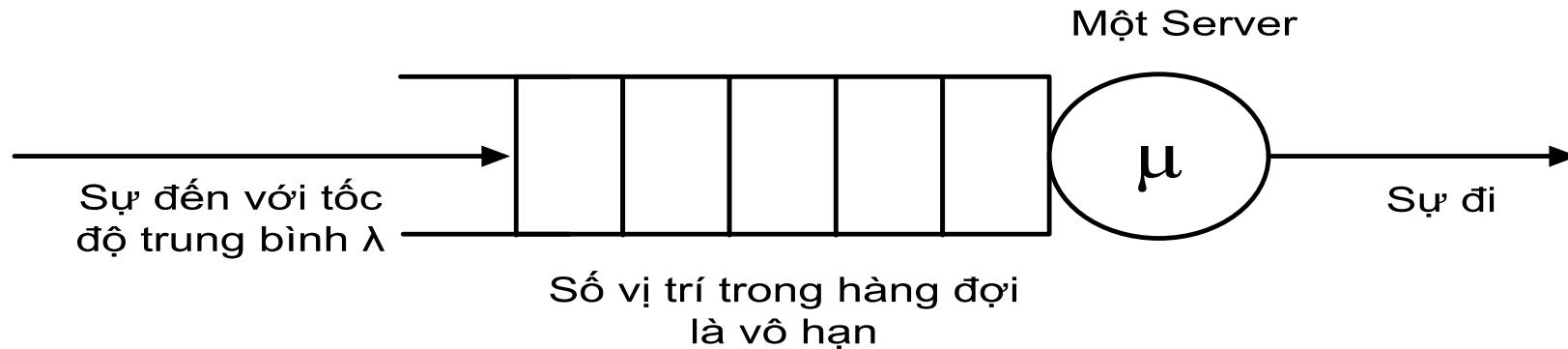
- **Phân tích giải tích**
- **Quá trình mô phỏng**
- **Cả hai phương pháp trên**

- Thời gian xếp hàng (trễ hàng đợi)
- Tổng trễ (bao gồm trễ hàng đợi và trễ phục vụ)
- Số lượng khách hàng trong hàng đợi
- Số lượng khách hàng trong hệ thống (gồm khách hàng chờ và khách hàng đang được phục vụ)
- Xác suất nghẽn mạng (khi kích thước bộ đệm hữu hạn)

Kết quả phân tích về phía hệ người phục vụ

- Khả năng sử dụng server
- Khả năng sử dụng bộ đệm
- Lợi ích thu được (thông số dịch vụ và các xem xét về kinh tế)
- Lợi ích bị mất (thông số dịch vụ và các xem xét về kinh tế)

Phân tích hàng đợi

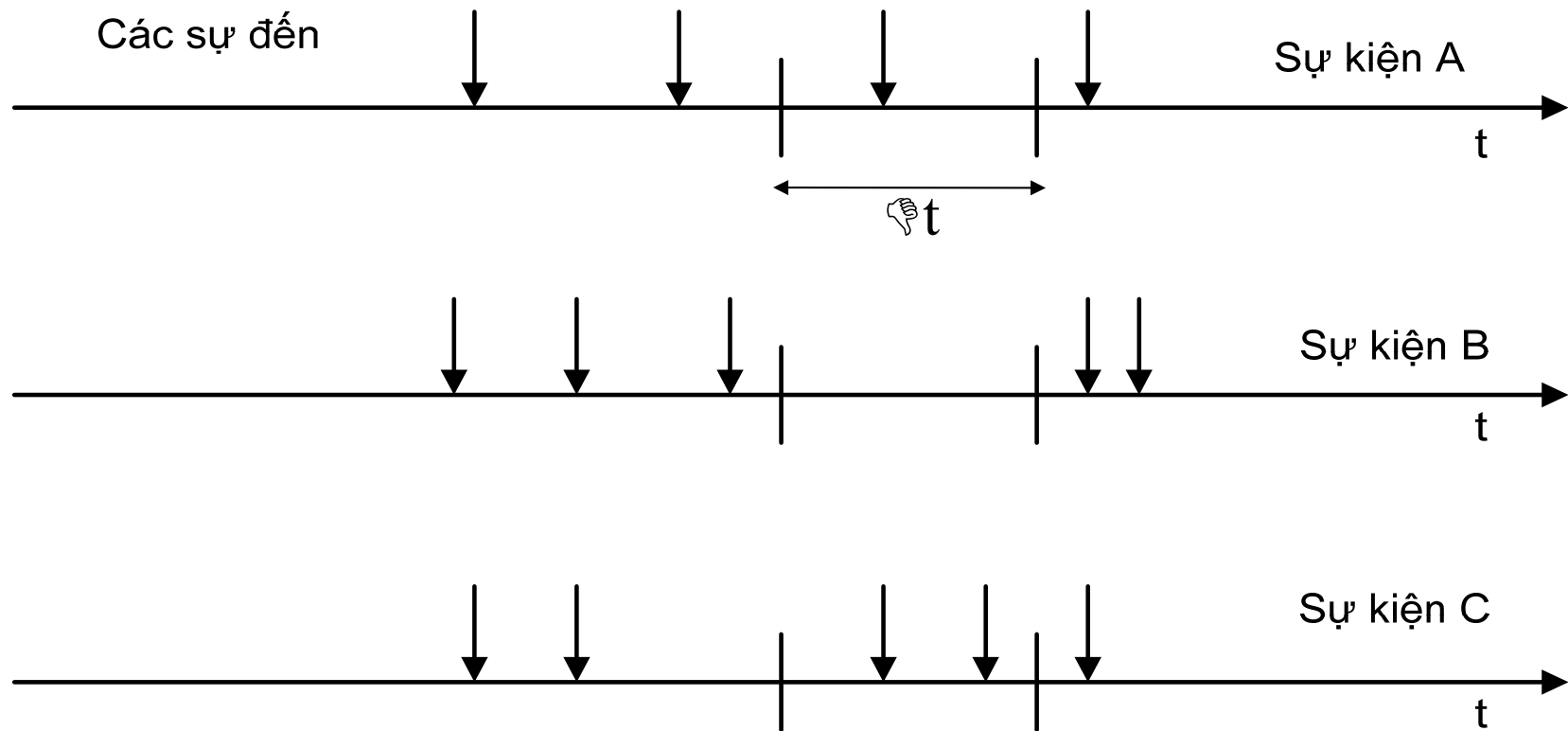


λ - tốc độ đến trung bình , thời gian đến trung bình $-1/\lambda$

μ - tốc độ phục vụ trung bình, thời gian phục vụ trung bình $1/\mu$

Với kích thước của bộ đệm là vô hạn, quy tắc phục vụ là FCFS

Phân tích hàng đợi



- Sự kiện A : có 1 sự đến trong Δt
- Sự kiện B : không có sự đến nào trong Δt
- Sự kiện C : có nhiều hơn 1 sự đến trong Δt
- Giả sử rằng $\Delta t \rightarrow 0$. Như vậy ta sẽ có:

