«Estudio de un sistema centralizado de redes de colas»

por J. ARANDA GALLEGO
Departamento de Estadística
Facultad de Ciencias Económicas y Empresariales
Alcalá de Henares

1. INTRODUCCION

Son numerosos los problemas reales que pueden modelarse como una red de colas, hasta el punto que en los últimos años el estudio de este tipo de sistemas ha experimentado un gran auge, basado, sin lugar a dudas, en la importancia y frecuencia con que se presentan en distintas áreas, tales como redes de comunicaciones y tráfico, sistemas de programación mediante ordenadores, operaciones de control de tráfico aéreo, sistemas hospitalarios, etc. Sin embargo, no se encuentran suficientemente divulgados ni son del todo conocidos los distintos modelos existentes, así como los problemas que la modelización basada en las redes de colas puede resolver.

En este artículo planteamos, mediante una red de colas, cuya estructura gráfica viene representada en la figura 1, lo que podría ser un sistema de decisión empresarial. En este sistema existen r centros a los que se producen llegadas de r tipos distintos de ítems (clientes, informes, etc.), estando asociado cada tipo a un único centro. Una vez realizado el procesamiento de un ítem en su centro respectivo (toma de datos de los clientes, verificación de los informes, etc.) son enviados a un centro principal (o centro de decisión) en el que se tratan todos los ítems, en orden de llegada, que han entrado al sistema por los r centros anteriores. Cuando han sido procesados (servicio solicitado por los clientes, estudio y decisión de los informes, etc.) en esta unidad central, los clientes de cada tipo pasan a ser tratados en un nuevo centro correspondiente a cada tipo (pago de facturas, envío y archivado de informes, etc.), abandonando posteriormente el sistema.

Para facilitar los cálculos posteriores, supondremos que se tienen $r + 1$ tipos distintos de clientes, con lo cual el número de estaciones o centros de servicio en la red es de $2(r + 1) + 1$.

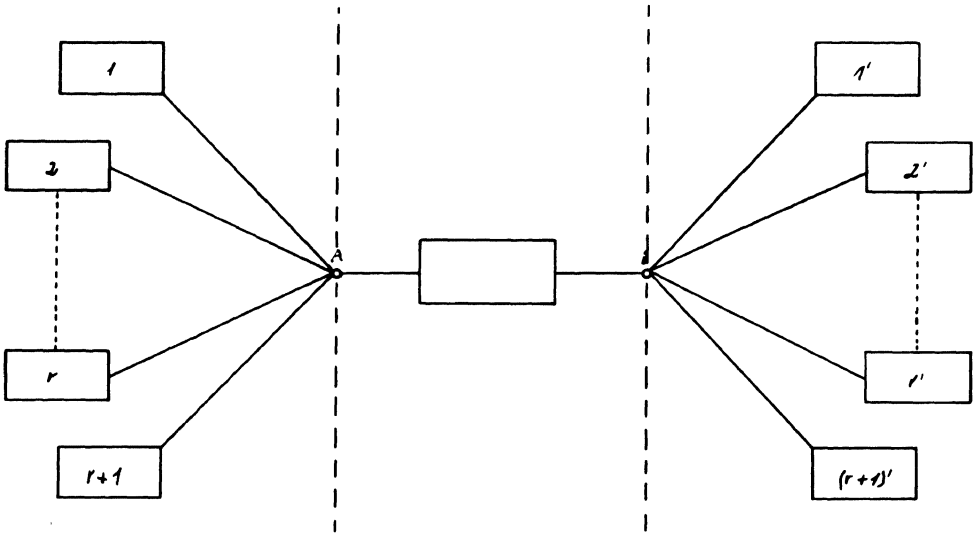


Fig. 1

2. DESCRIPCION DE LA RED

La red que planteamos se puede considerar estructurada en tres sistemas, cuya descripción realizamos a continuación.

