SISTEMAS DE COLAS CON INTERRUPCIÓN DE SERVICIOS SIN PRELACIÓN

MIGUEL MIRANDA
Facultad de Ingeniería. Universidad de Buenos Aires. ARGENTINA[1]
miguelmiranda@netscape.net

Fecha Recepción: Junio 2014 - Fecha Aceptación: Noviembre 2014

RESUMEN

En el presente trabajo se formula un modelo matemático para determinar expresiones cuantitativas de las típicas variables de eficiencia y de la distribución de probabilidades de estado en sistemas de colas en los cuales el servicio que se presta está sujeto a interrupciones aleatorias e instantáneas, con la particularidad adicional de que los clientes presentan características de tolerancia absoluta. Se describe el caso general de un sistema de un solo canal, en donde tanto los procesos de arribo y de servicio de clientes, como los de interrupción y restitución del funcionamiento son de tipo Poisson. El modelo puede asimismo ser utilizado para resolver un sistema multiclase con prioridad relativa que opere bajo las mismas hipótesis de trabajo.

PALABRAS CLAVE: Sistemas – Colas – Interrupción – Tolerancia – Prelación – MM1 – Prioridades.

ABSTRACT

This paper provides a mathematical model for quantitatively determining the typical variables of efficiency and the distribution of the steady-state probabilities for queueing systems with random, non-preemptive, breakdowns, wherein the customers present characteristics of absolute tolerance. In particular, an infinite population single server system with Markovian arrival and service processes for clients as well as Markovian frequency and duration for interruptions, are herein described through an analytical approach. The model further allow solving multiclass systems having non preemptive priority which operate under the same hypothesis of operation.

KEY WORDS: Systems – Queues – Breakdowns – Tolerance – Preemption – MM1 – Priorities.

[1] También Facultad de Ingeniería – Universidad Austral – Argentina; Facultad de Ingeniería – Universidad Católica Argentina – Argentina.

1. INTRODUCCIÓN

En la práctica es muy común que los sistemas de colas presenten interrupciones del trabajo que se realiza o del servicio que se presta. Estas suspensiones temporales de las actividades pueden ser debidas a diversas causas, ya sean exógenas (por ejemplo una falla en el suministro de energía, o un corte en las comunicaciones) o endógenas (tales como desperfectos o fallas de la máquina que realiza el trabajo en el canal de atención). Además, en la mayoría de los casos, tanto los tiempos entre desperfectos que generan las interrupciones como los tiempos de reparación del desperfecto son aleatorios, con distribución exponencial (Thiruvengadam, 1962).

En algunos sistemas, cuando se produce la falla se detiene inmediatamente el servicio. Esta situación se denomina interrupción absoluta, o también interrupción con prelación, o inmediata. En otros sistemas, por el contrario, el canal puede seguir funcionando hasta finalizar el trabajo que estaba realizando en el momento del desperfecto, por ejemplo mediante la utilización de una fuente auxiliar tal como un UPS. A esta última modalidad de operación se la llama interrupción relativa (o sin prelación, o no instantánea). Los modelos para describir sistemas con interrupción instantánea fueron descriptos tanto para situaciones de clientes tolerantes (Miranda, 2012) como intolerantes (Miranda, 2014). En el presente trabajo se analizará el caso de interrupción no inmediata, en donde el sistema puede encontrarse en alguna de las siguientes condiciones operativas:

F: Operativo
S: Temporalmente operativo
R: No operativo

La condición operativa F significa que el sistema está, o bien trabajando sobre un cliente o bien vacío pero en condiciones de funcionar. Una vez que se produce el desperfecto que lleva a la suspensión de la actividad, si el canal está atendiendo, el sistema pasa a la condición S, en donde permanece operativo pero pendiente de interrumpir la actividad tan pronto como finalice el trabajo que estaba realizando en el momento de la falla. Por último, la condición R es aquella en la cual el sistema está fuera de servicio (es decir, interrumpido) hasta tanto se repare el defecto.

A diferencia de los casos de interrupción con prelación, en donde la frecuencia y la duración de las interrupciones es independiente del proceso de atención a los clientes, en los sistemas que tienen la posibilidad de postergar la detención, como en este caso, las probabilidades de que se encuentren en alguna de las condiciones operativas arriba indicadas dependen de la duración del servicio del canal, lo que importa una dificultad adicional en la formulación de las expresiones cuantitativas para la resolución del modelo que los describe.

2. FORMULACIÓN DEL MODELO

Para el desarrollo del presente modelo matemático se adoptará la siguiente nomenclatura (Miranda, 2013):

n: Cantidad de clientes que se hallan en el sistema en un instante determinado.

