

**UN SISTEMA DE COLAS CON DEMANDAS REPETIDAS,
SERVIDOR POCO FIABLE Y
TIEMPO GENERAL DE VIDA DEL SERVIDOR**

G. Bouza¹, I. Fortes², P. Moreno²

¹Departamento de Matemática Aplicada
Universidad de La Habana

²Departamento de Matemática Aplicada
Universidad de Málaga

RESUMEN

Analizamos un sistema de colas $M/M/1$ con demandas repetidas, fallos del servidor y reparaciones. La demanda servida durante la avería del servidor espera hasta que finaliza la reparación para continuar su servicio. Tanto los tiempos de vida del servidor como los de reparación son generales. Estudiamos las probabilidades estacionarias, las distribuciones de los tamaños de la órbita y del sistema, además de algunas medidas de realización. Finalmente, hallamos la descomposición estocástica para el tamaño de nuestro sistema y como consecuencia proporcionamos estimaciones inferiores y superiores de la distancia entre las distribuciones estacionarias de nuestro modelo y el correspondiente modelo sin demandas repetidas.

Palabras y frases clave: Sistemas de colas, Demandas repetidas, Averías del servidor.
Clasificación AMS: 60K25.

1. INTRODUCCIÓN

Los sistemas de colas con demandas repetidas desempeñan un papel importante en el análisis de redes de comunicaciones computacionales, así como en sistemas de comunicaciones. La característica fundamental de estos sistemas es que las demandas que al llegar encuentran el servidor ocupado se unen a un grupo de demandas insatisfechas (llamado órbita) para, tras un tiempo aleatorio, intentar de nuevo obtener su servicio. Para una revisión detallada de los principales resultados y literatura sobre este tema pueden consultarse Artalejo (1999) y Falin y Templeton (1997).

Una circunstancia notable e inevitable en la función de servicio de un sistema de colas es su avería y consecuente reparación. Sin embargo, en la mayoría de los estudios existentes sobre modelos de colas, se supone que el servidor está siempre disponible, aunque esta suposición es evidentemente poco realista. De hecho, las interrupciones del servicio debido a averías del servidor ocurren en la práctica. Por supuesto, hasta que se recobra la función de servicio, los tiempos de espera de las demandas en el sistema aumentan.

Pese a que los sistemas de colas con demandas repetidas y averías del servidor tienen muchas aplicaciones prácticas de gran interés, hasta ahora básicamente sólo se han tratado modelos donde o bien la avería se produce justo al comenzar el servicio (vea por ejemplo Atencia et al. (2003)) o bien el tiempo de vida del servidor es exponencial (vea Wang et al.

(2001)). En este trabajo se analiza un sistema de colas $M/M/1$ con demandas repetidas en el cual el servidor está sujeto a fallos y su tiempo de vida sigue una distribución general. Para consultar la versión de nuestro sistema en tiempo discreto remitimos al lector a Moreno (2003).

En la siguiente sección, damos la descripción matemática del sistema de colas bajo estudio. En la sección 3, estudiamos la distribución estacionaria de nuestro sistema de colas, las distribuciones del número de demandas en la órbita y en el sistema, así como mostramos algunos índices de productividad. Finalmente, en la sección 4, presentamos la propiedad de descomposición estocástica para el tamaño de nuestro sistema y como consecuencia damos cotas de la proximidad entre las distribuciones estacionarias de nuestro sistema de colas y su correspondiente sin demandas repetidas.

2. DESCRIPCIÓN DEL SISTEMA

Consideremos un sistema de colas con un único servidor donde las demandas llegan según un flujo de Poisson de intensidad λ . Si el servidor está libre en el momento de una llegada externa, esta demanda comienza a servirse inmediatamente. Por el contrario, las demandas que al llegar encuentran el servidor ocupado, abandonan el área de servicio, pero después de un tiempo aleatorio repiten un intento para conseguir servicio. Esas demandas que están esperando para solicitar su servicio se consideran *en órbita*. Suponemos que los tiempos entre reintentos de cualquier demanda son independientes y distribuidos exponencialmente con intensidad γ (política de reintento clásica). Los tiempos de servicio son independientes y exponencialmente distribuidos con media μ^{-1} . Por supuesto, cualquier demanda abandona el sistema después de completar su servicio.

