Prácticas de teletráfico

marzo de 1999

1 Introducción

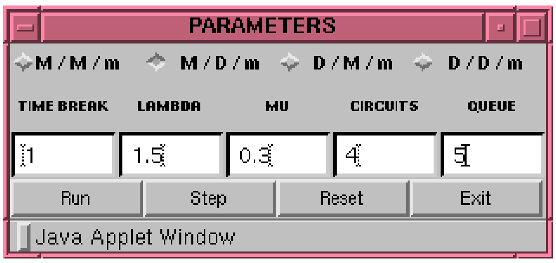
El objetivo de estas prácticas es familiarizar al estudiante con los conceptos y la terminología propias de la teoría de las colas.

El procedimiento utilizado se basa en una simulación gráfica mediante un *applet* de Java, que permite reproducir el funcionamiento del modelo con los parámetros fijados por el usuario. El *applet* consta de tres ventanas:

- *Parameters*
- *Statistics*
- *Simulation*

1.1 Ventana *Parameters*

En la parte superior del panel de entradas podemos observar cuatro opciones que determinan el modelo que simula el programa. La opción que está activada determina la naturaleza de generación de llamadas y el servicio del sistema ($M/M/m$, $M/D/m$, $D/M/m$, $D/D/m$). En cualquier momento se puede cambiar de modelo, simplemente con hacer clic en la opción deseada.



Debajo encontramos unas casillas con los siguientes significados:

TIME BREAK es el tiempo entre cada uno de los pasos de la simulación. La cifra representada en el cuadro de texto se multiplica por 100 ms. El rango del tiempo de proceso en la simulación de $\{1 \ ... \ 5\}$. Sólo se modificará si la simulación va excesivamente rápida.

LAMBDA es la tasa media de llegadas al sistema. Representa el número de llamadas que llegan al sistema por unidad de tiempo (llamada/tiempo). Su rango es de $\{0.1 \ldots 1.9\}$.

MU es la tasa media de servicio. Representa el número medio de llamadas que el sistema sirve por unidad de tiempo (llamada/tiempo). Su rango es de $\{0.1 \ldots 2.0\}$.

CIRCUITS indica el número de circuitos que tiene el servidor. Su rango es de $\{1 \ldots 4\}$.

QUEUE fija el número máximo de llamadas que pueden esperar en cola. Su rango es de $\{0 \ldots 8\}$.

Más abajo encontramos una serie de botones que nos permiten controlar la simulación.

RUN pone en marcha la simulación con los valores de los cuadros de texto superiores.

STEP nos permite detener la simulación, y, si la accionamos repetidamente, nos permite simular paso a paso y observar detenidamente el comportamiento del sistema. Este botón también actualiza los valores de los cuadros de texto superiores.

RESET termina la actual simulación y reinicia el sistema con una nueva simulación.

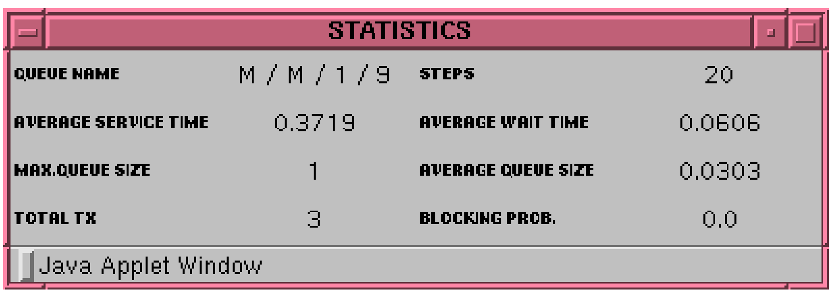
EXIT termina definitivamente con la simulación y elimina las ventanas.

1.2 Ventana *Statistics*

Esta ventana proporciona información sobre los parámetros de la propia simulación.

QUEUE NAME sigue la nomenclatura Kendall $A/B/C/k$, donde

- A representa la distribución de la llegada (M markoviano, D determinista)
- B representa la distribución del servicio (M marcoviano, D determinista)
- C representa el número de circuitos
- k es el número máximo de llamadas que puede contener el sistema (incluida la que se está sirviendo).



STEPS representa el número actual de pasos de la simulación.

AVERAGE SERVICE TIME es el tiempo medio de servicio del sistema.

AVERAGE WAIT TIME es el tiempo medio de espera en cola.

MAX. QUEUE SIZE es la ocupación máxima que ha sufrido la cola.

AVERAGE QUEUE SIZE es el número medio de llamadas que esperan en cola.

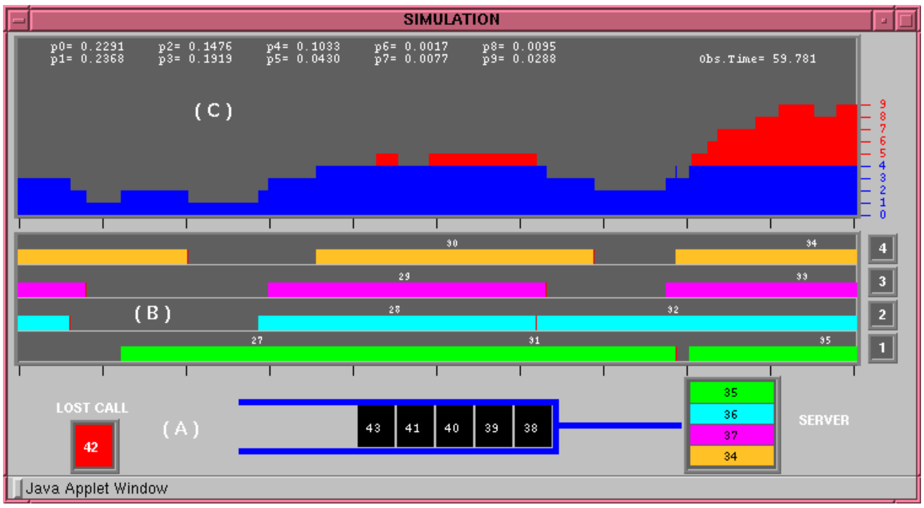
TOTAL TX es el número total de paquetes transmitidos con éxito.

BLOCKING PROB. es la probabilidad de boqueo del sistema.

1.3 Ventana *Simulation*

La ventana Simulation tiene tres partes:

- (A) Animación del servidor, cola y llamada perdida. Representa lo que hay en el sistema en cada momento; es decir, qué llamadas se están sirviendo, cuáles están esperando y cuál es la última llamada perdida.
- (B) Animación de la ocupación del servidor. Es una pantalla fraccionada en diez unidades de tiempo que representa el histórico de lo que sucede en el servidor, concretamente en cada uno de los circuitos que lo componen.
- (C) Animación de la ocupación global del sistema. Es una pantalla fraccionada en diez unidades de tiempo que representa la ocupación global del sistema (servidor más cola). La ocupación del servidor se representa en azul y la de la cola en rojo. Además, también se visualizan las probabilidades estimadas de ocurrencia de cada estado del sistema (estado 0 hasta el estado k).



Si comparamos las dos pantallas de animación, observamos que la suma de la ocupación del servidor es exactamente la saturación en azul que dibujamos en la animación del cambio de estado. La parte pintada en rojo representa gráficamente lo que sucede en la cola; es decir, cada flanco de bajada de esta zona corresponde a un final de servicio representado en la animación de la ocupación del servidor (se distingue con una línea roja al final de cada llamada) y significa que un circuito del servidor se ha quedado libre y, por tanto, la cola se vacía en una unidad.

2 Conceptos de tráfico

El modelo básico utilizado en teoría de colas corresponde a una población que genera llamadas con una tasa λ, estas llamadas llegan a un sistema de circuitos donde serán servidas con una tasa μ.

La primera parte podría corresponder a la modelización de generación aleatoria de llamadas correspondiente a una población grande. Es un parámetro fácil de medir, porque sólo es necesario observar durante un tiempo de observación (T_{ob}) el número de llamadas generadas (N). Entonces,

$$\lambda = \frac{N}{T_{ob}}$$

Este parámetro será tanto más fiable cuanto mayor sea T_{ob}. Las unidades de λ serán llamadas/tiempo.

A veces a la generación de llamadas también se le llama llegada de llamadas al sistema, porque visto desde el sistema es como si llegasen.

Una vez la llamada ha llegado al sistema (se ha generado), entonces, en el caso de que haya recursos suficientes, se sirve. En nuestro caso los recursos son circuitos, pero estos modelos son aplicables a campos muy diversos.

El tiempo de servicio de llamadas se modela como una variable aleatoria exponencial de tasa μ. Este parámetro también es fácil de medir, pues $1/\mu$ es el tiempo medio de servicio, que se puede estimar midiéndolo.

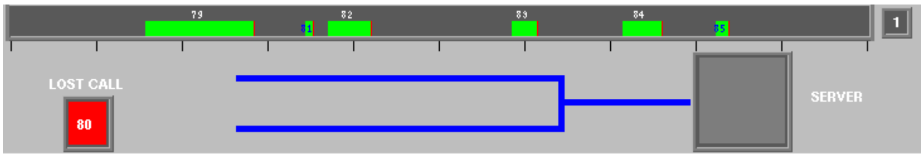
En caso de que no haya suficientes recursos para servir la llamada, hay dos políticas de funcionamiento, correspondientes al modelo de pérdida y al modelo de espera. El grado de servicio, la satisfacción del cliente, se mide a través de la probabilidad de bloqueo (PB). Esta probabilidad de bloqueo es diferente para cada modelo; es decir, en un modelo de pérdida la probabilidad de bloqueo es la probabilidad de pérdida, mientras que en un modelo de espera la probabilidad de bloqueo es la probabilidad de espera, aunque también es importante el tiempo medio de espera en cola.