λ: Número promedio de clientes que arriban al sistema por unidad de tiempo (llamado tasa promedio de arribos). La inversa de este parámetro es el intervalo promedio entre arribos de clientes: $T_A = 1/\lambda$.

μ: Velocidad promedio de atención (o tasa promedio de servicios a clientes cuando el canal está funcionando). La inversa es la duración promedio del servicio: $T_S = 1/\mu$.

ρ: Factor de tráfico del sistema: $\rho = \lambda/\mu$.

λ_R: Cantidad promedio de interrupciones por unidad de tiempo de funcionamiento del sistema. La inversa de este parámetro es el tiempo promedio que transcurre desde que se reanuda el funcionamiento hasta que se produce el próximo desperfecto: $T_R = 1/\lambda_R$

μ_R: Velocidad promedio de restitución del servicio. La inversa de este parámetro es la duración promedio de la interrupción. $T_{SR} = 1/\mu_R$

ρ_R: Factor de interrupción del sistema: $\rho_R = \lambda_R/\mu_R$.

L: Número promedio de clientes en el sistema.

L_C: Cantidad promedio de clientes en la cola.

W: Tiempo promedio de permanencia de un cliente en el sistema.

W_C: Tiempo promedio de espera de un cliente en la cola.

M: Cantidad de canales del sistema.

H: Número promedio de clientes recibiendo atención (o número promedio de canales atendiendo).

$p(H)$: Probabilidad de que el canal se encuentre atendiendo. Está dada por la relación H/M, por lo que para un sistema monocanal será $p(H) = H$.

I: Número promedio de clientes cuyo servicio se ha interrumpido y que se encuentran a la espera de que el servicio se restaure (o número promedio de canales ocupados pero en situación de interrupción).

$p(I)$: Probabilidad de que un canal se encuentre en situación de interrupción cuando hay clientes para atender ($n > 0$). Está dada por la relación I/M, por lo que para un sistema de un solo canal será $p(I) = I$.

λ_n: Tasa de ingresos, cuando hay “n” clientes en el sistema.

μ_n: Tasa de egresos, cuando hay “n” clientes en el sistema.

El estado del sistema queda definido por dos dimensiones: la cantidad de clientes que se encuentran en el sistema (n = 0, 1, 2, 3, …,) y la condición operativa (j = F, S, R). En consecuencia, llamaremos:

p(n,j): Probabilidad de que el sistema se encuentre en el estado n,j, siendo n ≥ 0 y j = F, S, R

También:

$p(F) = \sum_{0}^{\infty} p(n,F)$: Probabilidad de que el sistema se encuentre en la condición F,

$p(R) = \sum_{0}^{\infty} p(n,R)$: Probabilidad de que el sistema se encuentre en la condición R, y

$p(S) = \sum_{0}^{\infty} p(n,S)$: Probabilidad de que el sistema se encuentre en la condición S.

Supondremos que los clientes presentan tolerancia absoluta frente a las interrupciones. Esto significa que un cliente interrumpido no abandona el sistema sino que, por el contrario, espera a que se reanude la actividad.

Se analizarán aquí sistemas en régimen permanente de un solo canal, a los que arriban clientes provenientes de una población infinita y que pertenecen a una misma clase o categoría. La disciplina de atención se supone tipo FIFO (de manera que el canal selecciona para la próxima atención al cliente que haya arribado primero al sistema). Otra hipótesis de trabajo es que tanto los procesos de arribo y de atención de clientes, como los de interrupción y de restauración de los servicios son de tipo Poisson. Es decir, se asume que el tiempo entre arribos de clientes, la duración del servicio, el tiempo que transcurre hasta que se produce una interrupción y el tiempo requerido para restaurar el servicio son variables aleatorias de distribución exponencial.

Finalmente, consideraremos el caso en donde el tiempo de restauración empieza una vez que finaliza el servicio del cliente que se estaba atendiendo en el momento del desperfecto (en el caso de haya clientes en el sistema), o inmediatamente (cuando el sistema está vacío).

En la FIGURA 1 se muestra esta modalidad cuando se produce una intermisión en un sistema ocupado (n > 0). Al producirse el desperfecto, el sistema pasa de la condición F a la S y permanecerá así hasta que finalice el servicio que se estaba brindando. Teniendo en cuenta la propiedad de "falta de memoria" de los procesos poissonianos, el tiempo promedio de permanencia en la condición S es igual al tiempo de un servicio, cuyo promedio es T_S. Al terminar dicho trabajo, entonces, comienza la interrupción efectiva de las actividad (condición R) y permanecerá así mientras dure la restauración del servicio, cuyo tiempo promedio es T_{SR}. Finalmente, cuando se restituye el funcionamiento, el sistema pasa de nuevo a la condición operativa F, y permanecerá en ella un tiempo promedio T_R, hasta que se verifique la próxima falla.

