Ejercicio 2-12

Para dar servicio de consultas a base de datos en una red se dispone de dos ordenadores. El primero, que llamaremos A, tiene un tiempo medio de servicio de 100 ms., pero por construcción sólo admite 5 peticiones en espera de servicio. El segundo, que llamaremos B, tiene un tiempo medio de servicio de 500 ms., pero admite un número ilimitado de peticiones en cola de espera. Los tiempos de servicio de ambos servidores se pueden considerar distribuidos exponencialmente. La arquitectura que se decide para dar el servicio coloca los dos servidores en paralelo. Las peticiones serán procesadas por el servidor A, excepto en el caso de que éste se encuentre al máximo de su capacidad (5 peticiones en cola mas una en servicio), en cuyo caso la petición será pasada al servidor B. La llegada de consultas al sistema se puede considerar como un proceso de Poisson, con una tasa de llegadas de 9 peticiones por segundo. Existe un número muy grande de clientes, de modo que el número de peticiones pendientes de servicio no afecta al ritmo de llegada de nuevas peticiones.

1. *Calcular el tiempo medio de estancia en el sistema de las peticiones que son procesadas por el servidor A.*
2. *Calcular el tiempo medio de estancia en el sistema total compuesto por los servidores A y B.*

- La arquitectura descrita en el enunciado puede verse en la Figura 1.
- Dado que la entrada al sistema es de Poisson (con tasa de llegadas λ), y nos encontramos ante una bifurcación aleatoria de un proceso de Poisson, la entrada a los subsistemas A (con tasa de llegadas λ_A) y B (con tasa de llegadas λ_B) también será de Poisson.
- Analizando cada subsistema por separado tenemos:
- <u>**Subsistema A**</u>: se trata de un **M/M/1/6** puesto que:
  - Llegadas de Poisson con tasa λ_A (resultado de una bifurcación aleatoria de un proceso de Poisson).
  - Tiempo de servicio exponencial de media $T_S^A = 0.1$ segundos.
  - 1 servidor.
  - Tamaño de cola finito de tamaño 5 $\rightarrow$ En el sistema puede haber a lo sumo 6 clientes.
  - Número muy grande de clientes, de modo que el número de peticiones pendientes de servicio no afecta al ritmo de llegada de nuevas peticiones.
- <u>**Subsistema B**</u>: se trata de un **M/M/1** puesto que:
  - Llegadas de Poisson con tasa λ_B (resultado de una bifurcación aleatoria de un proceso de Poisson).
  - Tiempo de servicio exponencial de media $T_S^B = 0.5$ segundos.
  - 1 servidor.
  - Tamaño de cola infinito.

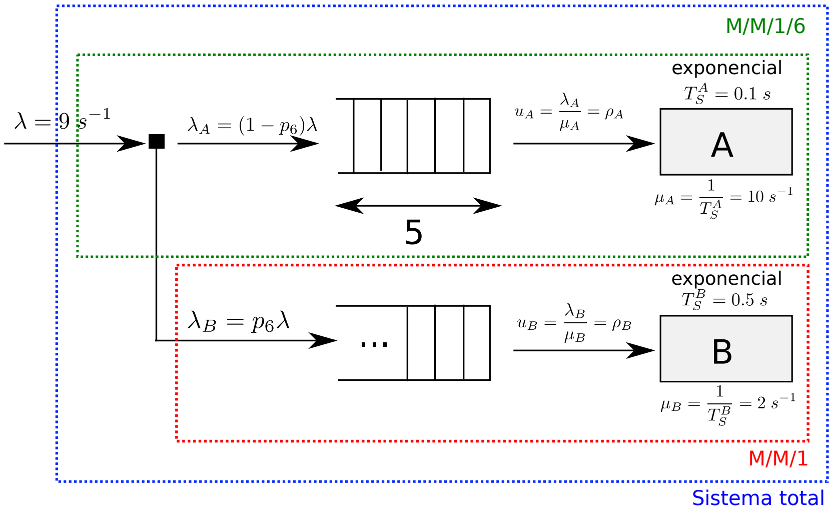

Figure 1: Red de colas definida por un modelo M/M/1/6 y un modelo M/M/1.