Sistema I:

Está formado por $(r + 1)$ estaciones de servicio, cada una de ellas con distribución exponencial del tiempo de servicio de parámetro μ_i, $i = 0, 1, \ldots, r$, y c_i canales (servidores). A cada una de ellas se producen llegadas únicamente desde el exterior del sistema según un proceso de Poisson de media λ_i, $i = 0, 1, \ldots, r$. Los procesos de llegadas son independientes entre sí e independientes de la distribución del tiempo de servicio. Asimismo, no se hace ninguna limitación sobre la capacidad de cada una de las colas que componen el Sistema I.

Sistema II:

El proceso de llegadas a este sistema está constituido por la superposición de los procesos de salidas de cada una de las estaciones del Sistema I, formando la entrada a

una única estación de servicio con distribución exponencial del tiempo de servicio de parámetro μ y un solo servidor.

El proceso de salidas, que caracterizamos posteriormente, se descompone en el punto B (conectador de descomposición) en $r + 1$ flujos distintos, cada uno de ellos formado por clientes del mismo tipo.

Sistema III:

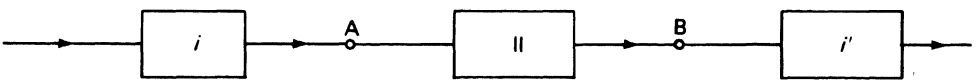
Los flujos de clientes del tipo i, $i = 0, 1, \ldots, r$, que abandonan el Sistema II se dirigen a la cola de la estación correspondiente i', $i' = 0, 1, \ldots, r$, donde son servidos con distribución exponencial del tiempo de servicio de parámetro μ_i, con un solo servidor en cada estación, abandonando posteriormente la red.

El estudio de esta red puede abordarse considerándola globalmente o bien descomponiéndola en sub-redes más sencillas, que será el procedimiento que seguiremos aquí.

3. DESCOMPOSICION DE LA RED

Podemos considerar una primera descomposición de la red por los puntos A y B, con lo que tendríamos tres sub-redes formadas por los tres sistemas anteriormente citados.

El Sistema II únicamente está formado por una sola cola, por lo que es el modelo más simple que podemos encontrar. Sin embargo, los Sistemas I y III están formados, cada uno de ellos, por $r + 1$ colas distintas, si bien para cada tipo de cliente en particular sólo es necesario conocer el comportamiento de una única cola de cada sistema, mientras que para conocer las medidas de efectividad globales de la red nos basamos en la independencia entre los distintos tipos de clientes. En consecuencia, para un tipo de cliente cualquiera, podemos esquematizar su recorrido como sigue:



por lo que en cada sistema sólo es necesario estudiar una sola cola.

4. ANALISIS DE LAS SUB-REDES COMPONENTES

Consideremos una cola cualquiera de las que componen el Sistema I. Como ya hemos indicado, es del tipo $M/M/c_i$, $i = 0, 1, \ldots, r$, por lo que sus medidas de

efectividad son conocidas (5), así como las distribuciones asociadas a cada una de ellas. Así, tendremos que para la cola i se tendrá:

$P_i[n]$ = Probabilidad de n clientes en la cola i, incluyendo los que están siendo servidos.

$$P_i[n] = \begin{cases} \dfrac{\lambda_n}{n!\,\mu_i^n} P_i(0) & 1 \leq n \leq c_i \\ \dfrac{\lambda_n}{c^{n-c_i} \cdot c_i!\,\mu_i^n} P_i(0) & n \geq c_i \end{cases}$$

$$P_i(0) = \left[\sum_{n=0}^{c_i-1} (r_i^n/n!) + \frac{c_i \cdot r_i^{c_i}}{c_i!(c_i - r_i)} \right]^{-1}$$

siendo

$$r_i = \frac{\lambda_i}{\mu_i} \; ; \; \text{para} \left(\frac{r_i}{c_i}\right) = P_i < 1 = 0, 1, \ldots, r$$

