4.- Técnicas analíticas

4.1. Introducción
4.2. Cadenas de Markov
4.3. Redes de colas
4.4. Redes de Petri

4.1. Introducción

- Las técnicas analíticas permiten el estudio de un sistema a partir de un modelo del mismo.
- Pueden utilizarse en cualquier momento del ciclo de vida del sistema
- Con las técnicas analíticas se puede realizar:
  - Análisis cualitativo del sistema:
    - » Corrección: ¿Funciona el sistema?
    - » Verificación de propiedades lógicas (ausencia de bloqueos, vivacidad, limitación, ergodicidad, ...)
  - Análisis cuantitativo: ¿Cómo de bien funciona?
    - » Velocidad del sistema en realizar una tarea
    - » Utilización del sistema
    - » Disponibilidad (tiempo durante el cual el sistema está operativo)
- Para obtener soluciones eficientes, los modelos no pueden tener un gran nivel de detalle.

Clasificación de las técnicas analíticas

- Por el formalismo utilizado
  - Cadenas de Markov
  - Redes de colas
  - Redes de Petri
- Por la técnica de resolución
  - Enumerativas: Basadas en los estados del modelo
  - Reducción-transformación
  - Estructurales: Basadas en el modelo del sistema
- Por el objeto de estudio
  - Análisis de la etapa transitoria
  - Análisis del estado estacionario
- Por la calidad de los resultados
  - Valores exactos
  - Cotas
  - Aproximaciones

Cadenas de Markov

- Formalismo basado en los estados del modelo
- Existen técnicas para:
  - Técnicas enumerativas
  - Análisis transitorio y estacionario
  - Análisis exacto y aproximado
- Principal problema: Bajo nivel de abstracción
- La dimensión del modelo es igual al número de estados del sistema
- Sólo en casos muy particulares se pueden agregar estados (técnicas de reducción-transformación). Las condiciones de agregación son muy fuertes.

Redes de colas

- Alto nivel de abstracción

  El número de estados considerados para caracterizar el sistema es exponencial en el tamaño del modelo.

- Técnicas de resolución:
  - –Enumerativas (basadas en Cadenas de Markov)
  - –Reducción-transformación (exactas y aproximadas)
  - –Estructurales (forma producto -> exacta)
- Existen técnicas para:
  - –Análisis transitorio y estacionario
  - –Análisis exacto, aproximado y cotas
- Principal problema: No existe primitiva para modelar sincronizaciones
- Existen extensiones del formalismo para modelar sincronizaciones, pero destruyen la forma producto y por lo tanto la eficiencia de la resolución.

Redes de Petri

- Nivel de abstracción parecido al de las redes de colas
- Incluyen primitiva de sincronización
- Si se añade temporización exponencial al modelo se tienen las redes de Petri estocásticas (SPN)

  SPN = PN + temporización

  = Redes de colas + sincronizaciones
- Amplia gama de técnicas de análisis cualitativo
  - Enumerativas
  - Reducción-transformación
  - Estructurales
- Técnicas de análisis cuantitativo
  - Análisis exacto: Métodos enumerativos basados en Cadenas de Markov
  - Cotas (técnicas estructurales).
  - Análisis aproximado: Tema de investigación (técnicas de reducción-transformación).

4.2. Cadenas de Markov

- Definiciones básicas
- Cadenas de Markov en tiempo discreto (CMTD)
  - Concepto y representación
  - Distribución en el paso *n*
  - Estado estacionario: Clasificación de estados y ecuaciones de balance
  - Ejemplo
- Cadenas de Markov en tiempo continuo (CMTC)
  - Concepto y representación
  - Estado estacionario
  - Ejemplo

Definiciones básicas

- **Proceso estocástico**: Es una sucesión de variables aleatorias $\{X_n,\ n \in \mathrm{N}\}$definidas sobre el mismo espacio probabilístico
  - El índice suele representar el tiempo
  - El conjunto de valores que pueden tomar las variables aleatorias es el espacio de estados *S*
  - El espacio de estados puede ser continuo o discreto (nos centraremos en los discretos)
  - El tiempo puede ser:
    - » Discreto -> CMTD (DTMC en inglés)
    - » Continuo -> CMTC (CTMC en inglés)
- **Propiedad de Markov:**

$$P\left\{X_t \leq x \middle| X_{t_n} = x_n, \ldots, X_{t_0} = x_o\right\} =$$

$$= P\left\{X_t \leq x \mid X_{t_n} = x_n\right\}$$

$$t > t_n > \ldots > t_o$$

- **Proceso de Markov:** Proceso estocástico que cumple la propiedad de Markov.

Cadenas de Markov en tiempo discreto

- **Definición**: Es un proceso de Markov $\{X_n, n \in \mathrm{N}\}$ tal que $\mathrm{P}\{X_{n+1} = x_{n+1} \mid X_n = x_n\}$ es independiente del tiempo n.
- **Representación**:
  - Por las probabilidades de transición entre estados:

    $p_{ij} = \mathrm{P}\{X_{n+1} = j \mid X_n = i\}$

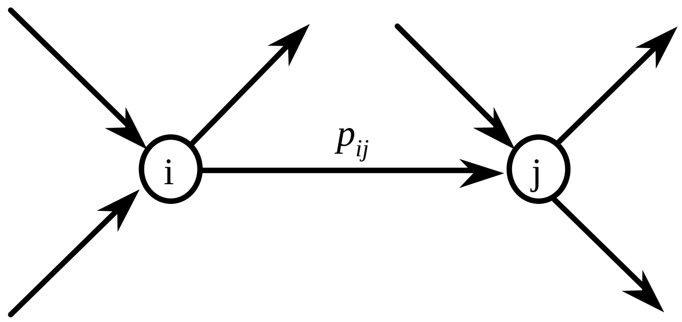
    $P=[p_{ij}]$ matriz de probabilidades de transición

    $\Pi^{(n)}$ distribución de probabilidad en el instante n
  - Mediante la ecuación de estado

$$P\{X_{n+1} = j\} = \sum_{i \in S} p_{ij} \cdot P\{X_n = i\} \Rightarrow$$

$$\Rightarrow p_j^{(n+1)} = \sum_{i \in S} p_{ij} \cdot p_i^{(n)} \Rightarrow \boxed{\left(\Pi^{(n+1)}\right)^t = \left(\Pi^{(n)}\right)^t \cdot P}$$

  - Mediante el grafo de transición entre estados



CMTD: Distribución en el paso *n*

$$\left(\Pi^{(n+1)}\right)^t = \left(\Pi^{(n)}\right)^t \cdot P \Rightarrow$$

$$\Rightarrow \left(\Pi^{(n)}\right)^t = \left(\Pi^{(n-1)}\right)^t \cdot P^1 \Rightarrow \left(\Pi^{(n)}\right)^t = \left(\Pi^{(0)}\right)^t \cdot P^n$$

Sea $p_{ij}^{(k)} = P\{X_{k+m} = j \mid X_m = i\}$ (probabilidad de ir del estado i al j en k pasos). Es obvio que:

Ecuaciones de Chapman-Kolmogorov:

$$\boxed{p_{ij}^{(k)} = \sum_{m \in S} p_{im}^{(k-l)} \cdot p_{mj}^{(l)}}$$

De aquí:

$$\left[p_{ij}^{(k)}\right] = \left[p_{ij}^{(k-l)}\right] \cdot \left[p_{ij}^{(l)}\right] \qquad \forall k \in N, l = 1, \ldots, k-1$$

Como $\left[p_{ij}^{(1)}\right] = P$, entonces $\left[p_{ij}^{(k)}\right] = P^k$

CMTD: Estado estacionario

- **Clasificación de estados**

  Sea $f_i^{(k)}$ la probabilidad de que el primer retorno a i ocurra k pasos después de dejarlo. Y sea

  $$f_i = \sum_{k=1}^{\infty} f_i^{(k)} \quad \text{(probabilidad de retorno a } i\text{)}$$

  Un estado puede ser:

  » **Transitorio**: Si f_i<1 (la CM puede no volver a i)

  » **Recurrente**: Si f_i=1 (la CM siempre retorna a i)

  » **Absorbente**: Si p_{ii}=1 (la CM se queda en i siempre)

  Tiempo medio de retorno $$q_i = \sum_{k=1}^{\infty} k \cdot f_i^{(k)}$$

  Un estado recurrente (f_i=1) puede ser:

  » **Recurrente nulo**: Si q_i =

  » **Recurrente no nulo**: Si q_i <•

  También es posible clasificar los estados estudiando las componentes fuertemente conexas del grafo de transición entre estados

CMTD: Estado estacionario

- Subconjunto cerrado de estados: Conjunto de estados fuertemente conexo (todos los estados conectados entre si)
- Una CMTD es **irreducible** si S es cerrado y no contiene subconjuntos propios cerrados. En este caso se cumple:
  - El grafo de transición entre estados es fuertemente conexo
  - Todos los estados son transitorios, recurrentes nulos o recurrentes no nulos
  - No hay estados absorbentes.
- **Probabilidades límite**:

$$p_j = \lim_{k\to\infty} p_j^{(k)}$$

- Si la CMTD es irreducible P_i son independientes de $p_i^{(0)} \quad \forall\ i \in S$
  - Si todos los estados son transitorios o recurrentes nulos $\Pi_i = 0 \quad \forall\ i \in S$
  - Si todos los estados son recurrentes no nulos $\Pi_i > 0 \quad \forall\ i \in S$ (la CM se dice ergódica y las probabilidades límite son únicas).

CMTD: Estado estacionario

- Cálculo de las probabilidades límite:

$$\left(\Pi^{(k+1)}\right)^t = \left(\Pi^{(k)}\right)^t P$$

- Haciendo tender k a •

$$\boxed{\begin{array}{ll} \Pi^t = \Pi^t P & \text{(ecuaciones de balance)} \\ \sum_{i \in S} p_i = 1 & \text{(normalización)} \end{array}}$$

- Otros índices de prestaciones:
  - Tiempo medio de retorno al estado i

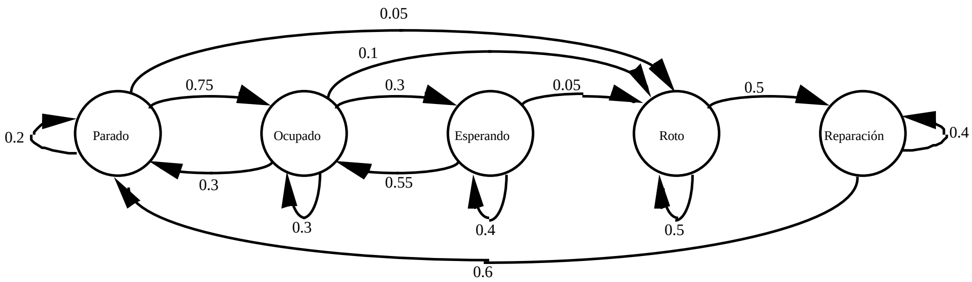
$$q_i = \frac{1}{p_i}$$

  - Ratio de visita entre dos estados i, j

$$v_{ij} = \frac{p_i}{p_j}$$

CMTD: Ejemplo

- Un procesador puede estar en los siguientes estados:
  - Parado: Nada que hacer
  - Ocupado: Trabajando en una tarea
  - Esperando: Parado a la espera de un recurso
  - Roto: No es operativo por más tiempo
  - Reparación del fallo correspondiente



$$
\begin{array}{cc}
 & \begin{array}{ccccc} \text{Parado} & \text{Ocupado} & \text{Esperando} & \text{Roto} & \text{Reparación} \end{array} \\
P = & \begin{bmatrix} 0.2 & 0.75 & 0.0 & 0.05 & 0.0 \\ 0.3 & 0.3 & 0.3 & 0.1 & 0.0 \\ 0.0 & 0.55 & 0.4 & 0.05 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.5 \\ 0.6 & 0.0 & 0.0 & 0.0 & 0.4 \end{bmatrix}
\end{array}
$$