El servidor puede que se averíe mientras sirve demandas y cuando ésto ocurre se envía directamente a reparar. La demanda que estaba sirviéndose durante la avería espera hasta que se repara el servidor para completar su servicio. Suponemos que los tiempos de vida del servidor siguen una ley general $A(x)$ con transformada de Laplace-Stieltjes $\alpha(s)$. Los tiempos de reparación están gobernados por una distribución arbitraria $B(x)$ con transformada de Laplace-Stieltjes $\beta(s)$ y momentos de n -ésimo orden respecto al origen β_n . Indudablemente, el servidor funciona correctamente después de la reparación.

Suponemos que el flujo de llegada de demandas, los intervalos entre reintentos, los tiempos de servicio, los tiempos de vida del servidor y los tiempos de reparación son mutuamente independientes.

En un momento arbitrario t , el sistema puede describirse por el proceso

$$X(t) = (C(t), N(t), \xi_1(t), \xi_2(t))$$

donde $C(t)$ denota el estado del servidor (0, 1 o 2, según que en el momento t el servidor esté libre, ocupado o averiado) y $N(t)$ es el número de demandas en la órbita. Si $C(t) = 1$, $\xi_1(t)$ representa el tiempo transcurrido de vida del servidor. Si $C(t) = 2$, $\xi_2(t)$ simboliza el tiempo transcurrido de reparación del servidor.

3. ANÁLISIS DE LAS PROBABILIDADES ESTACIONARIAS

En esta sección estudiamos la distribución conjunta del estado del servidor y el número de demandas repetidas en el régimen estacionario. Consideremos las siguientes probabilidades límite

$$p_{0,n} = \lim_{t \rightarrow \infty} P[C(t) = 0, N(t) = n], \quad n \geq 0$$

y las densidades de probabilidad límite

$$p_{i,n}(x) = \lim_{t \rightarrow \infty} P[C(t) = i, N(t) = n, x < \xi_i(t) \leq x + dx]; \quad i = 1, 2, \quad n \geq 0, \quad x \geq 0.$$

Por el método de la variable suplementaria, fácilmente obtenemos el sistema de ecuaciones de equilibrio

$$(\lambda + n\gamma)p_{0,n} = \mu \int_0^\infty p_{1,n}(x) dx; \quad n \geq 0 \quad (1)$$

$$\frac{d}{dx}p_{1,n}(x) = -[\lambda + \mu + a(x)]p_{1,n}(x) + (1 - \delta_{0,n})\lambda p_{1,n-1}(x); \quad n \geq 0, \quad x \geq 0 \quad (2)$$

$$\frac{d}{dx}p_{2,n}(x) = -[\lambda + b(x)]p_{2,n}(x) + (1 - \delta_{0,n})\lambda p_{2,n-1}(x); \quad n \geq 0, \quad x \geq 0 \quad (3)$$

donde las funciones $a(x)$ y $b(x)$ son las intensidades de completación condicional (en el momento x) para los tiempos de vida del servidor y los tiempos de reparación respectivamente, es decir,

$$a(x) = \frac{A'(x)}{1 - A(x)} \quad \text{y} \quad b(x) = \frac{B'(x)}{1 - B(x)}.$$

Las condiciones de frontera son

$$p_{1,n}(0) = \lambda p_{0,n} + (n+1)\gamma p_{0,n+1} + \int_0^\infty p_{2,n}(x)b(x)dx, \quad n \geq 0 \quad (4)$$

$$p_{2,n}(0) = \int_0^\infty p_{1,n}(x)a(x)dx, \quad n \geq 0. \quad (5)$$

Para resolver el sistema de ecuaciones (1) – (5), definimos las siguientes funciones generatrices

$$P_0(z) = \sum_{n=0}^\infty p_{0,n} z^n; \quad P_i(x, z) = \sum_{n=0}^\infty p_{i,n}(x) z^n, \quad i = 1, 2.$$

