Teoría de colas

Modelado y Análisis de Redes de Telecomunicaciones

IIE - Facultad de Ingeniería

Contenido

Proceso de conteo

- Un proceso estocástico $\{N_t : t \geq 0\}$ es un proceso de conteo si:
  - N_t es natural para todo t.
  - $N_0 = 0$.
  - Si $s < t$, entonces $N_s \leq N_t$.
- N_t cuenta el número total de eventos que ocurren en $[0, t]$ y $N_t - N_s$ cuenta el número de eventos en $(s, t]$.

Incrementos independientes y estacionarios

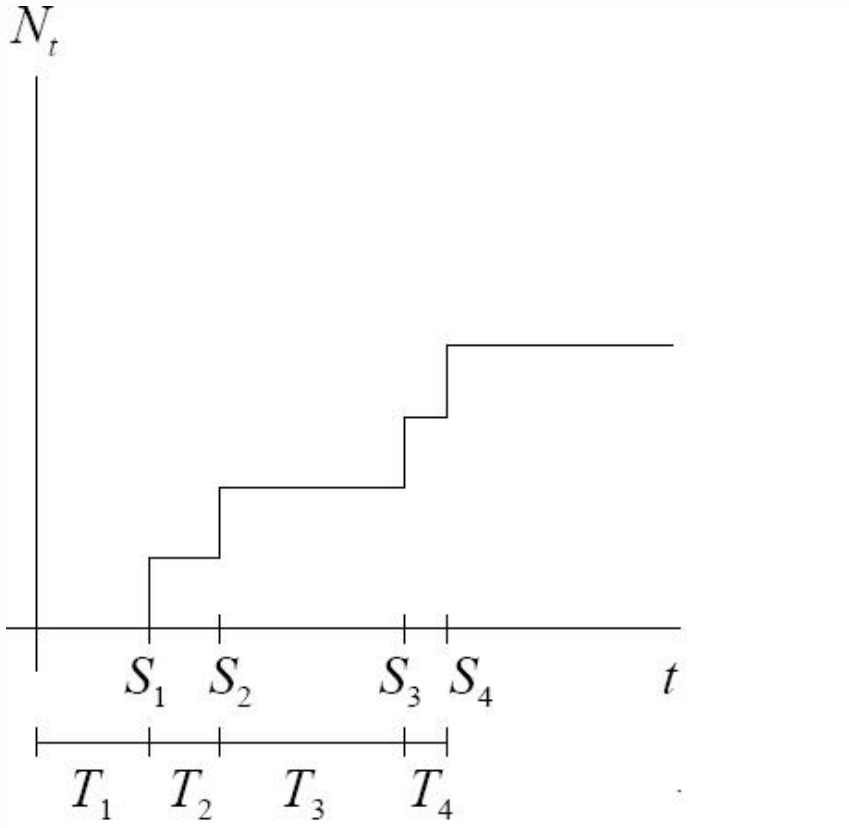
- Un proceso de conteo N tiene incrementos independientes si son independientes

$$N_t - N_s,\ N_v - N_u \ \forall\ 0 \leq s < t \leq u < v$$

  es decir que el número de eventos que ocurre en intervalos disjuntos son v.a. independientes.
- Un proceso de conteo N es estacionario si, para cualquier $s < t$, la distribución de $N_t - N_s$ depende sólo de $t - s$ (longitud del intervalo).

Proceso de Poisson: definición

- $T_1, T_2, \ldots$ i.i.d. con distribución exponencial de parámetro λ
- Se define la v.a. $S_n = \sum_{i=1}^{n} T_i$
- $N = \{N_t\}_{t \in \mathcal{R}}$ tal que $N_t = \#\{n : S_n \leq t\}$ es un proceso de Poisson de parámetro λ



Propiedades del proceso de Poisson

- Es un proceso de conteo.
- Es estacionario y tiene incrementos independientes.
- $N_{t+s} - N_s$ tiene la misma distribución que N_t y es Poisson de parámetro λ_t

$$P(N_{t+s} - N_s = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Definiciones equivalentes (1)

Un proceso N es un proceso de Poisson de parámetro λ si:

- es un proceso de conteo
- tiene incrementos independientes
- N_t tiene distribución de Poisson de parámetro λt.

Definiciones equivalentes (2)

Un proceso N es un proceso de Poisson de parámetro λ si:

- es estacionario
- es un proceso de conteo
- tiene incrementos independientes
- se cumple que:
  - $P(N_h = 0) = 1 - \lambda h + o(h)$
  - $P(N_h = 1) = \lambda h + o(h)$
  - $P(N_h \geq 2) = o(h)$

Suma de procesos de Poisson

Sean $N^1, N^2, \ldots N^k$ procesos de Poisson independientes de parámetros $\lambda_1, \lambda_2, \ldots, \lambda_k$. Entonces el proceso N definido por

$$N_t = \sum_{i=1}^{k} N_t^i$$

es un proceso de Poisson de parámetro

$$\lambda = \sum_{i=1}^{k} \lambda_i$$

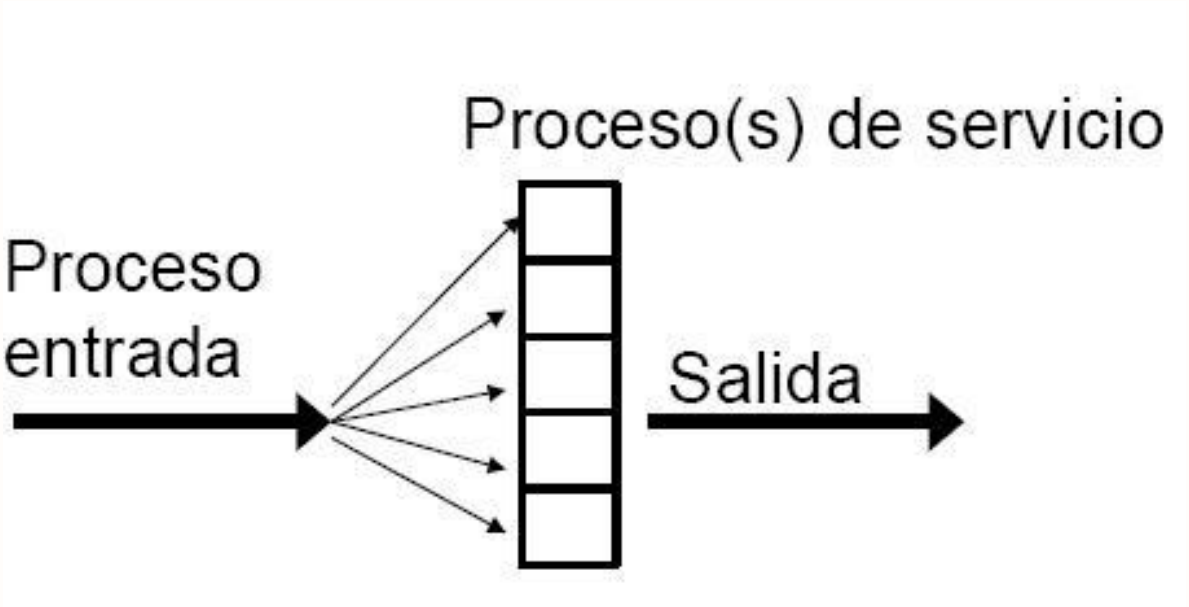
División de un proceso de Poisson

- Sea $\{N_t : t \geq 0\}$ un proceso de Poisson con parámetro λ.
- Supongamos que en cada tiempo que un evento ocurre, éste es clasificado como tipo I con probabilidad p y como tipo II con probabilidad $1 - p$, de manera independiente.
- Sean N_t^1 y N_t^2 respectivamente el número de eventos de tipo I y de tipo II hasta el tiempo t.
- Entonces, N^1 y N^2 son procesos de Poisson independientes con parámetros λp y $\lambda(1 - p)$.

Contenido

Introducción

La teoría de colas se ocupa de estudiar sistemas con la siguientes características:



Nos interesará estudiar la cantidad de clientes en el sistema y/o en la cola, el retardo, tiempo de espera, pérdidas en el sistema,etc.