CMTD: Ejemplo

- Solución en estado estacionario:

$$\left.\begin{array}{r}\Pi^t = \Pi^t P \\ \Pi^t 1 = 1\end{array}\right\}$$

$$\begin{array}{lccccc} & \text{parado} & \text{ocupado} & \text{esperando} & \text{roto} & \text{reparando} \\ P^t= & (0.2155, & 0.3804, & 0.1902, & 0.1167, & 0.0972) \end{array}$$

- Otros índices de prestaciones:
  - Disponibilidad = parado+ocupado+esperando = 0.7861

    (el 78.61 % del tiempo)
  - No disponibilidad = 1- disponibilidad = 0.2139

    (el 21.39 % del tiempo)
  - Tiempo trabajando = ocupado+esperando = 0.5706

    (el 57.06 % del tiempo)

Cadenas de Markov en tiempo continuo

- **Definición**:Es un proceso estocástico $\{X(t_n), n \in \mathrm{N},\, t_n \in \mathrm{R},\, t_n > t_{n-1} > ... > t_0\}$ tal que

  $$P\left\{X(t_{n+1}) = x_{n+1} \middle| X(t_n) = x_n,\, X(t_{n-1}) = x_{n-1}, ..., X(t_o) = x_o\right\} =$$
  $$= P\left\{X(t_{n+1}) = x_{n+1} \middle| X(t_n) = x_n\right\}$$

- Notación: $p_{ij}(t,q) = P\{X(q) = j | X(t) = i\}$
- CMTC **homogénea**: Si $t = q - t$, una CMTC se dice homogénea si $p_{ij}(t) = p_{ij}(t,q)\ \forall\ t \in \mathrm{R}$

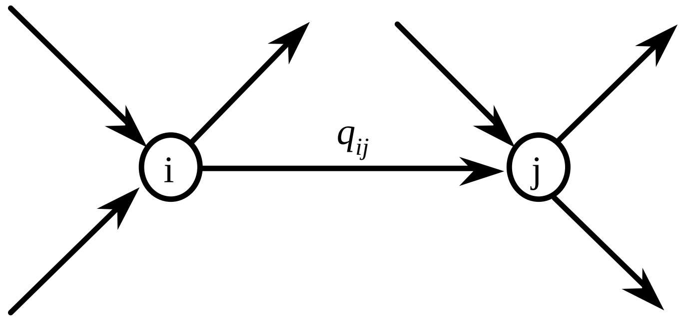
  En este caso, el tiempo de permanencia en un estado no depende del tiempo pasado ya en el estado (distribución sin memoria), es decir, está distribuido exponencialmente.
- Una CMTC puede verse como una CMTD con tiempos de permanencia en los estados distribuidos exponencialmente
- Sean:

  $$q_{ij} = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t} \qquad q_{ii} = \lim_{\Delta t \to 0} \frac{p_{ii}(\Delta t) - 1}{\Delta t}$$

  q_{ij} es la tasa de paso del estado i al j, $-q_{ii}$ es la tasa a la que se deja el estado i
- $Q=[q_{ij}]$ es el **generador infinitesimal** de la CMTC
- Es obvio que $Q1=0$

CMTC: Representación

- Por el grafo de tasas de transición entre estados



- Por las probabilidades de transición entre estados

$$P(t) = \left[p_{ij}(t)\right] \qquad\qquad P(0) = I$$

$\Pi(t)$ son las probabilidades en el instante t

$$\Pi(t)^t = \Pi(0)^t P(t)$$

Ecuaciones de Chapman-Kolmogorov:

$$P(t) = P(t-q)P(q)$$

CMTC: Estado estacionario

- La distribución de probabilidades en estado estacionario pueden obtenerse resolviendo el sistema

$$\boxed{\left.\begin{array}{r} \Pi^t Q = 0 \\ \Pi^t 1 = 1 \end{array}\right\}}$$

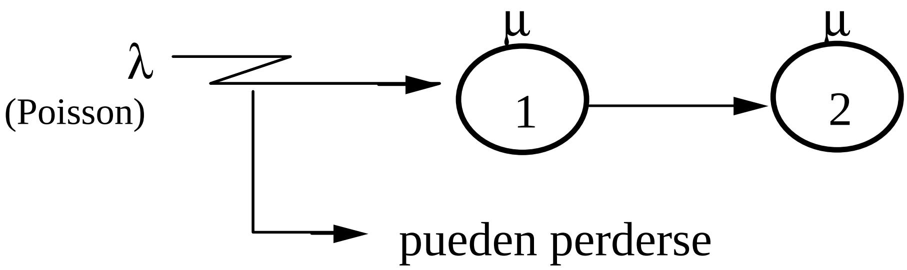
- Otros índices de prestaciones:
  - Tasa de salida del estado *i*:

$$r_i = \sum_{j \in S, j \neq i} q_{ij}$$

  - Tiempo medio de permanencia en el estado *i*:

$$\frac{1}{r_i}$$

CMTC: Ejemplo



- Los trabajos llegan al sistema con tiempo entre llegadas exponencial de tasa λ (proceso Poisson)
- Tiempo de servicio en las dos estaciones exponencial de tasa μ
- Si al acabar un trabajo en la estación 1, la 2 está ocupada, la 1 queda bloqueada
- Si la estación 1 está ocupada cuando llega un nuevo trabajo, éste se pierde.
- Determinar la proporción de trabajos perdidos, el número medio de estaciones trabajando y el número medio de trabajos que hay en el sistema

El conjunto de estados del sistema es:

S={(0,0), (1,0), (0,1), (1,1), (b,1)}

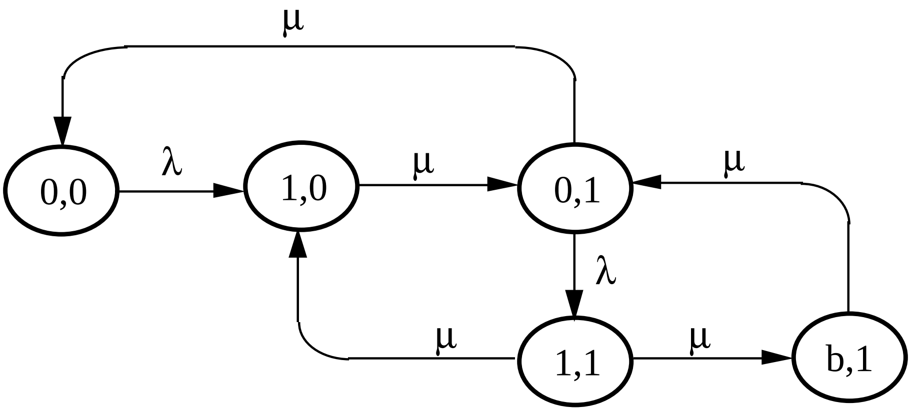
0 -> estación vacía

1- > estación trabajando

b -> estación bloqueada

CMTC: Ejemplo

Diagrama de tasas de transición entre estados



Generador infinitesimal

$$Q = \begin{array}{c} 00 \\ 10 \\ 01 \\ 11 \\ b1 \end{array} \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 \\ 0 & -\mu & \mu & 0 & 0 \\ \mu & 0 & -(\lambda+\mu) & \lambda & 0 \\ 0 & \mu & 0 & -2\mu & \mu \\ 0 & 0 & \mu & 0 & -\mu \end{bmatrix}$$

Solución en estado estacionario

$$\left.\begin{array}{r} \Pi^t Q = 0 \\ \Pi^t \mathbf{1} = 1 \end{array}\right\} \Rightarrow \Pi^t = \left(\frac{2}{\Delta}, \frac{2r + r^2}{\Delta}, \frac{2r}{\Delta}, \frac{r^2}{\Delta}, \frac{r^2}{\Delta} \right)$$

$$r = \frac{l}{m} \qquad \Delta = 3r^2 + 4 + 2$$

CMTC: Ejemplo

- Proporción de trabajos perdidos

  Es la probabilidad de que cuando llegue un trabajo nuevo, la primera estación no esté vacía, es decir:

$$p_{10} + p_{11} + p_{b1} = \frac{3r^2 + 2r}{3r^2 + 4r + 2}$$

- Número medio de estaciones trabajando

  En el estado (0,0) no hay ninguna estación trabajando y en el (1,1) hay dos. En el resto de estados sólo trabaja una estación, luego

$$B = p_{01} + p_{10} + p_{b1} + 2p_{11} = \frac{4r^2 + 4r}{3r^2 + 4r + 2}$$

- Número medio de trabajos en el sistema

  En el estado (0,0) no hay ninguno, en los estados (1,1) y (*b*,1) hay dos y en el resto uno, luego

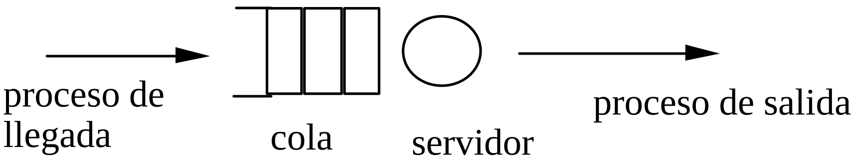
$$L = p_{01} + p_{10} + 2p_{b1} + 2p_{11} = \frac{5r^2 + 4r}{3r^2 + 4r + 2}$$

4.3. Redes de colas

- Colas simples
  - Introducción
  - Notación de Kendall
  - Ley de Little
  - Procesos de nacimiento y muerte
  - Colas básicas
    - » M/M/1
    - » M/M/c
    - » M/M/•
- Redes de colas
  - Introducción
  - Matriz de routing
  - Ratios de visita y utilizaciones relativas
  - Redes de colas forma producto
    - » Redes abiertas (Jackson)
    - » Redes cerradas (Gordon-Newell)

Colas simples

- **Cola**: Línea de espera de clientes para recibir un servicio determinado



- **Estado del modelo**: El número de clientes en la cola (incluidos los que están siendo servidos)
  - El espacio de estados puede ser infinito (si se admite una población infinita)
- **Notación de Kendall**

Las colas pueden clasificarse atendiendo a diversos criterios:

1.- Patrón de llegada de los clientes

» Distribución del tiempo entre llegadas

» Llegadas individuales o en grupos (lotes)

» Reacción de los clientes al llegar a la cola (clientes impacientes, etc.)