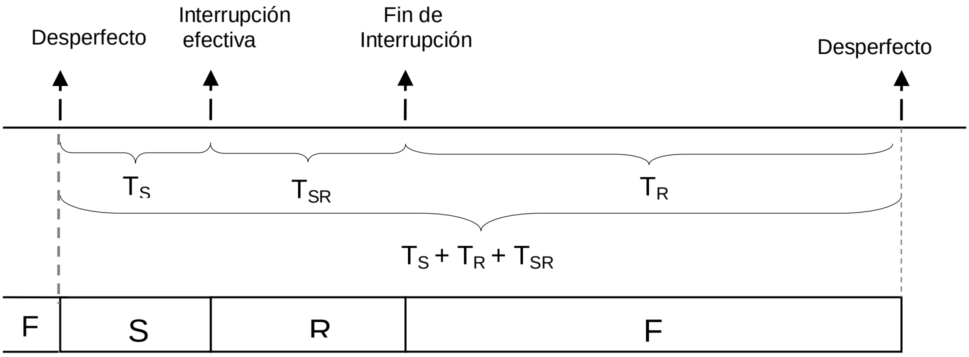

FIGURA 1
Fuente: Elaboración propia.

Si, en cambio, cuando se produce el desperfecto el canal se encuentra vacío (n = 0), el sistema pasa de F directamente a R, tal como se muestra en la Fig. 2.

Estando el sistema en funcionamiento, la probabilidad de que el canal esté ocioso (n = 0) cuando se produce el desperfecto es:

$$p(0/F) = \frac{p(0,F)}{p(F)}$$

mientras que el complemento es la probabilidad de que el canal se encuentre trabajando (n > 0):

$$p(n > 0/F) = 1 - \frac{p(0,F)}{p(F)}$$

En consecuencia, el porcentaje de funcionamiento del sistema estará dado por la siguiente expresión:

$$p(F) = \frac{T_R}{\left(T_R + T_{SR} + T_S\right)\cdot\left(1 - \frac{p(0,F)}{p(F)}\right) + \left(T_R + T_{SR}\right)\cdot\frac{p(0,F)}{p(F)}}$$

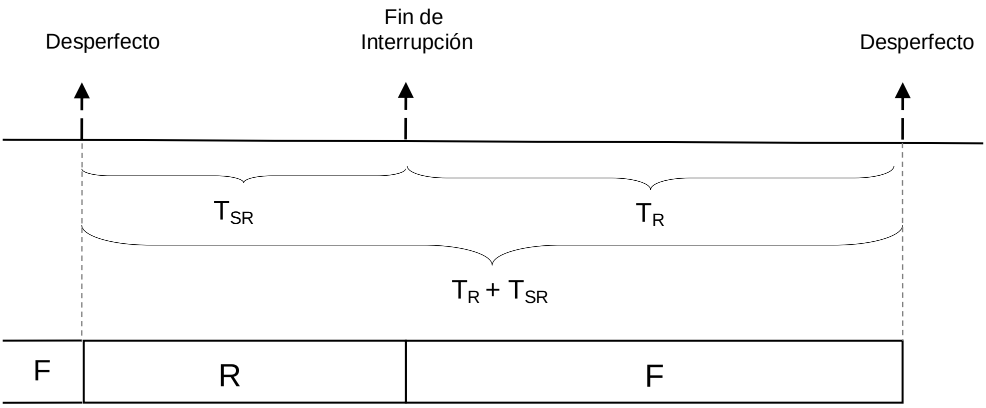

FIGURA 2
Fuente: Elaboración propia.