2.1 Modelo de pérdida

En el modelo de pérdida, en caso de que una llamada llegue al sistema y encuentre que todos los recursos estén ocupados, ésta no se sirve y se pierde.

En este caso un parámetro de medida del grado de servicio es la probabilidad de pérdida (PP), es decir, la probabilidad de que una llamada que llegue al sistema se encuentre con todos los recursos ocupados. Si el sistema está sobredimensionado, esta probabilidad es baja. Si el sistema está infradimensionado, esta probabilidad es alta.

1. ¿Cuál es la probabilidad de pérdida (PP) en el sistema de un circuito de la figura? Estimadlo con la información que veis. Fijaos que es el cociente entre el número de paquetes perdidos y el número de paquetes totales.



A la hora de dimensionar un sistema de pérdida se debe calcular el número de recursos suficientes para poder obtener una probabilidad de bloqueo adecuada.

2.2 Modelo de espera

En el modelo de espera, en caso de que una llamada llegue al sistema y no encuentre recursos libres para ser servida, ésta espera hasta que se libere algún recurso.

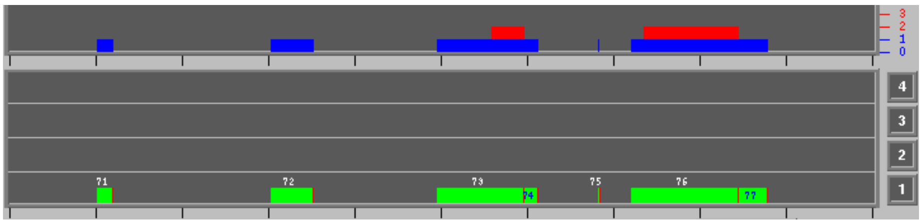
En este caso, el parámetro para medir el grado de servicio es la probabilidad de demora (PD), es decir, la probabilidad de que una llamada que llega al sistema tenga que esperar porque no hay recursos libres.

Otro parámetro interesante es el tiempo medio que una llamada tiene que esperar en cola hasta el momento de ser servida (WQ).

A la hora de dimensionar un sistema se calculan los recursos necesarios para poder cumplir con alguno de los parámetros anteriores.

1. ¿Cuál es la probabilidad de demora (PD) en el sistema de un circuito de la figura siguiente? Estimdalo con la información que veis. Fijaos que es el cociente entre el número de paquetes demorados y el número de paquetes totales.

2. ¿Y cuál es el tiempo medio que una llamada tiene que esperar en cola hasta ser servida, según la información de la figura? Fijaos que se debe hacer un promedio de los tiempos de espera en cola.



2.3 Conceptos generales de tráfico

Aunque consideremos en primer lugar el modelo de pérdida, los conceptos que ahora se definen se aplican también al modelo de espera.

Se define intensidad de tráfico instantáneo cursado, o tráfico instantáneo cursado, *N(t)*, como el número de circuitos ocupados a lo largo del tiempo.

En el caso de un circuito, esta función podría tener la siguiente forma,

Se define volumen de tráfico cursado (V_{tc}) de un circuito como la suma de los tiempos en que éste está ocupado en un tiempo de observación, T_{ob}. En el ejemplo, $V_{tc} = 5$s.

Si tenemos 2 circuitos, la función *N(t)* podría tener la siguiente forma:

Y V_{tc} sería en este caso la suma de los tiempos en que éstos están ocupados en un tiempo de observación T_{ob}. En el ejemplo, $V_{tc} = 6$s.

En general, para C circuitos,

$$V_{tc} = \int_{T_{ob}} N(t)dt$$

1. La función *N(t)* de un sistema de circuitos tiene la siguiente forma,

   (a) ¿Cuántos circuitos tiene nuestro sistema como mínimo?

   (b) Calculad el V_{tc} en un tiempo de observación T_{ob} de las *10* unidades de tiempo.

Hasta ahora hemos calculado el V_{tc}. Este parámetro no da mucha información. ¿Qué quiere decir que tenemos un V_{tc} de *20 s*? En cambio, la relación entre V_{tc} y T_{ob} sí que nos desvela el porcentaje de ocupación del sistema.

Se define intensidad de tráfico cursado, o simplemente tráfico cursado (TC), como el cociente entre V_{tc} y T_{ob}.

$$TC = \frac{V_{tc}}{T_{ob}}$$

En el ejemplo, $V_{tc} = 5$s y $T_{ob} = 10$s, entonces, *TC = 0.5*.

En general,

$$TC = \frac{V_{tc}}{T_{ob}} = \frac{1}{T_{ob}} \int_{T_{ob}} N(t)dt$$

Gráficamente, TC coincide con el valor que tiene una función constante que tenga el mismo valor que la integral de la función *N(t)*, área bajo la curva, en el tiempo de observación, ya que

$$T_{ob}TC = \int_{T_{ob}} N(t)dt$$

En el caso del último ejemplo, obtendríamos gráficamente un valor que tendría que aproximarse al que habéis calculado.

En el apéndice se puede ver que

$$TC = \frac{\lambda_c}{\mu}$$

donde λ_c es la tasa de llamadas llegadas al sistema y cursadas.

Y de ello se extraen diferentes características de TC. Para un sistema de un circuito:

- TC es adimensional y tiene un nombre, *Erlang*, en honor a un matemático danés. Así, en el ejemplo, TC = *0.5 E* (el tráfico cursado es 0.5 erlangs).
- Se cumple que $TC \leq 1$.
- TC nos da el porcentaje de ocupación del circuito.

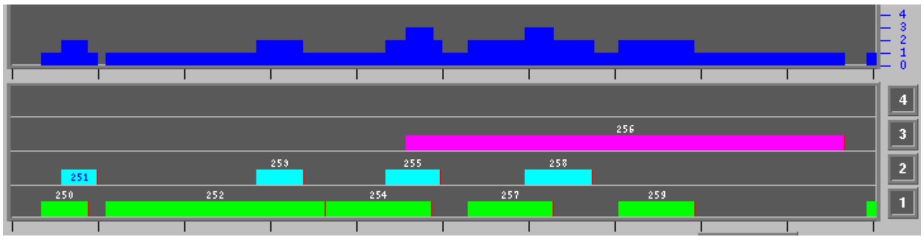
Si en lugar de un circuito, tenemos C,

- TC es adimensional y se da en Erlangs.
- Se cumple que $TC \leq C$.
- TC nos da el porcentaje de ocupación del sistema.

Si consideramos un modelo de pérdida, es posible que alguna llamada llegue al sistema y se pierda. Por eso es interesante considerar el caso en que todas las llamadas sean servidas.

Fijaos en que hasta ahora no nos preocupábamos de las llamadas que llegaban al sistema y no eran cursadas a causa de la falta de recursos libres para ser servidas. Pero es un parámetro importante del grado de servicio. Para saber qué perdemos, primero hemos de calcular qué tendríamos en el caso de que hubiera bastantes circuitos (matemáticamente, infinitos circuitos) para servir cualquier llamada que se generara.

Definimos tráfico ofrecido (TO) como el tráfico cursado por un sistema con un infinito número de circuitos. En el simulador, el tráfico ofrecido se podría ver en un sistema con suficientes circuitos para servir todas las lamadas que llegasen.



Se puede ver que,

$$TO = \frac{\lambda}{\mu}$$

Y como en el modelo de pérdida, no todo el tráfico ofrecido es cursado, y se define el tráfico perdido (TP) como el tráfico ofrecido que no se cursa.

$$TP = TO - TC$$

También se puede ver de la siguiente forma

$$\frac{\lambda_p}{\mu} = \frac{\lambda}{\mu} - \frac{\lambda_c}{\mu}$$

Donde λ_p es la tasa de llamadas que llegan al sistema y se pierden.

Se cumple que $C \geq TO \geq TC$. Comprobadlo.

Todos estos conceptos también se pueden aplicar al modelo de espera. El volumen de tráfico cursado (V_{tc}) de un circuito será la suma de los tiempos en que éste está ocupado durante un tiempo de observación T_{ob}. En el simulador, corresponde a la función dibujada en azul.

En el ejemplo, $V_{tc} = 4.7$s.

Como antes, nos interesa la relación entre V_{tc} y T_{ob}; ya que nos indica el porcentaje de ocupación del circuito.

Se define intensidad de tráfico cursado, o simplemente tráfico cursado (TC), como el cociente entre V_{tc} y T_{ob}.

$$TC = \frac{V_{tc}}{T_{ob}}$$

Intuitivamente, en un modelo de espera puro todas las llamadas que lleguen al sistema serán servidas tarde o temprano; el único inconveniente es que quizás, antes de ser servidas, tendrán que esperar en la cola. Por ello, según este razonamiento, no existirá tránsito perdido (no se pierden llamadas) y, por tanto, el tráfico cursado será el mismo que el tráfico ofrecido (esto lo veremos matemáticamente más adelante). En cambio, habrá una parte de las llamadas que lleguen al sistema que tendrán que esperar, λ_d, y esto nos llevará a calcular un tráfico demorado, que es otro parámetro del grado de servicio.