Notación de Kendall

La primera letra indica el proceso de entrada , la segunda la distribución del tiempo de servicio y la tercera el número de servidores.

- M/M/1 Poisson/Exponencial/1 servidor (M = Poisson, sin memoria)
- M/G/1 Poisson/General/1 servidor
- D/G/n Determinístico/General/n servidores
- E/G/∞ Erlang/General/Infinitos servidores

Se puede agregar tamaño del buffer, disciplina de servicio, etc..

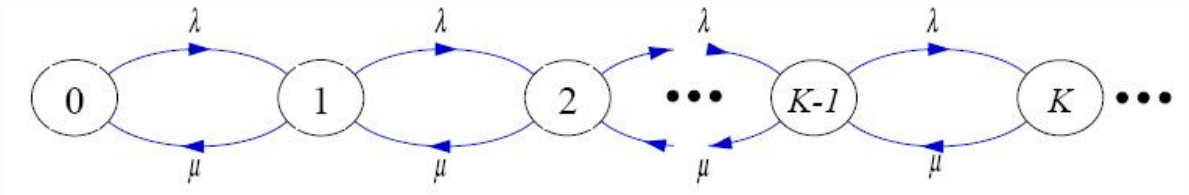
Contenido

Características

- Es el modelo más simple.
- Los clientes llegan de acuerdo a un proceso de Poisson de parámetro λ (tiempo entre llegadas exponencial).
- Son atendidos por un solo servidor donde el tiempo de servicio es exponencial de parámetro μ.
- La disciplina es FIFO (First In First Out).
- Se puede pensar como un enlace con capacidad C constante al que arriban paquetes de tamaño exponencial de parámetro μ/C, separados tiempos exponenciales de parámetro λ.

Cadena asociada

- Vimos que no hay arribos simultáneos. Se prueba que tampoco hay arribos y partidas simultáneas.
- Sea X_t la cantidad de clientes en el sistema en el instante t.
- X_t es una cadena de Markov homogénea en tiempo continuo con espacio de estados discreto (pero infinito).



Generador infinitesimal

- El generador infinitesimal de la cadena es

$$P = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & \cdots \\ \mu & -(\mu+\lambda) & \lambda & 0 & \cdots \\ 0 & \mu & -(\mu+\lambda) & \lambda & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

- Las probabilidades de transición de la cadena incluida son

$$\overline{p_{i,i+1}} = \frac{\lambda}{\lambda+\mu} \quad \overline{p_{i,i-1}} = \frac{\mu}{\lambda+\mu}$$

Ergodicidad

- E es infinito. La cadena es irreducible y recurrente positiva (y por lo tanto ergódica) siempre que $\lambda < \mu$.
- En ese caso se dice que el sistema es estable y $\rho = \frac{\lambda}{\mu}$ se interpreta como la carga del sistema.
- Existe una única distribución límite π_{est} que verifica $\pi\mathbf{Q} = 0$ (i.e. las ecuaciones de balance)

Ecuaciones de balance

Si π_{est} es la distribución invariante, planteando el sistema $\pi_{est} Q = 0$ (ecuaciones de balance) y definiendo $\rho = \frac{\lambda}{\mu}$ se tiene que

$$\pi_{est}(k) = \rho^k \pi_{est}(0)$$

y por normalización

$$\pi_{est}(0) = \frac{1}{\sum_{k=0}^{\infty} \rho^k}$$

Distribución invariante

- Si la utilización $\rho = \frac{\lambda}{\mu} < 1$ la serie converge y

$$\pi_{est}(0) = 1 - \rho$$

- La distribución invariante es

$$\pi_{est}(k) = (1 - \rho)\rho^k$$

- El número de clientes en estado estacionario tiene distribución Geométrica de parámetro ρ.

Número medio de clientes

- Si N_{sist} es el numero de clientes en el sistema en estado estacionario, el número medio de clientes en el sistema es

$$E(N_{sist}) = \sum_{k=0}^{\infty} k\pi_{est}(k) = \sum_{k=0}^{\infty} k(1-\rho)\rho^k$$

$$E(N_{sist}) = \frac{\rho}{1-\rho}$$

- Si N_{cola} es el numero de clientes en la cola en estado estacionario, el número medio de clientes en la cola es

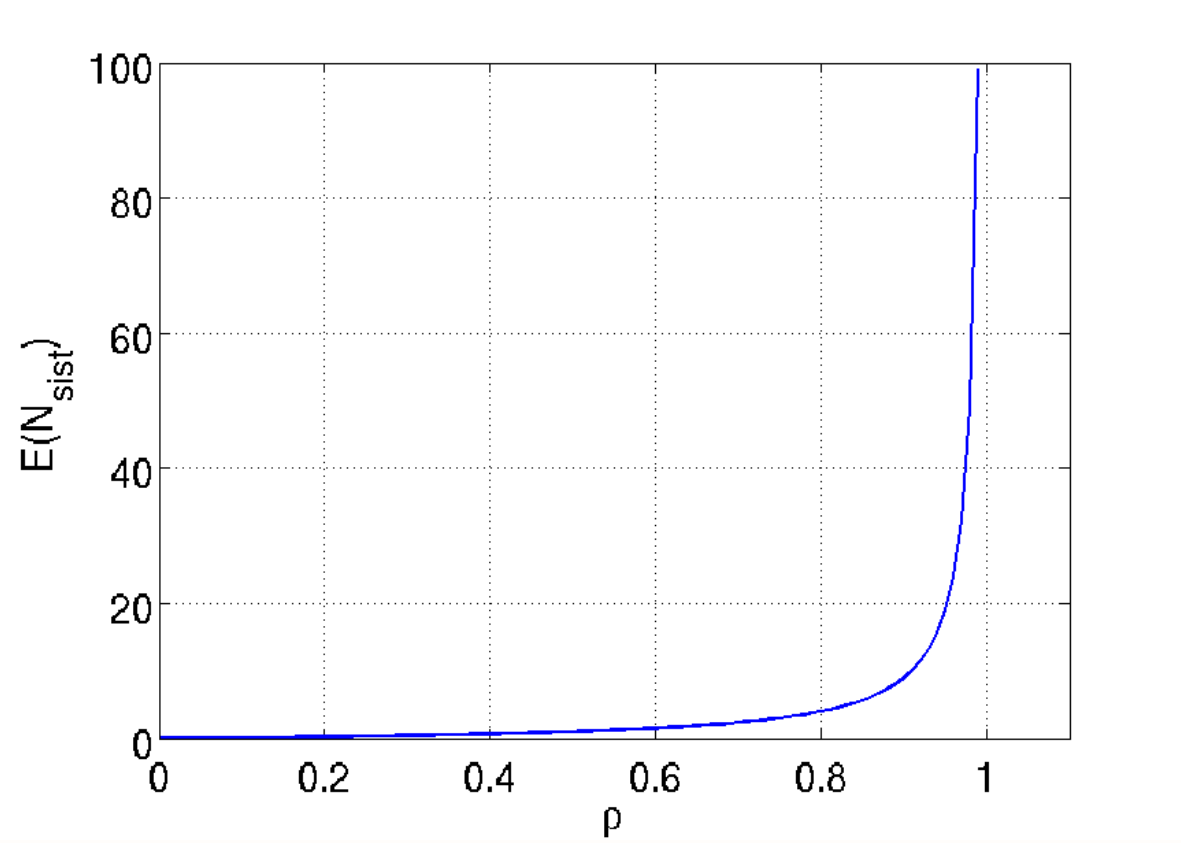
$$E(N_{cola}) = \sum_{k=1}^{\infty} (k-1)\pi_{est}(k) = \sum_{k=1}^{\infty} (k-1)(1-\rho)\rho^k$$

