III Congreso Matemática Aplicada
XIII C.E.D.Y.A. Madrid, 1993

Santana, A. Hernández, C. y Almeida, A.

POLÍTICA ÓPTIMA DE SERVICIO EN UN SISTEMA DE COLAS CON PROCESADOR UNICO, DOS CLASES DE CLIENTES Y TRÁFICO PESADO.

1.- Introducción.

Se considera un sistema de colas en que un único procesador atiende a dos distintas clases de clientes. Las llegadas y tiempos de servicio de ambas clases son independientes. En cada clase i (i=1,2), las llegadas se producen de acuerdo con un proceso de Poisson homogéneo de parámetro λ_i, y los tiempos de servicio son independientes e idénticamente distribuidos, con distribución exponencial de parámetro μ_i. Se asume que las llegadas de los clientes del tipo 1 son más frecuentes que las del tipo 2 ($\lambda_1 > \lambda_2$), así como que sus tiempos de servicio son, en media, más cortos ($\mu_1 > \mu_2$). Cada clase de cliente accede a un buffer de espera de capacidad finita N_i, de tal forma que las llegadas producidas cuando el buffer está lleno, resultan rechazadas. Asimismo, hay tráfico pesado ($\rho_i=\lambda_i / \mu_i$, próximo a 1, i=1,2, $\rho_1 + \rho_2 > 1$), por lo que el equilibrio del sistema se alcanza muy deprisa.

Cada cliente del tipo i reporta al sistema un beneficio β_i, siendo $\beta_2 > \beta_1$; cada cliente rechazado produce un coste γ_i, $(\gamma_2 > \gamma_1)$; y mantener un cliente en cola supone un costo δ_i $(\delta_2 > \delta_1)$ por cada unidad de tiempo de espera.

El interés primordial del sistema es atender a los clientes tipo 2; ahora bien, de atender sólo a estos clientes, por ser $\rho_2 < 1$ con seguridad se producirán periodos de desocupación; la política óptima de gestión es entonces la que aprovecha estos periodos para atender a los clientes del tipo 1. Obviamente, si se espera a que no haya ningún cliente 2 en el sistema para admitir a los clientes 1, seguirá habiendo un periodo de desocupación; así pues parece razonable permitir el acceso de los clientes 1 a su buffer antes de que se vacíe la cola 2; pero no mucho antes, porque se les podría hacer esperar demasiado. Con este objetivo se diseña la siguiente política:

Etapa 1. Permitir el acceso a los clientes de ambas clases, atendiendo exclusivamente a los del tipo 1, hasta que haya n_1 clientes del tipo 2 en cola.

Etapa 2. Bloquear el acceso a los clientes de la clase 1, y atender a todos los clientes 1 presentes hasta que el buffer 1 se vacíe. El acceso al buffer 2 sigue abierto, pero no se atiende a estos clientes.

Etapa 3. Permitir el acceso sólo a los clientes de la clase 2, atendiéndolos de acuerdo con una política FIFO, hasta que queden n_0 clientes en esa cola.

Etapa 4. Permitir el acceso a los clientes de ambas clases, atendiendo exclusivamente a los de la clase 2, hasta que el buffer 2 se vacíe. Ir a la etapa 1.

Nuestro objetivo en lo que sigue será evaluar el rendimiento a largo plazo de esta política, y determinar los valores de n_0 y n_1 que optimizan dicho rendimiento.

2.- Formulación del modelo.

Sea $C_i(t)$ el número de clientes de la clase i en cola en el instante t. El estado del sistema en el instante t, queda descrito mediante el par $(C_1(t),C_2(t))$, y es sencillo probar que el sistema pasa reiteradamente por el estado $(0,n_0)$. Puesto que de las condiciones del modelo se sigue que los tiempos entre pasos sucesivos por este estado son v.a.i.i.d., el proceso que cuenta el número de veces hasta t que el sistema ha pasado por ese estado, es un proceso de renovación.