Hemos visto que el sistema de colas que estamos estudiando se caracteriza por tres parámetros: λ, μ y C. Además, para relacionarlo con el grado de servicio, hemos definido una serie de conceptos que nos ayudarán a medirlo. Ahora nos faltará relacionar estos conceptos con los parámetros y con el grado de servicio. Todos estos sistemas se caracterizan por una serie de estados, caracterizados normalmente por el número de llamadas en el sistema. Así, el estado *0* corresponde al sistema vacío, el estado *1* al sistema con una llamada, el estado *2* con dos llamadas, y así sucesivamente.

Si se conoce la probabilidad de ocurrencia (ver apéndice) de cualquier estado del sistema, entonces se puede calcular cualquiera de sus parámetros *(TC, TP, PP, ...)*. En el caso del simulador, se estimará la probabilidad de ocurrencia de cualquier estado del sistema, entonces se puede calcular una estimación de cualquiera de sus parámetros (*TC, TP, PP, ...*).

3 Modelos de pérdida

Vamos a familiarizarnos con los conceptos que se han explicado en la sección anterior y con el simulador. Veremos cómo a partir del simulador podemos calcular (realmente estimar) los conceptos que hemos visto anteriormente. En el apéndice se dan los instrumentos matemáticos para calcularlos, y comprobaremos que los resultados coinciden.

Comenzaremos por el caso más sencillo, que es considerar que tanto las llegadas como el servicio no son aleatorios, sino determinísticos.

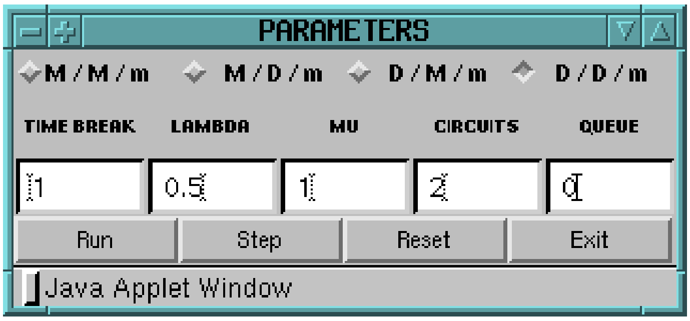
También comenzaremos por los modelos de bloqueo. Un modelo se considera de bloqueo cuando una llamada que llega al sistema y no tiene recursos, circuitos, para ser atendida, se pierde.

3.1 Modelo $D/D/C/C$

Este modelo compuesto de C circuitos y sin cola (Q = 0) se caracteriza por tener las llegadas y el servicio determinísticos; es decir, las llegadas se producen periódicamente cada $1/\lambda$ s y el servicio de cada llamada tarda exactamente $1/\mu$ s.

Un sistema sobredimensionado es aquel en el que hay más circuitos de los que se necesita para dar un grado de servicio (satisfacción del cliente) razonable. Un sistema infradimensionado es aquel en el que hay menos circuitos de los necesarios para dar un grado de servicio aceptable. En un modelo de pérdida el grado de servicio está íntimamente ligado a la probabilidad de bloqueo.

1. Simulad el modelo *D/D/2/2* con un caso sobredimensionado. Por ejemplo, que lleguen llamadas cada *2* s, y que tarden en servirse *1* s. Esto corresponderá a una $A = \frac{\lambda}{\mu} = \frac{1/\mu}{1/\lambda} = \frac{1}{2} = 0.5$.



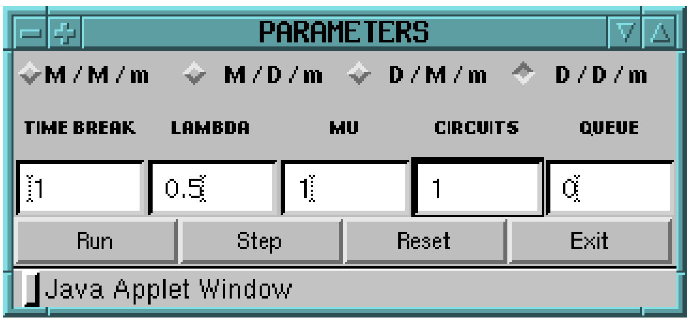
2. ¿Cuáles son las probabilidades que se obtienen por simulación de cada estado?

   (a) ¿Cuál es la ocupación media del circuito *1*? ¿Y del circuito *2*?

   (b) ¿Podríamos quitar el circuito *2* y tener el mismo sistema (aunque más económico)?

   (c) ¿Cuál es la probabilidad de bloqueo; es decir, de cada *10* llamadas que lleguen, cuántas perdemos como media? ¿Perdemos alguna llamada?

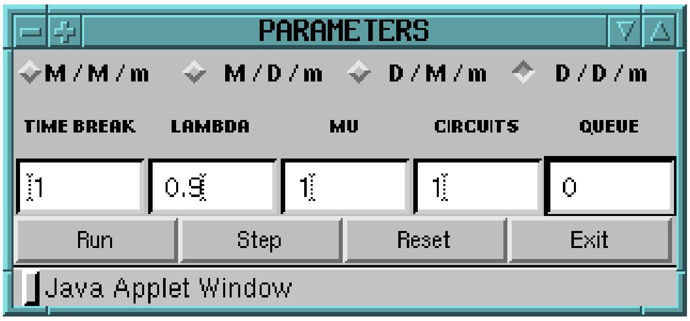
3. Simula el modelo *D/D/1/1* con A=*0.5*. Simplemente reducimosel número de circuitos en una unidad.



(a) ¿Cuáles son las probabilidades que se obtienen por simulación de cada estado?

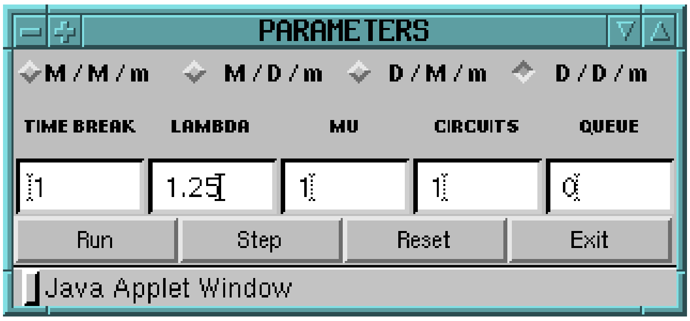
(b) ¿Cuál es el porcentaje de tiempo que el circuito está ocupado? ¿Coincide con algún parámetro del sistema? Relacionadlo con A, P_0 i P_1.

4. Vamos a ver si se cumplen estas relaciones en otros casos. Simulad el modelo *D/D/1/1* con A=*0.9*.



(a) Se mantienen las relaciones calculadas anteriormente?

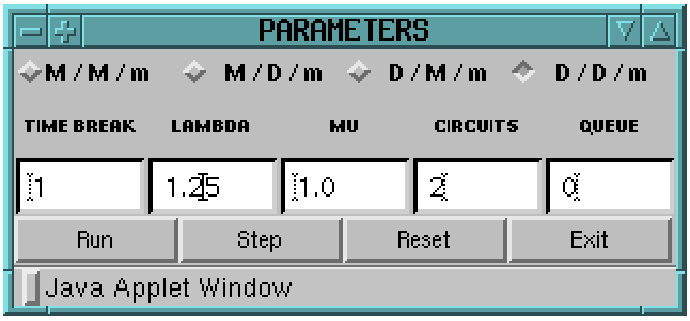
5. Parece que no queramos sobrepasar A=*1*, vamos a hacerlo. Simulad el modelo *D/D/1/1* con A=*1.25*.



(a) ¿Qué veis en la simulación? Fijaos en que se pierden llamadas. Coged una secuencia de numeración de llamadas perdidas. ¿Cuál es la probabilidad de pérdida? ¿Coincide con la que se da en el panel de estadísticas?

(b) ¿Cuál es el porcentaje de tiempo que el circuito está ocupado? ¿Es un sistema sobredimensionado? ¿Tiene este sistema un grado de servicio aceptable? ¿Cómo se puede mejorar de forma sencilla este grado de servicio?

6. Quizás la solución sea añadir un circuito más. Simulad un modelo $D/D/2/2$ con A=*1.25*. ¿Cuál es el porcentaje de tiempo que *1* circuito está ocupado, y el porcentaje de tiempo que *2* circuitos están ocupados? Relacionadlos con las probabilidades de estado.

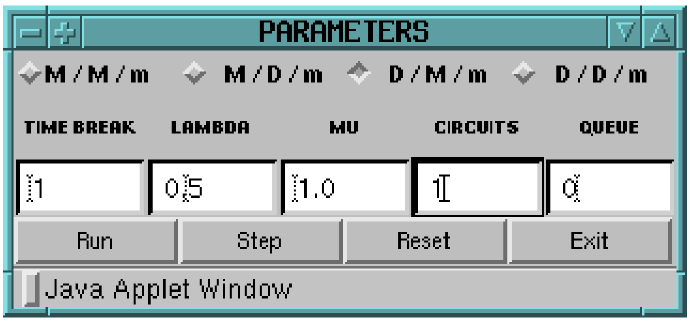


7. ¿Cuál creeis que es la relación que hay entre A y C para que no se pierda ninguna llamada?
8. En conclusión, si tenemos A=*3.1*. ¿Cuál es el número de circuitos necesarios para dar un grado de servicio aceptable?
9. Dado un modelo $D/D/2/2$, dibuja la gráfica de la probabilidad de bloqueo en función de λ para una a $\mu = 1$.