Notación	Distribución
M	exponencial
D	determinista
E_k	Erlang tipo k
H_k	hiperexponencial tipo k
G	general

Notación de Kendall

2.- Patrón de servicio

- Distribución del tiempo de servicio
- Individual o en lotes
- Dependiente del estado o no

3.- Número de servidores

- Uno, varios, infinitos, etc.

4.- Capacidad del sistema

Tamaño de la sala de espera -> espacio de estados finito

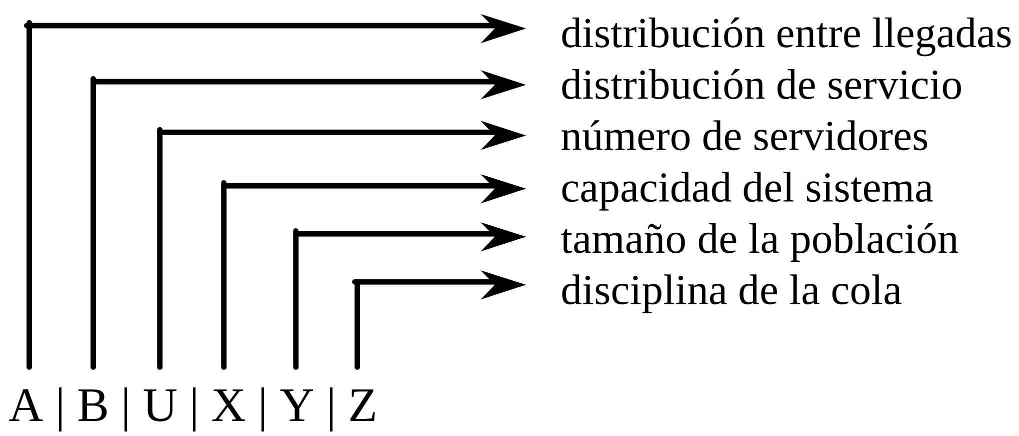
5.- Tamaño de la población

Número de clientes potenciales -> esp. de estados finito

6.- Disciplina de servicio

- FCFS (FIFO): El primero en entrar es el primero en salir (cola)
- LCFS (LIFO): El último en entrar es el primero en salir (pila)
- RR: (Round robin). Cada cliente recibe una pequeña cantidad fija de tiempo cíclicamente
- PS: (Processor sharing). Situación límite de la anterior cuando el tiempo fijo tiende a 0
- Aleatoria
- Prioridades: Algunos clientes tienen preferencia
  - Non preemptive: No interrumpe
  - Preemptive-resume: Interrumpe y continua
  - Preemptive-restart: Interrumpe y comienza

Notación de Kendall



- Ejemplos:

  M/M/1

  M/M/c

  M/M/•

  M/M/c/k

  M/M/1/k/P

  $M/E_k/1$

Utilización del sistema

- Uno de los índices de prestaciones más importantes
- Definición: Proporción de tiempo que el sistema está ocupado (se denota con ρ)
- En sistemas con muchos componentes el incremento de la carga se traduce en un aumento de la utilización de los componentes (no de forma uniforme)
- Cuello de botella (bottleneck): Componente del sistema con utilización cercana a 1 cuando el resto de componentes tienen utilizaciones muy inferiores. El cuello de botella es el componente que está ralentizando todo el sistema.

x_n = tiempo de servicio del cliente n

a_n = tiempo entre las llegadas n-1 y n

Para colas de un solo servidor:

$$r = \lim_{N\to\infty} \frac{\sum_{n=1}^{N} x_n}{\sum_{n=1}^{N} a_n} = \lim_{N\to\infty} \frac{\sum_{n=1}^{N} x_n / N}{\sum_{n=1}^{N} a_n / N} = l\,\bar{x}$$

λ = tasa de llegadas (llegadas por unidad de tiempo)

$\bar{x}$ = tiempo medio de servicio

Para colas con m servidores: $r = \dfrac{l\,\bar{x}}{m}$

Ley de Little

- Permite calcular el número medio de clientes en una cola observando el proceso de llegadas y el tiempo medio de estancia en el sistema de los clientes.
- Ejemplo: Un espía de Burguer King intenta calcular cuantos clientes tiene un McDonald´s. No puede entrar, pero observa lo siguiente:
  - Cada hora llegan 32 clientes
  - Cada uno está dentro una media de 12 minutos

  Aplicando la ley de Little se puede calcular el número medio N de clientes dentro del McDonald´s

  $N = l\,T$=0.53 clientes/min * 12 min = 6.4 clientes

Ley de Little

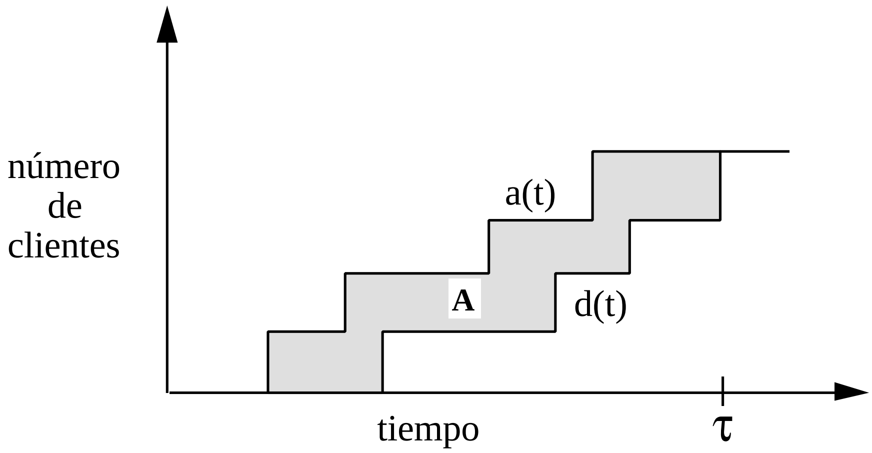
Demostración:

$a(t)$ = número de llegadas hasta el instante t

$d(t)$ = número de salidas hasta el instante t

$n(t) = a(t) - d(t)$ = número de clientes en el instate t

t_k = tiempo de permanencia del cliente k (espera + servicio)



$$l = \frac{a(t)}{t} = \text{ tasa media de llegada}$$

$$T = \frac{1}{a(t)} \sum_{k=1}^{a(t)} t_k = \text{ tiempo medio de permanencia}$$

$$N = \frac{1}{t} \int_0^t n(t)dt = \text{ número medio de clientes en el sistema}$$

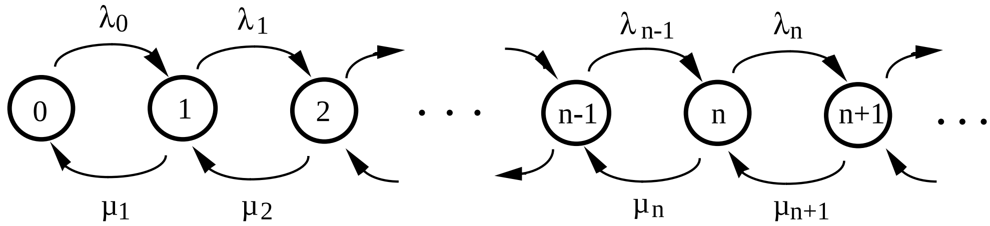
$$A = \int_0^t n(t)dt = \sum_{k=1}^{a(t)} t_k \Rightarrow$$

$$\Rightarrow t \cdot N = a(t) \cdot T = t \cdot l \cdot T \Rightarrow \boxed{N = l \cdot T}$$

Procesos de nacimiento y muerte

- El espacio de estados de una cola es N (o un subconjunto propio suyo) representando el número de clientes del sistema (esperando + servicio).
- Puede representarse por medio de una CMTC con estados 0, 1, 2, ...

En un tiempo infinitesimal el siguiente cambio de estado de la CMTC solo puede ser -1, 0 ó +1 (de ahí el nombre de proceso de nacimiento y muerte)



l_n = tasa de nacimiento en el estado n

m_n = tasa de muerte en el estado n

$$Q=\begin{bmatrix} -l_0 & l_0 & 0 & 0 & 0 & \cdots \\ m_1 & -(l_1+m_1) & l_1 & 0 & 0 & \cdots \\ 0 & m_2 & -(l_2+m_2) & l_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Solución: $$\left.\begin{aligned} \Pi^t Q &= 0 \\ \Pi^t 1 &= 1 \end{aligned}\right\}$$

Procesos de nacimiento y muerte

- Ecuaciones de balance

$$\left.\begin{array}{r} l_0 p_0 = m_1 p_1 \\ \left(l_n + m_n\right) p_n = l_{n-1} p_{n-1} + m_{n+1} p_{n+1} \quad n \geq 1 \end{array}\right\}$$

- La solución del sistema es:

$$\left.\begin{array}{r} p_n = \dfrac{l_0 \cdots l_{n-1}}{m_1 \cdots m_n} p_0 \quad n \geq 1 \\ \displaystyle\sum_{n=0}^{\infty} p_n = 1 \end{array}\right\} \Rightarrow$$

$$\Rightarrow \boxed{\begin{array}{l} p_0 = \dfrac{1}{1 + \displaystyle\sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{l_i}{m_{i+1}}} \\ p_n = \dfrac{\displaystyle\prod_{i=0}^{n-1} \frac{l_i}{m_{i+1}}}{1 + \displaystyle\sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{l_i}{m_{i+1}}} \quad n \geq 1 \end{array}}$$

- Condición de ergodicidad: $p_n > 0 \quad \forall \quad n \in \mathrm{N} \Leftrightarrow$

$$\Leftrightarrow \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{l_i}{m_{i+1}} < \infty$$

Colas básicas: M/M/1

- Las tasas de nacimiento y muerte son constantes:

$$l_n = l \qquad \forall n \geq 0$$

$$m_n = m \qquad \forall n \geq 1$$

Entonces la solución en estado estacionario es:

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \left(\frac{l}{m} \right)^n} \quad ; \quad p_n = \left(\frac{l}{m} \right)^n p_0$$

- Condición de ergodicidad:

$$\sum_{n=1}^{\infty} \left(\frac{l}{m} \right)^n < \infty \quad \Leftrightarrow \quad \boxed{l < m}$$

$$r = \frac{l}{m} \quad \text{(utilización)}$$

- Entonces: $p_0 = \dfrac{1}{1 + \dfrac{r}{1 - r}} = 1 - r \Rightarrow$

$$\Rightarrow \quad \boxed{\begin{array}{l} p_0 = 1 - r \\ p_n = r^n (1 - r) \quad n \geq 1 \end{array}}$$

(distribución geométrica)

Colas básicas: M/M/1

- Casos patológicos

  λ>μ: CM transitoria

  λ=μ: CM recurrente nula (tráfico pesado)

- Otros índices de prestaciones
  - Número medio de clientes en el sistema

$$N = \sum_{n=0}^{\infty} n \cdot p_n = \sum_{n=0}^{\infty} n(1-r)r^n = \frac{r}{1-r}$$

  - Número medio de clientes en servicio

$$N_s = 0 \cdot p_0 + 1 \cdot \sum_{n=1}^{\infty} p_n = r \quad \text{(utilización)}$$