L_q^i = Número medio de individuos esperando en la cola i.

$$L_q^i = \frac{r_i^{c_i+1}/c_i}{c_i!(1 - n_i/c_i)^2} P_i(0)$$

L_i = Número medio de individuos en la cola i.

$$L_i = r_i + \left[\frac{r_i^{c_i+1}/c_i}{c_i!(1 - n_i/c_i)^2} \right] P_i(0)$$

W_q^i = Tiempo medio de espera en la cola i.

$$W_q^i = L_q^i/\lambda_i$$

W_i = Tiempo medio de estancia en la cola i.

$$W_i = L_i/\lambda_i$$

Unicamente queda por conocer el proceso de salidas de dicha cola, resultado que viene dado por el siguiente teorema:

Teorema 4.1. (Burke, 1968)

«La distribución del intervalo de tiempo entre dos salidas consecutivas del servicio en una cola M/M/c, en estado estacionario, es la misma que la del tiempo entre dos llegadas consecutivas al sistema.»

En consecuencia, el proceso de salidas de la cola i será de Poisson con parámetro λ_i, $i = 0, 1, \ldots, r$.

Como ya hemos indicado, el flujo de salidas de cada una de las colas del Sistema I se superpone en el punto A para constituir un único proceso que es el de llegadas al Sistema II, por lo tanto necesitamos conocer este proceso, que viene dado por el siguiente resultado.

Teorema 4.2. (Cinlar, 1972)

«La superposición de k procesos de Poisson independientes de parámetros respectivos λ_i, $i = 1, \ldots, k$, es un proceso de Poisson de parámetro

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

Por lo tanto el proceso de llegadas, dado que los flujos de salidas de todas las colas del Sistema I son independientes, es de Poisson de parámetro

$$\lambda = \sum_{i=0}^{r} \lambda_i$$

Teniendo en cuenta este resultado, se tendrá que la cola del Sistema II es del tipo M/M/1 con $r + 1$ tipos de clientes, que notaremos por M_r/M/1.

Las medidas de efectividad y distribuciones asociadas para esta cola pueden verse en (1), y vienen dadas por:

$P[x_i]$ = Probabilidad de que hay x_i clientes del tipo i en el Sistema II.

$$P[x_i] = [(1 - \rho)(p_i \cdot \rho)^{xi}]/[(1 - \rho \cdot q_i)(1 - q_i \cdot \rho)^{xi}]$$

para $x_i = 0, 1, \ldots$

$i = 0, 1, \ldots, r$

$q_i = i - p_i$

$\rho = \lambda/\mu$

siendo $p_i = \lambda_i/_i\lambda$ la probabilidad de que se presente un cliente del tipo i.

L_i = Número medio de individuos del tipo i en el sistema.

$$L_i = p_i \cdot \frac{\rho}{1 - \rho}$$

L_q^i = Número medio de individuos del tipo i en cola.

$$L_q^i = p_i \left[\frac{2\rho - 1}{1 - \rho} + \frac{1 - \rho}{1 - q_i \rho} \right]$$

L = Número medio de individuos en el Sistema II.

$$L = \frac{\rho}{1 - \rho}$$

L_q = Número medio de individuos en la cola del Sistema II.

$$L = \frac{\rho^2}{1 - \rho}$$

W = Tiempo medio de estancia en el sistema (para cualquier tipo de individuo).