$$Pr\{A\} = \lambda \Delta t$$

$$Pr\{B\} = 1 - \lambda \Delta t$$

$$\text{Giả thiết } Pr\{C\} = 0$$

- Số lượng sự kiện đến tuân theo phân bố Poisson

Phân tích hàng đợi

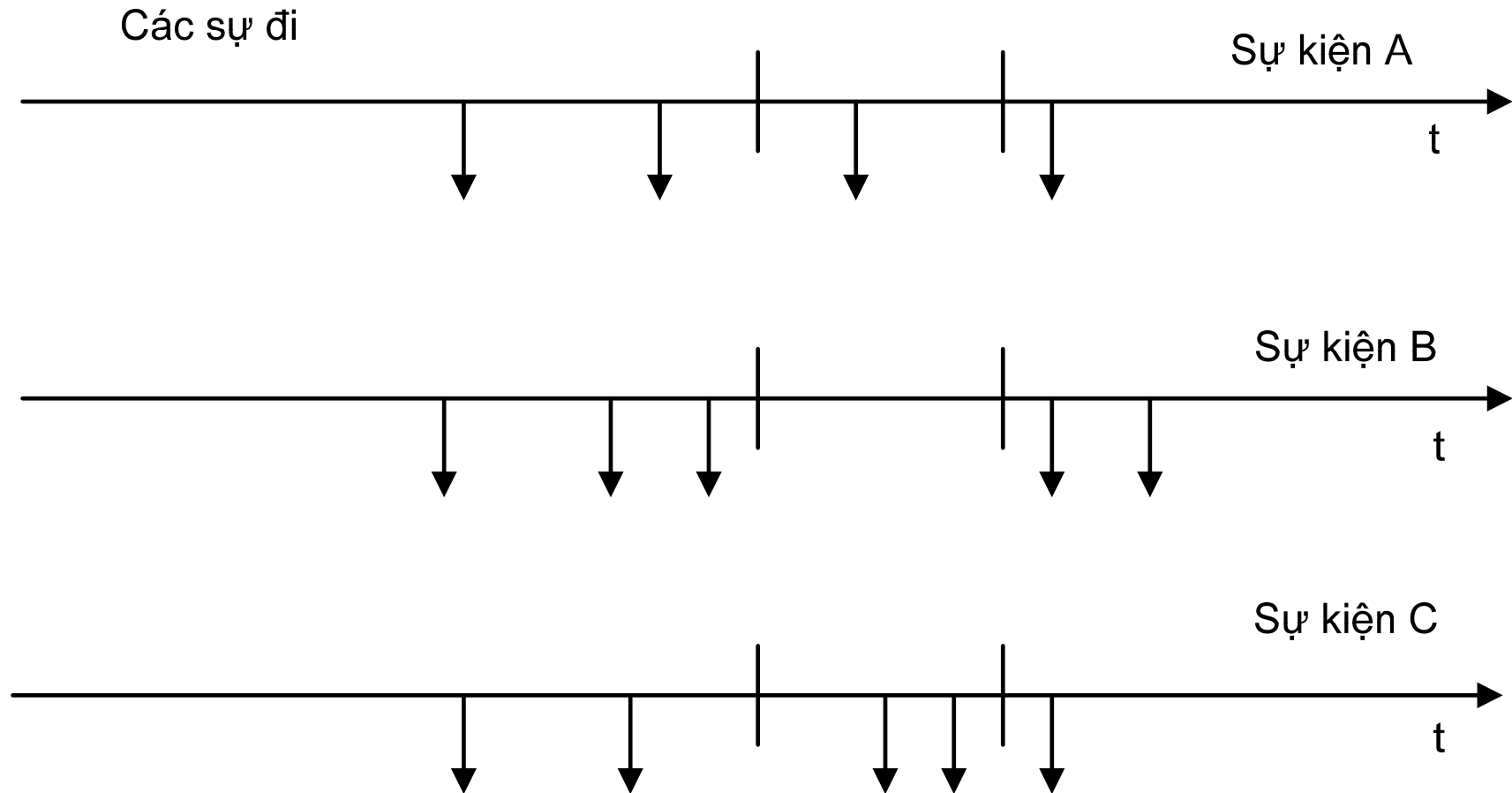
- Định nghĩa luật phân bố Poisson*

$$P(N=n) = \frac{e^{-\lambda} \lambda^n}{n!} \quad (n=0,1,2,\dots)$$

- Đồng thời, *khoảng thời gian đến* (được tính giữa hai sự đến liên tiếp) tuân theo luật phân bố mũ* với tham số λ

(*) Trong MS Excel có hàm POISSON và hàm
 $a(t) = \lambda e^{-\lambda t}$
EXPONDIST

Phân tích hàng đợi



Phân tích hàng đợi

- Sự kiện A : có 1 sự kiện đi trong Δt
- Sự kiện B : không có sự kiện đi nào trong Δt
- Sự kiện C : có nhiều hơn 1 sự kiện đi trong Δt
- Giả sử $\Delta t \rightarrow 0$. Như vậy ta sẽ có:

$$Pr\{A\} = \mu \Delta t$$

$$Pr\{B\} = 1 - \mu \Delta t$$

- D là sự kiện của 1 hoặc nhiều sự đến AND với sự kiện của 1 hoặc nhiều sự đi trong khoảng Δt . Giả sử $Pr\{D\} = 0$

- Định nghĩa $p_N(t)$ là xác suất mà hệ thống có N khách hàng tại thời điểm t
- Khi đó có:

$$p_0(t+\Delta t) = p_0(t)(1-\lambda\Delta t) + p_1(t)\mu\Delta t, \quad N=0$$

$$p_N(t+\Delta t) = p_N(t)(1-\lambda\Delta t - \mu\Delta t) + p_{N-1}(t)\lambda\Delta t + p_{N+1}(t)\mu\Delta t,$$

$$N>0$$

Ở thời điểm $t+\Delta t$ có N khách hàng nếu ở t có N khách hàng và không có sự đến/ sự đi...