Sean $N_i(t)$, $R_i(t)$,$\delta_i(t)$, respectivamente, el número de clientes atendidos, rechazados, y el coste de la espera de la clase i hasta el instante t. Entonces el rendimiento (beneficio por unidad de tiempo) obtenido hasta t por el sistema es:

$$R(n_0,n_1,t)=\frac{\sum_{i=1}^{2}[\beta_i N_i(t)-\gamma_i R_i(t)-\delta_i(t)]}{t}$$

Considerando como ciclo de renovación el tiempo transcurrido entre dos pasos sucesivos por el estado $(0,n_0)$, y siendo T su duración, resultados conocidos de la teoría de la renovación garantizan que el rendimiento a largo plazo del sistema coincide con el rendimiento medio del mismo durante un ciclo de renovación, esto es:

$$R(n_0,n_1)=\lim_{t\to\infty} R(n_0,n_1,t)=\frac{\sum_{i=1}^{2}\{\beta_i E[N_i(T)]-\gamma_i E[R_i(T)]-E[\delta_i(T)]\}}{E[T]} \quad (1)$$

Al término de cada una de las etapas descritas en nuestra política de gestión, los buffers de las colas muestran los siguientes tamaños:

I	$0\ldots m_0$	$\ldots m_1$	$\ldots 0$	$\ldots 0$
II	$\underbrace{n_0\ldots 0}_{T_1}$	$\underbrace{\ldots n_1}_{T_2}$	$\underbrace{\ldots n_2}_{T_3}$	$\underbrace{\ldots n_0}_{T_4}$

donde T_1 corresponde a la etapa 4, T_2 a la etapa 1, T_3 a la etapa 2 y T_4 a la etapa 3; n_0 y n_1 son fijos, mientras que n_2, m_0 y m_1 son aleatorios. La descripción de la política de gestión del sistema, y el hecho de que las llegadas y salidas siguen distribuciones exponenciales, permiten simplificar (1) de la forma:

$$R(n_0,n_1)=\frac{k_1(E[T_1]+E[T_4])+k_2(E[T_2]+E[T_3])+\sum_{i=1}^{2}E[\delta_i(T)]}{\sum_{i=1}^{4}E[T_i]} \quad (2)$$

donde las k_i, (i=1,2) son constantes dependientes de los valores de $\lambda_i,\mu_i,\beta_i,\gamma_i$.

3.- Cálculo del rendimiento del sistema

Para calcular $E[T_1]$ definimos τ_n como el tiempo medio que tarda en vaciarse la cola 2 si inicialmente tiene n clientes. Condicionando por el siguiente suceso en producirse, puede probarse que las τ_n verifican el siguiente sistema de ecuaciones en diferencias:

$$\begin{cases} \tau_1=\dfrac{1}{\lambda_2+\mu_2}+\dfrac{\lambda_2}{\lambda_2+\mu_2}\tau_2 \\ \tau_k=\dfrac{1}{\lambda_2+\mu_2}+\dfrac{\lambda_2}{\lambda_2+\mu_2}\tau_{k+1}+\dfrac{\mu_2}{\lambda_2+\mu_2}\tau_{k-1} \quad 1<k<N_2 \\ \tau_{N_2}=\dfrac{1}{\mu_2}+\tau_{N_2-1} \end{cases} \quad (3)$$

Resolviendo el sistema, se tiene:

$$E[T_1] = \tau_{n_0} = \frac{1}{\mu_2}\frac{n_0 - n_0\rho_2 - \rho_2^{N_2-n_0+1} - \rho_2^{N_2+1}}{(1-\rho_2)^2}.$$

Resultados elementales de la teoría de colas permiten obtener:

$$E[T_2] = n_1\frac{1}{\lambda_2}, \qquad E[T_3] = m_1\frac{1}{\mu_2}.$$

Ahora bien, $m_1 = m_0 + A(T_2) - S(T_2) - R(T_2)$, donde:

m_0= "Número medio de clientes del tipo 1 al final de T_1"

$A(T_2)$= "Número medio de clientes 1 llegados durante" T_2 = $\lambda_1 E[T_2]$

$S(T_2)$ = "Número medio de clientes 1 atendidos durante T_2" = $\mu_1 E[T_2]$

$R(T_2)$ = "Número medio de clientes 1 rechazados durante T_2"