$$E(N_{cola}) = \frac{\rho^2}{1-\rho}$$

Retardo vs utilización

El retardo medio en el sistema depende de la cantidad media de clientes

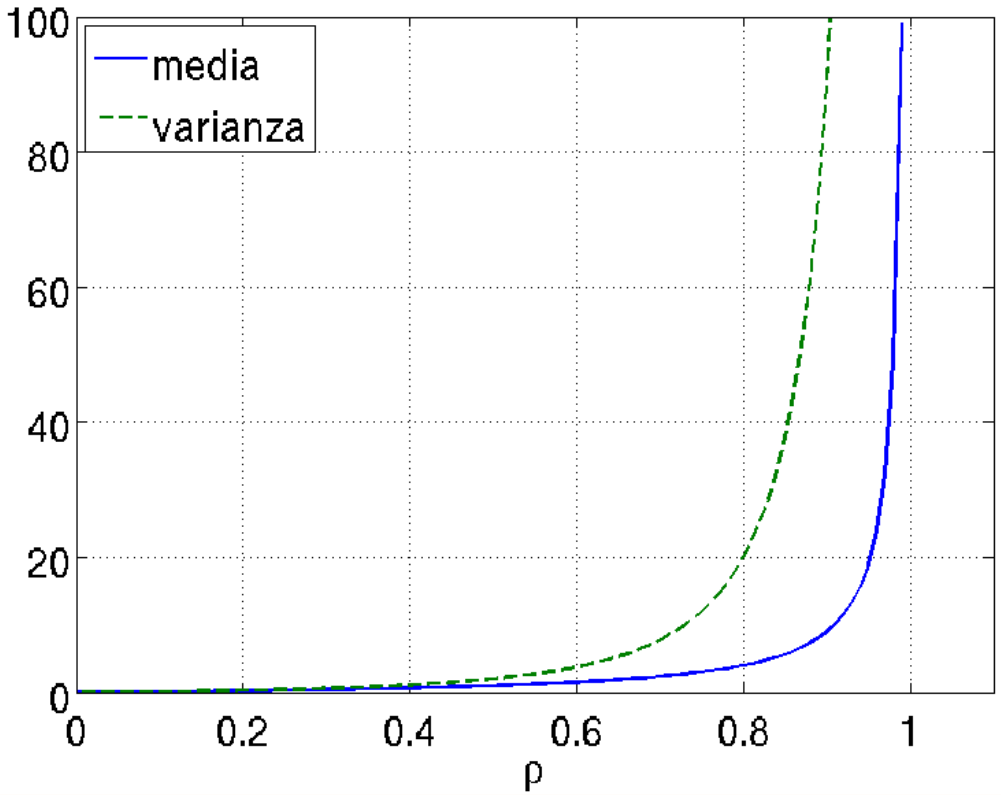
$$E(N_{sist}) = \frac{\rho}{1-\rho}$$



Jitter vs utilización

La varianza del número de clientes en el sistema es una medida del jitter

$$Var(N_{sist}) = \frac{\rho}{(1-\rho)^2}$$



Tiempo de espera en la cola

- Sea T_n el tiempo de espera en la cola del cliente n (tiempo entre que llega y comienza a ser atendido).
- Sea Y_n la cantidad de clientes en el sistema al momento del arribo del cliente n.
- Si $Y_n \geq 1$ (si $Y_n = 0,\ T_n = 0$) entonces

$$T_n = \sum_{k=1}^{Y_n - 1} S_k + \widehat{S}_{n-Y_n}$$

donde S_k es el tiempo de servicio del cliente k y $\widehat{S}_{n-Y_n}$ es el tiempo residual del cliente que está siendo servido.

Tiempo de espera en la cola

- Los tiempos de servicio y el tiempo residual son exponenciales de parámetro μ, de donde:

$$T_n = \sum_{k=1}^{Y_n} S'_k$$

con $S'_k \sim exp(\mu)$ e independiente de Y_n, entonces:

$$E(T_{cola}) = E(T_n) = E(Y_n)E(S'_1) = \frac{\rho}{1-\rho}\frac{1}{\mu}$$

- Observar que el tiempo medio en la cola y el número medio de clientes en la cola verifican

$$E(N_{cola}) = \lambda E(T_{cola})$$

Tiempo en el sistema (Media)

- Sea T_{sist} el tiempo de espera en el sistema (tiempo entre que un cliente llega y sale del sistema).
- El tiempo medio en el sistema es:

$$E(T_{sist}) = E(T_{cola}) + E(T_{servicio})$$

Entonces:

$$E(T_{sist}) = \frac{\rho}{1-\rho}\frac{1}{\mu} + \frac{1}{\mu} = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu-\lambda}$$

- Como antes, el tiempo medio en el sistema y el número medio de clientes en el sistema verifican:

$$E(N_{sist}) = \lambda E(T_{sist})$$

- Esta relación se conoce como **fórmula de Little** y es válida en contextos mucho más generales (tipo caja negra).
- Es fácil obtener la distribución del tiempo de estadía en el sistema.

Tiempo en el sistema (Distribución)

- Además del valor medio se puede hallar la distribución del tiempo de espera:

$$T_{sist} = T_n + S_n = \sum_{k=1}^{Y_n+1} S'_k$$

donde $Y_n \sim Geo(p)$ y $S'_k \sim exp(\mu)$ independientes

- Entonces:

$$P(T_{sist} \leq t) = \sum_{N=0}^{\infty} P\left(\sum_{k=1}^{N+1} S_k \leq t\right) P(Y_n = N)$$

donde $\sum_{k=1}^{N+1} S_k$ tiene distribución $Erlang(N+1, \mu)$

- Se obtiene que $T_{sist} \sim exp(\mu - \lambda)$
- Con las mismas ideas es posible obtener la distribución del tiempo de espera en la cola.

Tamaño de paquetes vs eficacia

- El mismo volumen de información puede ser transmitido por paquetes de tamaño medio $1/\mu$ y tasa de envío de paquetes λ o con paquetes de tamaño medio $1/m\mu$ y tasa $m\lambda$ ($m > 1$).
- La utilización será la misma (mismo ρ) y en consecuencia el número medio de paquetes en la cola será el mismo.
- Sin embargo los retardos son diferentes: paquetes de mayor tamaño implican retardos mayores.
- Paquetes de menor tamaño implican menor eficiencia (volumen de información útil vs. control).

"PASTA" Poisson Arrivals See Time Averages.

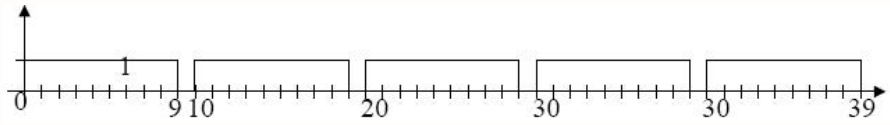
- Se considera un sistema de colas en estado estacionario, con arribos según un proceso de Poisson.
- La probabilidad de que un cliente al llegar encuentre al sistema en el estado i es igual a la probabilidad estacionaria de que el sistema se encuentre en dicho estado.
- Sean $\{A_n\}_{n=0,\ldots,\infty}$ el conjunto de los momentos en que llegan los clientes. Entonces

$$p(i) = P(N_{A_n^-} = i) = P(N_{sist} = i) = \pi_{est}(i)$$

- Ejemplos
  - La probabilidad de que al llegar un cliente encuentre el sistema vacío es la probabilidad de que el sistema esté vacío.
  - En un sistema con colas finitas, la probabilidad de que al llegar un cliente sea rechazado es la probabilidad de que el sistema se encuentre en el estado correspondiente a esa cantidad de clientes en espera.

¿Vale PASTA en otros casos?

- Supongamos arribos determinísticos cada 10 s y tiempo de servicio determinístico cada 9 s.



- Cuando un cliente arriba el sistema está siempre vacío, por lo tanto la probabilidad de que este observe un cliente en el sistema es $p(1) = 0$.
- Sin embargo la probabilidad de que exista un cliente en el sistema es $\pi_{est}(1) = 0,9$
- Los promedios que ven los clientes no son necesariamente el promedio temporal.