  - Número medio de clientes en cola (esperando)

$$N_q = N - N_s = \frac{r}{1-r} - r = \frac{r^2}{1-r}$$

  - Espera media (cola + servicio): Por la Ley de Little

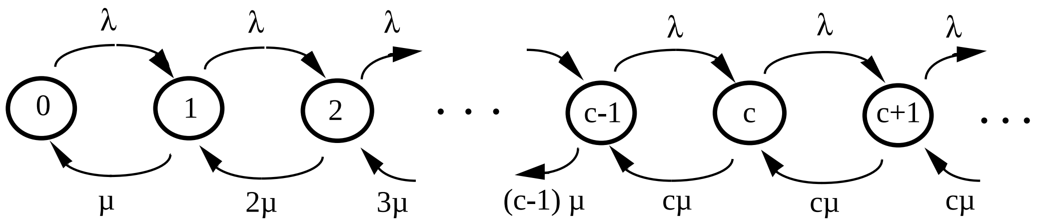
$$N = l \cdot T \Rightarrow T = \frac{N}{l} = \frac{r}{1-r} \cdot \frac{1}{l} = \frac{1/m}{1-r}$$

  - Tiempo medio de espera en la cola

$$W = T - \frac{1}{m} = \frac{1/m}{1-r} - \frac{1}{m} = \frac{l/m^2}{1-r}$$

Colas básicas: M/M/c

- Hay *c* servidores trabajando en paralelo (como en una peluquería)



- Tasas de nacimiento iguales y de muerte dependientes del estado

$$l_n = l \qquad \forall n \in \mathrm{N}$$

$$m_n = \begin{cases} nm & si \quad 1 \le n \le c \\ cm & si \quad n \ge c \end{cases}$$

- Solución en estado estacionario

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

$$p_n = \begin{cases} \dfrac{1}{n!}\left(\dfrac{l}{m}\right)^n p_0 & si \quad 1 \le n \le c \\ \dfrac{1}{c^{n-c}c!}\left(\dfrac{l}{m}\right)^n p_0 & si \quad n \ge c \end{cases}$$

- Condición de ergodicidad: $\boxed{\lambda<c\mu}$

Colas básicas: M/M/c

- Otros índices de prestaciones
  - Número medio de clientes en la cola (esperando)

$$N_q = \sum_{n=c}^{\infty} (n-c) p_n = \frac{\left(l/m \right)^c l m}{(c-1)!(cm - l)^2} p_0$$

  - Tiempo medio de espera en cola (Little)

$$W = \frac{N_q}{l} = \frac{\left(l/m \right)^c m}{(c-1)!(cm - l)^2} p_0$$

  - Espera media (cola + servicio)

$$T = W + \frac{1}{m}$$

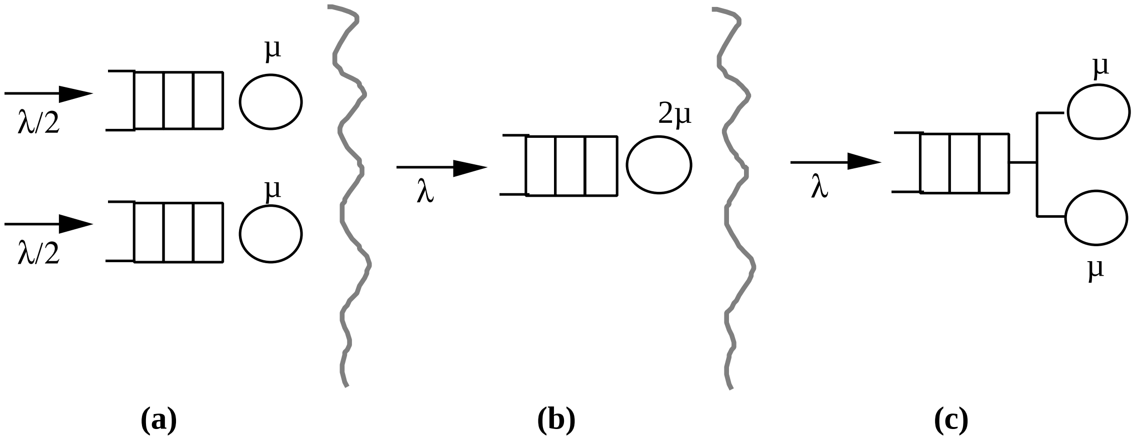
  - Número medio de clientes en el sistema: (Little)

$$N = l T$$

Ejercicio

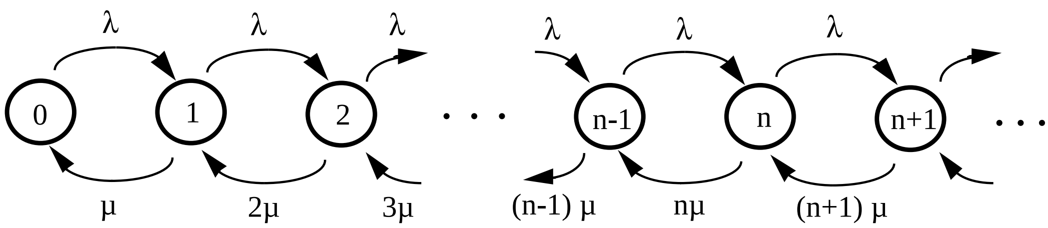
Comparación de colas simples

- Comparar la espera media (cola más servicio) de los clientes en estos tres sistemas:
  - a) Dos colas M/M/1 independientes con tasa de llegada $\lambda/2$ y tasa de servicio μ.
  - b) Una cola M/M/1 con tasa de llegada λ y tasa de servicio 2μ.
  - c) Una cola M/M/2 con tasa de llegada λ y tasa de servicio μ para cada servidor.



Colas básicas: M/M/•

- Hay tantos servidores como clientes lleguen al sistema.



- Tasas de nacimiento iguales y de muerte dependientes del estado

$$\lambda_n = \lambda \qquad \forall n \in \mathrm{N}$$

$$\mu_n = n\mu \qquad \forall n \in \mathrm{N}$$

- Solución en estado estacionario

$$p_n = \frac{\left(\lambda/\mu\right)^n}{n!} e^{-\lambda/\mu} \qquad \text{(distribución de Poisson)}$$

- Otros índices de prestaciones
  - Número medio de clientes en el sistema

$$N = \sum_{n=0}^{\infty} n p_n = \frac{\lambda}{\mu}$$

  - Espera media (cola+servicio): (Little)

$$T = \frac{N}{\lambda} = \frac{1}{\mu}$$

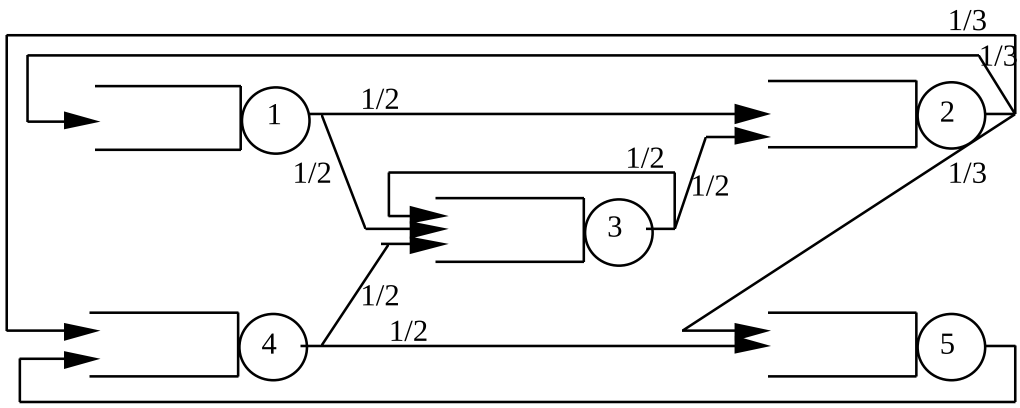
Obviamente
el tiempo medio de servicio

Otras colas básicas

- Se pueden resolver otras colas, pero no las detallaremos.
  - M/M/c/K
  - M/M/1/K/P
  - $M/E_k/1$
  - M/G/1

Redes de colas

- Muchos sistemas informáticos (y otros) se pueden ver como una serie de estaciones en las que se ofrecen unos servicios. Cada estación tiene su propia cola de espera. Esto nos lleva a estudiar las redes de colas.



- Los modelos de redes de colas son (de las técnicas analíticas) los más ampliamente utilizados por el momento para evaluación de prestaciones de sistemas informáticos.
- El problema de encontrar soluciones para estos modelos es, en general, muy duro.
- Existen subclases de redes para las que pueden encontrarse soluciones exactas de forma eficiente por medio de técnicas numéricas.

Matriz de routing

- Es la matriz formada por las probabilidades de transición entre las diferentes colas de la red. Por lo tanto describe las rutas que siguen los clientes a través de la red.
- Un cliente, cuando completa el servicio en una cola, se dirige a otra de forma aleatoria, siguiendo una distribución discreta de probabilidad.

  p_{ij}= probabilidad de que un cliente vaya de la cola i a la j
- **Matriz de routing**:P=[p_{ij}]. Es una matriz MxM, siendo M el número de colas de la red.
- Ejemplo anterior

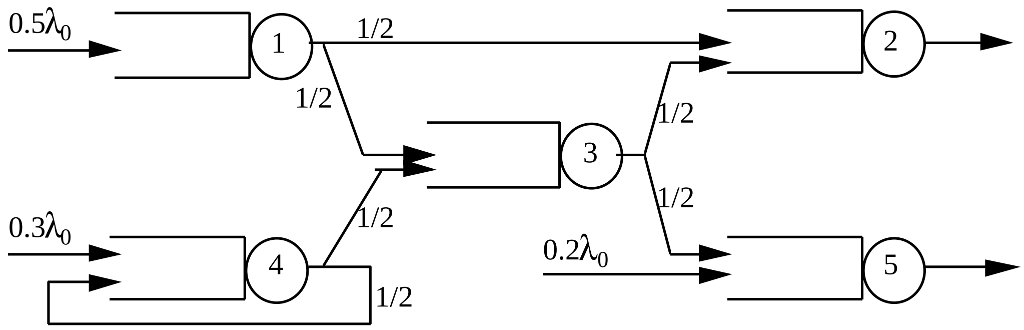
$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

- La matriz de routing de una red de colas es una matriz estocástica (P`=1)

Matriz de routing

- **Redes de colas abiertas**: Son redes en las que pueden llegar clientes del exterior a cualquier cola de la red. De la misma manera, desde cualquier cola un cliente puede abandonar el sistema.
- Este tipo de redes tiene un espacio de estados infinito
- Ejemplo: Llegadas tipo Poisson (tiempo entre llegadas distribuido exponencialmente)
  - 5 clientes por hora a la cola 1
  - 3 clientes por hora a la cola 4
  - 2 clientes por hora a la cola 5

total: 10 clientes por hora a la red



λ_0 = 10 clientes por hora

- El exterior de la red se representa en la cola 0, obteniendo una red de colas cerrada.