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

W_q = Tiempo medio de permanencia en cola (para cualquier tipo de individuo).

$$W_q = \frac{\rho}{\mu(1 - \rho)}$$

Tenemos así determinadas las características operativas de la sub-red II. Sin embargo, para pasar al estudio del Sistema III necesitamos conocer el proceso de salidas, para cada tipo de clientes, del Sistema II, para lo cual necesitamos el siguiente resultado:

Teorema 4.3

«En una cola $M_r/G/1$, donde todos los clientes tienen igual distribución, general, del tiempo de servicio, la transformada de Laplace de la distribución del tiempo entre dos salidas consecutivas de clientes del mismo tipo viene dada por:

$$D_i^*(s) = p_i \cdot B^*(s) \cdot \frac{1}{1 - q_i B^*(s)} \cdot \frac{\lambda}{\mu} \cdot \left(\frac{s + \mu}{\lambda + s} \right) \quad i = 0, 1, \ldots, r$$

con $\mathrm{Re}(s) \geq 0$, $|q_q \cdot B^*(s)| < 1$. Siendo $q_i = 1 - p_i$ y $B^*(s)$ la transformada de Laplace de la función de densidad del tiempo de servicio.»

Demostración:

Al estudiar el sistema en los instantes de salida del servicio de un cliente se pueden presentar dos casos.

a) Que el sistema se encuentre vacío al abandonar el servicio el cliente que estamos considerando, lo que ocurrirá con probabilidad $(1 - \rho)$, siendo

$$\rho = \lambda/E \text{ [tiempo de servicio]}$$

o bien

b) Que el sistema no se encuentre vacío en ese instante, lo que ocurrirá con probabilidad ρ.

En el caso *a*) tendremos que la transformada de Laplace buscada será:

$$\sum_{n=0}^{\infty} \frac{\lambda}{\lambda + s} p_i \cdot q_i^n \cdot B^{*(n+1)}(s)$$

mientras que el caso *b*) se tiene:

$$\sum_{n=0}^{\infty} p_i q_i^n \cdot B^{*(n+1)}(s)$$

por lo tanto, la transformada de Laplace buscada será:

$$D_i^*(s) = \rho \cdot \sum_{n=0}^{\infty} p_i q_i^n \cdot B^{*(n+1)}(s) +$$

$$+ (1 - \rho) \sum_{n=0}^{\infty} \frac{\lambda}{\lambda + s} \cdot p_i q_i^n \cdot B^{*(n+1)}(s) =$$

$$= \rho \cdot p_i \cdot B^*(s) \cdot \frac{1}{1 - q_i B^*(s)} +$$

$$+ (1 - \rho) p_i \frac{\lambda}{\lambda + s} \cdot \frac{1}{1 - q_i B^*(s)} =$$

$$= p_i B^*(s) \cdot \frac{1}{1 - q_i B^*(s)} \cdot \frac{\lambda}{\mu} \cdot \frac{(s + \mu)}{(\lambda + s)}$$

En el caso que estamos considerando, el sistema es del tipo $M_r/M/1$, por lo que tendremos:

Corolario 4.1

«El tiempo entre dos salidas consecutivas del sistema para clientes del tipo i, en el modelo $M_r/M/1$ tiene por función de densidad:

$$g_i(t) = \alpha_1 \cdot \lambda \cdot \exp(-\lambda t) + \alpha_2 \cdot \mu \cdot p_i \cdot \exp(-\mu p_i t)$$

para $t \geq 0$, con $\alpha_1 + \alpha_2 = 1$ y $\rho \neq p_i$».

Demostración:

Hemos de calcular la transformada de Laplace del tiempo de servicio, que en este caso será:

$$B^*(s) = \int_0^{\infty} \mu \cdot \exp[-(\mu + s)t]\, dt = \mu/(\mu + s)$$

con lo que $D_i^*(s)$ quedará:

$$D_i^*(s) = p_i \cdot \frac{\lambda}{\lambda + s} \cdot \frac{\mu + s}{s + \mu p_i} = \overline{TL[g_i(t)]}$$

por lo que, si descomponemos $D_i^*(s)$ en fracciones simples y aplicamos transformada de Laplace inversa, se tiene:

$$g_i(t) = p_i \cdot \frac{(\mu - \lambda)}{\mu p_i - \lambda} \cdot \lambda \cdot \exp(-\lambda t) + \\ + \frac{(P_i - 1)}{\mu p_i - \lambda} \cdot \mu p_i \cdot \exp(-\mu p_i t)$$