- Từ đó suy ra:

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t), N = 0$$

$$\frac{dp_N(t)}{dt} = -(\lambda + \mu)p_N(t) + \lambda p_{N-1}(t) + \mu p_{N+1}(t), N > 0$$

Phân tích hàng đợi

- Ở điều kiện ổn định, khi $t \rightarrow \infty$, ta có

$$\frac{dp_0(t)}{dt} = 0, N = 0$$

$$\frac{dp_N(t)}{dt} = 0, N > 0$$

- Hay: $p_0(t) = p_0$, với $N = 0$

$$p_N(t) = p_N, \text{ với } N > 0$$

Tức là xác suất hệ thống rơi vào một trạng thái nào đó không phụ thuộc thời gian nữa...

Phân tích hàng đợi

- Hệ phương trình vi phân trở thành (đặt $\rho = \lambda / \mu < 1$):

$$p_1 = \rho p_0$$

$$p_{N+1}(t) = (1 + \rho)p_N - \rho p_{N-1} = \rho p_N = \rho^{N+1} p_0, N > 0$$

- Theo điều kiện phân bố chuẩn:

$$\sum_{\forall i} p_i(t) = 1, t \geq 0$$

- Suy ra:

$$p_i = \rho^i (1 - \rho), i = 0, 1, \dots$$

Số lượng khách hàng trung bình

- Xét trong một khoảng thời gian đủ lớn, số lượng khách hàng lưu trong *hệ thống* được tính theo công thức:

$$E[N] = \sum_{i=0}^{\infty} i p_i = \sum_{i=0}^{\infty} i \rho^i (1-\rho) = \frac{\rho}{1-\rho}$$

* Để chứng minh, tách thành 2 tổng rồi thay thế $i+1$ sang i để rút gọn...

Số lượng trung bình

- Số lượng khách hàng lưu trong *hàng đợi* được tính bằng:

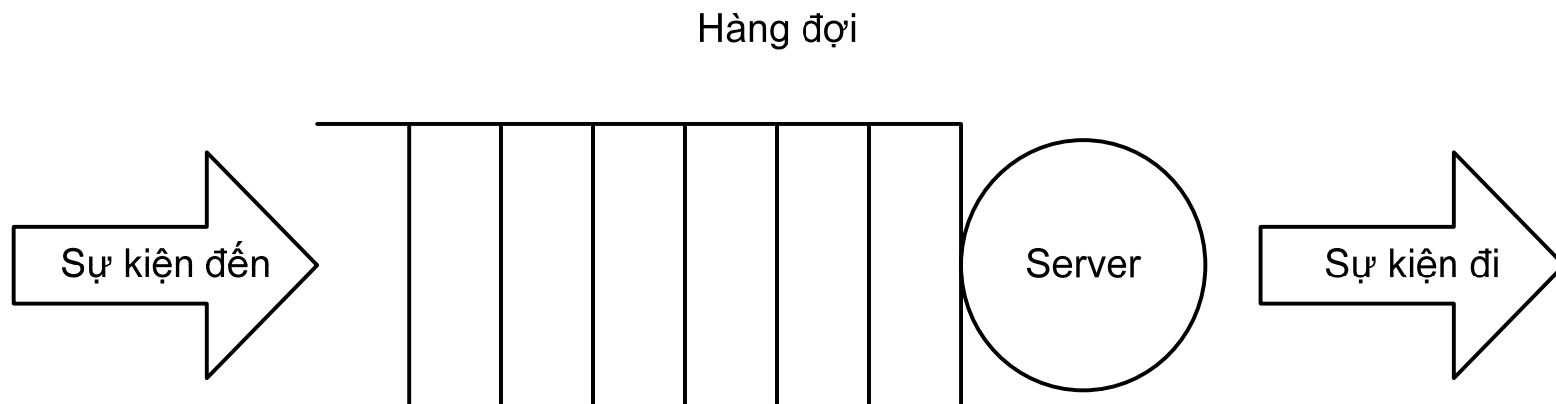
$$E[N_Q] = \sum_{i=1}^{\infty} (i-1)p_i = \sum_{i=1}^{\infty} ip_i - \sum_{i=1}^{\infty} p_i = \frac{\rho}{1-\rho} - (1-p_0) = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}$$

* Tổng xuất phát từ $i = 1$, nghĩa là công thức chỉ đúng khi có ít nhất một khách hàng trong hệ thống

- Thời gian một khách hàng lưu lại trong hệ thống bao gồm:

- Thời gian chờ xếp hàng

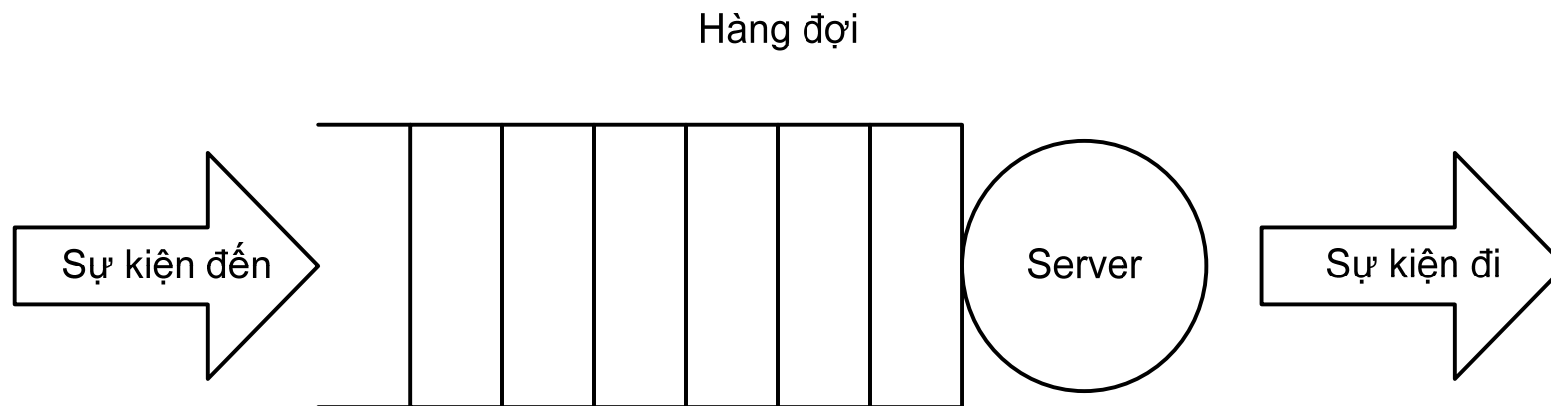
- Thời gian phục vụ



- Giả thiết:

- Hệ thống ở trạng thái ổn định tức $\lambda \leq \mu$

- Quy tắc phục vụ là FCFS



- Tổng thời gian lưu trong hệ thống:

$$E[W] = \sum_{k=0}^{\infty} \frac{k}{\mu} p_k + \sum_{k=0}^{\infty} \frac{1}{\mu} p_k = \sum_{k=0}^{\infty} \frac{k+1}{\mu} p_k = \frac{1}{\mu(1-\rho)}$$

Thời gian trung bình

- Tổng thời chờ trong hàng đợi:

$$E[W_Q] = \sum_{k=0}^{\infty} \frac{k}{\mu} p_k = \frac{\rho}{\mu(1-\rho)}$$

$$E[W_Q] = E[W] - \frac{1}{\mu} = \frac{1}{\mu(1-\rho)} - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}$$

- Sự kiện một khách hàng đến phải đợi chính là khi trong hệ thống có ít nhất 1 khách hàng:

$$P_{wait} = 1 - p_0 = \rho$$

- Đây cũng chính là xác suất hệ thống ở trạng thái bận P_{busy}

- Tính dừng (stationary-time). Diễn tiến của các sự kiện trong một khoảng thời gian không phụ thuộc vào thời điểm bắt đầu quan sát. Xác suất có k cuộc gọi đến trong khoảng thời gian $[t_1, t_1+t_2]$ là độc lập với t_1 , nghĩa là với mọi t, k ta có:

$$p\{(N_{t_1+t_2} - N_{t_1}) = k\} = p\{(N_{t_1+t_2+t} - N_{t_1+t}) = k\}$$

- Tính độc lập (independence) hay còn gọi là tính không nhớ. Trạng thái tiếp theo chỉ phụ thuộc vào trạng thái hiện tại, nhưng độc lập với việc nó đã có được như thế nào. Đây chính là đặc tính của tiến trình Markov

$$p\{(N_{t_2} - N_{t_1}) = k | N_{t_1} - N_{t_0} = n\} = p\{(N_{t_2} - N_{t_1}) = k\}$$

- Tính đều đặn (regularity). Xác suất xảy ra với nhiều hơn một sự kiện ở cùng một thời điểm bằng không:

$$p\{(N_{t+\Delta t} - N_t) \geq 2\} = o(\Delta t); \text{ khi } : \Delta t \rightarrow 0, o(\Delta t) \rightarrow 0$$

- Một ví dụ điển hình của tiến trình điểm là tiến trình tuân theo luật Poisson. Xác suất để có k sự kiện trong khoảng thời gian T:

$$p(k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!}$$

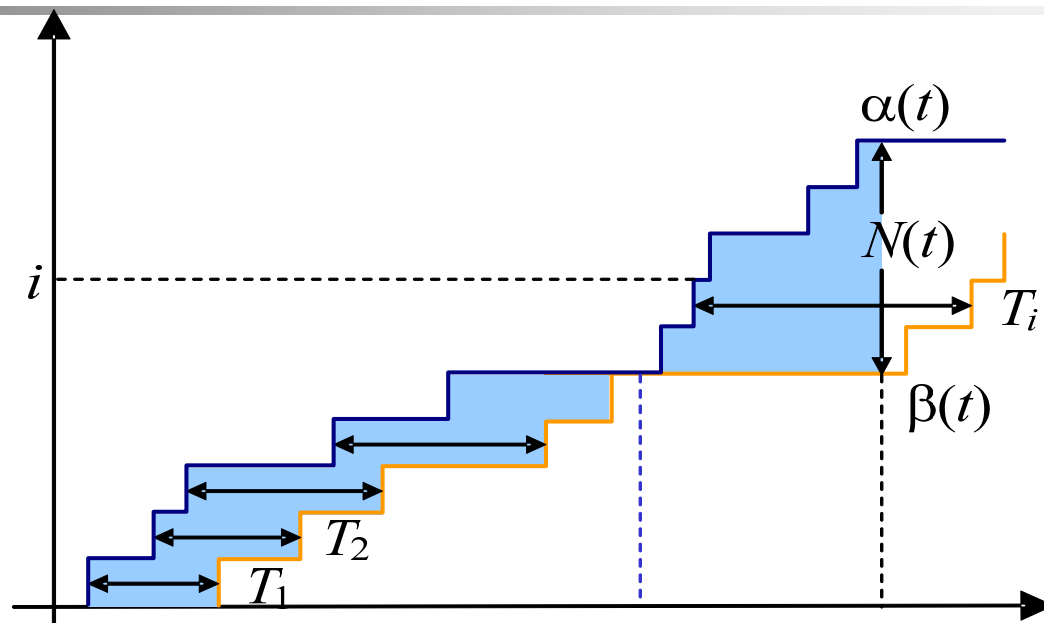
- Phân bố chuẩn: $\sum_{k=0}^{\infty} p(k) = 1$

- Kỳ vọng:

$$E(k) = \sum_{k=0}^{\infty} kp(k) = \lambda T$$

Tính trung bình, số lượng khách hàng nằm trong hệ thống được tính bằng tích của tốc độ đến và thời gian phục vụ

- Ký hiệu:
 - $N(t)$: số lượng khách hàng nằm trong hệ thống tại thời điểm t
 - α_t : số lượng khách hàng đến hệ thống khoảng thời gian $(0,t)$
 - β_t : số lượng khách hàng rời khỏi hệ thống trong $(0,t)$
 - T_i : thời gian chiếm dùng của khách hàng i



