

## SISTEMAS DE COLAS CON INTERMISIÓN DE SERVICIOS Y CLIENTES SIN TOLERANCIA

MIGUEL MIRANDA

Facultad de Ingeniería. Universidad de Buenos Aires. ARGENTINA<sup>1</sup>  
*miguelmiranda@netscape.net*

*Fecha Recepción: Diciembre 2013 - Fecha Aceptación: Abril 2014*

### RESUMEN

En el presente trabajo se desarrollan modelos matemáticos para determinar cuantitativamente las variables típicas de eficiencia y la distribución de probabilidades de estado en sistemas de colas con interrupciones aleatorias del servicio que brindan, con la particularidad de que los clientes presentan características de intolerancia y, en consecuencia, pueden abandonar el sistema sin recibir o completar la atención. Se describen situaciones de sistemas PP1, es decir con un único servidor con población infinita y con procesos poissonianos tanto para los clientes como para las interrupciones, mediante un enfoque analítico. Estos modelos permiten resolver adicionalmente sistemas de clases múltiples con prioridad absoluta que operen bajo las mismas hipótesis de intolerancia.

**PALABRAS CLAVE:** Sistemas. Colas. Interrupción. Abandono. Intolerancia. PP1. Prioridades.

### ABSTRACT

This paper provides mathematical models for quantitatively determining the typical variables of efficiency and the distribution of the steady-state probabilities for queueing systems with random breakdowns, wherein the customers present characteristics of intolerance. In particular, MM1 single server systems with infinite population and Markovian processes for both, clients and interruptions, are herein described through an analytical approach. These models further allow solving multiclass systems with preemptive priority operating under the same hypothesis of intolerance.

**KEY WORDS:** Systems – Queues – Breakdowns – Intolerance – MM1 – Preemptive priorities.

---

<sup>1</sup> También Facultad de Ingeniería – Universidad Austral – Argentina; Facultad de Ingeniería – Universidad Católica Argentina – Argentina.

## **1. INTRODUCCIÓN**

La mayoría de los sistemas de colas que observamos en el mundo real están sujetos a situaciones de intermisión (es decir, interrupciones temporales) del servicio que se brinda en ellos (Thiruvengadam, 1962). Estos sistemas están operativos durante un período de tiempo hasta que se produce la interrupción. A partir de ese instante, quedarán fuera de servicio durante un cierto período (llamado tiempo de interrupción o tiempo muerto) hasta que se restaure la actividad. Típicamente, tanto el tiempo de funcionamiento como la duración de la interrupción son variables aleatorias y, en muchos casos, responden a distribuciones exponenciales (procesos Poisson).

El corte del servicio puede ser provocado por muy diversas causas, endógenas o exógenas al sistema, tales como desperfectos en el funcionamiento de máquinas, recesos o descansos del personal de atención, pérdida de comunicación, fallas en el suministro de energía, etc., que impiden que algunos o todos los canales puedan prestar el servicio hasta que se restauren las condiciones de funcionamiento.

Esta intermitencia operativa tiene normalmente un impacto económico de importancia, ya que la deserción de clientes puede generar lucro cesante y desaprovechamiento de recursos, además del efecto psicológico en términos de insatisfacción que provoca en los clientes. Por ejemplo, si el sistema percibe ingresos por los servicios prestados, los usuarios que se retiran sin recibir la atención completa no los abona.

Adicionalmente, en los denominados sistemas “no conservantes” del trabajo (que son aquellos en los cuales no puede recuperarse el trabajo que se realizó parcialmente hasta que se produzco el corte) se incurre en costos hundidos por la dedicación de tiempo y recursos físicos, humanos, materiales y operativos dedicados a un trabajo inconcluso que se pierde.

Llamaremos “tolerancia” a la actitud que tienen los clientes frente a las interrupciones, ya sea que ellos estén arribando al sistema o que se encuentren esperando en cola o recibiendo el servicio. Si se produce una interrupción, un cliente con un alto grado de tolerancia, típicamente decide esperar a que se reanude el servicio, mientras que un cliente muy poco tolerante, probablemente opte por abandonar el sistema.

Se analizarán aquí sistemas de un solo canal, en régimen permanente, con disciplina de atención tipo FIFO (de manera que el canal selecciona para la próxima atención al cliente que haya arribado primero al sistema), con una población infinita de clientes pertenecientes a una misma clase o categoría.

Asimismo se considerarán casos de interrupción absoluta, que son aquellos en los cuales el servicio que se está brindando a un cliente se detiene inmediatamente cuando se produce la interrupción; es decir, estos sistemas no tiene capacidad para finalizar una atención en proceso antes de quedar efectivamente fuera de servicio (como sería, por ejemplo, el caso en el que se dispusiera de un UPS). 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 arribo de clientes, la duración del servicio, el tiempo que transcurre hasta que se produce una interrupción y el tiempo del corte hasta que se activa nuevamente el servicio son variables aleatorias de distribución exponencial. Para el desarrollo de los modelos se adoptará la siguiente nomenclatura (Miranda, 2013):

- $\lambda$ : Número promedio de clientes que arriban al sistema por unidad de tiempo (tasa promedio de arribos). La inversa de este parámetro es el intervalo promedio entre arribos de clientes:  $T_A = 1/\lambda$ .
- $\mu$ : Tasa promedio de servicios a clientes cuando el canal está funcionando. La inversa es la duración promedio del servicio:  $T_S = 1/\mu$ .
- $\rho$ : Factor de tráfico del sistema:  $\rho = \lambda/\mu$ .
- $\lambda_R$ : 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 la próxima interrupción:  $T_R = 1/\lambda_R$ .
- $\mu_R$ : Velocidad promedio de restitución del servicio. La inversa de este parámetro es la duración promedio de la interrupción.  $T_{SR} = 1/\mu_R$ .
- $\rho_R$ : Factor de interrupción del sistema:  $\rho_R = \lambda/\mu_R$ .
- L: Longitud promedio de clientes en el sistema.
- $L_C$ : Longitud 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.