Mediante las funciones generatrices previas, el sistema de ecuaciones de equilibrio (1)–(5) se transforma en

$$\lambda P_0(z) + \gamma z P'_0(z) = \mu \int_0^\infty P_1(x, z) dx \quad (6)$$

$$\frac{\partial}{\partial x}P_1(x, z) = -[\lambda - \lambda z + \mu + a(x)]P_1(x, z), \quad x \geq 0 \quad (7)$$

$$\frac{\partial}{\partial x}P_2(x, z) = -[\lambda - \lambda z + b(x)]P_2(x, z), \quad x \geq 0 \quad (8)$$

$$P_1(0, z) = \lambda P_0(z) + \gamma P'_0(z) + \int_0^\infty P_2(x, z)b(x)dx \quad (9)$$

$$P_2(0, z) = \int_0^\infty P_1(x, z)a(x)dx. \quad (10)$$

Las soluciones de las ecuaciones diferenciales (7) – (8) son

$$\begin{aligned} P_1(x, z) &= P_1(0, z) [1 - A(x)] e^{-(\lambda - \lambda z + \mu)x} \\ P_2(x, z) &= P_2(0, z) [1 - B(x)] e^{-(\lambda - \lambda z)x}. \end{aligned}$$

Sustituyendo las ecuaciones precedentes en las ecuaciones (6) y (9) – (10) llegamos a

$$\lambda P_0(z) + \gamma z P'_0(z) = \mu \frac{1 - \alpha^*(z)}{\lambda - \lambda z + \mu} P_1(0, z) \quad (11)$$

$$P_1(0, z) = \lambda P_0(z) + \gamma P'_0(z) + \beta^*(z) P_2(0, z) \quad (12)$$

$$P_2(0, z) = \alpha^*(z) P_1(0, z) \quad (13)$$

donde hemos definido las funciones auxiliares

$$\alpha^*(z) = \alpha(\lambda - \lambda z + \mu) \quad \text{y} \quad \beta^*(z) = \beta(\lambda - \lambda z).$$

Sustituyendo las ecuaciones (12) – (13) en la ecuación (11) obtenemos la siguiente ecuación diferencial

$$P'_0(z) = \frac{\lambda}{\gamma} \frac{\rho [1 - \alpha^*(z) \beta^*(z)] + \alpha^*(z) \frac{1 - \beta^*(z)}{1 - z}}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1 - z \beta^*(z)}{1 - z}} P_0(z) \quad (14)$$

donde $\rho = \lambda/\mu$ es la carga del sistema. Resolviendo esta ecuación diferencial se obtiene

$$P_0(z) = P_0(1) \exp \left\{ \frac{\lambda}{\gamma} \int_1^z \frac{\rho [1 - \alpha^*(u) \beta^*(u)] + \alpha^*(u) \frac{1 - \beta^*(u)}{1 - u}}{1 - \rho u [1 - \alpha^*(u) \beta^*(u)] - \alpha^*(u) \frac{1 - u \beta^*(u)}{1 - u}} du \right\}$$

donde la constante $P_0(1)$ se hallará más adelante.

Eliminando $P'_0(z)$ al combinar las ecuaciones (12) – (14) nos queda

$$P_1(0, z) = \frac{\rho (\lambda - \lambda z + \mu) P_0(z)}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1 - z \beta^*(z)}{1 - z}}$$

$$P_2(0, z) = \frac{\rho (\lambda - \lambda z + \mu) \alpha^*(z) P_0(z)}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1 - z \beta^*(z)}{1 - z}}.$$

El siguiente lema nos proporciona una condición bajo la cual las funciones generatrices están bien definidas en $[0,1]$ y en $z = 1$ se extienden por continuidad.

Lema 1 *La desigualdad $\rho z [1 - \alpha^*(z) \beta^*(z)] + \alpha^*(z) \frac{1 - z \beta^*(z)}{1 - z} < 1$ se verifica para $0 \leq z < 1$ si y sólo si $\rho < 1 - \lambda \beta_1 \frac{\alpha(\mu)}{1 - \alpha(\mu)}$.*

Finalmente, a partir de la condición de normalización, se determina la incógnita

$$P_0(1) = 1 - \lambda \beta_1 \frac{\alpha(\mu)}{1 - \alpha(\mu)} - \rho.$$

Puesto que $P_0(1) > 0$, tenemos que $\rho < 1 - \lambda \beta_1 \frac{\alpha(\mu)}{1 - \alpha(\mu)}$ es una condición necesaria para la estabilidad de nuestro sistema.