Contenido

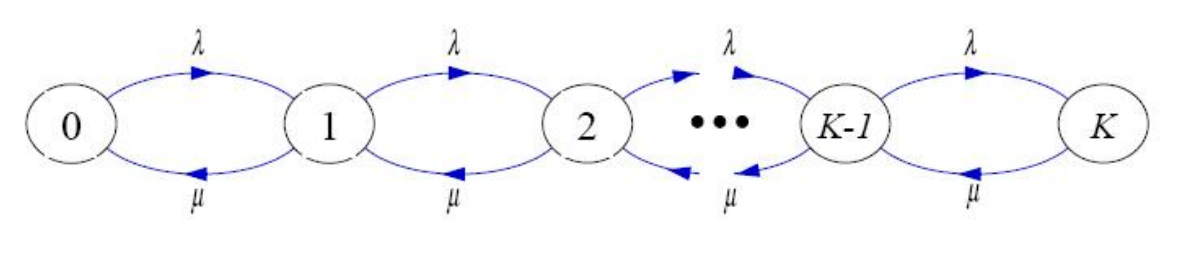
Características

- Proceso de arribos Poisson de parámetro λ.
- Tiempos de servicio exponenciales de parámetro μ.
- Tiempos de servicio y entre arribos independientes.
- $N_{sist}(t)$ número de clientes en el sistema en tiempo t.
- $\{N_{sist}(t) : t \geq 0\}$ es una cadena de Markov de tiempo continuo.
- Las transiciones entre estados dependen de λ y μ.
- Vale la fórmula de Little: si T_{sist} y N_{sist} son el tiempo en el sistema y el número de clientes en estado estacionario

$$E(N_{sist}) = \lambda E(T_{sist})$$

El proceso M/M/1/K

- Es un proceso M/M/1 con tamaño de buffer finito.
- A lo sumo hay K clientes en el sistema.
- Los clientes que arriban cuando el sistema está lleno (K clientes) son descartados.



Distribución invariante y probabilidad de pérdida

- Aplicando las ecuaciones de balance del sistema (coinciden con las del sistema M/M/1), se obtiene la distribución invariante:

$$\pi_{est}(n) = \rho^n \pi_{est}(0), \;\; n = 0, \ldots, K$$

- Siempre existe distribución invariante (Geométrica truncada):

$$\pi_{est}(0) = \frac{1-\rho}{1-\rho^{K+1}}$$

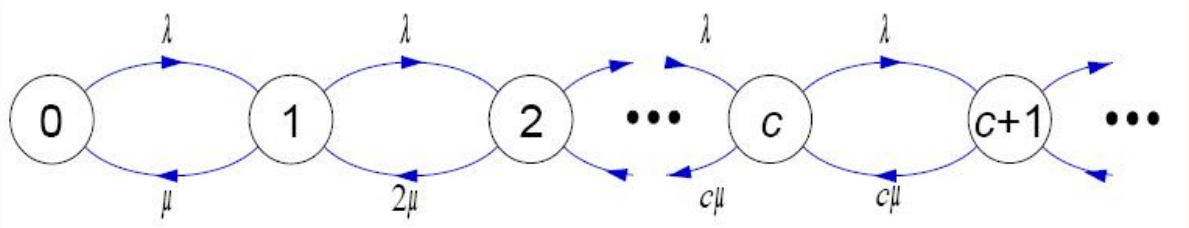
- La probabilidad de pérdida (PASTA) es:

$$P(\text{pérdida}) = \pi_{est}(K) = \frac{(1-\rho)\rho^K}{1-\rho^{K+1}}$$

- Ejercicio: relación con la probabilidad de sobrepasar un umbral en un sistema M/M/1.

El proceso M/M/C

- C servidores.
- Si el cliente que arriba encuentra n clientes en el sistema entonces
  - si $n < c$ es encaminado a cualquier servidor libre
  - si $n \geq c$ va a la cola de espera



Distribución invariante

- Distribución invariante: aplicando las ecuaciones de balance con $\rho = \lambda/\mu$ se tiene

$$\pi_{est}(i) = \frac{\rho^i}{i!}\pi_{est}(0) \text{ si } i < C$$

$$\pi_{est}(i) = \left(\frac{\rho}{C}\right)^{i-C} \frac{\rho^C}{C!}\pi_{est}(0) \text{ si } i \geq C$$

- La distribución invariante existe si $\lambda/C\mu < 1$, con

$$\pi_{est}(0) = \left(\left[\sum_{i=0}^{C-1} \frac{\rho^i}{i!}\right] - \frac{\rho^C}{C!}\left(\frac{1}{1-\rho/C}\right)\right)^{-1}$$

- Expresión analítica pero no muy tratable...

Fórmula Erlang-C

La probabilidad de, en estado estacionario, encontrar todos los servidores ocupados (Erlang-C) es:

$$\begin{aligned} P(\text{ocupado}) &= P(N_{sist} \geq C) \\ &= \sum_{n=C}^{\infty} \pi_{est}(n) \\ &= \frac{\rho^C}{C!} \left(\frac{1}{1-\rho/C} \right) \pi_{est}(0) \end{aligned}$$

Desempeño de la cola

- Si N_{cola} es el número de clientes en la cola

$$E(N_{cola}) = \sum_{i=C}^{\infty}(i - C)\pi_{est}(i) = \frac{\rho^{C+1}}{(C-\rho)^2(C-1)!}\pi_{est}(0)$$

- Si T_{cola} es el tiempo de espera en la cola, usando Little

$$E(T_{cola}) = \frac{\rho^{C+1}}{\lambda(C-\rho)^2(C-1)!}\pi_{est}(0)$$

- ¿Distribución del tiempo de espera en la cola?

Desempeño del sistema

- Si T_{sist}, T_{cola} y S son los tiempos de espera en el sistema, en la cola y el tiempo de servicio se tiene

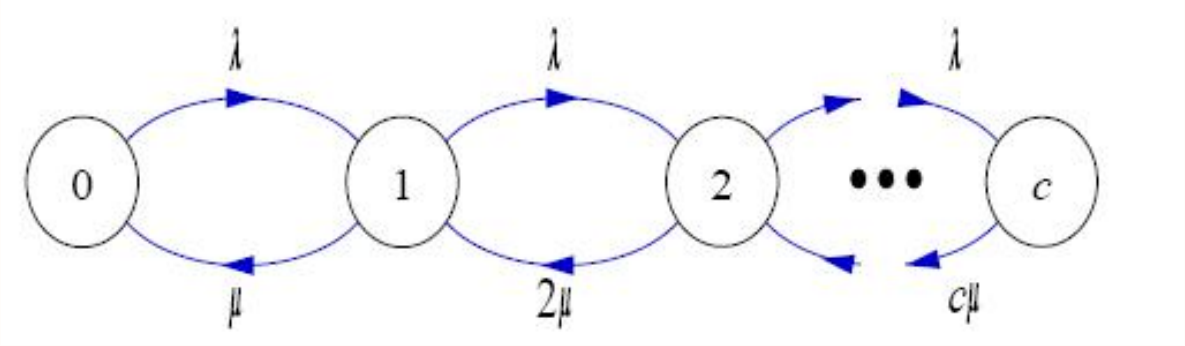
$$E(T_{sist}) = E(T_{cola}) + E(S) = \frac{\rho^{C+1}}{\lambda(C-\rho)^2(C-1)!}\pi_{est}(0) + \frac{1}{\mu}$$

- El número medio de clientes en el sistema se obtiene usando Little

$$E(N_{sist}) = \frac{\rho^{C+1}}{(C-\rho)^2(C-1)!} + \rho$$

El proceso M/M/C/C

- Es un proceso M/M/C con pérdidas.
- No hay espera, si el cliente llega y el sistema está lleno (C servidores ocupados) se descarta.



Distribución invariante

- Aplicando las ecuaciones de balance del sistema (coinciden con las del sistema M/M/C), se obtiene la distribución invariante:

$$\pi_{est}(n) = \frac{\rho^n}{n!}\pi_{est}(0), \;\; n = 0, \ldots, C$$

- Siempre existe distribución invariante:

$$\pi_{est}(0) = \left(\sum_{k=0}^{C} \frac{\rho^k}{k!} \right)^{-1}$$

- Erlang B: la probabilidad de bloqueo o pérdida es $\pi_{est}(C)$ (estamos usando PASTA).

El proceso M/M/∞

- ¡Ejercicio!

Multiplexado de recursos

- Comparar el desempeño de los siguientes sistemas:
  1. Arribos Poisson de intensidad λ y dos servidores con tiempo de servicio exponencial de parámetro μ y una sola cola.
  2. Arribos Poisson de intensidad λ y un servidor con tiempo de servicio exponencial de parámetro 2μ y una sola cola.
  3. Dos colas separadas con arribos de intensidad $\lambda/2$ y tiempo de servicio exponencial de parámetro μ.
- ¿Cuál es la carga de cada uno de los sistemas?
- ¿Número medio de clientes?
- ¿Tiempos de espera?