Es decir:

$$p(F) = \frac{T_R}{\left(T_R + T_{SR} + T_S - T_S \cdot \frac{p(0,F)}{p(F)}\right)}$$

Despejando F:

$$p(F) = \frac{T_R + p(0,F) \cdot T_S}{T_R + T_{SR} + T_S} \tag{1}$$

Del mismo modo, el porcentaje de interrupción real del sistema es:

$$p(R) = \frac{T_{SR}}{\left(T_R + T_{SR} + T_S - T_S \cdot \frac{p(0,F)}{p(F)}\right)} \tag{2}$$

Finalmente, el porcentaje de tiempo que el canal está en funcionamiento pero pendiente de interrupción es:

$$p(S) = \frac{T_S \cdot \left(1 - \frac{p(0,F)}{p(F)}\right)}{T_R + T_{SR} + T_S \cdot \left(1 - \frac{p(0,F)}{p(F)}\right)}$$

Operando:

$$p(S) = \frac{p(F)\cdot T_S - T_S \cdot p(0,F)}{p(F)\cdot \left(T_R + T_{SR} + T_S\right) - T_S \cdot p(0,F)}$$

Reemplazando el valor de p(F) de (1):

$$p(S) = \frac{\dfrac{T_R + p(0,F)\cdot T_S}{T_R + T_{SR} + T_S}\cdot T_S - T_S \cdot p(0,F)}{\dfrac{T_R + p(0,F)\cdot T_S}{T_R + T_{SR} + T_S}\cdot \left(T_R + T_{SR} + T_S\right) - T_S \cdot p(0,F)}$$

$$p(S) = \frac{\dfrac{T_R + p(0,F)\cdot T_S \cdot T_S - \left(T_R + T_{SR} + T_S\right)\cdot T_S \cdot p(0,F)}{\left(T_R + T_{SR} + T_S\right)}}{T_R + p(0,F)\cdot T_S - T_S \cdot p(0,F)}$$

$$p(S) = \frac{T_S}{T_R}\cdot \frac{T_R + p(0,F)\cdot T_S - \left(T_R + T_{SR} + T_S\right)\cdot p(0,F)}{\left(T_R + T_{SR} + T_S\right)}$$

$$p(S) = \frac{T_S}{T_R}\cdot \frac{T_R - p(0,F)\cdot T_R - p(0,F)\cdot T_{SR}}{\left(T_R + T_{SR} + T_S\right)} \qquad (3)$$

En la parte superior de la FIGURA 3 se muestra esquemáticamente el porcentaje de tiempo que el canal se encuentra atendiendo p(H) = H, el porcentaje de tiempo que se encuentra sin clientes p(0) y el porcentaje de tiempo que, si bien hay clientes esperando ser atendidos, el sistema está interrumpido p(I) = I. En la parte inferior de la figura se muestran las probabilidades de que el sistema se encuentre en alguna de las posibles condiciones operativas (F, S, R). La probabilidad de que el sistema esté vacío, a su vez, se muestra desagregada en las probabilidades de cada una de sus posibles condiciones (en funcionamiento o en intermisión). Puede observarse claramente que:

$$p(R) = p(0,R) + I$$

y también:

$$p(F) + p(S) = p(0,F) + H \qquad (4)$$

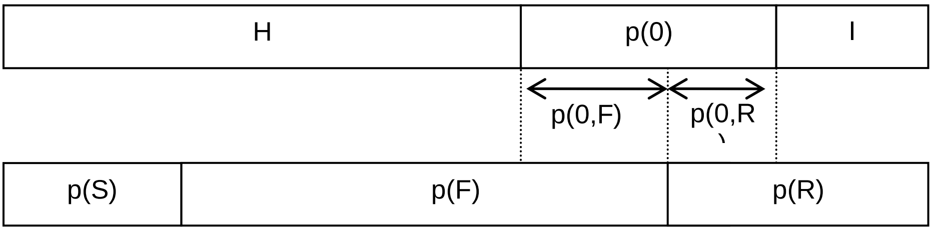

FIGURA 3
Fuente: Elaboración propia.

Reemplazando (1) y (3) en (4):

$$\frac{T_R + p(0F)\cdot T_S}{\left(T_R + T_{SR} + T_S\right)} + \frac{T_S}{T_R}\cdot\frac{T_R - p(0,F)\cdot T_R - p(0,F)\cdot T_{SR}}{\left(T_R + T_{SR} + T_S\right)} = p(0,F) + H$$

Operando:

$$\frac{T_R^{\,2} + T_R\cdot T_S - p(0,F)\cdot T_{SR}\cdot T_S}{T_R\cdot\left(T_R + T_{SR} + T_S\right)} = p(0,F) + H$$

$$T_R^{\,2} + T_R\cdot T_S - p(0,F)\cdot T_{SR}\cdot T_S = p(0,F)\cdot T_R\cdot\left(T_R + T_{SR} + T_S\right) + H\cdot T_R\cdot\left(T_R + T_{SR} + T_S\right)$$

$$p(0,F) = \frac{T_R^{\,2} + T_R\cdot T_S - H\cdot T_R\cdot(T_R + T_{SR} + T_S)}{T_R\cdot\left(T_R + T_{SR} + T_S\right) + T_{SR}\cdot T_S}$$