Los resultados de esta sección quedan resumidos en el siguiente teorema.

Teorema 1 *La distribución estacionaria del proceso $\{X(t), t \geq 0\}$ viene dada por las funciones generatrices*

$$P_0(z) = \left[1 - \lambda \beta_1 \frac{\alpha(\mu)}{1 - \alpha(\mu)} - \rho \right] \exp \left\{ \frac{\lambda}{\gamma} \int_1^z \frac{\rho [1 - \alpha^*(u) \beta^*(u)] + \alpha^*(u) \frac{1 - \beta^*(u)}{1 - u}}{1 - \rho u [1 - \alpha^*(u) \beta^*(u)] - \alpha^*(u) \frac{1 - u \beta^*(u)}{1 - u}} du \right\}$$

$$P_1(x, z) = \frac{\rho (\lambda - \lambda z + \mu) P_0(z)}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1 - z \beta^*(z)}{1 - z}} [1 - A(x)] e^{-(\lambda - \lambda z + \mu) x}$$

$$P_2(x, z) = \frac{\rho (\lambda - \lambda z + \mu) \alpha^*(z) P_0(z)}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1 - z \beta^*(z)}{1 - z}} [1 - B(x)] e^{-(\lambda - \lambda z) x}$$

donde

$$\alpha^*(z) = \alpha(\lambda - \lambda z + \mu) \quad y \quad \beta^*(z) = \beta(\lambda - \lambda z).$$

Si en los casos $C(t) \in \{1, 2\}$ no tenemos en cuenta respectivamente el tiempo de vida del servidor transcurrido $\xi_1(t)$ ni el tiempo de reparación transcurrido $\xi_2(t)$, entonces las correspondientes funciones generatrices vienen dadas por

$$\begin{aligned} P_1(z) &= \frac{\rho(\lambda - \lambda z + \mu) P_0(z)}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1-z\beta^*(z)}{1-z}} \frac{1 - \alpha^*(z)}{\lambda - \lambda z + \mu} \\ P_2(z) &= \frac{\rho(\lambda - \lambda z + \mu) \alpha^*(z) P_0(z)}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1-z\beta^*(z)}{1-z}} \frac{1 - \beta^*(z)}{\lambda - \lambda z}. \end{aligned}$$

Corolario 1

- (1) La función generatriz de probabilidad del tamaño del sistema (es decir, de la variable L) viene dada por

$$Q(z) = P_0(z) + z P_1(z) + z P_2(z) = \frac{[1 - \alpha^*(z)] P_0(z)}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1-z\beta^*(z)}{1-z}}.$$

- (2) La función generatriz de probabilidad del tamaño de la órbita (es decir, de la variable N) viene dada por

$$R(z) = P_0(z) + P_1(z) + P_2(z) = \frac{(\rho - \rho z + 1) [1 - \alpha^*(z) \beta^*(z)] P_0(z)}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1-z\beta^*(z)}{1-z}}.$$

Corolario 2

- (1) La probabilidades de que el servidor esté desocupado, ocupado o averiado son respectivamente

$$P_0(1) = 1 - \lambda \beta_1 \frac{\alpha(\mu)}{1 - \alpha(\mu)} - \rho, \quad P_1(1) = \rho, \quad P_2(1) = \lambda \beta_1 \frac{\alpha(\mu)}{1 - \alpha(\mu)}.$$

- (2) Los tamaños medios del sistema y de la órbita son respectivamente

$$\begin{aligned} E[L] &= \frac{\lambda + \gamma}{\gamma} \frac{\rho}{P_0(1)} + \lambda \alpha(\mu) \frac{2 \beta_1 [\lambda + \gamma (1 - \rho)] + \lambda \gamma \beta_2}{2 \gamma [1 - \alpha(\mu)] P_0(1)} - \frac{\lambda^2 \beta_1 \alpha'(\mu)}{[1 - \alpha(\mu)]^2 P_0(1)} \\ E[N] &= E[L] - [1 - P_0(1)]. \end{aligned}$$

- (3) El tiempo medio que una demanda pasa en el sistema viene dado por $W = E[L]/\lambda$.