Contenido

Definición

- Proceso de arribos: Poisson de parámetro λ.
- Tiempos de servicio i.i.d. siguiendo una distribución cualquiera con media $1/\mu$.
- Los tiempos de servicio son independientes del proceso de arribos.
- Caso particular: M/D/1 (tiempo de servicio determinístico).

Características

- La cantidad de clientes en el sistema en tiempo t ($N(t)$) no es un proceso markoviano (por ejemplo si el tiempo de servicio es determinístico).
  - para pasar del estado n al estado $n-1$ (salida de un cliente) depende de cuánto tiempo hace que el cliente está en el sistema (no hay pérdida de memoria)
- No se puede calcular la distribución estacionaria en general.
- Calcularemos el tiempo medio de espera en la cola sin conocer la distribución estacionaria (fórmula de Pollaczek-Khinchin).

Notación

- W_i es el tiempo de espera del cliente i.
- X_i es el tiempo de servicio del cliente i.
- A_i es el número de arribos durante el tiempo de servicio X_i del cliente i.
- Q_i es el número de clientes en la cola cuando arriba el cliente i (excluido el cliente en servicio).
- R_i es el tiempo residual del cliente i.
- D_t es el número de partidas del sistema en $(0, t]$.
- A_t es el número de arribos al sistema en $(0, t]$.

Fórmula de Pollaczek-Khinchin

- El tiempo de espera para el cliente i es

$$W_i = R_i + X_{i-1} + X_{i-2} + \cdots + X_{i-Q_i} = R_i + \sum_{k=i-Q_i}^{i-1} X_k$$

- Tomando valor esperado de ambos lados

$$E(W_i) = E(R_i) + E\left(\sum_{k=i-Q_i}^{i-1} X_k\right) = E(R_i) + E(X)E(Q_i)$$

- En estado estacionario y por PASTA

$$E(W) = E(R) + E(X)E(Q)$$

Fórmula de Pollaczek-Khinchin

- Aplicando Little para la cola

$$E(Q) = \lambda E(W)$$

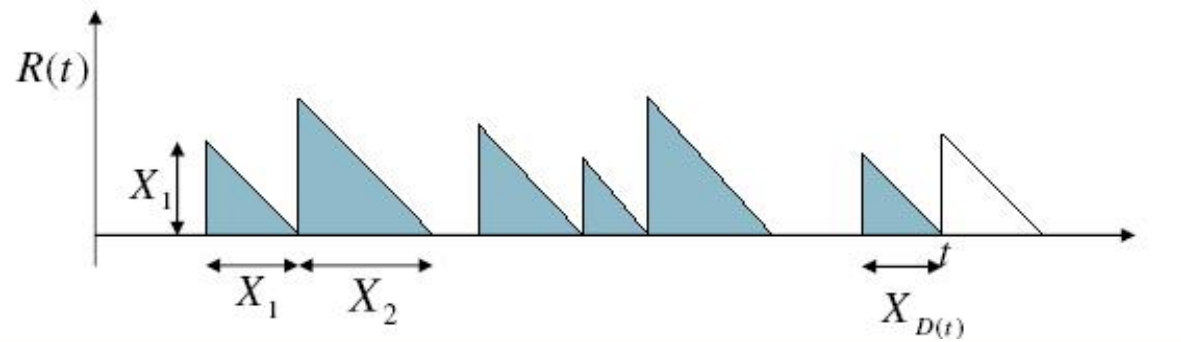
- Sustituyendo en la ecuación $E(W) = E(R) + E(X)E(Q)$ se tiene

$$E(W) = E(R) + E(X)\lambda E(W) = E(R) + \rho E(W)$$

de donde

$$E(W) = \frac{E(R)}{1-\rho}$$

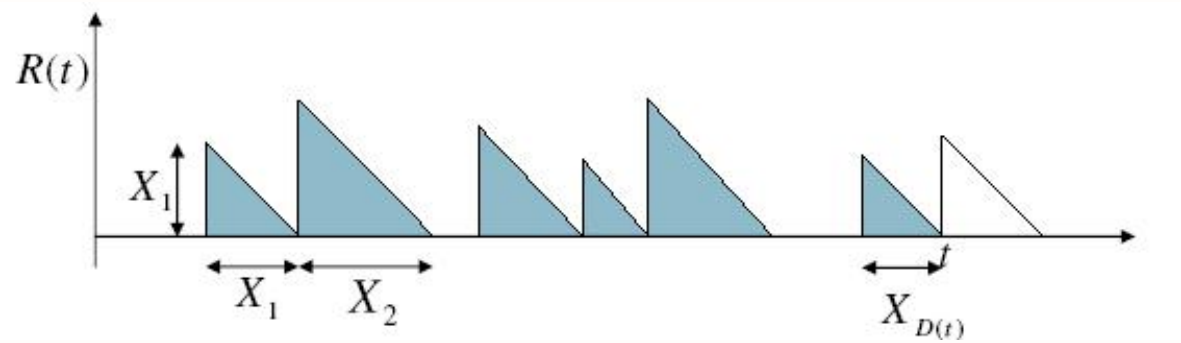
Tiempo residual medio



- R_t tiempo residual en función del tiempo.
- Consideremos t tal que $R_t = 0$ (cola vacía). Sea D_t el número de partidas en $(0, t]$ y asumimos que $R_0 = 0$.
- En estado estacionario se cumple que

$$E(R) = \lim_{t \to \infty} \frac{1}{t} \int_0^t R_s ds$$

Tiempo residual medio



- De la gráfica

$$\frac{1}{t}\int_0^t R_s ds = \frac{1}{t}\sum_{i=1}^{D_t}\frac{X_i^2}{2}$$

- Entonces

$$E(R) = \frac{1}{2}\lim_{t\to\infty}\frac{D_t}{t}\frac{1}{D_t}\sum_{i=1}^{D_t}X_i^2$$

Tiempo residual medio

- Por conservación de la masa se cumple:

$$\lim_{t\to\infty} \frac{D_t}{t} = \lim_{t\to\infty} \frac{A_t}{t} = \lambda$$

- Además

$$\lim_{t\to\infty} \frac{1}{D_t} \sum_{i=1}^{D_t} X_i^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i^2 = E(X^2)$$

- Entonces el tiempo residual medio es

$$E(R) = \frac{1}{2}\lambda E(X^2)$$

- De donde:

$$E(W) = \frac{\lambda E(X^2)}{2(1-\rho)}$$

- No depende solo de la media sino también de la varianza: tiempos de servicio muy dispares producen mayores retardos

Desempeño del sistema y de la cola: tiempos

- Sea W el tiempo en la cola y X el tiempo de servicio. Probamos que:

$$E(W) = \frac{\lambda E(X^2)}{2(1-\rho)} = \frac{\rho}{1-\rho}\frac{1}{\mu}\frac{1+C_v^2}{2}$$

- Sea T es el tiempo en el sistema. Entonces el tiempo medio en el sistema es

$$E(T) = E(W) + E(X) = \frac{\lambda E(X^2)}{2(1-\rho)} + \frac{1}{\mu}$$

- Comparar con los resultados obtenidos para una M/M/1.

Desempeño del sistema y de la cola: número de clientes

- Se obtienen a partir de la fórmula de Little
- Sea Q el número de clientes en la cola.