y notando por

$$\alpha_1 = p_i \cdot \frac{\mu - \lambda}{\mu p_i}; \qquad \alpha_2 = \lambda \cdot \frac{(p_i - 1)}{\mu p_i - \lambda}$$

que verifican

$$\alpha_1 + \alpha_2 = 1$$

obtenemos que la función de densidad de los intervalos de tiempo entre salidas consecutivas de clientes del tipo i es:

$$g_i(t) = \alpha_1 \cdot \lambda \exp(-\lambda t) + \alpha_2 \cdot \mu p_i \cdot \exp(-\mu p_i t); \qquad t \geq 0$$

Obviamente esta función así calculada ha de ser función de densidad, sin embargo, no lo es para algunos valores de p_i. Para que $g_i(t)$ sea función de densidad ha de cumplir:

1) $\int_0^t g_i(t)dt = 1$

2) $g_i(t) \geq 0; \quad \forall t \geq 0$

La primera condición se cumple de forma inmediata, mientras que para la segunda condición se tiene que

$$\alpha_1 \cdot \lambda \exp(-\lambda t) + \alpha_2 \mu p_i \exp(-\mu p_i t) \geq 0$$

que es equivalente a

$$t \geq \frac{1}{\lambda - \mu p_i} \cdot \log[(1 - \rho)/(1 - p_i)]$$

pero como $t \geq 0$, se ha de cumplir que

$$\frac{1}{\mu p_i - \lambda} \cdot \log[(1 - \rho)/(1 - p_i)] \geq 0$$

y para que esto se verifique pueden ocurrir dos casos:

a) $\mu p_i - \lambda < 0$

$(1 - \rho)/(1 - p_i) \leq 1$

de donde se deduce que ha de ser $p_i < \rho$, y

b) $\mu p_i - \lambda > 0$

$(1 - \rho)/(1 - p_i) \geq 1$

obteniéndose la condición $p_i > \rho$. Siendo en ambos casos $p_i \neq \rho$.

En consecuencia se tiene que $g_i(t)$ es función de densidad salvo para el caso $p_i = \rho$, si bien los posibles valores de p_i afectan al signo de los coeficientes α_1 y α_2. En efecto:

$$p_i < \rho \Rightarrow \alpha_1 < 0 \qquad \alpha_2 \geq 0$$

$$p_i > \rho \Rightarrow \alpha_1 > 0 \qquad \alpha_2 \leq 0$$

Notemos que si ambos coeficientes hubieran sido del mismo signo (positivo), la función de densidad $g_i(t)$ hubiera sido la de una distribución hiperexponencial.

Una vez conocida la distribución del tiempo entre dos salidas consecutivas de clientes del mismo tipo en el Sistema II estamos en disposición de estudiar el Sistema III, ya que esta distribución, para los clientes del tipo i, $i = 0, 1, \ldots, r$ constituye la distribución del tiempo entre llegadas a la cola i; $i' = 0, 1, \ldots, r$ del Sistema III. En consecuencia y dado que todas las colas del Sistema III tienen un comportamiento análogo, estudiaremos las medidas de efectividad de una cualquiera de las colas que lo componen.

Todas las colas que componen el Sistema III son del tipo GI/M/1, con distribución del tiempo entre llegadas conocidas y tiempo de servicio exponencial de media $\mu_{i'}$. En consecuencia tendremos:

Teorema 4.4

«Las medidas de efectividad de la cola i', $i' = 0, 1, \ldots, r$, son:

$$L_{i'} = \beta_{i'}/(1 - \beta_{i'})$$
$$L_q^i = \beta_{i'}^2/(1 - \beta_{i'})$$
$$W_{i'} = 1/\mu_{i'}(1 - \beta_{i'})$$
$$W_q^i = \beta_{i'}/\mu_{i'}(1 - \beta_{i'})$$

siendo $\beta_{i'}$, la única solución de la ecuación

$$\beta^2 - (\rho_{i'} + 1 + \delta_{i'}p_i)\beta + (\rho_{i'} \cdot \delta_{i'} \cdot p_i + p_i\rho_i) = 0$$

que verifica que $0 < \beta_{i'} < 1$, siendo $\rho_{i'} = \lambda/\mu_{i'}$ y $\delta_{i'} = \mu/\mu_{i'}$.