Matriz de routing

- La matriz de routing en este ejemplo sería:

$$P = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 3/10 & 1/5 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Ratios de visita

- **Throughput**: Clientes por unidad de tiempo (tasa) que abandonan una determinada cola. Se representa con l_i.
- En estado estacionario, la tasa de clientes que entran a una cola (procedentes de cualquier otra) debe ser igual a la tasa de clientes que la abandona, es decir:

$$\sum_{i=1}^{M} l_i p_{ik} = l_k \Rightarrow \vec{l}^{\,t} P = \vec{l}^{\,t}$$

- Los throughputs "pueden calcularse" resolviendo un sistema de M ecuaciones lineales. Como P es una matriz estocástica, rg(P-Id)=M-1, luego falta una ecuación.
  - **Redes abiertas**: Conocemos el throughput de la estación 0 (la exterior), luego se pueden calcular los throughputs reales de todas las estaciones.
  - **Redes cerradas**: No conocemos el throughput de ninguna estación, luego sólo podemos calcular las relaciones entre los throughputs de las estaciones.
- **Ratios de visita**: Throughputs relativos normalizados con respecto a una cola (por ejemplo la primera). Se denotan con v.

$$\left.\begin{array}{c} v_1 = 1 \\ \vec{v}^{\,t} P = \vec{v}^{\,t} \end{array}\right\}$$

Ratios de visita

- Ejemplo redes abiertas (el anterior):

$$\left.\begin{aligned} \tfrac{1}{2}\, l_0 &= l_1 \\ \tfrac{1}{2}\, l_1 + \tfrac{1}{2}\, l_3 &= l_2 \\ \tfrac{1}{2}\, l_1 + \tfrac{1}{2}\, l_4 &= l_3 \\ \tfrac{3}{10}\, l_0 + \tfrac{1}{2}\, l_4 &= l_4 \\ \tfrac{1}{5}\, l_0 + \tfrac{1}{2}\, l_3 &= l_5 \end{aligned}\right\} \qquad \begin{aligned} & l_0 = 10 \text{ clientes por hora} \\ & \vec{l}^{\,t} = (5, \tfrac{21}{4}, \tfrac{11}{2}, 6, \tfrac{19}{4}) \end{aligned}$$

- Ejemplo redes cerradas (el anterior): Sólo podemos calcular los ratios de visita.

$$\left.\begin{aligned} \tfrac{1}{3} v_2 &= v_1 \\ \tfrac{1}{2} v_1 + \tfrac{1}{2} v_3 &= v_2 \\ \tfrac{1}{2} v_1 + \tfrac{1}{2} v_3 + \tfrac{1}{2} v_4 &= v_3 \\ \tfrac{1}{3} v_2 + v_5 &= v_4 \\ \tfrac{1}{3} v_2 + \tfrac{1}{2} v_4 &= v_5 \\ v_1 &= 1 \end{aligned}\right\} \Rightarrow \vec{v}^{\,t} = (1, 3, 5, 4, 3)$$

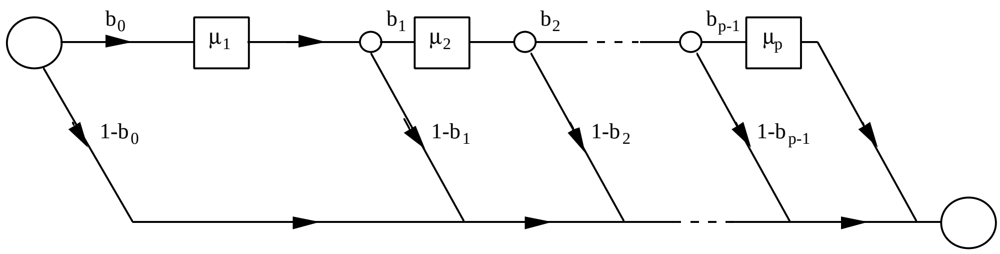
Utilizaciones relativas

- **Utilizaciones relativas** (o demandas de servicio): Combinan la información del routing con los tiempos de servicio de las colas. Se denota con *u*.
  - En redes abiertas se pueden calcular las utilizaciones reales (igual que los ratios de visita)
- Las colas con utilizaciones relativas más altas son las que proporcionalmente más están trabajando. Si hay una cola con una utilización relativa muy superior al resto es un cuello de botella.
- Utilización=tiempo medio de servicio*throughput relativo
- Formas de calcular las utilizaciones relativas:
  - A partir de los ratios de visita
  - Resolviendo el sistema (μ_i=1/tiempo medio servicio)

$$(\vec{m}\,\vec{u})^t P = (\vec{m}\,\vec{u})^t$$

Redes de colas forma producto

- Son redes de colas que tienen como solución en estado estacionario el producto (adecuadamente ponderado) de las soluciones en estado estacionario de las colas que la forman.
- Redes de colas con forma producto: Todas las que se formen con colas de alguno de los siguientes tipos:
  - FCFS con tiempos de servicio exponenciales
  - LCFS preemptive-resume con tiempos de servicio tipo Cox
  - PS con tiempos de servicio tipo Cox
  - Infinitos servidores con tiempos de servicio tipo Cox



Distribución de Cox

Redes de colas forma producto

- Solución en estado estacionario (forma producto)
  - Colas con un único servidor

$$p_{n_1,n_2,\ldots,n_M} = \frac{r_1^{n_1} r_2^{n_2} \ldots r_M^{n_M}}{C}$$

Donde r_i es la utilización de la cola i y C es una constante de normalización para que P sea una distribución de probabilidad.

  - Colas con varios servidores

$$p_{n_1,n_2,\ldots,n_M} = \frac{r_1^{n_1} r_2^{n_2} \ldots r_M^{n_M}}{f_1(n_1) f_2(n_2) \cdots f_n(n_M) C}$$

Donde:

$$f_i(n_i) = \begin{cases} \dfrac{n_i!}{c_i^{n_i}} & si \quad n_i < c_i \\ \dfrac{c_i!}{c_i^{c_i}} & si \quad n_i \geq c_i \end{cases}$$

c_i = número de servidores

Redes abiertas forma producto

Jackson, 1963

- La constante de normalización es, simplemente, el inverso de la probabilidad de que la red esté vacía

$$C = \frac{1}{p_{0,0,\ldots,0}} = \frac{1}{(1-r_1)(1-r_2)\cdots(1-r_M)}$$

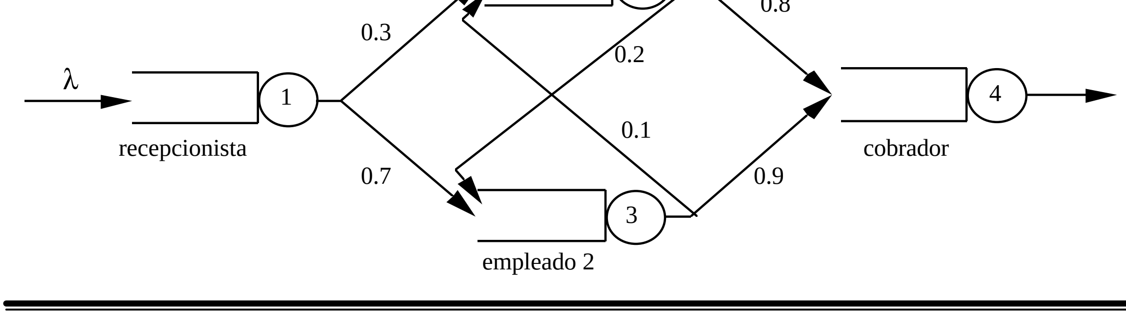
- Luego la solución en estado estacionario queda:

$$p_{n_1,n_2,\ldots,n_M} = (1-r_1)r_1^{n_1}(1-r_2)r_2^{n_2}\cdots(1-r_M)r_M^{n_M}$$

- Ejemplo:(Molloy, 1989)

  Proceso de registro de vehículos a motor en un ayuntamiento.

  - Recepción: Envía al usuario al empleado que corresponda (según la tarea). 20 segundos
  - Empleados: Realizan las tareas (10’ y 5’ resp.)
  - Cobrador: Cobra al usuario (1’)



Redes abiertas forma producto

- Throughputs:

$$\left.\begin{aligned} l_1 &= l \\ l_2 &= 0.3\,l + 0.1\,l_3 \\ l_3 &= 0.7\,l + 0.2\,l_2 \\ l_4 &= 0.8\,l_2 + 0.9\,l_3 \end{aligned}\right\} \Rightarrow \begin{aligned} l_1 &= l \\ l_2 &= 0.38\,l \\ l_3 &= 0.78\,l \\ l_4 &= l \end{aligned}$$

- Utilizaciones:

$$\begin{aligned} r_1 &= s_1\,l_1 = && 20\,l \\ r_2 &= s_2\,l_2 = 600 * 0.38\,l = && 228\,l \\ r_3 &= s_3\,l_3 = 300 * 0.78\,l = && 234\,l \\ r_4 &= s_4\,l_4 = && 60\,l \end{aligned}$$

- Ergodicidad si y solo si todas las utilizaciones son menores que 1, es decir:

  $\lambda<0.00427$ clientes por segundo (15.38 por hora)

- Solución en estado estacionario:

$$p_{n_1,n_2,n_3,n_4} = (1-20\,l)(20\,l)^{n_1} \cdot (1-228\,l)(228\,l)^{n_2} \cdot$$

$$\cdot (1-234\,l)(234\,l)^{n_3} \cdot (1-60\,l)(60\,l)^{n_4}$$

Redes abiertas forma producto

Otros índices de prestaciones

- Cotas del número medio de clientes en cada cola (usando la cota de λ<0.00427)

$$E[n_1] = \frac{r_1}{1 - r_1} = \frac{20\,l}{1 - 20\,l} \leq 0.093$$

recepcionista ocupado máximo un 8,5% del tiempo (r_1)

$$E[n_2] = \frac{r_2}{1 - r_2} = \frac{228\,l}{1 - 228\,l} \leq 36.82$$

$$E[n_3] = \frac{r_3}{1 - r_3} = \frac{234\,l}{1 - 234\,l} \leq \infty$$

cuello de botella

$$E[n_4] = \frac{r_4}{1 - r_4} = \frac{60\,l}{1 - 60\,l} \leq 0.34$$

cobrador ocupado máximo un 25,6% del tiempo (r_4)

- Tiempo medio de permanencia de un cliente en el sistema (Little)

$$T = \frac{N}{l} = \frac{E[n_1] + E[n_2] + E[n_3] + E[n_4]}{l} =$$

$$= \frac{20}{1 - 20\,l} + \frac{228}{1 - 228\,l} + \frac{234}{1 - 234\,l} + \frac{60}{1 - 60\,l}$$

Redes cerradas forma producto

Gordon-Newell, 1967

- Forma producto en términos de las utilizaciones relativas:

$$p_{n_1,n_2,\ldots,n_M} = \frac{1}{G(N)} \prod_{i=1}^{M} \frac{u_i^{n_i}}{b_i(n_i)}$$

con $b_i(n_i) = \begin{cases} n_i! & si \quad n_i \le c_i \\ c_i!\, c_i^{n_i - c_i} & si \quad n_i \ge c_i \end{cases}$ c_i = número de servidores

y $G(N) = \sum_{\forall n, \sum_{i=1}^{M} n_i = N} \prod_{i=1}^{M} \frac{u_i^{n_i}}{b_i(n_i)}$