3.2 Modelo $D/M/C/C$

Este modelo compuesto de C circuitos y sin cola ($Q = 0$) se caracteriza por tener las llegadas determinísticas y el servicio aleatorio; es decir, las llegadas se producen exactamente cada $1/\lambda$ s y el servicio de cada llamada tarda como media $1/\mu$ s.

1. Simulad un modelo $D/M/1/1$ con A=*0.5*. Comprobad que cada *2* s llega un llamada y que, como media, se tarda *1* s en servirse (unas veces más y otras menos).



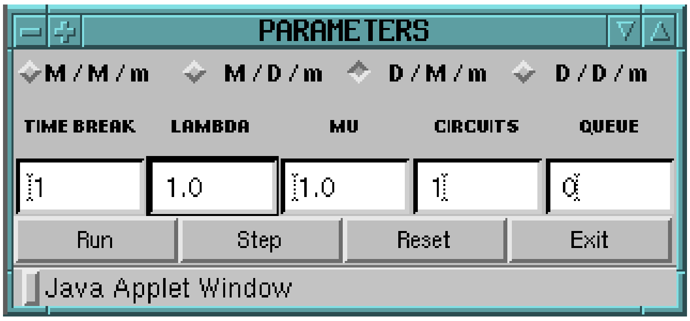
2. ¿Se pierden llamadas? ¿Por qué? Anotad la probabilidad de bloqueo que se obtiene. ¿Creeis que proporciona un grado de servicio aceptable?

   (a) ¿Podríais dar un valor de A para garantizar que no se pierdan llamadas? ¿Por qué?

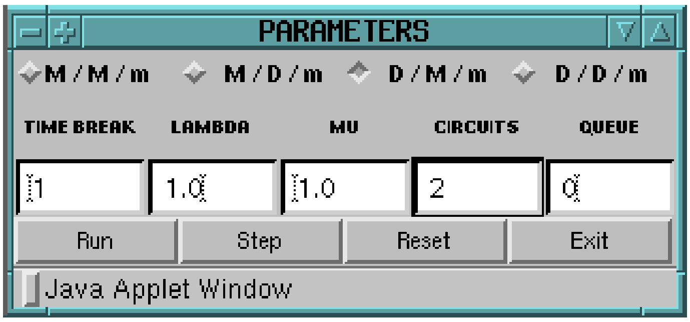
   (b) ¿Podríais dar un valor de C para garantizar que no se pierdan llamadas?

   (c) Ya se ve que se puede ir reduciendo la probabilidad de bloqueo con un coste económico. Hay dos vertientes que tener en cuenta: con un coste económico fijo tengo un grado de servicio determinado, y para conseguir un grado de servicio aceptable tendré un coste determinado. A la hora de diseñar quizás tengamos fijo un parámetro y hayamos de calcular el otro.

3. Intuitivamente se ve que si aumentamos A, aumentaremos la probabilidad de pérdida. Simulad un modelo *D/M/1/1* con A=*1* y anotad la probabilidad de bloqueo.



4. También se ve que si aumentamos C, disminuimos la probabilidad de bloqueo. Simulad un modelo *D/M/1/1* con A=*1* y C=*2* y anotad la probabilidad de bloqueo.

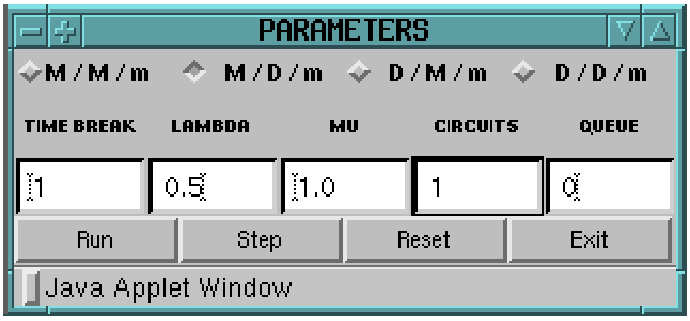


En conclusión, para unas necesidades de tráfico (A), tenemos que poner un número de circuitos (C) tales que tengamos un grado de servicio aceptable.

3.3 Modelo $M/D/C/C$

Este modelo compuesto de C circuitos y sin cola (Q=*0*) se caracteriza por tener las llegadas aleatorias y el servicio determinístico; es decir, las llegadas se producen, como media, cada $1/\lambda$ s y el servicio de cada llamada tarda exactamente $1/\mu$ s.

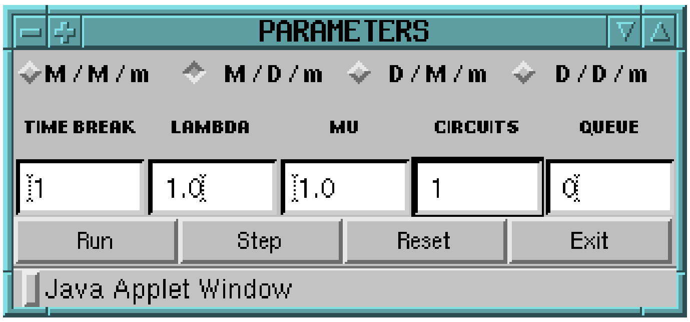
1. Simulad un modelo $M/D/1/1$ con A=*0.5*. Comprobad que llega una llamada cada *2* s como media (unas veces más y otras menos) y que se tarda exactamente *1* s en servirse.



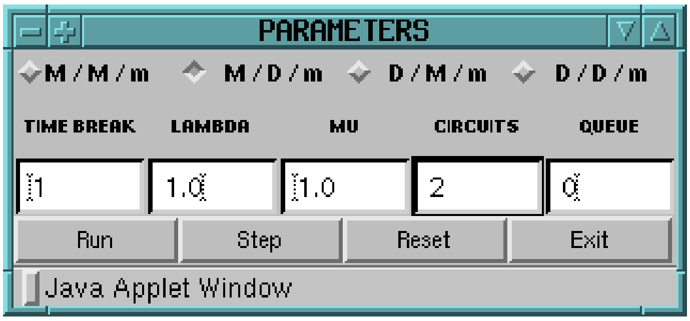
2. ¿Se pierden llamadas? ¿Por qué? Anotad la probabilidad de bloqueo que se obtiene. ¿Creeis que proporciona un grado de servicio aceptable?

   (a) ¿Podríais dar un valor de A y/o C para garantizar que no se pierden llamadas? ¿Por qué?

3. Intuitivamente se ve que si aumentamos A, aumentaremos la probabilidad de pérdida. Simula un modelo $M/D/1/1$ con A=*1* y anotad la probabilidad de bloqueo.



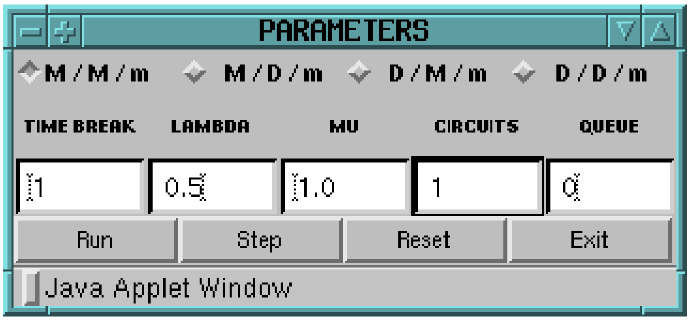
4. También se ve que si aumentamos C, disminuiremos la probabilidad de bloqueo. Simulad un modelo $M/D/2/2$ con A = *1* y anotad la probabilidad de bloqueo.



3.4 Modelo $M/M/C/C$

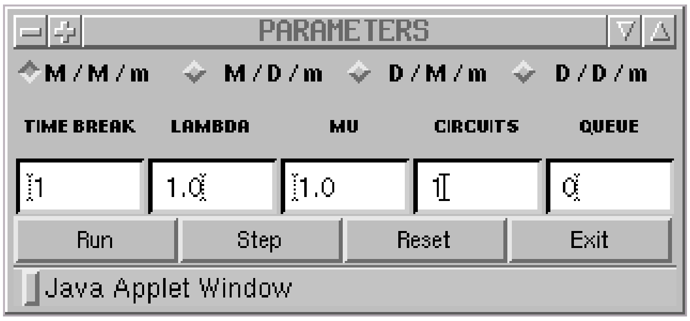
Este modelo compuesto de C circuitos y sin cola (Q=*0*) se caracteriza por tener las llegadas aleatorias y el servicio aleatorio; es decir, las llegadas se producen como media cada $1/\lambda$ s y el servicio de cada llamada tarda una media de $1/\mu$ s.

1. Simulad un modelo *M/M/1/1* con A=*0.5*. Comprobad que cada *2* s, como media, llega una llamada (a veces más y otras menos) y que se tarda una media de *1* s en servirse (unas veces más y otras menos).