Demostración:

Sabemos que una cola GI/M/1 (6), la probabilidad de que el n-ésimo cliente que llega a la cola se encuentre a m clientes en el sistema, sigue una ley geométrica de parámetro β, tal que

$$\beta = L^*(s)_{\,s=\mu(1-\beta)} \qquad [1]$$

siendo L*(s) la transformada de Laplace de la función de densidad de los tiempos entre llegadas. Además, la función de densidad del tiempo de espera en el sistema, para los instantes de tiempo en que se producen llegadas al sistema, puede expresarse como

$$w(t) = \mu_{i'}(1 - \beta) \cdot \exp[-\mu_{i'}(1 - \beta)t]$$

en consecuencia, la obtención de las medidas de efectividad se realiza de la siguiente forma:

$$L_{i'} = \sum_{m=1}^{\infty} m(1 - \beta_{i'}) \cdot \beta_{i'}^{m} = \bar{\beta}_{i'}/(1 - \beta_{i'})$$

$$L_q^i = \sum_{m=2}^{\infty} (m - 1)(1 - \bar{\beta}_{i'}) \cdot \beta_{i'}^{m} = \beta^2/(1 - \bar{\beta}_{i'})$$

$$W_{i'} = \int_0^{\infty} t \cdot \mu_{i'}(1 - \beta_{i'}) \cdot \exp[-\mu_{i'}(1 - \beta_{i'})t]dt = 1/\mu_{i'}(1 - \beta_{i'})$$

$$W_q^i = W_{i'} \cdot \mu(1/\bar{\mu}_{i'}) = \beta_{i'}/\mu_{i'}(1 - \beta_{i'})$$

Como ya hemos indicado, el valor de $\beta_{i'}$ se obtiene de la condición [ ], pero L*(s) coincide con la transformada de Laplace de $g_i(t)$, por tanto

$$\beta = p_i \cdot \frac{\lambda}{\lambda + \mu_{i'}(1 - \beta)} \cdot \frac{\mu + \mu(1 - \beta)}{\bar{\mu}_{i'}(1 - \beta) + \mu p_i}$$

de donde se obtiene

$$(\beta - 1)[\beta^2 - \beta(\rho_{i'} + 1 + \delta_{i'}p_i) + \rho_{i'} \cdot \delta_{i'} \cdot p_i + \rho_{i'}p_i] = 0$$

y dado que la solución $\beta_{i'}$ ha de ser menor que la unidad, la raíz de la ecuación anterior que nos interesa será aquella raíz de

$$\beta^2 - \beta(\rho_{i'} + 1 + \delta_{i'}p_i) + \rho_{i'} \cdot \delta_{i'} \cdot p_i + \rho_{i'}p_i = 0$$

que verifique $0 < \beta < 1$, que siempre existe (4).

Notemos que estas medidas de efectividad están referidas a los instantes de llegadas de los clientes a la cola i'.