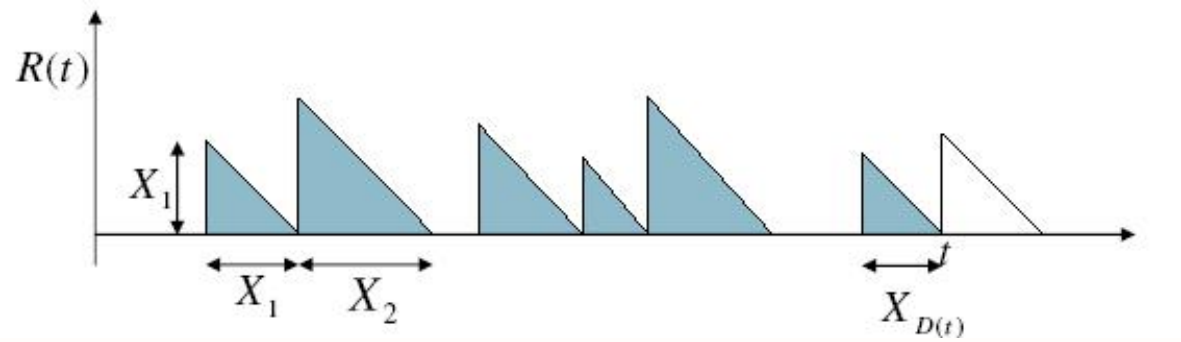
$$E(Q) = \lambda E(W) = \frac{\lambda^2 E(X^2)}{2(1-\rho)}$$

- Sea N el número de clientes en el sistema.

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

Distribución estacionaria: Método de la cadena incluida

- Sea N_t el número de clientes en la cola: N_t no es un proceso de Markov.
- Sea τ_k el instante de partida del sistema del k-ésimo cliente y $N_k = N_{\tau_k}$.
- El número de clientes N_k en el sistema inmediatamente después de la partida del k-ésimo cliente es una cadena de Markov en tiempo discreto (depende de N_k pero no de los que se fueron antes).



Distribución estacionaria de la cadena incluida

- La distribucion estacionaria de N_t coincide con la distribucion estacionaria π_{est} de N_k (arribos y partidas vienen de a pares).
- $\pi_k(n) = P(N_k = n)$ es la probabilidad de que haya n clientes en el sistema al partir el k-ésimo cliente y

$$\pi_{est} = \lim_{k \to \infty} \pi_k(n)$$

- Se tiene que $N_{k+1} = \begin{cases} N_k - 1 + V_{k+1} & N_k \geq 1 \\ V_{k+1} & N_k = 0 \end{cases}$ donde V_k es la cantidad de arribos durante el servicio del cliente k.
- Se pueden calcular las probabilidades de transición (o al menos estimarlas numéricamente).
- También se puede obtener la transformada z de π_{est}:

$$\Pi(z) = \frac{(1-\rho)L(\lambda - \lambda z)(1-z)}{L(\lambda - \lambda z) - z}$$

donde $\rho = \lambda/\mu$ y L es la transformada de la Laplace del tiempo de servicio.

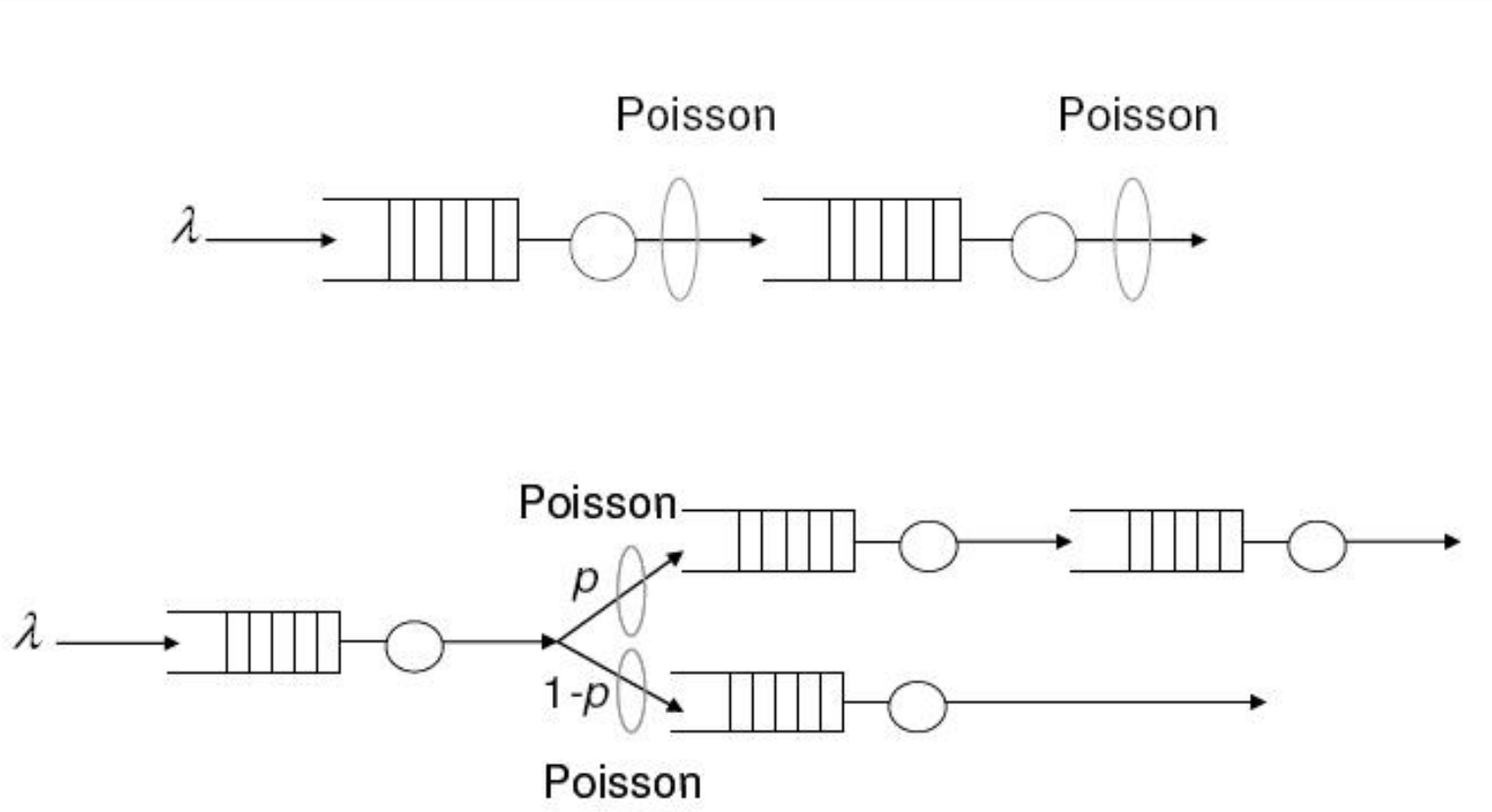
Processor Sharing

- El servidor divide su capacidad equitativamente entre todos los clientes presentes.
- Versión idealizada de Round Robin: el servidor reparte slots de tiempo de tamaño δ a cada cliente
  - si hay N clientes el tiempo residual de cada uno de ellos decrece a tasa $1/N$
- PS y FIFO tienen la misma capacidad (ρ), mismo número medio de clientes y el mismo tiempo medio de espera
- Sin embargo la distribución de los tiempos de espera son diferentes
  - El PS penaliza clientes con tiempos de servicio largos y la penalidad es proporcional al tiempo de servicio (*fairness*)