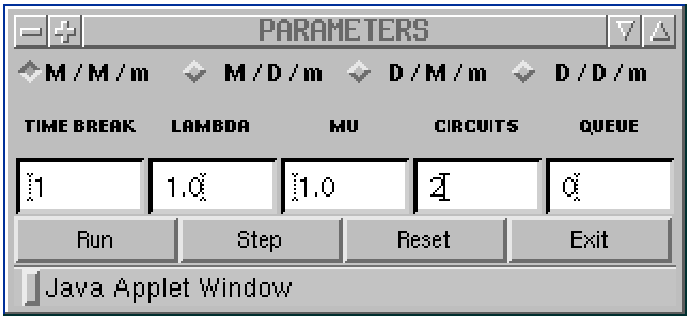


2. Anotad la probabilidad de bloqueo y comparadla con la probabilidad de bloqueo que hay en las tablas (*ErlangB[1,O.5]*).

3. Intuitivamente se ve que si aumentamos A, aumentaremos la probabilidad de pérdida. Simula un modelo *M/M/1/1* con A = *1* y anotad la probabilidad de bloqueo. Calculadla matemáticamente también.



4. También se ve que si aumentamos *C*, disminuiremos la probabilidad de bloqueo. Simulad un modelo *M/M/2/2* y anotad la probabilidad de bloqueo. Calculadla matemáticamente también.



4 Modelos de espera

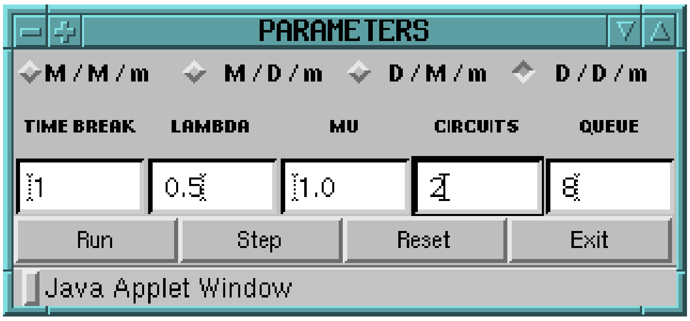
Un modelo se considera de espera cuando una llamada que llega al sistema y no tiene recursos, circuitos, para ser atendida, espera en una cola que puede ser finita o infinita. En el caso de que sea finita, cuando esta cola está llena y llega una llamada, la llamada se pierde. En el caso de que sea infinita, no se pierde ninguna llamada, pues todas esperan su turno en una cola que no tiene límite. Notad que un modelo de bloqueo es un caso particular de modelo de espera con una cola nula ($Q = 0$).

El modelo de espera puro es aquel en el que la cola es infinita, es decir, en el que no se pierden llamadas, pues si llega una llamada al sistema y encuentra todos los circuitos ocupados, espera en una cola a ser atendida.

4.1 Modelo $D/D/C/C{+}Q$

Este modelo compuesto de C circuitos y una cola de Q elementos se caracteriza por tener las llegadas y el servicio determinístico; es decir, las llegadas se producen periódicamente cada $1/\lambda$ s y el servicio de cada llamada tarda exactamente $1/\mu$ s. Un sistema sobredimensionado es aquel en el que hay más circuitos de los que se necesita para dar un grado de servicio (satisfacción del cliente) razonable. Un sistema infradimensionado es aquel en el que hay menos circuitos de los necesarios para dar un grado de servicio aceptable. Querríamos calcular el número de circuitos necesarios para dar un grado de servicio aceptable.

1. Simulad un modelo $D/D/2/10$ con un caso de sobredimensionado. Por ejemplo, que lleguen llamadas cada *2* s y que tarden en servirse *1* s. Esto corresponderá a una $A = \frac{\lambda}{\mu} = \frac{1/\mu}{1/\lambda} = \frac{1}{2} = 0.5$.

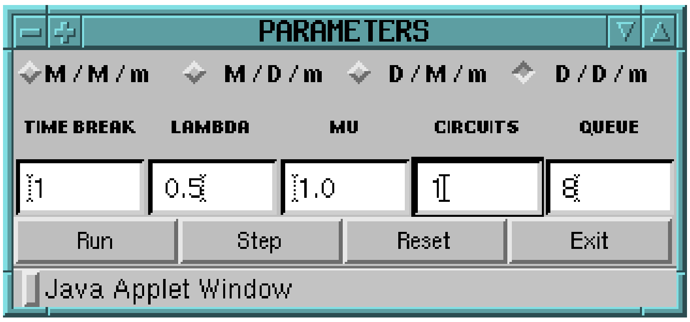


   (a) ¿Cuáles son las probabilidades que se obtienen por simulación de cada estado?

   (b) ¿Qué utilidad tiene la cola en este modelo? ¿Podríamos quitarla? Es decir, ¿tiene el mismo comportamiento el modelo $D/D/2/10$ que el modelo $D/D/2/2$? Si no lo veis claro, simuladlo.

   (c) ¿Es un sistema sobredimensionado?

2. Simulad un modelo $D/D/1/9$ con A=*0.5* . Simplemente reducimos en un circuito.



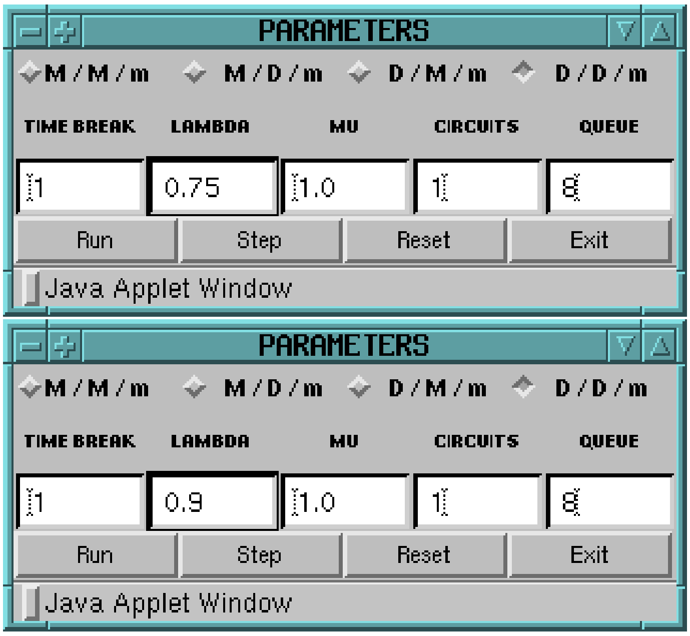
   (a) ¿Cuáles son las probabilidades que se obtienen por simulación de cada estado?

(b) ¿Qué utilidad tiene la cola en este modelo? ¿Podríamos quitarla? Es decir, ¿tiene el mismo comportamiento el modelo *D/D/1/9* que el modelo *D/D/1/1*?

(c) ¿Es un sistema sobredimensionado?

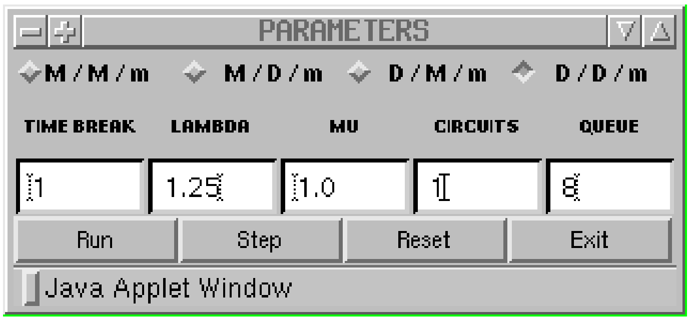
(d) ¿Cuál es el porcentaje de tiempo que el circuito está ocupado? ¿Coincide con algún parámetro del sistema? Relacionadlo con A, P_0, P_1,...

3. Veamos si lo antedicho se cumple. Simulad un modelo *D/D/1/9* con A=*0.75* y A=*0.9*.



(a) ¿Se mantienen las relaciones calculadas anteriormente?

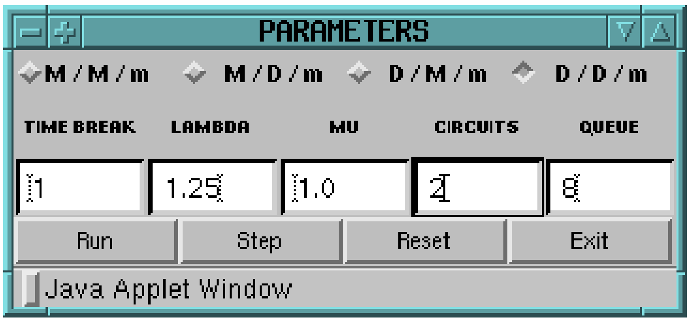
4. Parece que no queramos pasar de A = *1*; hagámoslo. Simulad el modelo *D/D/1/9* con A=*1.25*.



(a) ¿Qué veis en la simulación? ¿Donde se estabiliza el sistema? ¿Cuáles creeis que serían las probabilidades de estado al cabo de un buen rato?

(b) ¿Cuál es ahora el porcentaje de tiempo que el circuito está ocupado? ¿Es un sistema sobredimensionado? ¿Tiene este sistema un grado de servicio aceptable? ¿Cómo se puede aportar de forma más sencilla este grado de servicio?