5. MEDIDAS DE EFECTIVIDAD DE LA RED

Las medidas de efectividad de la red quedan determinadas a partir de las obtenidas para cada una de las sub-redes estudiadas, por lo que se tendrá, para cada tipo de cliente:

1. Número medio de individuos del tipo i en la red:

$$L(i) = r_i + P_i(0)\left[\frac{r_i^{c_i+1}/c_i}{c_i!\left(1 - \frac{r_i}{c_i}\right)^2}\right] + p_i \cdot \frac{\rho}{1 - \rho} + \beta_{i'} \cdot \frac{1}{1 - \beta_{i'}}$$

2. Número de individuos del tipo i, esperando para ser servidos en la red:

$$L_q(i) = P_i(0) \cdot \frac{r_i^{c_i+1}/c_i}{c_i!\left(1 - \frac{r_i}{c_i}\right)^2} + p_i \cdot \left[\frac{2\rho - 1}{1 - \rho} + \frac{1 - \rho}{1 - q_i\rho}\right] + \beta_i^2 \cdot \frac{1}{1 - \beta_i}$$

3. Tiempo medio de estancia en el sistema para un cliente del tipo i:

$$W(i) = (1/\mu_i) + P_i(0) \cdot \left[\frac{r_i^{c_i+1}/\lambda_i C_i}{\lambda_i \cdot c_i!\left(1 - \frac{r_i}{c_i}\right)^2}\right] + \frac{1}{\mu(1 - \rho)} + \frac{1}{\mu_i(1 - \beta_i)}$$

4. Tiempo medio de espera en cola para un cliente del tipo i:

$$W_q(i) = P_i(0) \cdot \frac{r^{c_i+1}/\lambda_i c_i}{c_i!\left(1 - \frac{r_i}{c_i}\right)^2} + \frac{\rho}{\mu(1 - \rho)} + \beta_i \cdot \frac{1}{\mu_i(1 - \beta_i)}$$

5. Número medio de individuos en la red:

$$L = \sum_{i=0}^{r} r_i + \sum_{i=0}^{r} P_i(0) \cdot \left[\frac{r_i^{c_i+1}/c_i}{c_i!\left(1 - \frac{r_i}{c_i}\right)^2}\right] + \frac{\rho}{1 - \rho} + \sum_{i=0}^{r} \frac{\beta_i}{1 - \beta_i}$$

6. Número medio de individuos en cola:

$$L_q = \sum_{i=0}^{r} P_i(0) \frac{r_i^{c_i+1}/c_i}{c_i!\left(1 - \frac{r_i}{c_i}\right)^2} + \frac{2\rho - 1}{1 - \rho} + (1 - \rho)\sum_{i=0}^{r} \frac{1}{1 - q_i\rho} + \sum_{i=0}^{r} \beta_i^2/(1 - \beta_i)$$

BIBLIOGRAFIA

(1) ARANDA, J., PASCUAL, A.: «Cálculo de las medidas de efectividad y nuevas cotas para una cola M/M/1 con distintos tipos de clientes». *Actas de las V Jornadas Luso-Españolas de Matemáticas*. Aveiro, 1978.

(2) BURKE, P. J.: «The output process of a sationary M/M/s queuering systems». *Annals of Mth. St.*, 29, 1144-1152, 1968.

(3) CINLAR, E.: «Superposition of point processes». En Stochastic point processes: *Statistical Analysis Theory and Applications*. Wiley, 1972.

(4) COX, D. R., SMITH, W. L.: *Estudio Matemático de las colas*. Uteha, 1964.

(5) GROOS, D., HARRIS, C. M.: *Fundamentals of queueing theory*. Wiley 1974.

(6) KLEINROCK, L.: *Queueing Systems*. Vol 1. Wiley, 1975.

RESUMEN

En este artículo se estudia una red de colas de un tipo particular en la que se pueden presentar distintos tipos de clientes, siguiendo cada tipo un camino predeterminado a través de la red. Para calcular sus medidas de efectividad se aplica el principio de descomposición de la red en subredes más sencillas, estudiando los flujos de clientes que aparecen en ellas.

Palabras clave: colas, redes.

SUMMARY

In this paper we study a queueing network of a particular type in which there are several classes of customers. Each class of customer proceed through a fixed sequence of queues before leaving the system. We apply the decomposition principle for to calculate the measures of efectivesness, studing the streams of customers in the system.

Key words: queue, network.
AMS Subject classification: 90B15.