Contenido

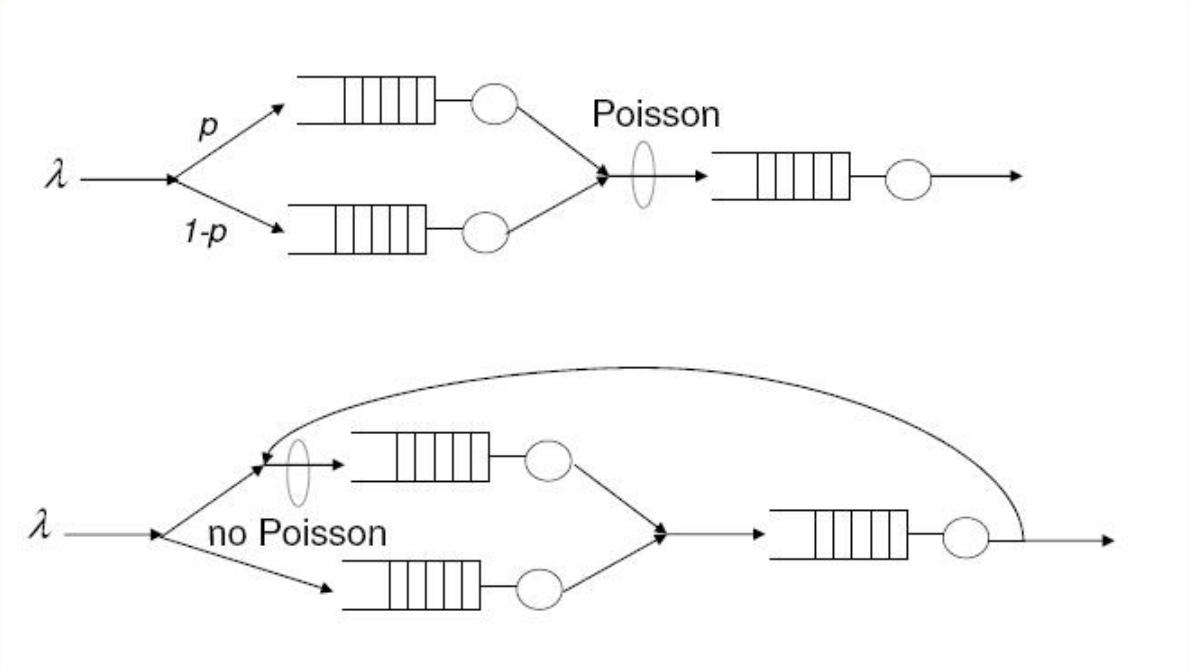
Redes de Colas

- Se considera un sistema M/M/1 o M/M/C, con tasa de arribos λ, en estado estacionario. Entonces:
  - El proceso de partida es Poisson con tasa λ.
  - En cada tiempo t, el número de clientes en el sistema es independiente de la cantidad de partidas anteriores a t.



Redes de Colas

- Se considera un sistema M/M/1 o M/M/C, con tasa de arribos λ, en estado estacionario. Entonces:
  - El proceso de partida es Poisson con tasa λ.
  - En cada tiempo t, el número de clientes en el sistema es independiente de la cantidad de partidas anteriores a t.



Reversibilidad

- Para el análisis hay que observar el proceso $\overline{X}_t = X_{T-t}$ para un T fijo.
- Se dice que X_t es reversible si $\overline{X}_t$ tiene la misma distribución que X_t.
- Las probabilidades de transición para $\overline{X}_t$ son:

$$\overline{p}_{ij}(t) = p_{ji}(t)\frac{\pi_j}{\pi_i}$$

$$\overline{q}_{ij}(t) = q_{ji}\frac{\pi_j}{\pi_i}$$

- Observar que no depende de T: $\overline{X}_t$ se puede definir $\forall \in \mathbf{R}$.
- X_t es reversible si y solo si $\pi_i q_{ij} = \pi_j q_{ji}$ (ecuaciones de balance detalladas).

Reversibilidad

- Ecuaciones de balance global ($\pi \mathbf{Q} = 0$):

$$\sum_{j \neq i} \pi_j q_{ji} = \pi_i \sum_{j \neq i} q_{ij}$$

- Si se verifican las ecuaciones de balance detalladas, se verifican las globales.
- Los procesos M/M/*/* son reversibles y se cumple el siguiente resultado:
  Teorema de Burke: el proceso de partidas $D = \{D_t\}$ donde D_t es el número de clientes que abandonan el sistema en el intervalo $[0, t]$ es un proceso de Poisson de parámetro λ. Además X_t es independiente de D_s para todo $s < t$.

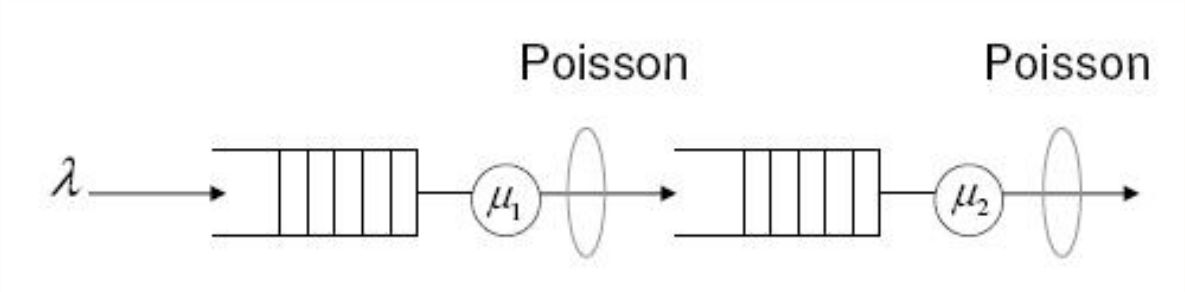
Reversibilidad

- La clave para estudiar las redes de colas es el siguiente resultado:
  - Sea X CMTC con generador infinitesimal $\mathbf{Q} = (q_{ij})_{ij \in E}$ y π un vector de probabilidad en E. Se define $\overline{Q} = (\overline{q}_{ij})$ como $\overline{q}_{ij} = q_{ij} \frac{\pi_j}{\pi_i}$. Si se cumple:

$$\sum_{i \neq j} \overline{q}_{ij} = \sum_{i \neq j} q_{ij} \tag{1}$$

    entonces:
    - π es la distribución estacionaria de X
    - $\overline{Q}$ es el generador infinitesimal del proceso $\overline{X}$
- Si se tiene un candidato a distribución, basta verificar la ecuación (1).

M/M/1 en tándem



- Proceso de arribos: Poisson de parámetro λ.
- Tiempos de servicio: exponenciales independientes de parámetro μ_i, $i = 1, 2$.
- Condición de estabilidad: $\rho_i = \lambda/\mu_i < 1$.
- N número de clientes en el sistema, $N = (N_1, N_2)$ donde N_i es el número de clientes en la cola i.

M/M/1 en tándem

- ¿Cuál es la distribución en estado estacionario de N_1 y N_2?

$$\pi_{est}(n_1, n_2) = (1 - \rho_1)\rho_1^{n_1}(1 - \rho_2)\rho_2^{n_2} = \pi_{est}^1(n_1)\pi_{est}^2(n_2)$$

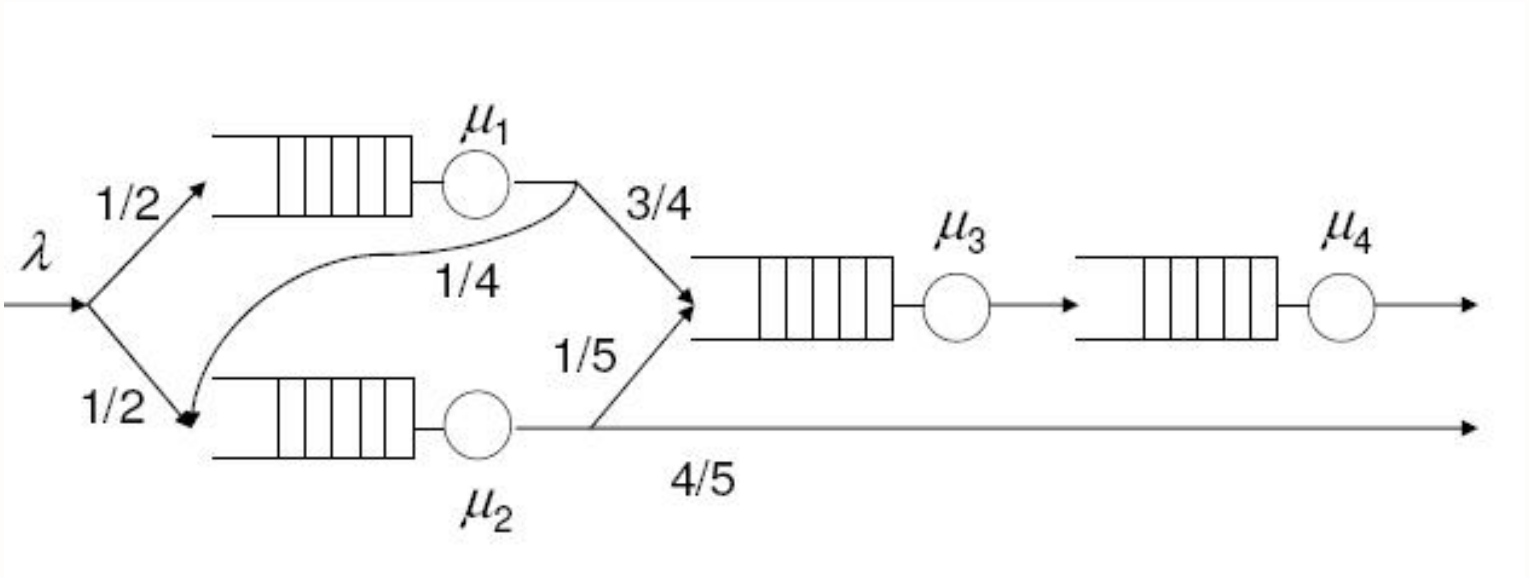
- $N_2(t)$ depende de los arribos a la cola 2 anteriores a t (i.e. las salidas de la cola 1 antes de t) y por lo tanto son independientes de $N_1(t)$.
- En general, la distribución estacionaria para el caso de m colas en cascada de tipo M/M/1, con tiempos de servicio independientes entre sí, es:

$$\pi_{est}(n_1, n_2, \ldots, n_m) = \Pi_{i=1}^m(1 - \rho_i)\rho_i^{n_i} = \Pi_{i=1}^m \pi_{est}^i(n_i)$$