5. Puede que la solución sea añadir un circuito más. Simulad un modelo *D/D/2/10* con A=*1.25*. ¿Qué porcentaje de tiempo está ocupado *1* circuito, y qué porcentaje de tiempo están ocupados *2* circuitos? Relacionadlos con las probabilidades de estado.



6. ¿Cuál creeis que es la relación que existe entre A y C para que se proporcione un grado de servicio aceptable?

7. En conclusión, si tenemos una A = *3.1*, ¿qué número de circuitos es necesario para dar un grado de servicio aceptable?

4.2 Modelo $D/M/C/C+Q$

Este modelo compuesto de C circuitos y una cola de Q elementos se caracteriza por tener las llegadas determinísticas y el servicio aleatorio; es decir, las llegadas se producen exactamente cada $1/\lambda$ s y el servicio de cada llamada tarda, como media, $1/\mu$ s.

1. Simula un modelo *D/M/1/9* con A=*0.5*. Comprobad que cada *2* s llega una llamada y que se tarda una media de *1* s en servirse.

   (a) ¿Cuál es el porcentaje de ocupación del circuito? ¿Concuerda con las fórmulas anteriores?

   (b) ¿Se utiliza la cola durante la simulación?

   (c) ¿Qué debería pasar para que un elemento fuera a la cola?

2. Simulad un modelo D/M/1/9 con A=1. Explicad qué observais.

4.3 Modelo $M/M/C/C+Q$

Este modelo compuesto de C circuitos y con cola de longitud máxima Q se caracteriza por tener las llegadas aleatorias y el servicio aleatorio; es decir, las llegadas se producen, como media, cada $1/\lambda$ s, y el servicio de cada llamada tarda una media de $1/\mu$ s.

1. Simulad un modelo *M/M/1/10* con A=*0.5* . Comprobad que llega como media una llamada cada 2 s (unas veces más y otras menos) y que se tarda una media de 1 s en servirse (unas veces más y otras menos).

   (a) ¿Se podría decir que corresponde a un modelo de espera puro? ¿Por qué?

   (b) Haced una tabla con los valores teóricos, considerando un modelo de espera puro, y con los simulados de los *10* primeros estados. ¿Se parecen?

   (c) Calculad de forma teórica y simulada la probabilidad de pérdida y la probabilidad de demora, así como el tiempo medio de espera en cola. ¿Se parecen? ¿A qué conclusión podríais llegar?

2. Consideremos otro caso. Simula un modelo *M/M/1/2* con A=*1* .

   (a) ¿Se podría decir que corresponde a un modelo de espera puro? ¿Y de pérdida puro? ¿Por qué?

   (b) Haced una tabla con los valores teóricos y simulados de los estados. ¿Se parecen?

   (c) Calculad de forma teórica y simulada la probabilidad de pérdida y la probabilidad de demora, así como el tiempo medio de espera en cola. ¿Se parecen? ¿A qué conclusión podríais llegar?

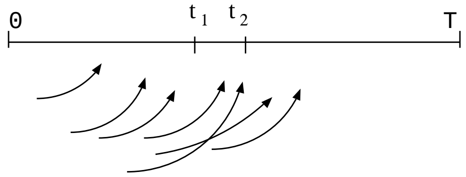
   (d) Parece que si aumentamos C, disminuiremos la probabilidad de pérdida. Comprobadlo simulando el sistema *M/M/2/3* con A = *1* y comparadlo con el valor teórico.

(e) Parece también que si aumentamos A, aumentamos la probabilidad de pérdida. Comprobadlo simulando el sistema *M/M/2/3* con $A = 2$ y comparadlo con el valor teórico.

5 Apéndice

5.1 Modelo de generación de llamadas (Procesos de Poisson)

Comenzaremos solucionando un problema paralelo. Supongamos en primer lugar que lanzamos al azar n puntos sobre un intervalo *(0,T)*. ¿Qué probabilidad hay de que k de estos n puntos caigan en un subintervalo $t = (t_1, t_2)$ de *(0, T)*?



La probabilidad de que, si se lanza un punto, caiga en t, es

$$p = \frac{t_2 - t_1}{T} = \frac{t}{T}$$

Si el experimento se repite n veces, la probabilidad que k de estos n puntos caigan en un subintervalo $t = (t_1, t_2)$ de *(0,T)* es:

$$P\{k\, de\, los\, n\, en\, t\} = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}$$

Si suponemos que $n \gg 1$ y $T \gg t$, pero $n/T = \lambda$ se mantiene constante, entonces se podría demostrar que

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-np} \frac{np^k}{k!}$$

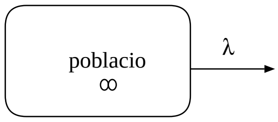
y como $np = \lambda t$,

$$P\{k\, en\, t\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

Si en vez de hablar de puntos, hablamos de generación de llamadas, entonces la probabilidad de que se generen k llamadas en un intervalo t, por parte de una población infinita que genera una tasa de llamadas λ, es

$$P\{k\,en\,t\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

Se dice que la generación de llamadas por parte de esta población infinita viene modelada por un proceso estocástico de Poisson.

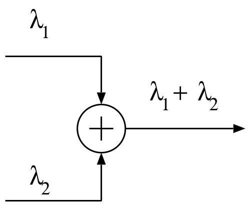


Para ver el significado físico de λ, calcularemos cuántas llamadas se generan como media en un tiempo de observación T_{ob}.

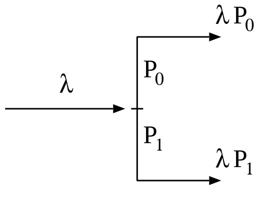
$$\bar{N} = \sum_{i=0}^{\infty} iP\{ienT_{ob}\} = \sum_{i=0}^{\infty} ie^{-\lambda T_{ob}}\frac{(\lambda T_{ob})^i}{i!} = e^{-\lambda T_{ob}}\lambda T_{ob}\sum_{i=1}^{\infty}\frac{(\lambda T_{ob})^{i-1}}{(i-1)!} = \lambda T_{ob}$$

Propiedades de los procesos de Poisson

Aditiva La suma de procesos de Poisson es otro proceso de Poisson, cuya tasa es la suma de las tasas de los procesos que se suman.



Multiplicativa La división de un proceso de Poisson a través de un conjunto de probabilidades P_i son procesos de Poisson cuyas tasas son, respectivamente, la tasa del proceso inicial multiplicada por la probabilidad correspondiente P_i.



Caracterización del tiempo entre llegadas consecutivas en un proceso de Poisson

El próximo paso será caracterizar estadísticamente el tiempo entre llegadas consecutivas en un proceso de Poisson, es decir, calcularemos su función de densidad de probabilidad.

Sea t_0 el instante en que se produce una llegada y sea t_1 el instante en que se produce la siguiente llegada.

Definimos la variable aleatoria § como la distancia $t_1 - t_0$. Queremos calcular la función de distribución de §, $F(x)$.

$F(x)$ es la probabilidad de que $\S \leq x$, pero $\S \leq x$ quiere decir que hay al menos *1* llegada entre t_0 i $t_0 + x$, y este valor lo conocemos porque, debido a ser un proceso de Poisson,

$$P\{0\, en\, x\} = e^{-\lambda x}\frac{(\lambda x)^0}{0!} = e^{-\lambda x}$$

Por tanto,

$$F(x) = 1 - e^{-\lambda x}$$

Y la función de densidad de probabilidad será:

$$f(x) = \frac{d}{dx}F(x) = \lambda e^{-\lambda x}u(x)$$

Es decir, el tiempo entre llegadas consecutivas en un proceso de Poisson es una variable aleatoria exponencial.

Tiempo medio entre llegadas consecutivas Sólo es necesario calcular la esperanza de esta variable aleatoria.

$$E(x) = \int_0^\infty f(x)dx = \int_0^\infty \lambda e^{-\lambda x}dx = \frac{1}{\lambda}$$

5.2 Modelo de servicio de llamadas (Servicio exponencial)

Debido a medidas experimentales, se ha comprobado que el comportamiento del servicio de llamadas se puede modelar como una variable aleatoria exponencial, de tasa μ. Sólo alrededor del origen hay discrepancia.

$$f(x) = \mu e^{-\mu x}u(x)$$

Significado físico de μ

El tiempo medio de duración del servicio (la duración de una llamada) se puede calcular a partir de este modelo mediante su esperanza.

$$E(x) = \int_0^\infty f(x)dx = \int_0^\infty \mu e^{-\mu x} dx = \frac{1}{\mu}$$

Entonces, el inverso de μ es el tiempo medio de servicio.

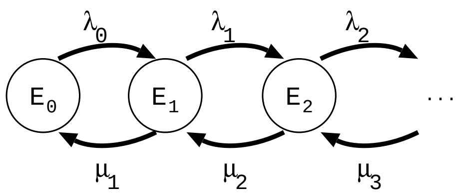
5.3 Procesos de Markov de nacimiento y muerte

El próximo paso será caracterizar estocásticamente el modelo de pérdidas y el modelo de espera, por lo que utilizaremos la parte de la teoría de colas correspondiente a los procesos de Markov.

Un sistema se puede modelar a través de un proceso de Markov si éste presenta un número finito o infinito de estados (E_i) por los cuales va pasando aleatoriamente, y tiene memoria nula, es decir, no depende del pasado. Cada estado (E_i) tiene asociado una probabilidad de ocurrencia (P_i), que querremos calcular porque nos caracterizará estocásticamente el sistema.