$$p(0,F) = \frac{T_R\cdot\left[T_R + T_S - H\cdot\left(T_R + T_{SR} + T_S\right)\right]}{T_R^{\,2} + T_R\cdot T_{SR} + T_R\cdot T_S + T_{SR}\cdot T_S} = \frac{T_R\cdot\left[T_R + T_S - H\cdot\left(T_R + T_{SR} + T_S\right)\right]}{T_R\cdot\left(T_R + T_{SR}\right) + T_S\cdot\left(T_R + T_{SR}\right)}$$

$$p(0,F) = \frac{T_R\cdot\left[T_R + T_S - H\cdot\left(T_R + T_{SR} + T_S\right)\right]}{\left(T_R + T_{SR}\right) + \left(T_R + T_S\right)} \qquad (5)$$

En definitiva, reemplazando (5) en (1), (2) y (3) y (5) se determinan las variables P(F), P(R) y P(S) en función de los parámetros del problema.

Suponiendo que no hay impaciencia, la tasa de ingreso es igual a la tasa de arribos:

$$\bar{\lambda} = \lambda$$

La tasa de egresos es igual a la velocidad de atención del canal por el número promedio de clientes recibiendo el servicio:

$$\bar{\mu} = \mu \cdot H$$

Por lo tanto, para un sistema en régimen permanente:

$$H = \rho \tag{6}$$

lo que resulta obvio ya que no hay pérdida de trabajo.

Llamaremos

$$\lambda_C = \frac{1}{T_R + T_{SR} + T_S} \tag{7}$$

a la frecuencia de interrupciones del sistema cuando el canal se encuentra activo.

A continuación se plantea la expresión cuantitativa del tiempo total esperado de permanencia para un cliente que arriba al sistema, objeto principal del presente trabajo, teniendo en consideración tres principios básicos: la propiedad PASTA (*Poisson Arrivals See Time Averages*) de los sistemas de colas (Bolch, Greiner, De Meer y Trivedi, 2006), la propiedad de falta de memoria de los procesos poissonianos y la ley de John Little (Little, 1961), como agregación de los siguientes tiempos componentes:

- Tiempo de espera debido a la atención de los clientes que estaban en el sistema en el momento del arribo:
  $$L \cdot T_S$$
- Tiempo de espera por el hecho de encontrar al sistema en el estado R, es decir fuera de servicio, o bien en el estado S, o sea pendiente de finalizar el trabajo:
  $$T_{SR} \cdot \left(p(R) + p(S)\right)$$
- Tiempo de espera debido a las interrupciones que se producen mientras el cliente está esperando en cola: La cantidad de interrupciones promedio que se producen durante el tiempo de espera en cola cuando la condición operativa es R en el momento del arribo es $(W_C – T_{SR})\cdot\lambda_C$; es decir, se debe descontar el tiempo de espera de la interrupción en el momento del arribo. Cuando el estado es S, esta cantidad es: $(W_C – T_{SR} – T_S)\cdot\lambda_C$, en donde se debe descontar, además, el tiempo de espera de la atención del cliente que se estaba atendiendo previo a la interrupción. Finalmente, cuando el estado es F, la cantidad de interrupciones es simplemente $W_C\cdot\lambda_C$. Considerando que el tiempo de espera promedio por cada interrupción es T_{SR}, el tiempo promedio de espera por las interrupciones producidas durante la permanencia en cola del cliente será:

$$(W_C - T_{SR})\cdot \lambda_C \cdot T_{SR} \cdot p(R) + (W_C - T_{SR} - T_S)\cdot \lambda_C \cdot T_{SR} \cdot p(S) + W_C \cdot \lambda_C \cdot T_{SR} \cdot p(F)$$

- Tiempo de espera de su propia atención:
  T_S

De modo que, la expresión del tiempo promedio de permanencia en el sistema es:

$$\begin{aligned} W = {} & L \cdot T_S + T_{SR} \cdot \big(p(R) + p(S)\big) + (W_C - T_{SR})\cdot \lambda_C \cdot T_{SR} \cdot p(R) + \\ & (W_C - T_{SR} - T_S)\cdot \lambda_C \cdot T_{SR} \cdot p(S) + W_C \cdot \lambda_C \cdot T_{SR} \cdot p(F) + T_S \end{aligned} \qquad (8)$$

y considerando que:

$$L = L_C + H = W_C \cdot \lambda + \rho$$

y que:

$$W = W_C + T_S$$

reemplazando en (8):

$$\begin{aligned} W_C + T_S = {} & W_C \cdot \lambda \cdot T_S + \rho \cdot T_S + T_{SR} \cdot \big(p(R) + p(S)\big) + W_C \cdot \lambda_C \cdot T_{SR} \cdot p(R) \\ & - T_{SR} \cdot \lambda_C \cdot T_{SR} \cdot p(R) + W_C \cdot \lambda_C \cdot T_{SR} \cdot p(S) - T_{SR} \cdot \lambda_C \cdot T_{SR} \cdot p(S) \\ & - T_S \cdot \lambda_C \cdot T_{SR} \cdot p(S) + W_C \cdot \lambda_C \cdot T_{SR} \cdot p(F) + T_S \end{aligned}$$