- Número muy grande de clientes, de modo que el número de peticiones pendientes de servicio no afecta al ritmo de llegada de nuevas peticiones.

Es importante tener en cuenta que la probabilidad que regula si una petición pasa al sistema A o B viene determinada por la probabilidad de rechazo de una petición en el sistema A, puesto que el enunciado nos dice que todas las peticiones van a A salvo que A se encuentre lleno, en cuyo caso pasan a B. La probabilidad de que una petición sea rechazada por el sistema A será la probabilidad de que en A ya haya 6 clientes y, por tanto, el sistema esté lleno. Está probabilidad viene dada por p_6 para el modelo M/M/1/6. Es decir, una petición pasará a B con probabilidad p_6 y será atendida por A con probabilidad $(1 - p_6)$.

1 Tiempo medio de estancia en el sistema de las peticiones que son procesadas por el servidor A

Nos están pidiendo W para el sistema A (M/M/1/6); la denotaremos por W_A. Para calcular W_A podemos seguir los siguientes pasos:

1. Calcular el número medio de clientes en el sistema L_A.
2. Obtener W_A a partir de L_A usando el Teorema de Little sobre el sistema A (recuadro verde de la figura).

$$
\begin{aligned}
L_A &= \frac{\frac{\lambda}{\mu_A}}{1-\frac{\lambda}{\mu_A}}\left[\frac{1-(K+1)\left(\frac{\lambda}{\mu_A}\right)^K+K\left(\frac{\lambda}{\mu_A}\right)^{K+1}}{1-\left(\frac{\lambda}{\mu_A}\right)^{K+1}}\right]= \\
&= \frac{\frac{9}{10}}{1-\frac{9}{10}}\left[\frac{1-7\left(\frac{9}{10}\right)^6+6\left(\frac{9}{10}\right)^7}{1-\left(\frac{9}{10}\right)^7}\right]\approx 2.588\ \textit{clientes}
\end{aligned}
$$

Ahora calculamos el tiempo medio de estancia en el sistema (latencia) en función de L y la tasa **efectiva** de llegadas[1]. Para ello, vamos a calcular primero la tasa efectiva de llegadas λ_A. Para ello necesitamos calcular primero la probabilidad de rechazo.

Como ya se mencionó antes, la probabilidad de rechazar una solicitud viene dada por la probabilidad de que haya actualmente 6 clientes en el sistema (1 siendo servido y 5 en cola). Por tanto, la probabilidad de rechazo viene dada por p_6. Podemos calcularla de acuerdo a la expresión de la distribución estacionaria de Markov para el modelo M/M/1/K:

$$p_6 = p_0 \cdot \left(\frac{\lambda}{\mu_A}\right)^6$$

Conocemos λ y μ_A pero necesitamos calcular p_0. Usamos la expresión para p_0 derivada del segundo axioma de probabilidad:

$$p_0 = \frac{1-\frac{\lambda}{\mu_A}}{1-\left(\frac{\lambda}{\mu_A}\right)^{K+1}} = \frac{1-\frac{9}{10}}{1-\left(\frac{9}{10}\right)^7} \approx 0.192$$

Con este valor de p_0 podemos calcular p_6 como sigue:

$$p_6 = p_0 \cdot \left(\frac{\lambda}{\mu_A}\right)^6 = 0.192 \cdot \left(\frac{9}{10}\right)^6 \approx 0.102$$

Entonces la tasa efectiva de llegadas λ_A será $\lambda_A = \lambda \cdot (1-p_6) = 9 \cdot (1-0.102) = 8.082\ s^{-1}$, que es menor que la tasa de servicio $\mu_A = 10\ s^{-1}$ y, por tanto, existirá el estado estacionario.