- Una vez se calcula la constante de normalización, podemos obtener los throughputs, utilizaciones reales de todas las colas y las probabilidades en estado estacionario
- Otros índices de prestaciones

$$P\{n_i \ge n\} = u_i^n \frac{G(N-n)}{G(N)}$$

En particular las utilizaciones reales

$$r_i = P\{n_i \ge 1\} = u_i \frac{G(N-1)}{G(N)}$$

Redes cerradas forma producto

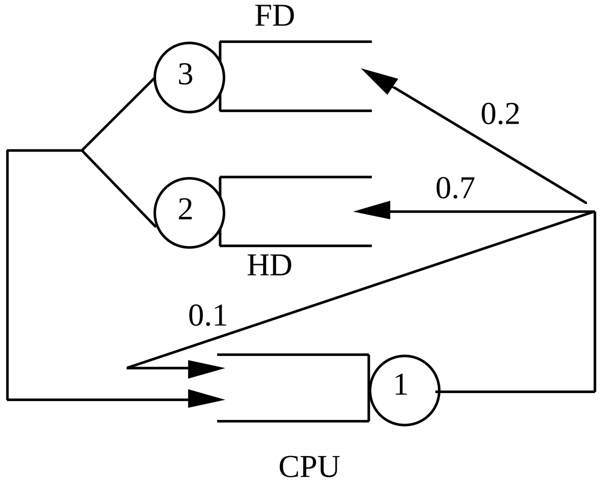
$$E[n] = \sum_{n=1}^{\infty} nP\{n\} = \sum_{n=1}^{\infty}\sum_{k=n}^{\infty} P\{k\} = \sum_{\mathrm{n}=1}^{\infty} P\{k \geq n\}$$

luego

$$E[n_i] = \sum_{n=1}^{N} u_i^n \frac{G(N-n)}{G(N)}$$

- El mayor problema es calcular la constante de normalización
- Ejemplo: (Molloy, 1989)

  Pequeño sistema computador con:
  - Disco blando: Tiempo medio de acceso 280 ms. FCFS
  - Disco duro: Tiempo medio de acceso 40 ms. FCFS
  - CPU: Tiempo medio cómputo 28 ms. RR (1 ms)



Redes cerradas forma producto

- Utilizaciones relativas:

$$v_2 = 0.7 v_1 \Rightarrow \frac{1}{40} u_2 = 0.7 \frac{1}{28} u_1 \Rightarrow u_2 = u_1$$

$$v_3 = 0.2 v_1 \Rightarrow \frac{1}{280} u_3 = 0.2 \frac{1}{28} u_1 \Rightarrow u_3 = 2 u_1$$

- Constante de normalización para diferente número de clientes (aplicando la fórmula):

G(0)	G(1)	G(2)	G(3)	G(4)	G(5)	G(6)
1	4	11	26	57	120	247

- Solución en estado estacionario para 6 trabajos:

$$p_{n_1, n_2, 6-n_1-n_2} = \frac{1}{247} \cdot 1^{u_1} \cdot 1^{u_2} . \cdot 2^{6-u_1-u_2} = \frac{2^{6-u_1-u_2}}{247}$$

- Utilizaciones reales:

$$r_i = u_i \frac{G(N-1)}{G(N)} \Rightarrow r_1 = r_2 = 1 \cdot \frac{120}{247} = 0.4858$$

$$r_3 = 2 \cdot \frac{120}{247} = 0.9717$$

Redes cerradas
forma producto

- Throughputs reales (en trabajos por segundo):

$$l_1 = r_1 \cdot m_1 = \frac{0.4858}{0.028} = 17.35$$

$$l_2 = r_2 \cdot m_2 = \frac{0.4858}{0.040} = 12.145$$

$$l_3 = r_3 \cdot m_3 = \frac{0.9717}{0.280} = 3.47$$

- Número medio de clientes en cada cola:

$$E[n_1] = E[n_2] = \frac{1+4+11+26+57+120}{247} = 0.8866$$

$$E[n_3] = \frac{2^6 \cdot 1 + 2^5 \cdot 4 + 2^4 \cdot 11 + 2^3 \cdot 26 + 2^2 \cdot 57 + 2 \cdot 120}{247} = 4.2267$$

- Esperas medias (cola+servicio) en cada cola (Little):

$$T_1 = \frac{E[n_1]}{l_1} = 0.0511 \quad \text{segundos}$$

$$T_2 = \frac{E[n_2]}{l_2} = 0.073 \quad \text{segundos}$$

$$T_3 = \frac{E[n_3]}{l_3} = 1.218 \quad \text{segundos}$$

4.4. Redes de Petri

- Introducción
- Propiedades básicas
- Ergodicidad
- Análisis exacto
  - Redes de Petri estocásticas (SPN)
  - Redes de Petri generalizadas (GSPN)

Introducción

- En sistemas informáticos reales hay relaciones de cooperación y competencia entre sus componentes
- Para ello se necesitan primitivas de sincronización en los formalismos de modelado que utilicemos
- Las redes de colas con forma producto no permiten modelar sincronizaciones
- Ejemplos:
  - Se necesita un drive de disco y un canal antes de comenzar una transferencia de E/S
  - Se necesita una partición de memoria antes de poder ejecutar un programa
  - Se necesita tener libre el medio de transmisión antes de poder enviar datos por una red
- Soluciones posibles:
  - Añadir primitivas de sincronización a las redes de colas (destruyen la forma producto)
  - Utilizar otro formalismo que incluya de forma natural estas primitivas (redes de Petri)

Introducción

- Redes de colas =

  = estructura (colas, servidores y routing)

  + distribución de clientes (estado del modelo)

  + interpretación estocástica

  (routing aleatorio, tiempos de servicio, número de servidores , etc.)

  ⇒ comportamiento dinámico: movimiento de clientes

- Redes de Petri (RdP) =

  = estructura (lugares, transiciones y arcos)

  + distribución de marcas (tokens):estado del modelo

  + interpretación estocástica

  (routing aleatorio, tiempos de servicio, número de servidores , etc.)

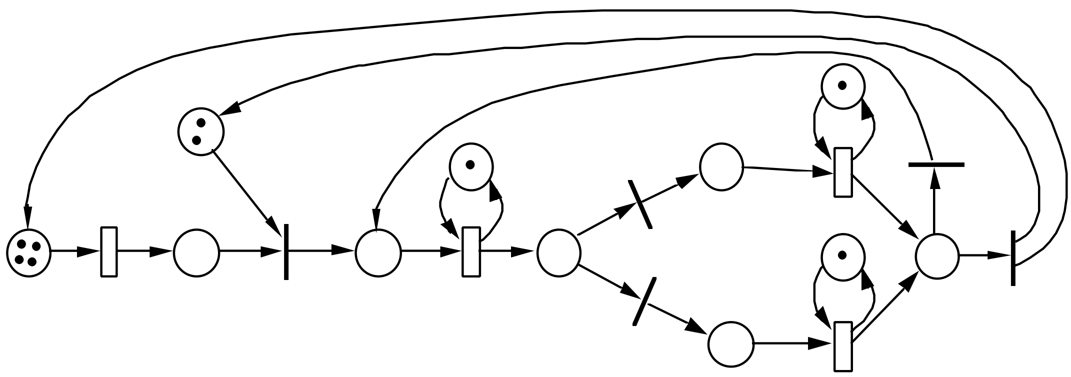
  ⇒ comportamiento dinámico: movimiento de clientes

  (regla de disparo de transiciones)

- Redes de Petri = redes de colas + sincronizaciones

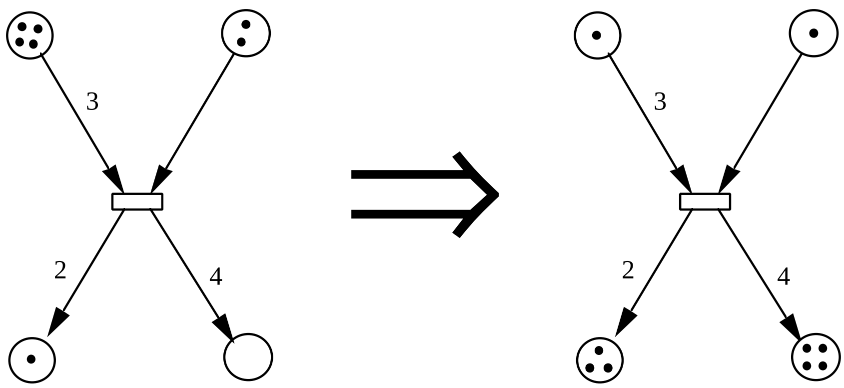
Propiedades básicas

- Estructura de una red de Petri
  - **Lugares**: Salas de espera. Pueden modelar recursos, clientes, etc. Se representan con circunferencias.
  - **Transiciones**: Actividades. Se representan con cajas rectangulares.
  - **Arcos**: Unen lugares con transiciones o transiciones con lugares. Se representan con flechas. Marcan el flujo de las marcas. Pueden tener un peso (que indica la cantidad de marcas que se mueven).
  - **Marcas**: Se representan con puntos dentro de los lugares. El conjunto de marcas asociadas en un instante dado a cada lugar constituye el **marcado** de la RdP (estado).



Propiedades básicas

- Evolución de una RdP (regla de disparo)
  - **Lugar de entrada** de un transición: Es un lugar tal que hay un arco que une el lugar con la transición.
  - **Lugar de salida** de una transición: Es un lugar tal que hay un arco que une la transición con el lugar.
  - **Transición sensibilizada**: Es una transición que tiene en cada lugar de entrada tantas marcas como indica el peso del arco que le une a la transición. En cada marcado de una RdP puede haber varias transiciones sensibilizadas lo que permite modelar concurrencia.
  - Una transición sensibilizada puede dispararse. Si se dispara se producen los siguientes cambios:
    - » De cada lugar de entrada se eliminan tantas marcas como indique el arco que lo une a la transición.
    - » En cada lugar de salida se ponen tantas marcas como indique el peso del arco que lo une a la transición disparada.



Propiedades básicas

- Espacio de estados de una RdP

  Es el conjunto de todos los estados alcanzables por la RdP. En general el tamaño del espacio de estados de una red es exponencial en el tamaño de la red

- Interpretación temporal

  En una RdP se puede asociar tiempos a sus actividades de diferentes maneras.

  – Tiempo asociado a transiciones o a lugares:

    » Transiciones: Tiempo que pasa desde la sensibilización de una transición hasta su disparo. Es la forma habitual y la que nosotros emplearemos.

    » Lugares: Tiempo que pasa desde que una marca llega a un lugar hasta que esta marca puede sensibilizar transiciones.

  – Disparo de transiciones:

    » Atómico: Sucede en tiempo cero. Las marcas que sensibilizan la transición permanecen en los lugares de entrada durante la actividad de la transición. Cuando ésta se dispara, se mueven las marcas de los lugares de entrada a los de salida en tiempo 0. Es el que emplearemos.