Pero

$$p(R) + p(S) + p(F) = 1$$

y

$$\lambda \cdot T_S = \rho$$

Luego:

$$\begin{aligned} W_C = {} & W_C \cdot \rho + \rho \cdot T_S + T_{SR} \cdot \big(p(R) + p(S)\big) + W_C \cdot \lambda_C \cdot T_{SR} - T_{SR} \cdot \lambda_C \cdot T_{SR} \cdot p(R) \\ & - T_{SR} \cdot \lambda_C \cdot T_{SR} \cdot p(S) - T_S \cdot \lambda_C \cdot T_{SR} \cdot p(S) \end{aligned}$$

$$W_C \cdot (1 - \rho - \lambda_C \cdot T_{SR}) =$$

$$\rho \cdot T_S + T_{SR} \cdot \big[\big(p(R) + p(S)\big) - T_{SR} \cdot \lambda_C \cdot \big(p(R) - p(S)\big) - \lambda_C \cdot T_S \cdot p(S)\big]$$

$$W_C = \frac{\rho \cdot T_S + T_{SR} \cdot \big[p(R) + p(S) - T_{SR} \cdot \lambda_C \cdot \big(p(R) + p(S)\big) - \lambda_C \cdot T_S \cdot p(S)\big]}{1 - \rho - \lambda_C \cdot T_{SR}} \qquad (9)$$

Entonces:

$$\boxed{W = \frac{\rho \cdot T_S + T_{SR} \cdot \left[p(R) + p(S) - T_{SR} \cdot \lambda_C \cdot \left(p(R) + p(S)\right) - \lambda_C \cdot T_S \cdot p(S)\right]}{1 - \rho - \lambda_C \cdot T_{SR}} + T_S} \qquad (10)$$

A partir de esta fórmula surgen inmediatamente el resto de las variables características L, Lc y Wc, considerando la ley de Little:

$$L = \lambda \cdot W \qquad (11)$$

$$L_C = \lambda \cdot W_C \qquad (12)$$

En la FIGURA 4 se muestra la cadena markoviana correspondiente a este sistema, a partir de la cual pueden determinarse las ecuaciones de estado correspondientes para una cantidad de clientes n > 0 en el sistema:

$$p(n,F) \cdot \left(\lambda_n + \lambda_R + \mu_n\right) = p(n-1,F) \cdot \lambda_{n-1} + p(n,R) \cdot \mu_R + p(n+1,F) \cdot \mu_{n+1} \qquad (13)$$

$$p(n,S) \cdot \left(\lambda_n + \mu_n\right) = p(n,F) \cdot \lambda_R \qquad (14)$$

$$p(n,R) \cdot \left(\lambda_n + \mu_R\right) = p(n-1,R) \cdot \lambda_{n-1} + p(n+1,S) \cdot \mu_n \qquad (15)$$

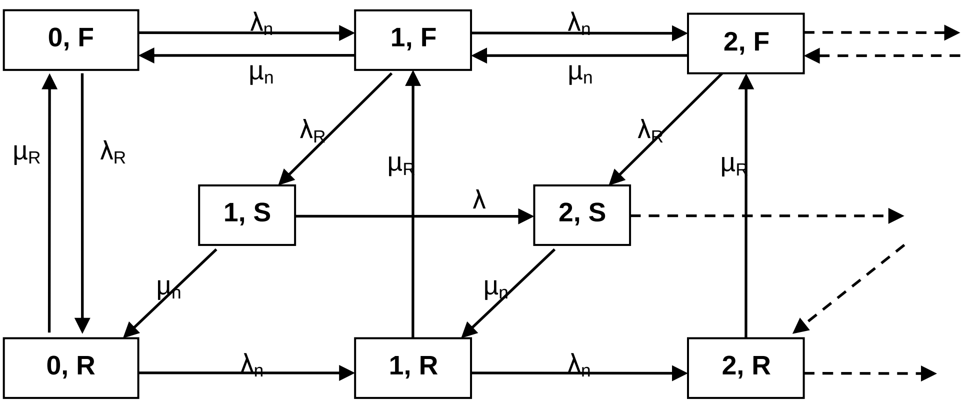

FIGURA 4
Fuente: Elaboración propia.