Un proceso de Markov se dice que es de nacimiento y muerte si los saltos desde un estado (E_i) sólo pueden darse hacia el estado inmediatamente superior (E_{i+1}) o el estado inmediatamente inferior (E_{i-1}).

Se puede dibujar un diagrama de estados del sistema de la siguiente forma



donde λ_i es la tasa de probabilidad de nacimiento en el estado E_i, y μ_i es la tasa de probabilidad de muerte en el estado E_i.

En régimen permanente hay un balance de flujos entrantes y salientes de cualquier superficie del diagrama de estados. El flujo es el producto de la tasa de probabilidad y la probabilidad de ocurrencia.

En la superficie S_0,

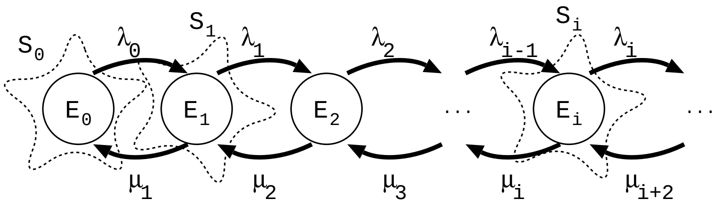
$$\lambda_0 P_0 = \mu_1 P_1$$

En la superficie S_1,

$$(\lambda_1 + \mu_1)P_1 = \lambda_0 P_0 + \mu_2 P_2$$

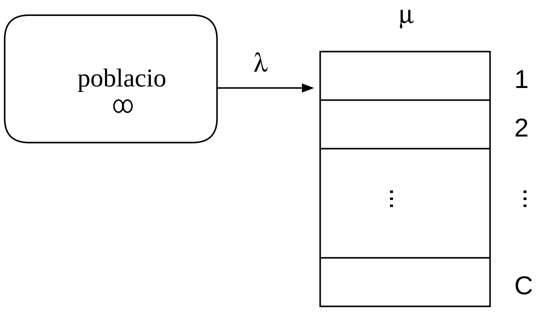
Y en la superficie S_i,

$$(\lambda_i + \mu_i)P_i = \lambda_{i-1}P_{i-1} + \mu_{i+1}P_{i+1}$$



5.4 Modelo de pérdida puro

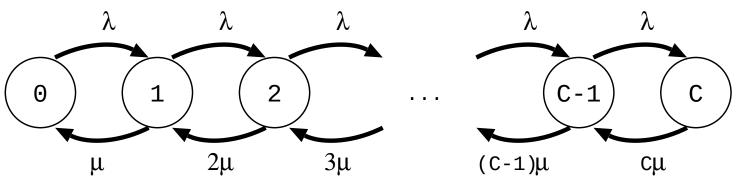
Consideremos una población infinita que llega a un sistema compuesto por C circuitos o canales. La tasa de generación de llamadas es λ y la tasa de servicio de llamadas es μ.



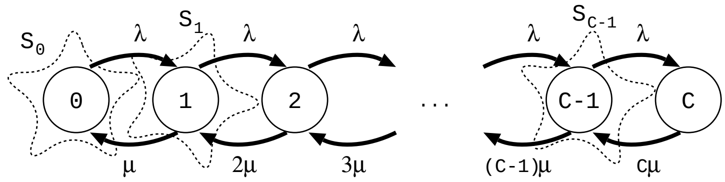
Consideremos un modelo de pérdidas, es decir, una llamada que llegue cuando el sistema esté saturado, con los C circuitos ocupados, se perderá y no será servida. La probabilidad de encontrar el sistema en este estado será de gran interés.

Un sistema con estas características es un proceso de Markov de nacimiento y muerto, con $\lambda_i = \lambda$ y $\mu_i = i\mu$.

Por tanto, el diagrama de estados será



Vamos a calcular las probabilidades de ocurrencia P_i. Igualaremos flujos por diferentes superficies.



En la superficie S_0,

$$\lambda P_0 = \mu P_1$$

y si $A = \lambda/\mu$,

$$P_1 = \frac{\lambda}{\mu} P_0 = A P_0$$

En la superficie S_1,

$$(\lambda + \mu) P_1 = \lambda P_0 + 2\mu P_2$$

y desarrollando,

$$P_2 = \frac{A}{2} P_1 = \frac{A^2}{2!} P_0$$

Y en la superfície S_{C-1},

$$(\lambda + (C-1)\mu) P_{C-1} = \lambda P_{C-2} + C\mu P_C$$

y desarrollando,

$$P_C = \frac{A}{C} P_{C-1} = \frac{A^C}{C!} P_0$$

Tenemos, pues, P_n en función de P_0, que nos falta calcular. Para hacerlo, aprovechamos que se ha de cumplir que

$$\sum_{i=0}^{C} P_i = 1$$

entonces,

$$\sum_{i=0}^{C} P_i = P_0 \left(1 + A + \frac{A^2}{2!} + \cdots + \frac{A^C}{C!}\right) = 1$$

y se obtiene,

$$P_0 = \frac{1}{1 + A + \frac{A^2}{2!} + \cdots + \frac{A^C}{C!}} = \frac{1}{\sum_{i=0}^{C} \frac{A^i}{i!}}$$

Y, en conclusión,

$$P_n = \frac{\frac{A^n}{n!}}{\sum_{i=0}^{C} \frac{A^i}{i!}} \text{ }, per\ 0 \leq n \leq C$$

Para medir el grado de servicio, calcularemos la probabilidad de bloqueo (PB). En el caso del modelo de pérdida, corresponde a la probabilidad de pérdida (PP).

Como ya hemos apuntado antes, ésta será la probabilidad de encontrar el sistema en estado de saturación, que es equivalente a

- Encontrarlo con los C circuitos ocupados
- Encontrarlo en el estado E_C,
- Si una llamada llega al sistema en este estado, se pierde y no se sirve.

Esta probabilidad es P_C, y es, por definición, la probabilidad de pérdida (PP).

$$P_C = PP = \frac{\frac{A^C}{C!}}{\sum_{i=0}^{C} \frac{A^i}{i!}}$$

Esta es un función de A y C y recibe el nombre de d'$Erlang_B(A, C)$.

$$Erlang_B(A, C) = P_C = PP = \frac{\frac{A^C}{C!}}{\sum_{i=0}^{C} \frac{A^i}{i!}}$$

De aquí podemos obtener los diferentes tráficos del sistema.

Vamos a calcular el TC. Para esto, como conocemos las probabilidades de cada estado y sabemos las llamadas que se está cursando en cada estado, el TC será el promedio de las llamadas cursadas a cada estado por la probabilidad de cada estado.

$$TC = 0P_0 + 1P_1 + \cdots + CP_C = \sum_{i=1}^{C} iP_i = \sum_{i=1}^{C} iP_0 \frac{A^i}{i!} =$$

$$= AP_0 \sum_{i=1}^{C} \frac{A^{i-1}}{(i-1)!} = AP_0 \sum_{j=0}^{C-1} \frac{A^j}{j!} = A\left(1 - Erlang_B(A, C)\right) = A\left(1 - PP\right)$$

Para calcular el TO, lo que se hace es calcular el TC de un sistema con los mismos parámetros λ y μ e infinitos circuitos. Si lo hacéis, hallaréis que TO=A.

A partir de aquí, se puede calcular el tráfico perdido TP,

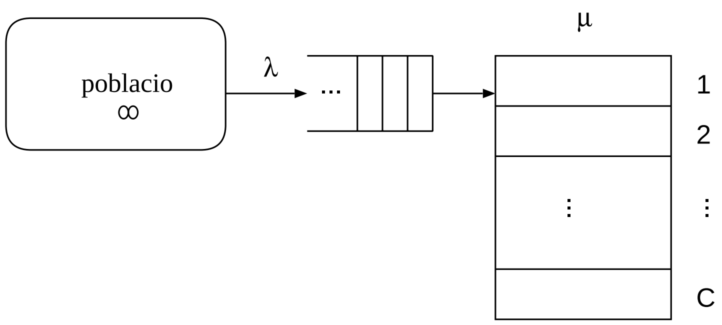
$$TP = TO - TC = A - A\left(1 - PP\right) = APP = AErlang_B(A, C)$$

De aquí se puede extraer que el TC es la relación entre la tasa de llamadas cursadas y la tasa de llamadas servidas

$$TC = TO(1 - PP) = \frac{\lambda}{\mu}(1 - PP) = \frac{\lambda_c}{\mu}$$

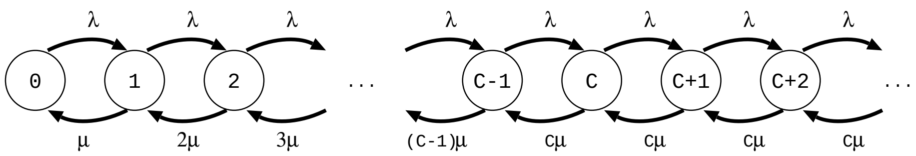
5.5 Modelo de espera puro

Consideremos una población infinita que llega a un sistema compuesto por C circuitos o canales. La tasa de generación de llamada es λ y la tasa de servicio de llamadas es μ.