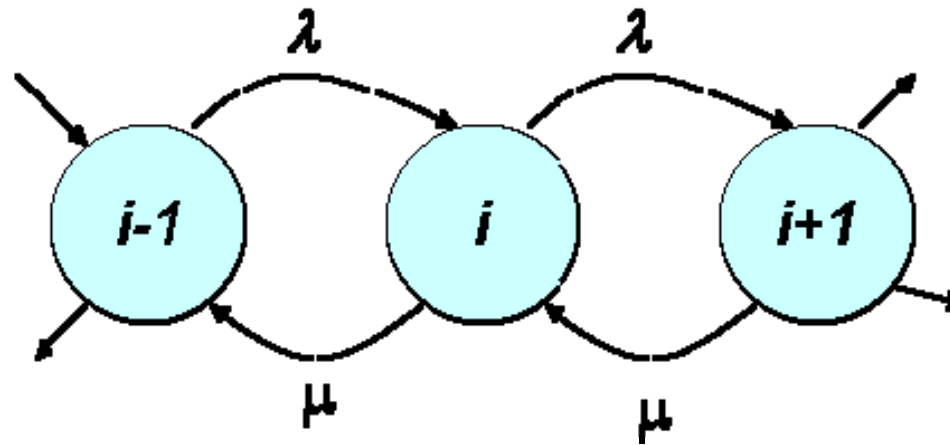
- Đẳng thức sau đây đúng vì 2 vế đều biểu diễn diện tích phần màu xanh:

$$\int_0^t N(t) = \sum_{i=1}^{\alpha_t} T_i$$

- Tương đương với:

$$\frac{1}{t} \int_0^t N(t) = \frac{\alpha_t}{t} \frac{\sum_{i=1}^{\alpha_t} T_i}{\alpha_t} \rightarrow dpcm$$

Quá trình sinh tử



- Trong đó λ_i, μ_i là tốc độ sự kiện đến và đi xét tại trạng thái i
- Với một hệ thống dừng và ổn định thì tổng các dòng đi vào một trạng thái bằng tổng các dòng đi ra

Các mô hình hàng đợi: ký hiệu kendal

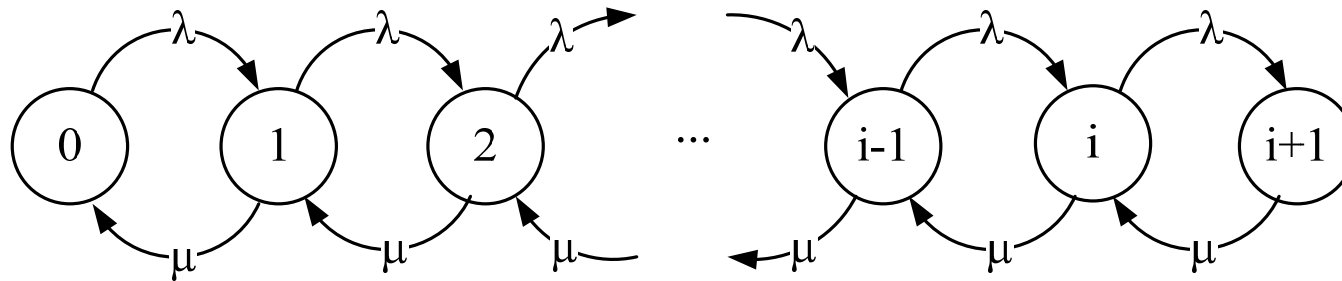
- Ký hiệu tổng quát cho một hệ thống hàng đợi:
 $A/S/c/B/R/SD$
- A: mô tả tiến trình đến, thường quan tâm đến khoảng thời gian giữa hai lần đến liên tiếp:
 - M: Tiến trình mũ (là tiến trình Markov hay tiến trình không nhớ). Tức trạng thái tiếp theo chỉ phụ thuộc vào trạng thái hiện tại, nhưng độc lập với việc nó đã có được như thế nào
 - Er: Tiến trình Erlang bậc r
 - Hr: Tiến trình siêu số mũ bậc r
 - D: Tiến trình tất định (deterministic)

Các mô hình hàng đợi: ký hiệu kendal

- S: mô tả tiến trình phục vụ, thường quan tâm đến khoảng thời gian cần thiết để phục vụ mỗi sự đến. Cũng có thể tuân theo 1 số luật như đối với tiến trình đến
- c: Số lượng server
- B: Kích thước bộ đệm. Đôi khi kí hiệu này bao gồm cả sự kiện đang được phục vụ
- R: Số nguồn phát ra sự kiện. Nếu các nguồn là tiến trình Markov thì tổng hợp của chúng cũng là 1 tiến trình Markov với:

$$\lambda = \sum \lambda_i$$

- SD: Quy tắc phục vụ
- Ví dụ : M/M/c/ ∞



- Tiến trình đến và tiến trình phục vụ có thuộc tính Markov. Thời gian giữa hai lần đến và thời gian phục vụ tuân theo phân bố mũ
- Hệ thống có 1 server, dung lượng đệm là vô hạn
- Mật độ lưu lượng: $\rho = \frac{\lambda}{\mu}$
- Số lượng khách hàng trung bình trong hệ thống: $L = \frac{\rho}{1-\rho}$
- Số lượng khách hàng trung bình trong server:

$$L_s = P(N \geq 1) = 1 - P(N = 0) = 1 - (1 - \rho) = \rho$$

- Số lượng khách hàng trung bình trong hàng đợi:

$$L_q = L - L_s = \frac{\rho}{1 - \rho} - \rho = \frac{\rho^2}{1 - \rho}$$

- Thời gian trung bình lưu trong hệ thống của mỗi khách hàng:

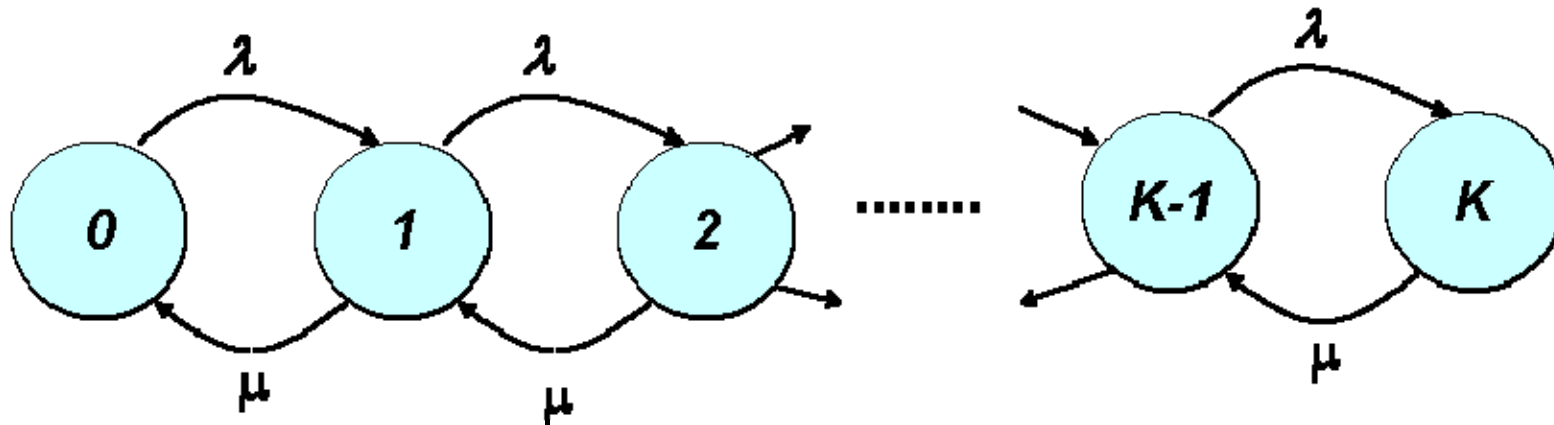
$$W = \frac{L}{\lambda} = \frac{\rho}{\lambda(1 - \rho)} = \frac{1}{\mu - \lambda}$$