Para calcular m_0, definimos $M_{n,m}$ como el número medio de clientes del tipo 1 almacenados en el bufffer 1 durante el tiempo transcurrido desde que el sistema alcanza el estado (n,m) hasta el final de la etapa 4. De manera análoga a como se hizo en (3), puede plantearse el siguiente sistema de ecuaciones en diferencias parciales:

$$\begin{cases} M_{n,m} = (1+M_{n+1,m})\dfrac{\lambda_1}{\lambda_1+\lambda_2+\mu_2} + M_{n,m+1}\dfrac{\lambda_2}{\lambda_1+\lambda_2+\mu_2} + M_{n,m-1}\dfrac{\mu_2}{\lambda_1+\lambda_2+\mu_2} \quad 0 \le n \le N_1,\ 0 < m < N_2 \\ M_{n,0} = 0 \quad 0 \le n \le N_1 \quad M_{N_1,m} = 0 \quad 0 \le m \le N_2 \\ M_{n,N_2} = (1+M_{n+1,N_2})\dfrac{\lambda_1}{\lambda_1+\mu_2} + M_{n,N_2-1}\dfrac{\mu_2}{\lambda_1+\mu_2} \quad 0 \le n < N_1 \end{cases}$$

Resolviendo este sistema se obtiene:

$$m_0 = M_{0,n_0} = \left(\frac{1}{A\theta_1^{n_0} + B\theta_2^{n_0}} + 1\right)\left[(1 + A\theta_1^{n_0} + B\theta_2^{n_0})^{N_1} - 1\right]$$

donde θ_1 y θ_2 son las raíces de la ecuación característica:

$\theta = \dfrac{\lambda_2}{\lambda_1+\lambda_2+\mu_2}\theta^2 + \dfrac{\mu_2}{\lambda_1+\lambda_2+\mu_2}$, y A , B se obtienen de las condiciones de frontera.

Para calcular $R(T_2)$, observemos que durante el periodo T_2 se atiende sólo a clientes del tipo 1; así, el sistema se comporta según un modelo M/M/1 con buffer finito; como ρ_1 es próximo a 1, el equilibrio se alcanza muy deprisa, y estamos en

condiciones de utilizar los resultados conocidos para este modelo en equilibrio. Entonces:

$$R(T_2) = \lambda_1 p_{N_1} E[T_2] = \rho_1^{N_1} \frac{1-\rho_1}{1-\rho_1^{N_1+1}} \lambda_1 n_1 \frac{1}{\lambda_2}$$

Sustituyendo en $E[T_3]$ y simplificando:

$$E[T_3] = (m_0 + \frac{n_1}{\lambda_2}[\lambda_1 - \mu_1 - \rho_1^{N_1} \frac{1-\rho_1}{1-\rho_1^{N_1+1}} \lambda_1]) \frac{1}{\mu_2}$$

$E[T_4]$ es el tiempo medio que tarda la cola 2 en pasar de n_2 clientes a n_0. Este tiempo medio es el mismo que tarda esta cola en pasar de $n_2 - n_0$ a 0 clientes. Luego, $E[T_4] = \tau_{n_2-n_0}$, donde $\tau_{n_2-n_0}$ se obtiene de la ecuación (3); por tanto:

$$E[T_4] = \frac{1}{\mu_2} \frac{(n_2 - n_0 - (n_2 - n_0)\rho_2 - \rho_2^{N_2-(n_2-n_0)+1} + \rho_2^{N_2+1}}{(1-\rho_2)^2}$$

donde $n_2 = n_1 + \lambda_2 E[T_3] = n_1 + \rho_2 m_1$

Calculemos ahora el coste medio de espera para ambas clases de clientes. Se tiene: $E[\delta_i(T)] = \delta$ x tiempo medio de espera de los clientes tipo i en la cola = $\delta \times \omega^{(i)}$