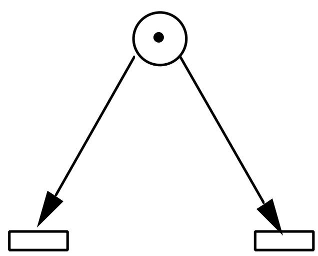
    » En tres fases: Cuando una transición se sensibiliza, retira las marcas de los lugares de entrada, los guarda mientras dura su actividad y después las pone en los lugares de salida.

Propiedades básicas

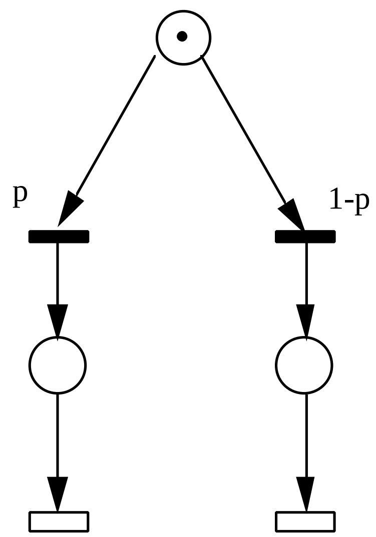
- Resolución de conflictos:

  Cuando dos o más transiciones están sensibilizadas y el disparo de una de ellas desensibiliza a las otras, se dice que están en **conflicto**. Los conflictos pueden resolverse de diferentes formas:

  + Race policy (carrera): Se dispara primero la que acabe antes su actividad.
  + Distribución discreta de probabilidad: Se asocia una probabilidad a cada transición y se efectua un sorteo. La que lo gane se dispara. Estas probabilidades se suelen modelar con **transiciones** especiales (**inmediatas**) que se disparan en tiempo 0.



race policy



distribución de probabilidad

Propiedades básicas

- Tiempos de servicio:

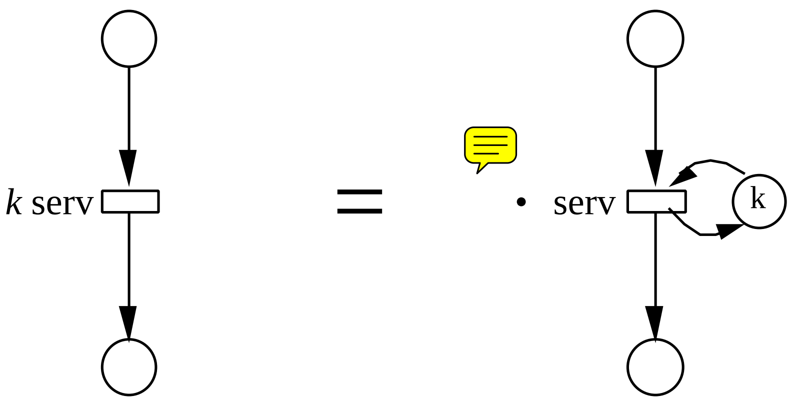
  Los tiempos asociados a las actividades de las transiciones pueden ser:

  + Exponenciales: Variable aleatoria exponencial. Es la que utilizaremos.
  + Determinista: Un tiempo fijo
  + General: Cualquier distribución de probabilidad

- Número de servidores:

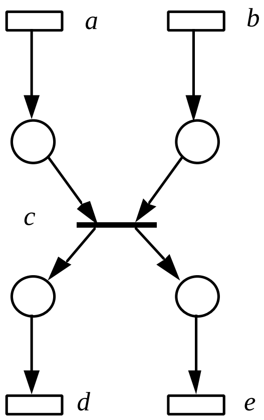
  + Un servidor (single server)
  + Infinitos servidores
  + k servidores

  Cambiando la estructura de la red es posible cambiar el funcionamiento de una transición en cuanto al número de servidores.

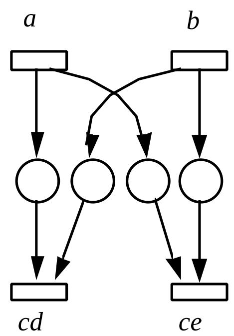


Propiedades básicas

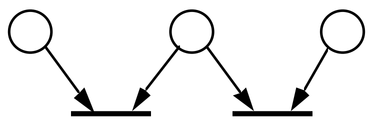
- Algunos modelos simples:



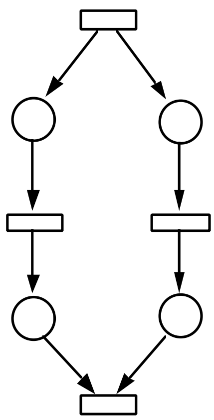
sincronización
por rendez-vous



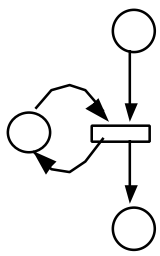
sincronización
por buffers



recurso
compartido



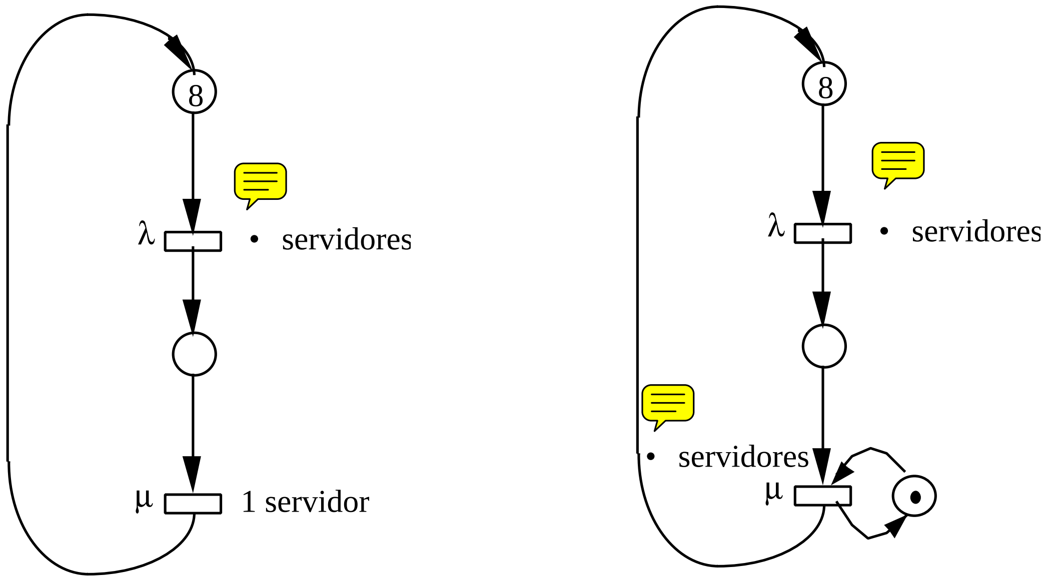
fork-joing



recurso no
consumible

Ejemplos

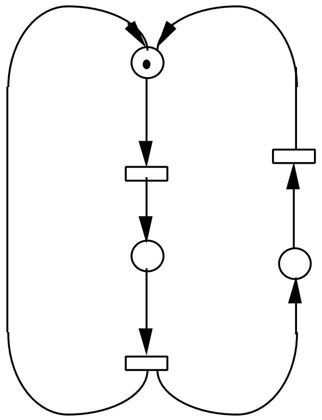
- Multiprocesador de un bus (el de simulación)



Dos modelos equivalentes del mismo sistema.

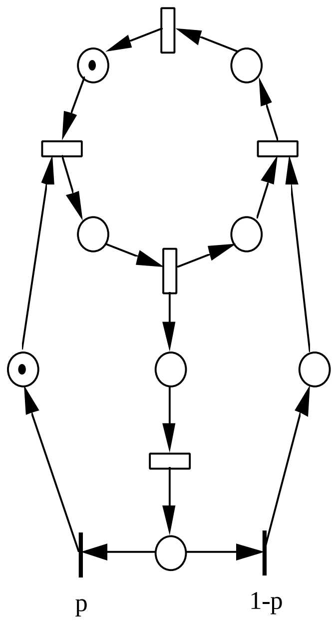
Casos patológicos

- Debido a la complejidad de las sincronizaciones, se puede llegar a casos patológicos



Red no acotada

Todos los lugares pueden tener cualquier número de marcas



Bloqueo total

Independientemente del valor de *p*, la red llegará a un marcado en el que no hay sensibilizada ninguna transición

- Es necesario realizar análisis cualitativo previamente para evitar estas situaciones

Análisis cualitativo

- Existen numerosas técnicas para chequear ciertas propiedades cualitativas de las redes (no las veremos en detalle)
- Propiedades que pueden definirse formalmente y chequearse:
  - Cotas de marcado de lugares
  - Lugares en exclusión mutua, ...
  - Ausencia de bloqueos
  - Vivacidad: Todas las transiciones pueden dispararse a partir de cualquier marcado alcanzable.
  - Existencia de home states: Un estado es home state si es alcanzable desde cualquier marcado alcanzable de la red.
- Tipos de técnicas para análisis cualitativo
  - Enumerativas: Basadas en el estudio del espacio de estados del modelo
  - Reducción/transformación: Consisten en aplicar reglas locales para reducir o transformar la red preservando las propiedades requeridas, simplificando el estudio.
  - Técnicas estructurales: Consisten en chequear una determinada propiedad basándose únicamente en la estructura de la red. Se usa teoría de grafos, álgebra lineal, geometría convexa, programación lineal, etc.

Ergodicidad

- Problema del estado estacionario:
  - Cadenas de Markov: Las CM irreducibles recurrentes no nulas tenían distribución en estado estacionario.
  - Redes de colas: Las RdC con tiempos de servicio exponenciales tenían distribución en estado estacionario si no estaban saturadas (condiciones de ergodicidad).
- En redes de Petri:

  A partir del espacio de estados puede construirse una CM con tantos estados como los que tenga la red. La existencia de distribución en estado estacionario para la red depende de las ergodicidad de esta CM.

  Condiciones suficientes para la existencia de solución en estado estacionario:
  - Limitación: Si la red es no acotada, la CM subyacente tiene • estados y, salvo en casos especiales, no se sabe resolver.
  - Existencia de home state: Asegura que la CM es irreducible (salvo quizás algunos estados transitorios). En cualquier caso, asegura la unicidad de la solución en estado estacionario.