- Thời gian phục vụ trung bình cho mỗi khách hàng: $W_s = \frac{1}{\mu} = \frac{\rho}{\lambda}$
- Thời gian đợi trung bình: $W_q = W - W_s = \frac{\rho^2}{\lambda(1 - \rho)}$

- Ví dụ: một luồng gói đến một thiết bị chuyển mạch gói với tốc độ trung bình là 180 packets/minute. Chiều dài gói trung bình là 1000 byte. Tốc độ của dòng số đầu ra là 20000byte/s. Giả thiết dung lượng đệm là đủ lớn. Hãy tính thời gian trễ trung bình tại hệ thống, số bản tin trung bình trong mỗi hệ thống, chiều dài hàng đợi và thời gian đợi trung bình.

- Tốc độ đến: $\lambda = \frac{240}{60} = 4$
- Thời gian phục vụ trung bình: $\mu = \frac{500}{100} = 5$
- Mật độ lưu lượng: $\rho = \frac{\lambda}{\mu} = \frac{4}{5} = 0.8$
- Số bản tin trong hệ thống L: $\frac{\rho}{1-\rho} = \frac{0.8}{1-0.8} = 4$
- Thời gian trong hệ thống: $\frac{L}{\lambda} = \frac{4}{4} = 1$
- Chiều dài hàng đợi: $\frac{\rho^2}{1-\rho} = \frac{0.8 \cdot 0.8}{1-0.8} = 3.2$
- Thời gian đợi trung bình: $\frac{\rho^2}{\lambda(1-\rho)} = \frac{L_q}{\lambda} = \frac{3.2}{4} = 0.8$

Hàng đợi M/M/1/K



$$\left\{ \begin{array}{l} \lambda p_0 = \mu p_1 \\ \lambda p_{n-1} + \mu p_{n+1} = (\lambda + \mu) p_n \\ \lambda p_{K-1} = \mu p_K; n = K \end{array} \right\}; n \in \{1, \dots, K-1\}$$

* K đã bao gồm khách hàng đang được phục vụ

- Do đó: $p_n = \rho^n p_0; n \in \{0, K\}$
- Mặt khác, cũng có: $\sum_{n=0}^K p_n = 1$
- Nên:

$$p_0 = \frac{1}{\sum_{n=0}^K \rho^n} = \frac{1-\rho}{1-\rho^{K+1}}$$

và:

$$p_n = \frac{1-\rho}{1-\rho^{K+1}} \times \rho^n$$

- Xác suất để một yêu cầu bị từ chối:

$$P_{loss} = p_K = \frac{1 - \rho}{1 - \rho^{K+1}} \times \rho^K$$

- Số sự kiện trung bình lưu trong hệ thống:

$$N = \sum_{n=1}^K n p_n = \frac{1 - \rho}{1 - \rho^{K+1}} \times \sum_{n=1}^K n \rho^n$$

- Đề ý:

$$f(\rho) = \sum_{n=0}^K \rho^n = \frac{1 - \rho^{K+1}}{1 - \rho}$$

$$f'(\rho) = \sum_{n=1}^K n \rho^{n-1} = \frac{1 - (K+1)\rho^K + K\rho^{K+1}}{(1-\rho)^2}$$

- Cuối cùng:

$$N = \frac{\rho}{1 - \rho} - (K + 1) \frac{\rho^{K+1}}{1 - \rho^{K+1}}$$

- Khi : $\rho \rightarrow 1$ thì: $p_n = p_0$ và:
$$p_n = \frac{1}{K+1}$$
- Mặt khác, cũng có:
$$\sum_{n=0}^K p_n = 1$$

- Suy ra:

$$N = \sum_{n=1}^K n p_n = \frac{K}{2}$$

- Số sự kiện trung bình trong hàng đợi:

$$N_q = \sum_{n=1}^K (n-1) p_n = \sum_{n=1}^K n p_n - \sum_{n=1}^K p_n = N - (1 - p_0)$$

- Hay:

$$N_q = \frac{1}{1-\rho} + \frac{\rho - (K+1)\rho^{K+1}}{1-\rho^{K+1}}$$

- Thời gian đợi trung bình được tính:

$$E(W_q) = \frac{1}{\mu} \frac{1}{1-p_k} [E(N) - kp_k] = \frac{1}{\mu} \frac{1}{1-p_k} \left[\frac{k}{2} - kp_k \right]$$

- Thời gian lưu trong hệ thống trung bình:

$$E(W) = \frac{E(N)}{\lambda} = \frac{k}{2\lambda}$$

- Tiến trình đến Markov, tham số λ
- Tiến trình phục vụ là bất kỳ
- Chiều dài hàng đợi không giới hạn
- Số sự kiện trung bình trong hệ thống:

$$E(N) = \frac{2\rho - \rho^2 + \lambda^2 \sigma_y^2}{2(1-\rho)} = \rho + \frac{\rho^2 + \lambda^2 \sigma_y^2}{2(1-\rho)} \quad (1)$$

- Trong đó σ_y^2 là phương sai của biến ngẫu nhiên $y = \frac{1}{\mu}$
- Thời gian trung bình lưu lại trong hệ thống:

$$E(W) = \frac{1}{\mu} + \frac{\rho + \lambda \mu \sigma_y^2}{2\mu(1-\rho)} = E \frac{(N)}{\lambda} \quad (2)$$

Trang web môn học

- <https://sites.google.com/site/fetphamvantien/home/fund-data-net>
- Bài tập dài và hướng dẫn đã được tải lên