- Los tiempos de espera en cada fila también son independientes

Redes de Jackson abiertas

- Los clientes arriban desde el exterior.
- Una sola clase de clientes.
- Proceso de arribos: Poisson de parámetro λ.
- Cantidad de nodos: m.
- Un solo servidor en cada nodo con tiempos de servicio exponenciales independientes, sin pérdidas (cola infinita).
- Ruteo probabilístico.



Ecuación de flujo

- Sea p_{jk} la probabilidad de un cliente que termina su servicio en el nodo j se dirija al nodo k, $j, k = 1, \ldots, m$.
- Sea p_{0k} la probabilidad de un cliente que arriba lo haga al nodo k.
- La tasa de arribos al nodo i es:

$$\lambda_i = \lambda p_{0i} + \sum_{k=1}^{m} \lambda_k p_{ki}$$

- La ecuación anterior ("ecuación de flujo") tiene una única solución positiva.

Teorema de Jackson

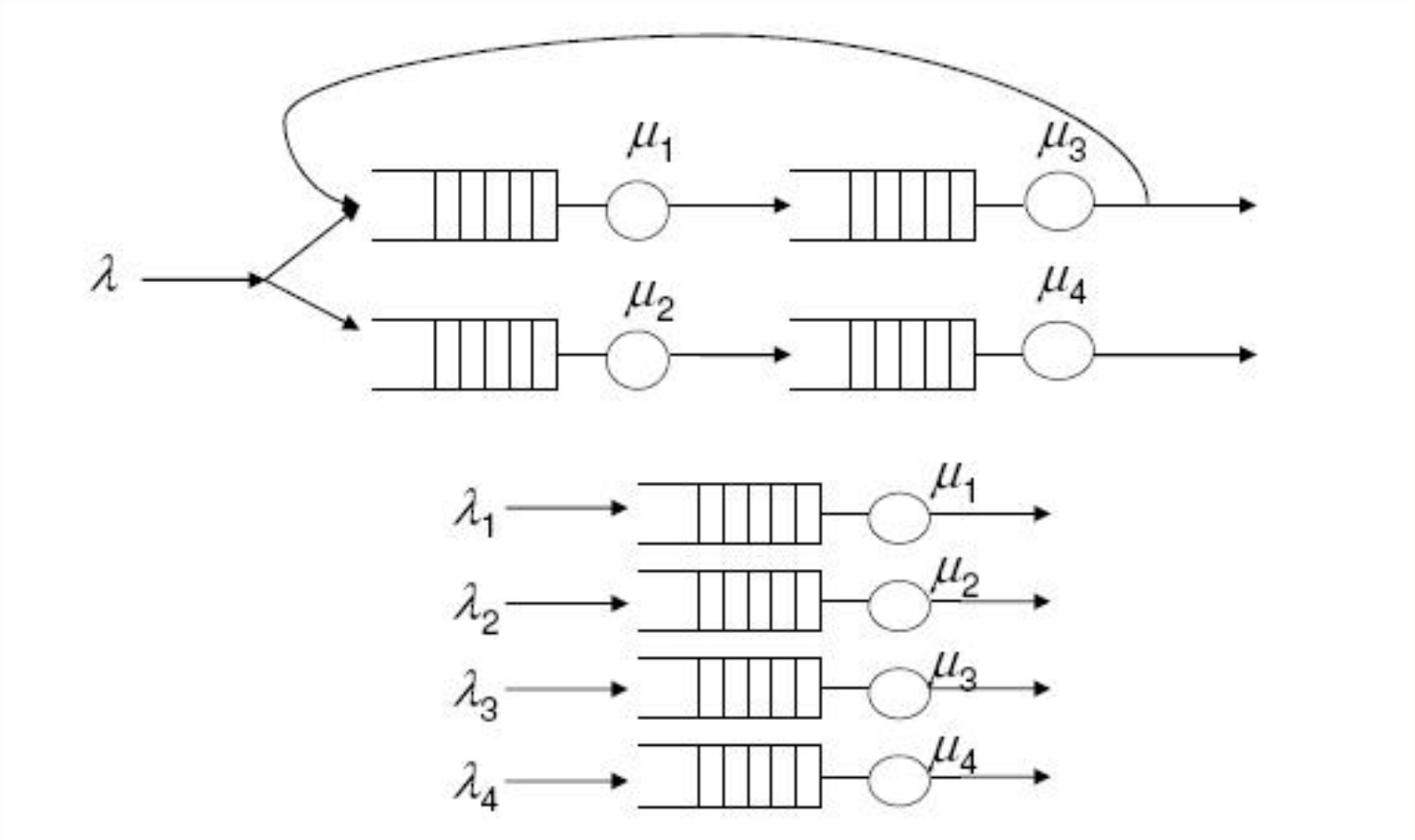
- Sea N_t^i el número de clientes en el instante t en el nodo i.
- El proceso de Markov N, tal que $N_t = (N_t^1, \ldots, N_t^m)$ describe la cantidad de clientes en el sistema.
- Teorema de Jackson: Si $(\lambda_1, \ldots, \lambda_m)$ es la solución de la ecuación de flujo, y el sistema es estable $(\lambda_i < \mu_i)$, el proceso N tiene una distribución invariante dada por:

$$\pi_{est}(n_1, n_2, \ldots, n_m) = \Pi_{i=1}^m (1 - \rho_i)\rho_i^{n_i} = \Pi_{i=1}^m \pi_{est}^i(n_i)$$

donde π_{est}^i es la distribución invariante de una fila M/M/1 con tasa de arribo λ_i y tasa de servicio μ_i.

Teorema de Jackson

El sistema es equivalente a un conjunto de colas M/M/1.



Desempeño de cada nodo

- El tiempo medio de espera en el nodo i es

$$E(T^i) = \frac{1}{(1-\rho_i)\mu_i} = \frac{1}{\mu_i - \lambda_i}$$

- Luego en el sistema es

$$E(N) = \sum_{i=1}^{m} E(N^i) = \sum_{i=1}^{m} \frac{\rho_i}{1-\rho_i}$$

- Por Little número medio de clientes en el nodo i es

$$E(N^i) = \lambda_i E(T^i) = \frac{\rho_i}{1-\rho_i}$$

- Y en el sistema es

$$E(T) = \frac{E(N)}{\lambda} = \frac{1}{\lambda} \sum_{i=1}^{m} \frac{\rho_i}{1-\rho_i}$$

Otros ejemplos

- Redes de Jackson cerradas: el número de clientes en el sistema es constante (no hay arribos ni partidas del exterior)
  - Hay solución de la ecuación de flujo pero no es única. Las filas no se pueden tratar como independientes.
  - Mean Value Analysis: expresar los parámetros de performance (en media) del sistema con N clientes en función de los mismos parámetros para un sistema con $N-1$ clientes.
    - Se basa en que la distribución del estado de la red en el momento de una transición coincide con la distribución de la misma red con un cliente menos.
- Redes de Whittle: insensibilidad a la distribución del tiempo de servicio.