siendo para este sistema monocanal:

$$\mu_n = \begin{cases} 0 & \text{para } n = 0 \\ \mu & \text{para } n > 0 \end{cases}$$

$$\lambda_n = \lambda \qquad \forall\, n$$

Para el caso particular en que n = 0, reemplazando en las ecuaciones de estado y operando algebraicamente, se puede despejar p(0,R) en función de p(0,F) y de los parámetros del problema:

$$p(0,R) = p(0,F) \cdot \frac{(2 \cdot \lambda + \lambda_R + \mu) \cdot \lambda_R}{\lambda_R \cdot \mu_R + (\mu_R + \lambda) \cdot (\lambda + \mu)} \tag{16}$$

La probabilidad de que el sistema se encuentre vacío es:

$$p(0) = p(0,F) + p(0,R) \tag{17}$$

El resto de las probabilidades de estado para cada una de las condiciones de operación, F, R y S, se obtienen con las ecuaciones generatrices de estado (13), (14) y (15).

Obviamente, la probabilidad de que el sistema se encuentre con una cantidad “n” de clientes es:

$$p(n) = p(n,F) + p(n,R) + p(n,S) \tag{18}$$

A título de ejemplo, tomando los siguientes datos: λ = 10, μ = 20, ambos parámetros expresados en clientes por hora, T_R = 1/3 en horas de funcionamiento y T_{SR} = 1/9, en horas de interrupción, y aplicando las expresiones arriba indicadas se obtiene la siguiente solución:

p(F)	0,701087
p(R)	0,233696
p(S)	0,065217
p(0)	0,323853
p(0,F)	0,266309
p(0,R)	0,057544
W	0,231652 h
W_C	0,181652 h
L_C	1,816516 cl
H	0,500000 cl
L	2,316516 cl

El modelo desarrollado puede utilizarse también para resolver sistemas multiclases con prioridades (Gross y Harris, 1988), por ejemplo aquellos de categorías de clientes de alta prioridad (A) y de baja prioridad (B), en donde no exista prelación, es decir que cuando un cliente de clase A arriba al sistema de clase B tiene prioridad sobre todos los clientes de clase B que se encuentren esperando para recibir el servicio, pero debe esperar a que termine la atención del cliente de clase B que estaba siendo atendido cuando arribó al sistema.

En efecto, el subsistema de clientes de baja prioridad B puede interpretarse como un sistema con interrupción en donde el evento generador de la intermisión (‘falla’) se produce cuando arriba un cliente de categoría A mientras se está atendiendo a un cliente de clase B, y la duración de la ‘interrupción’ es el tiempo del servicio del cliente A y de todos los de su categoría que hayan ingresado hasta que ya no quede ninguno de ellos en el sistema y pueda reanudarse el servicio a los clientes B de baja categoría. Este tiempo es el tiempo de ocupación (*busy time*) del subsistema de clientes A. Por otra parte, el tiempo que transcurre entre la ‘reparación’ del sistema y la próxima ‘falla’ es el tiempo entre arribos de los clientes de alta prioridad. Es decir, dados los parámetros:

λ_A: Tasa promedio de arribos de clientes A
λ_B: Tasa promedio de arribos de clientes B
μ_A: Tasa promedio de servicios de clientes A
μ_B: Tasa promedio de servicios de clientes B

la analogía con el modelo de interrupción se podría establecer considerando:

$$\frac{1}{\lambda_A - \mu_A} = \left(T_{SR} + T_S\right) \cdot \left(1 - \frac{p(0,F)}{p(F)}\right) + T_{SR} \cdot \frac{p(0,F)}{p(F)}$$

y

$$T_R = \frac{1}{\lambda_A}$$

Para los clientes de tipo A, la existencia de clientes de menor categoría es transparente, excepto para aquél que al ingresar al sistema encuentra que el canal está atendiendo a un cliente de categoría inferior, en cuyo caso deberá esperar a que termine dicho servicio.