- Tiến trình đến Markov, tham số λ
- Tiến trình phục vụ là bất biến, tức $\sigma_y^2 = 0$
- Từ (1) và (2) trong hàng đợi M/G/1 suy ra số sự kiện :

$$E(N)_D = E(N | \sigma_y^2 = 0) = \frac{\rho(2 - \rho)}{2(1 - \rho)}$$

- Số sự kiện trung bình trong hàng đợi:

$$E(N_q)_D = E(N) - \rho = \frac{\rho^2}{2(1 - \rho)}$$

- Thời gian trung bình lưu lại trong hệ thống:

$$E(W)_D = \frac{E(N)}{\lambda} = \frac{2 - \rho}{2\mu(1 - \rho)}$$

- Thời gian đợi trung bình:

$$E(W_q)_D = E(W) - \frac{1}{\mu} = \frac{\rho}{2\mu(1-\rho)}$$

Hàng đợi M/M/c/ ∞

- Tốc độ đến của các sự kiện là λ
- Tốc độ phục vụ của các server là μ_i với $i = 1, 2, \dots, c$. Để đơn giản, giả thiết $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_c = \mu$
- Mật độ lưu lượng : $\rho = \frac{\lambda}{c\mu}$
- Giả thiết thứ tự phục vụ của các server là từ 1, 2, 3, ..., c
- Trong sơ đồ sinh-tử ?

- Hệ phương trình cân bằng chuyển đổi trạng thái:

$$\left\{ \begin{array}{l} \lambda p_0 = \mu p_1 \\ \lambda p_{n-1} + (n+1)\mu p_{n+1} = (\lambda + n\mu) p_n; n \in \{1, \dots, c-1\} \\ \lambda p_{n-1} + c\mu p_{n+1} = (\lambda + c\mu) p_n; n > c-1 \\ \sum_{n=0}^{\infty} p_n = 1 \end{array} \right.$$

- Đặt $a = \frac{\lambda}{\mu} = c\rho$
- Xác suất để có n sự kiện trong hệ thống

$$p_n = \frac{1}{n!} a^n p_0 \quad \text{với } n < c$$

$$p_n = \frac{a^c}{c!} \rho^{n-c} p_0 \quad \text{với } n \geq c$$

- Tương tự như hàng đợi M/M/1/ ∞ ta có:

$$p_0 = \left[\sum_{n=0}^{c-1} a^n \frac{1}{n!} + \frac{a^c}{c! (1 - \rho)} \right]^{-1}$$

- Xác suất để một yêu cầu phải đợi (công thức Erlang C):

$$p\{W_q > 0\} = p\{N \geq c\} = p_q = 1 - \sum_{n=0}^{c-1} p_n = \frac{p_0 a^c}{c! (1 - \rho)}$$

- Hay:

$$p_q = \frac{a^c}{c! (1 - \rho)} \left[\sum_{n=0}^{c-1} a^n \frac{1}{n!} + \frac{a^c}{c! (1 - \rho)} \right]^{-1}$$

- Chiều dài hàng đợi trung bình:

$$E(N_q) = \sum_{n=c}^{\infty} (n-c) p_n = \frac{\rho}{1-\rho} p_q$$

- Số yêu cầu đang được phục vụ trung bình:

$$E(N_s) = \sum_{n=0}^c n p_n + \sum_{n=c+1}^{\infty} c p_n = \frac{\lambda}{\mu} = a$$

- Xác suất để thời gian đợi không lớn hơn giá trị t là:

$$P\{W_q \leq t\} = 1 - p_q e^{-\mu(c-a)t}$$

- Thời gian đợi trung bình:

$$E(W_q) = \frac{1}{\lambda} E(N_q) = \frac{1}{c\mu} \frac{1}{1-\rho} p_q$$

- Số yêu cầu trung bình trong hệ thống:

$$E(N) = E(N_q) + E(N_s) = \frac{\rho}{1-\rho} p_q + a$$

- Thời gian lưu lại trung bình trong hệ thống:

$$E(W) = \frac{1}{\lambda} E(N)$$

- Tốc độ đến của các sự kiện là λ
- Tốc độ phục vụ của các server là μ_i với $i = 1, 2, \dots, c$. Để đơn giản, giả thiết
$$\mu_1 = \mu_2 = \mu_3 = \dots = \mu_c = \mu$$
- Giả thiết thứ tự phục vụ của các server là từ 1, 2, 3, ..., c
- Sơ đồ quá trình sinh-tử ?

- Hệ phương trình cân bằng chuyển đổi trạng thái:

$$\left\{ \begin{array}{l} \lambda p_0 = \mu p_1 \\ \lambda p_{n-1} + (n+1)\mu p_{n+1} = (\lambda + n\mu) p_n \\ \lambda p_{c-1} = c\mu p_c; \\ \sum_{n=0}^c p_n = 1 \end{array} \right\}; n \in \{1, \dots, c-1\}$$

- Đặt $a = \frac{\lambda}{\mu} = c\rho$

- Xác suất để có n sự kiện trong hệ thống

$$p_n = \frac{1}{n!} a^n p_0 \quad \text{với } n \leq c$$

- Xác suất để một yêu cầu nào đó đến bị từ chối

$$p_c = \frac{\rho^c}{c!} \left[\sum_{i=0}^c \frac{\rho^i}{i!} \right]^{-1}$$

Đây chính là công thức Erlange B

Công thức Erlang

- Công thức Erlang C cho biết xác suất để một yêu cầu đến phải chờ trong hàng đợi:

$$p_q = \frac{a^c}{c! (1 - \rho)} \left[\sum_{n=0}^{c-1} a^n \frac{1}{n!} + \frac{a^c}{c! (1 - \rho)} \right]^{-1}$$