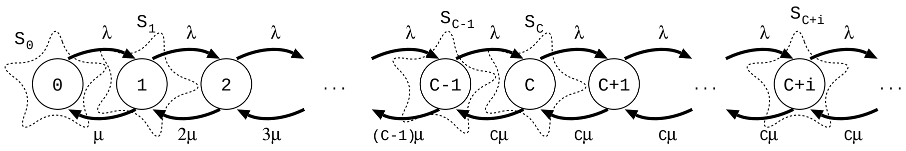


Consideremos un modelo de espera, es decir, una llamada que llegue cuando el sistema esté saturado, con los C circuitos ocupados, esperará a que uno de los circuitos se vacíe para ser servida; en caso de que lleguen más llamadas y el sistema esté lleno, éstas se pondrán en una cola de capacidad infinita.

Un sistema con estas características es un proceso de Markov de nacimiento y muerte, con $\lambda_i = \lambda$ y $\mu_i = i\mu$ para $0 \leq i \leq C$ y $\mu_i = C\mu$ para $i \geq C$.



Vamos a calcular las probabilidades de ocurrencia P_i. Igualaremos flujos por diferentes superficies.



En la superficie S_0,

$$\lambda P_0 = \mu P_1$$

y si $A = \lambda/\mu$,

$$P_1 = \frac{\lambda}{\mu} P_0 = A P_0$$

En la superficie S_1,

$$(\lambda + \mu) P_1 = \lambda P_0 + 2\mu P_2$$

y desarrollando,

$$P_2 = \frac{A}{2} P_1 = \frac{A^2}{2!} P_0$$

En la superficie S_{C-1},

$$(\lambda + (C-1)\mu) P_{C-1} = \lambda P_{C-2} + C\mu P_C$$

y desarrollando,

$$P_C = \frac{A}{C} P_{C-1} = \frac{A^C}{C!} P_0$$

En la superficie S_C,

$$(\lambda + C\mu) P_C = \lambda P_{C-1} + C\mu P_{C+1}$$

y desarrollando,

$$P_{C+1} = \frac{A}{C} P_C = \frac{A^C}{C!} \frac{A}{C} P_0$$

En la superficie S_{C+i},

$$(\lambda + C\mu)P_{C+i} = \lambda P_{C+i-1} + C\mu P_{C+i+1}$$

y desarrollando,

$$P_{C+i+1} = \frac{A}{C}P_{C+i} = \frac{A^C}{C!}\left(\frac{A}{C}\right)^{i+1} P_0$$

Tenemos, pues, P_n en función de P_0, que nos falta calcular. Para hacerlo, aprovechamos que se ha de cumplir que

$$\sum_{i=0}^{C} P_i = 1$$

Entonces,

$$\sum_{i=0}^{C} P_i = P_0\left(1 + A + \frac{A^2}{2!} + \cdots + \frac{A^C}{C!}\left(1 + \frac{A}{C} + \left(\frac{A}{C}\right)^2 + \cdots\right)\right) = 1$$

y se obtiene

$$P_0 = \frac{1}{1 + A + \frac{A^2}{2!} + \cdots + \frac{A^C}{C!}\left(1 + \frac{A}{C} + \left(\frac{A}{C}\right)^2 + \cdots\right)} = \frac{1}{\sum_{i=0}^{C}\frac{A^i}{i!} + \frac{A^C}{C!}\frac{A}{C-A}}$$

$$= \frac{1}{\sum_{i=0}^{C-1}\frac{A^i}{i!} + \frac{A^C}{C!}\frac{C}{C-A}}$$

Y, en conclusión,

$$P_n = \begin{cases} \dfrac{\frac{A^n}{n!}}{\sum_{i=0}^{C}\frac{A^i}{i!} + \frac{A^C}{C!}\frac{A}{C-A}}, para\, 0 \le n \le C \\ \dfrac{\frac{A^C}{C!}\left(\frac{A}{C}\right)^{n-C}}{\sum_{i=0}^{C}\frac{A^i}{i!} + \frac{A^C}{C!}\frac{A}{C-A}}, para\, C \le n \end{cases}$$

Para medir el grado de servicio, calcularemos la probabilidad de bloqueo (PB). En el caso del modelo de espera, corresponde a la probabilidad de demora (PD). De la misma manera que antes, nos interesará calcular la probabilidad de encontrar el sistema en estado de saturación, que es equivalente a

- Encontrarlo con los C circuitos ocupados

- Encontrarlo en los estados E_C, E_{C+1}, E_{C+2}, ...
- Si una llamada llega al sistema en este estado, se pone a la cola y espera a que se sirvan todas las llamadas que estén por delante en la cola, y no se pierde.

Esta probabilidad, por definición, es la probabilidad de demora (PD).

$$PD = P_C + P_{C+1} + P_{C+2} + \cdots = \sum_{n=C}^{\infty} \frac{A^C}{C!}\left(\frac{A}{C}\right)^{n-C} = P_0 \frac{A^C}{C!}\frac{C}{C-A}$$

Esta es la función de A y C y recibe el nombre de $Erlang_C(A, C)$.

$$Erlang_C(A,C) = P_C + P_{C+1} + P_{C+2} + \cdots = PD = P_0 \frac{A^C}{C!}\frac{C}{C-A}$$

De aquí podemos obtener los tráficos del sistema.

De la misma manera que para el modelo de pérdidas, se puede ver que $TO{=}A$.

Para calcular el TC haremos como en el modelo de pérdida. Para esto, como conocemos las probabilidades de cada estado y sabemos las llamadas que se están cursando en cada estado, el TC será el promedio de las llamadas cursadas a cada estado por la probabilidad de cada estado.

$$TC = 0P_0 + 1P_1 + \cdots + CP_C + CP_{C+1} + CP_{C+2} + \cdots =$$

$$= \sum_{i=1}^{C} iP_i + \sum_{i=C+1}^{\infty} CP_i = \sum_{i=1}^{C} iP_0\frac{A^i}{i!} + \sum_{i=C+1}^{\infty} CP_0\frac{A^C}{C!}\left(\frac{A}{C}\right)^{i-C} =$$

$$= AP_0 \sum_{i=1}^{C} \frac{A^{i-1}}{(i-1)!} + CP_0\frac{A^C}{C!}\sum_{j=1}^{\infty}\left(\frac{A}{C}\right)^{j} =$$

$$= AP_0\left(\sum_{i=1}^{C} \frac{A^{i-1}}{(i-1)!} + \frac{A^C}{C!}\frac{C}{C-A}\right) = A\frac{P_0}{P_0} = A$$

De aquí se puede calcular el tráfico perdido TP,

$$TP = TO - TC = A - A = A\,PB = 0$$

Estos resultados los podríamos haber obtenido inicialmente, puesto que en un modelo de espera, todo el tráfico que se ofrece se cursa. Por tanto, $TO = TC = A$.

Igualmente, no se pierde ningún tráfico, $TP = 0$.

También podríamos definir un tráfico demorado (TD) como la parte del tráfico ofrecido que se demora, es decir, que llega al sistema y ha de esperar.

$$TD = TO\,PD = A Erlang_C(A, C)$$

5.6 Relación de Little

La relación de Little relaciona el tiempo medio de permanencia en un sistema con la tasa de entrada al sistema y el número medio de elementos que hay dentro.

$$\bar{N} = \lambda \bar{W}$$

La relación de Little sirve para calcular el otro gran parámetro de medida del grado de servicio en modelos de espera: el tiempo medio de espera en cola.

En el modelo de pérdidas, el número medio de llamadas al sistema coincide con el número medio de circuitos ocupados, es decir, TC. Por tanto,

$$\bar{W} = \frac{\bar{N}}{\lambda} = \frac{A\,(1 - Erlang_B(A, C))}{\lambda} = \frac{1}{\mu}\,(1 - Erlang_B(A, C))$$

En el modelo de espera es un poco más complejo, porque hemos de calcular primero, $\bar{N}$ el número medio de llamadas al sistema, que será la suma del número medio de llamadas en cola ($\bar{N}_Q$) y el número medio de llamadas en los circuitos ($\bar{N}_C$) (o sea, el número medio de circuitos ocupados).

$$\bar{N} = \bar{N}_Q + \bar{N}_C = \bar{N}_Q + TC = \bar{N}_Q + A$$

Hemos de calcular, por tanto, N_Q.

$$\bar{N}_Q = 0P_0 + 0P_1 + \cdots + 0P_C + 1P_{C+1} + 2P_{C+2} + \cdots = \sum_{i=C+1}^{\infty} (i - C)P_i =$$

$$\sum_{i=C+1}^{\infty} (i - C)P_0 \frac{A^C}{C!} \left(\frac{A}{C}\right)^{i-C} = P_0 \frac{A^C}{C!} \sum_{j=1}^{\infty} j \left(\frac{A}{C}\right)^j$$

$$= P_0 \frac{A^C}{C!} \frac{A}{C} \frac{1}{\left(1 - \frac{A}{C}\right)^2} = Erlang_C(A, C) \frac{A}{C - A}$$

Entonces,

$$\bar{N} = \bar{N}_Q + A = Erlang_C(A,C)\frac{A}{C-A} + A = A\left(\frac{Erlang_C(A,C)}{C-A} + 1\right)$$

Y, aplicando la relación de Little,

$$\bar{W} = \frac{\bar{N}}{\lambda} = \frac{1}{\mu}\left(\frac{Erlang_C(A,C)}{C-A} + 1\right)$$