3. CONCLUSIONES

La formulación del tiempo de permanencia de un cliente como integración de los tiempos de espera por la atención efectiva de los clientes que se encuentran en el sistema en el momento del ingreso de un cliente, más la duración de todas las interrupciones producidas desde el ingreso hasta el egreso del mismo, sumado al tiempo de su propia atención, lo que conduce a la expresión analítica simple (10), combinada con (5), (1), (2) y (3), y que permite deducir inmediatamente el resto de las variables características de eficiencia, proporciona una solución cuantitativa sin la necesidad de intentar la solución a través del planteo de las siguientes series geométricas:

$$L = \sum_{0}^{\infty} n \cdot p(n)$$

$$L_C = \sum_{0}^{\infty} (n-1) \cdot p(n)$$

$$p(F) = \sum_{0}^{\infty} p(n,F) = p(0,F) + \sum_{1}^{\infty} p(n,F)$$

$$p(R) = \sum_{0}^{\infty} p(n,R) = p(0,R) \cdot \sum_{1}^{\infty} p(n,R) = p(0,R) + I$$

$$p(S) = \sum_{1}^{\infty} p(n,S)$$

$$H = \sum_{1}^{\infty} p(n,F) + \sum_{1}^{\infty} p(n,S)$$

$$I = \sum_{1}^{\infty} p(n,R)$$

sobre funciones generadoras de probabilidad de estado para obtener las expresiones cuantitativas, por ejemplo mediante la transformada de Laplace de la función de densidad de probabilidades (Stewart, 2009). Ese enfoque resultaría muy dificultoso en vista de la compleja vinculación existente entre las probabilidades de estado, tal como puede observarse en las ecuaciones de régimen permanente (12), (13) y (14).

Asimismo, la metodología puede aplicarse para abordar los complejos casos de sistemas multi-clases con prioridades (Cobham, 1954), por su analogía con los modelos de sistemas con interrupciones (White y Christie, 1958). En efecto, en un sistema de prioridades relativas, en donde un cliente de mayor prioridad que arriba al sistema no tiene prelación para recibir el servicio con respecto a un cliente de menor prioridad que esté siendo atendido pero sí la tiene con relación a cualquier otro de menor prioridad que esté esperando en cola, se puede considerar que para el subsistema de clientes de baja prioridad la presencia de un cliente de mayor prioridad significa una interrupción del servicio.

En base al enfoque propuesto, podrán abordarse otros modelos matemáticos para resolver cuantitativamente diferentes situaciones de sistemas con interrupciones, tales como el de interrupción con prelación pero en donde el tiempo de restauración empieza inmediatamente luego del desperfecto (haya o no clientes en el sistema), y el sistema análogo multi-clases con prioridades.

4. REFERENCIAS

- BOLCH, G.; GREINER, S.; DE MEER, H.; TRIVEDI, K. S. (2006): "QUEUEING NETWORKS AND MARKOV CHAINS". Editorial: John Wiley & Sons, Inc., USA, 878 págs.

- COBHAM, ALAN (1954): "PRIORITY ASSIGNMENT IN WAITING LINE PROBLEMS". Editorial INFORMS, Journal of Operations Research Vol. 2 págs. 70-76.

- GROSS, DONALD; HARRIS, CARL M. (1988): "FUNDAMENTALS OF QUEUEING THEORY". Editorial: John Wiley & Sons, Inc, USA.

- LITTLE, J. (1961): "A PROOF O THE THEOREM L = λ·W". Editorial: ORSA, Operations Research 9, 383-387.

- MIRANDA, MIGUEL. (2012): "SISTEMAS DE COLAS CON INTERRUPCIÓN DE SERVICIOS". Anales XXV ENDIO- XXIII EPIO/ENDIO, pág. 389.

- MIRANDA, MIGUEL. (2013): "TEORÍA DE COLAS". 2ª edición. Editorial: EDUCA, Buenos Aires.

- MIRANDA, MIGUEL. (2014): "SISTEMAS DE COLAS CON INTERMISIÓN DE SERVICIOS Y CLIENTES SIN TOLERANCIA". Revista "INVESTIGACIÓN OPERATIVA" de la Escuela de Perfeccionamiento en Investigación Operativa, Nro. 35, Mayo 2014, año XXII.

- STEWART, WILLIAM J. (2009): "PROBABILITY, MARKOV CHAINS, QUEUES, AND SIMULATION". Editorial: Princeton University Press, USA, 747 págs.

- THIRUVENGADAM, K. (1962): “QUEUEING WITH BREAKDOWNS”. Editorial: INFORMS, Journal of Operations Research, Vol. 11 págs. 62-71.

- WHITE, H. C.; CHRISTIE; L.S. (1958): “QUEUEING WITH PREEMPTIVE PRIORITIES OR WITH BREACKDOWNS”. Editorial: INFORMS, Journal of Operations Research, Vol. 6 págs. 79-95.