- Xác suất để một yêu cầu phải chờ trong khoảng thời gian t trước khi được phục vụ:

$$P\{W_q \leq t\} = 1 - p_q e^{-\mu(c-a)t}$$

- Công thức Erlang B cho biết xác suất một yêu cầu đến bị từ chối phục vụ ngay:

$$p_c = \frac{\rho^c}{c!} \left[\sum_{i=0}^c \frac{\rho^i}{i!} \right]^{-1}$$

- Lưu lượng phát sinh từ một người dùng (thời gian chiếm dùng H)

$$A_u = \lambda_u . H$$

- Lưu lượng phát sinh tổng cộng từ nhiều nguồn tổng hợp lại:

$$A = \lambda . H$$

- Lưu lượng mang A_c (không vượt quá số kênh) là giá trị lưu lượng thực tế được phục vụ
- Lưu lượng tổn thất là hiệu số giữa lưu lượng phát sinh và lưu lượng mang

$$A_l = A - A_c$$

- Đơn vị đo lưu lượng là Erlang, thường tính theo giờ

- Ví dụ: một người dùng có tốc độ phát sinh cuộc gọi là 3 cuộc/giờ, thời gian chiếm kênh trung bình là 10 phút. Vậy tải phát sinh từ người dùng này là $(30 \times 10) / 60 = 0,5E_r$
- Nếu có 100 người dùng như vậy thì lưu lượng phát sinh tổng cộng ?

$$A = \lambda.H = 100 \times 0,5 = 50E_r$$

- Hệ thống không hàng đợi (M/M/c/0) (blocked calls cleared):
 - Cuộc nối bị từ chối khi tất cả các kênh phục vụ đều bận
 - GoS được đánh giá qua công thức Erlang B

$$p(\text{blocking}) = \frac{A^c}{c!} \left[\sum_{i=0}^c \frac{A^i}{i!} \right]^{-1}$$



Grade of Service (GoS) Bậc phục vụ (GoS)

- Hệ thống có hàng đợi M/M/c/∞ (blocked calls delayed)
 - GoS được tính bằng xác suất một cuộc gọi phải chờ lâu hơn một khoảng thời gian chờ nào đó
 - Sử dụng công thức Erlang C để đánh giá

$$p(\text{delay} > 0) = \frac{A^c}{A^c + c!(1 - A/c) \sum_{n=0}^{c-1} A^n \frac{1}{n!}}$$

$$p(\text{delay} > t) = p(\text{delay} > 0) e^{-\frac{c - A}{H} t}$$

Bậc phục vụ (GoS)

- Ví dụ: Cho hệ thống trẽ có tốc độ cuộc gọi là 10 cuộc/giờ, thời gian chiếm kênh mỗi cuộc là 4 phút, số kênh phục vụ là 3. Tính lưu lượng phát sinh. Xác suất để một đẽ một cuộc gọi phải vào hàng đợi, xác suất để đợc phục vụ ngay là bao nhiêu ?

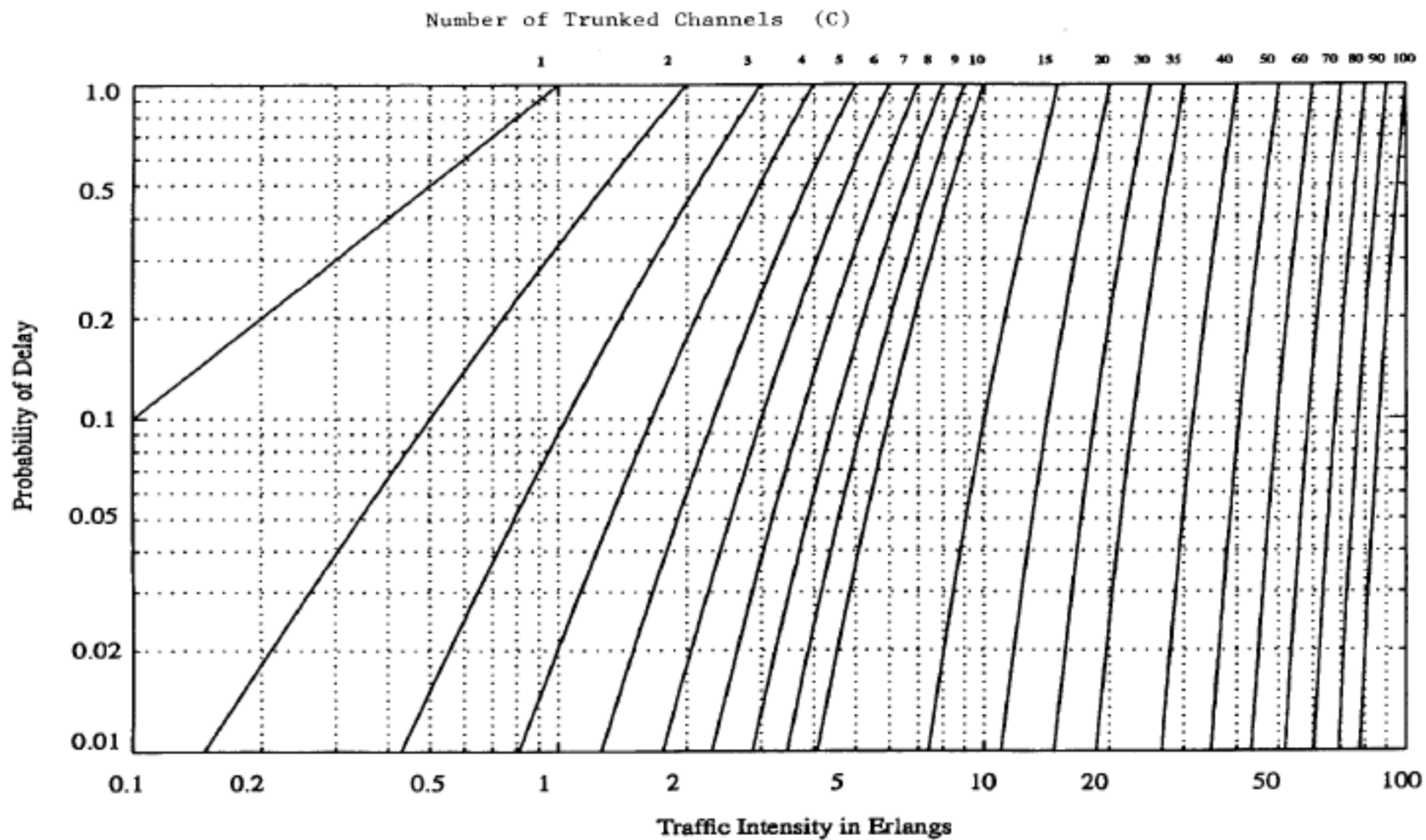
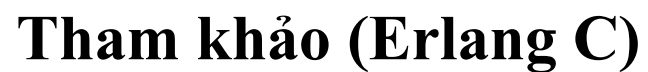


Bậc phục vụ (GoS)



VIỆN ĐIỆN TỬ - VIỄN THÔNG
School of Electronics and Telecommunications

ĐS: 4/9



Tham khảo (Erlang B)