Análisis exacto (SPN)

- Estudiaremos la forma de obtener de forma exacta la distribución en estado estacionario para:
  - Redes de Petri estocásticas (SPN)
  - Redes de Petri generalizadas (GSPN)
- Definición formal: una red de Petri estocástica es una 6-tupla (P, T, I, O, M_0, Λ), donde:
  - $P = \{p_1,\ldots,p_n\}$ es el conjunto de lugares
  - $T = \{t_1,\ldots,t_m\}$ es el conjunto de transiciones
  - I 1 $P \times T$ es el conjunto de arcos de entrada
  - O 1 $T \times P$ es el conjunto de arcos de salida
  - $M_0 = (M_0(p_1),\ldots, M_0(p_n))$ es el marcado inicial
  - $\Lambda = (\lambda_1,\ldots,\lambda_m)$ es el conjunto de tasas de disparo de las transiciones
- Asumimos tiempo asociado a transiciones distribuido exponencialmente, single server, disparo atómico y resolución de conflictos por race policy.
- El tiempo de permanencia en un estado está distribuido exponencialmente con tasa

  $\sum_{t \in E} l_t$ E = conjunto de transiciones sensibilizadas en el estado

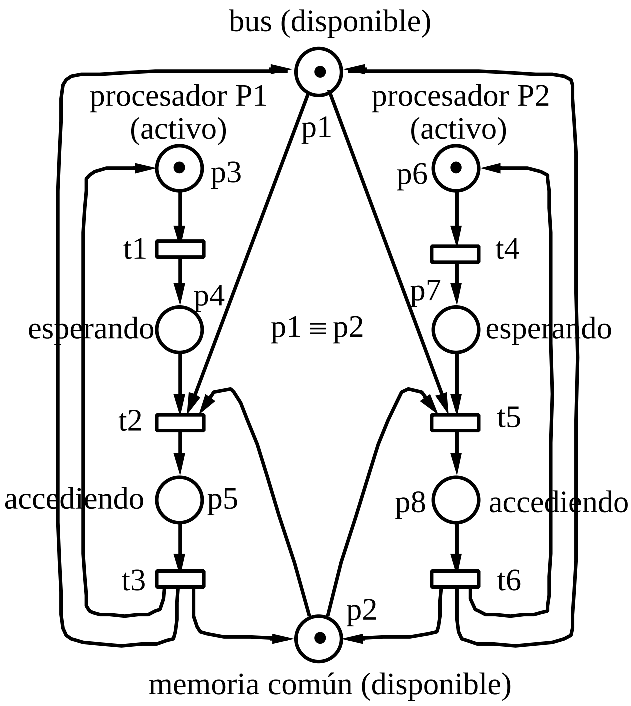
Análisis exacto (SPN)

- El proceso estocástico que marca la evolución de una SPN es una CMTC:
  - Los estados de la CM son los estados alcanzables de la red
  - La tasa de paso del estado *i* (marcado M_i) al *j* (M_j) es:

$$q_{ij} = \sum_{k \in H_{ij}} l_k$$

  Donde H_{ij} es el conjunto de transiciones sensibilizadas en el marcado M_i cuyo disparo produce el estado M_j

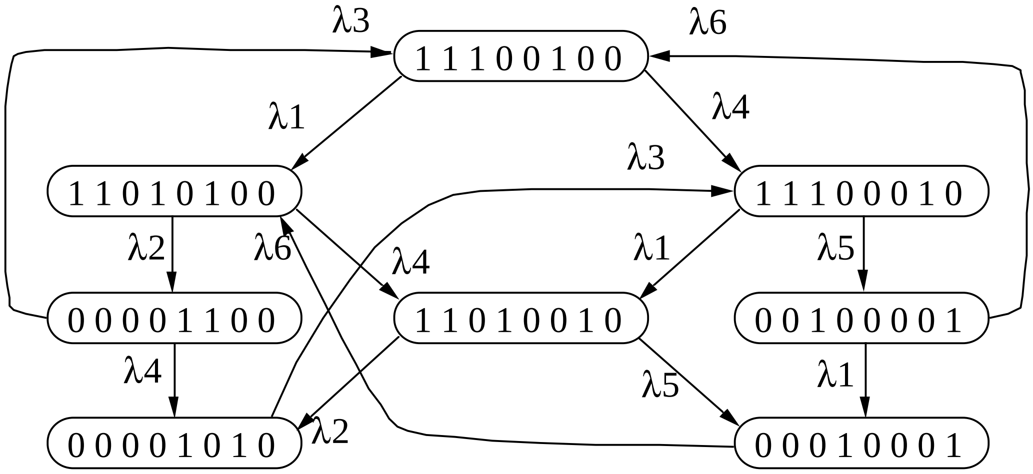
- Ejemplo:



Modelo SPN de un biprocesador de memoria compartida

Análisis exacto (SPN)

- Diagrama de tasas de transición entre estados de la CM subyacente



- Si en vez de single server, tenemos • servidores, sólo hay que cambiar las tasas de transición entre ciertos estados.
- La distribución en estado estacionario se obtiene resolviendo la CM (solución única si la red es limitada y tiene home state).

$$\left.\begin{aligned} \Pi^t Q &= 0 \\ \Pi^t 1 &= 1 \end{aligned}\right\}$$

- A partir de la solución en estado estacionario, pueden calcularse otros índices de prestaciones.

Análisis exacto (SPN)

- Índices de prestaciones:
  - Throughput de una transición *t*

$$c_t = \sum_{M_i \in A_t} l_t p_i$$

Donde A_t es el conjunto de marcados de la red en los que *t* está sensibilizada

  - Marcado medio de un lugar *p*

$$E[M(p)] = \sum_{i=1}^{N} M_i[p] p_i$$

Donde N es el número de estados de la red

  - Probabilidad de una condición C cualquiera sobre la red

$$P\{C\} = \sum_{M_i \in A} p_i$$

Donde A es el conjunto de estados alcanzables de la red que satisfacen la condición C.

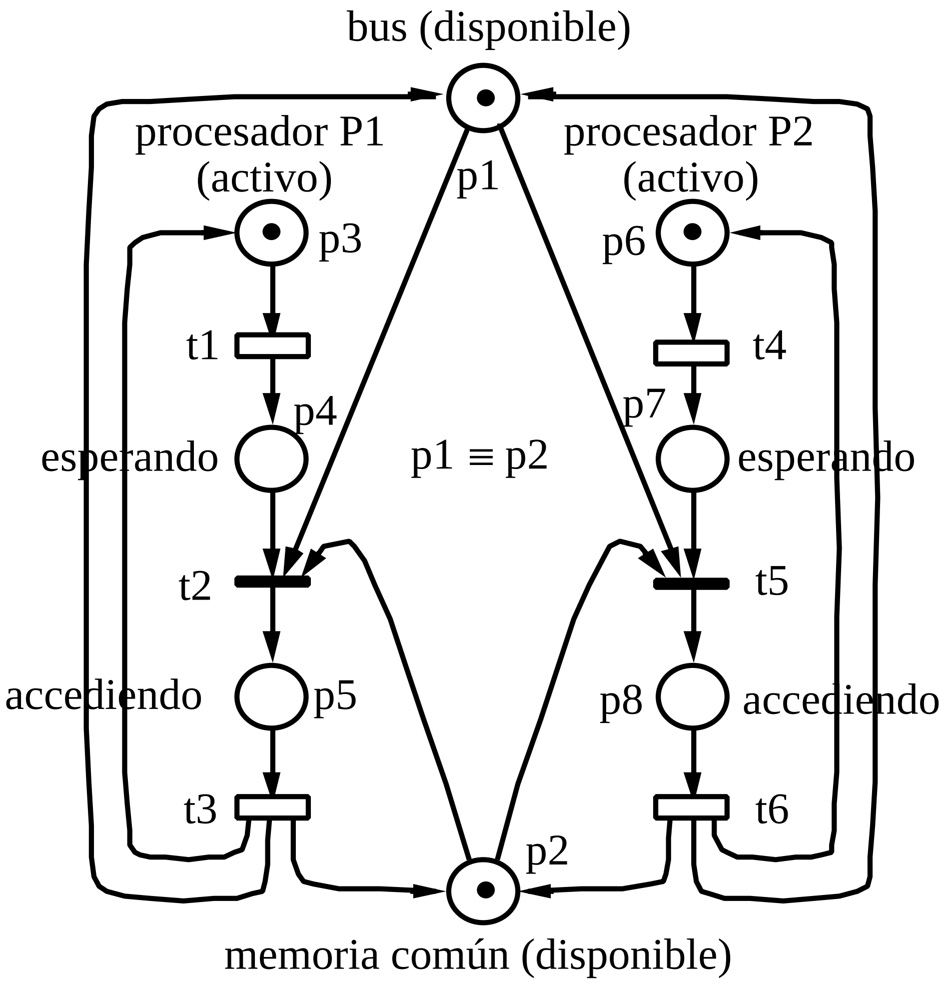
Análisis exacto (GSPN)

- En SPN todas las transiciones son del mismo tipo (temporizadas)
- Cuando se construyen modelos de sistemas complejos es conveniente separar el comportamiento lógico de los aspectos temporales
- Las GSPN incluyen un tipo especial de transiciones, llamadas inmediatas, que se disparan en tiempo 0 y con mayor prioridad que el resto de las transiciones temporizadas. Estas transiciones inmediatas permiten el modelado de los aspectos lógicos.

  Resolución de conflictos entre transiciones inmediatas:
  - Asignando distintas prioridades a las transiciones
  - Por medio de una distribución discreta de probabilidad (probabilidades de routing, como en redes de colas).
- Los estados de una GSPN pueden dividirse en dos tipos:
  - Vanishing: Aquellos que sensibilizan transiciones inmediatas. No consumen tiempo, luego pueden ser eliminados al generar la CM subyacente.
  - Tangibles: Los que sensibilizan únicamente transiciones temporizadas. Son los únicos que aparecen en la CM subyacente.
- Por ello los modelos basados en GSPN permiten soluciones más eficientes

Análisis exacto (GSPN)

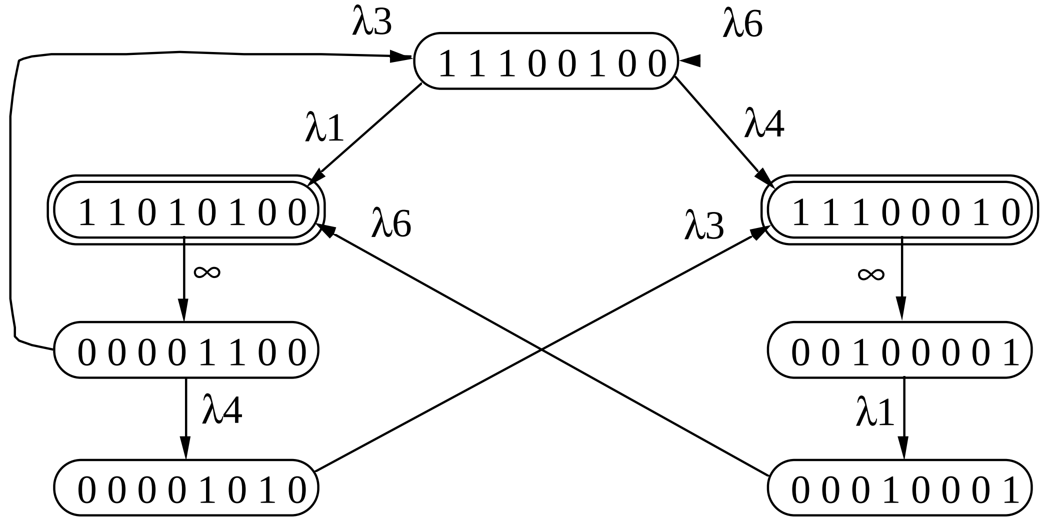
- Ejemplo anterior (biprocesador de memoria compartida) modelado con GSPN.



- Ahora modelamos los aspectos lógicos con transiciones inmediatas

Análisis exacto (GSPN)

- Ahora el espacio de estados es más reducido, debido a la mayor prioridad de las transiciones inmediatas (se eliminan algunas transiciones entre estados).



- Se pueden eliminar de la CM los estados vanishing. Así la CM queda (sólo los estados tangibles):