El tiempo medio de espera de los clientes de la clase i es $\omega^{(i)} = \sum_{j=1}^{4} \omega_{T_j}^i$, donde los $\omega_{T_j}^i$ son los tiempos de espera durante los distintos subperiodos. Se tiene:

$$\omega_{T_1}^1 = \frac{1}{\lambda_1} \frac{(m_0 - 1)m_0}{2} \qquad \omega_{T_3}^1 = \frac{1}{\mu_1} \frac{(m_1 - 1)m_1}{2}, \qquad \omega_{T_4}^1 = 0.$$

$$\omega_{T_2}^{(1)} = \mu_1 E[T_2] \left(\left(\frac{\rho_1}{(1-\rho_1} - \frac{(N_1+1)\rho_1^{N_1+1}}{1-\rho_1^{N_1+1}} \right) \frac{1}{\lambda_1} - \frac{1}{\mu_1} \right) + \frac{m_1(m+1)(N_1+1)}{\lambda_1 N_1}$$

Si llamamos τ_n al tiempo medio de espera global de los clientes de la cola 2 desde que en ésta quedan n clientes hasta que se vacía, los τ_n verifican la siguiente ecuación en diferencias:.

$$\begin{cases} \tau_k = \dfrac{k}{\lambda_2 + \mu_2} + \dfrac{\lambda_2}{\lambda_2 + \mu_2} \tau_{k+1} + \dfrac{\mu_2}{\lambda_2 + \mu_2} \tau_{k-1} & 1 < k < N_2 \\ \tau_1 = \dfrac{1}{\lambda_2 + \mu_2} + \dfrac{\lambda_2}{\lambda_2 + \mu_2} \tau_2 \qquad \tau_{N_2} = \dfrac{N_2}{\mu_2} + \tau_{N_2-1} \end{cases}$$

que, al resolverla, permite obtener:

$$\omega_{T_1}^2 = \tau_{n_0} = \frac{1}{\mu_2}\left[\frac{n_0\rho_2 - \dfrac{\rho_2^{N_2-n_0+1} - \rho_2^{N_2+1}}{1-\rho_2}}{(1-\rho_2)^2} - \frac{N_2\dfrac{\rho_2^{N_2-n_0+1} - \rho_2^{N_2+1}}{1-\rho_2}}{1-\rho_2} + \frac{n_0(n_0+1)}{2(1-\rho_2)}\right].$$

Asimismo:

$$\omega_{T_2}^2 = \frac{1}{\lambda_2}\frac{(n_1-1)n_1}{2}, \qquad \omega_{T_1}^2 = n_1\frac{n_2-n_1}{\lambda_2} + \frac{1}{\lambda_2}\frac{(n_2-n_1-1)(n_2-n_1)}{2},$$

y procediendo de modo análogo al caso $\omega_{T_1}^2$:

$$\omega_{T_1}^2 = \tau_{n_1-n_0} = \frac{1}{\mu_2}\left[\frac{(n_2-n_0)\rho_2 - \dfrac{\rho_2^{N_2-(n_2-n_0)+1} - \rho_2^{N_2+1}}{1-\rho_2}}{(1-\rho_2)^2} - \frac{N_2\dfrac{\rho_2^{N_2-(n_2-n_0)+1} - \rho_2^{N_2+1}}{1-\rho_2}}{1-\rho_2} + \frac{(n_2-n_0)(n_2-n_0+1)}{2(1-\rho_2)}\right]$$

5.- Optimización

Como puede apreciarse, las expresiones obtenidas para los distintos términos de la ecuación (2) no permiten obtener analíticamente los valores de n_0 y n_1 que optimizan el rendimiento del sistema; ello significa que dicha optimización ha de realizarse necesariamente utilizando métodos numéricos. Esta tarea se ha llevado a cabo en algunos casos particulares, comprobándose mediante simulación que los valores n_0 y n_1 así obtenidos corresponden realmente a un óptimo de $R(n_0, n_1)$.

6.- Bibliografía

FELLER, W. [1973] Introducción a la Teoría de Probabilidades y sus Aplicaciones. Limusa

KLEINROCK, L. [1975] Queuing Systems. Vol I. Wiley & Sons

MEDHI, J. [1991] Stochastic Models in Queuing Theory. Academic Press

SANTANA DEL PINO, A. [1992] Procesos Puntuales. Problemas de Control y Ramificaciones. Tesis Doctoral. Universidad de La Laguna.

SANTANA DEL PINO, A. SAAVEDRA SANTANA, P. [1989]. Un proceso de dos colas con prioridades. *XVIII Reunión Nacional de la SEIO.*