4. DESCOMPOSICIÓN ESTOCÁSTICA

En esta sección damos una ley de descomposición estocástica para el tamaño del sistema, cuya función generatriz de probabilidad puede escribirse como

$$Q(z) = \frac{\left[1 - \lambda \beta_1 \frac{\alpha(\mu)}{1 - \alpha(\mu)} - \rho\right] [1 - \alpha^*(z)]}{1 - \rho z [1 - \alpha^*(z) \beta^*(z)] - \alpha^*(z) \frac{1-z\beta^*(z)}{1-z}} \times \frac{P_0(z)}{P_0(1)}.$$

Como consecuencia de esta propiedad tenemos el siguiente teorema.

Teorema 2 *El número total de demandas en el sistema bajo estudio puede expresarse como la suma de dos v.a.i.: una de las cuales es el número total de demandas en el sistema de colas $M/M/1/\infty$ con servidor poco fiable y tiempo general de vida del servidor, y la otra es el número de demandas repetidas en nuestro sistema dado que el servidor está desocupado.*

Notamos que esperábamos que nuestro sistema tuviera esta propiedad, ya que los períodos de desocupación del servidor pueden interpretarse como vacaciones del servidor. Por tanto, la ley de descomposición estocástica para nuestro sistema es consistente con la presentada por Fuhrmann y Cooper (1985) para los modelos de colas $M/G/1$ con vacaciones del servidor.

Como consecuencia de la propiedad de descomposición estocástica, derivamos una medida de la proximidad entre las distribuciones estacionarias de nuestro sistema y el sistema de colas $M/M/1/\infty$ con servidor poco fiable y tiempo general de vida del servidor (L_∞). La importancia de las siguientes cotas es proporcionar estimaciones inferiores y superiores de la distancia entre ambas distribuciones.

Teorema 3 *Se verifican las siguientes desigualdades*

$$2 [P_0(1) - p_{0,0}] \leq \sum_{j=0}^{\infty} |P[L = j] - P[L_\infty = j]| \leq 2 \frac{P_0(1) - p_{0,0}}{P_0(1)}.$$

La idea de la demostración de este teorema puede encontrarse en Artalejo (1998).

Finalmente, observamos que la distancia $\sum_{j=0}^{\infty} |P[L = j] - P[L_\infty = j]|$ entre las distribuciones de las variables L y L_∞ decrece cuando $\gamma \rightarrow \infty$ (alta intensidad de reintentos). La alta intensidad de reintentos es de especial interés práctico en la vida real, ya que las demandas que encuentran el servidor ocupado repiten sus “llamadas” casi inmediatamente.

5. REFERENCIAS

- Artalejo, J.R. (1998). “Some results on the $M/G/1$ queue with N -policy”. *Asia-Pacific Journal of Operational Research* 15, 147-157.
- Artalejo, J.R. (1999). “Accessible Bibliography on Retrial Queues”. *Mathematical and Computer Modelling* 30, 1-6.
- Atencia, I., Fortes, I., Moreno, P. and Sánchez, S. (2003). “A retrial queue with starting failures, feedback and linear retrial policy”. *6th Hellenic European Conference on Computer Mathematics and its Applications*. Athens.
- Falin, G.I. and Templeton, J.G.C. (1997). *Retrial Queues*. Chapman & Hall, London.
- Fuhrmann, S.W. and Cooper, R.B. (1985). “Stochastic Decompositions in the $M/G/1$ Queue with Generalized Vacations”. *Operations Research* 33, 1117-1129.
- Moreno, P. (2003). “A discrete-time retrial queue with unreliable server and general server lifetime”. *XXIII Seminar on Stability Problems for Stochastic Models*. Pamplona.
- Wang, J., Cao, J. and Li, Q. (2001). “Reliability Analysis of the Retrial Queue with Server Breakdowns and Repairs”. *Queueing Systems* 38, 363-380.