Ahora sí podemos calcular el tiempo de estancia en el sistema A, W_A, aplicando el teorema de Little sobre el sistema A:

$$W_A = \frac{L_A}{\lambda_A} = \frac{2.588}{8.082} \approx 0.320\ \textit{segundos}$$

[1]IMPORTANTE: el Teorema de Little se calcula sobre la tasa efectiva de llegadas del sistema bajo estudio. En este caso, estamos trabajando sobre el sistema A - recuadro verde- y la tasa efectiva de llegadas a ese sistema es λ_A.

2 Tiempo medio de estancia en el sistema total compuesto por los servidores A y B.

El tiempo medio en el sistema total (sistema compuesto por los servidores A y B) podemos calcularlo de dos formas:

1. Calcular el tiempo medio de estancia en el sistema de las peticiones procesadas por A (W_A) y el tiempo medio de estancia en el sistema de las peticiones procesadas por B (W_B) y obtener la media del tiempo de estancia en el sistema total como la media ponderada de A y B teniendo en cuenta la probabilidad de que una petición sea procesada por uno u otro sistema.
2. Aplicar el Teorema de Little sobre el sistema total, obteniendo para ello el número total de clientes en el sistema total como la suma del número de clientes en el sistema A (M/M/1/6) y en el sistema B (M/M/1).

2.1 Opción 1: Media ponderada de los tiempos de estancia en los sistemas A y B.

El tiempo medio de estancia en el sistema total puede calcularse como:

$$W = (1 - p_6)W_A + p_6 W_B$$

p_6 y W_A se han calculado en el primer apartado. Para calcular W_B se aplican las ecuaciones del modelo M/M/1 para obtener el número medio de clientes en el sistema B, L_B, y luego se aplica el teorema de Little sobre el sistema B (recuadro rojo de la figura) para obtener $W_B = \frac{L_B}{\lambda_B}$.

$$L_B = \frac{\rho_B}{1 - \rho_B} = \frac{\frac{\lambda_B}{\mu_B}}{1 - \frac{\lambda_B}{\mu_B}} = \frac{\lambda_B}{\mu_B - \lambda_B},$$

donde λ_B viene dado por la probabilidad de que el sistema A haya rechazado la petición: $\lambda_B = p_6\lambda$. Sustituyendo: $\lambda_B = 0.102 \cdot 9 = 0.918\ s^{-1}$. Entonces:

$$L_B = \frac{\lambda_B}{\mu_B - \lambda_B} = \frac{0.918}{2 - 0.918} \approx 0.848\ clientes$$

Aplicando ahora el Teorema de Little sobre el sistema B:

$$W_B = \frac{L_B}{\lambda_B} = \frac{0.848}{0.918} \approx 0.924\ s.$$

Entonces, el tiempo medio de estancia en el sistema total será:

$$W = (1 - p_6)W_A + p_6 W_B = (1 - 0.102) \cdot 0.320 + 0.102 \cdot 0.924 \approx 0.382\ s$$

2.2 Opción 2: Teorema de Little sobre el sistema total.

La otra opción para obtener el tiempo medio de estancia en el sistema total es aplicar el Teorema de Little sobre todo el sistema (recuadro azul de la figura); es decir:

$$W = \frac{L}{\lambda}.$$

L es el número medio de clientes en el sistema total que calcularemos como la suma del número de clientes en A y en B. λ es la tasa efectiva de llegadas al sistema sobre el que estamos aplicando el Teorema de Little, que en este caso es el sistema global y, por tanto, la tasa efectiva de llegadas al sistema total es $\lambda = 9$ peticiones/s. El número de clientes en el sistema A, L_A, y el número de clientes en el sistema B, L_B, ya los hemos calculado anteriormente, así que solo hay que sustituir en el Teorema de Little:

$$W = \frac{L}{\lambda} = \frac{L_A + L_B}{\lambda} = \frac{2.588 + 0.848}{9} \approx 0.382\ s$$