La suma  $T_C = T_R + T_{SR}$  constituye el tiempo de un ciclo entre dos interrupciones sucesivas: la inversa de esta variable ( $\lambda_C$ ) indica la frecuencia efectiva de interrupciones, es decir la cantidad de interrupciones promedio que se dan sobre el tiempo total (funcionamiento más interrupción) en el sistema, tal como puede observarse en la FIGURA 1.

El sistema puede encontrarse en alguno de los dos metaestados siguientes: en "Funcionamiento" (operativo) o en "Restauración" (fuera de servicio). Llamaremos "F" al porcentaje de tiempo en funcionamiento, o probabilidad de que el sistema se encuentre operativo:

$$F = \frac{T_R}{T_R + T_{SR}} = \frac{\mu_R}{\lambda_R + \mu_R} \tag{1}$$

y “R” a su complemento, es decir al porcentaje de tiempo en restauración, o probabilidad de que el sistema se encuentre fuera de servicio.

$$R = \frac{T_{SR}}{T_R + T_{SR}} = \frac{\lambda_R}{\lambda_R + \mu_R} \quad (2)$$

Puede observarse que:

$$F = \lambda_C \cdot T_R \quad (3)$$

$$R = \lambda_C \cdot T_{RS} \quad (4)$$

y que:

$$\rho_R = \frac{R}{F} \quad (5)$$

Un estado del sistema queda definido por dos dimensiones: la cantidad de clientes (“n”) y la condición operativa (“F” o “R”). Llamaremos:

$p(n,F)$ : Probabilidad de que el sistema se encuentre en el estado “n” y operativo (es decir, funcionando o en condiciones de funcionar).

$p(n,R)$ : Probabilidad de que el sistema se encuentre en el estado “n” y fuera de servicio.

$p(n)$ : Probabilidad de que el sistema se encuentre en el estado “n”.

Obviamente:

$$p(n) = p(n,F) + p(n,R) \quad (6)$$

A su vez, el porcentaje de funcionamiento del sistema será la sumatoria de todas las probabilidades  $p(n,F)$ , para n variando entre 0 y la capacidad máxima del sistema (que en este caso es infinita):

$$F = \sum_0^{\infty} p(n,F) \quad (7)$$

Llamando “H” al número promedio de clientes recibiendo servicio en el sistema, dado que hay un único canal de atención, tendremos que es también el porcentaje del tiempo en que el canal está funcionando, es decir:

$$H = \sum_1^{\infty} p(n,F) \quad (8)$$

Tendremos que:

$$F = p(0,F) + H \quad (9)$$

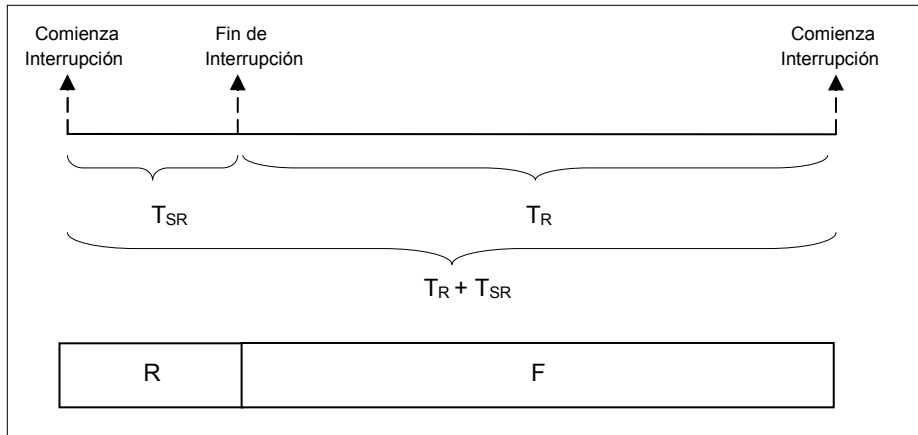


FIGURA N° 1

De esta forma, conocido  $H$ , se puede determinar  $p(0,F)$ , que es la probabilidad de que el sistema esté vacío pero en condiciones de operar. Por su parte, el porcentaje de interrupción del sistema será igual a la sumatoria de todas las probabilidades  $p(n,F)$  para "n" variando entre 0 y la capacidad máxima del sistema:

$$R = \sum_0^{\infty} p(n,R) \quad (10)$$

Denominaremos "interferencia" al estado para el cual, estando el sistema fuera de servicio, hay por lo menos un cliente para ser atendido. Al porcentaje del tiempo en que el sistema se encuentra en estado de interferencia lo llamaremos "I".

$$I = \sum_1^{\infty} p(n,R) \quad (11)$$

Además,

$$R = p(0,R) + I \quad (12)$$

De modo que, determinándose  $I$ , se puede calcular  $p(0,R)$ , que es la probabilidad de que el sistema se encuentre sin clientes y que, además, esté fuera de servicio.

Parte del tiempo el sistema estará vacío (probabilidad  $p(0)$ ), parte del tiempo el canal estará atendiendo ( $H$ ) y parte del tiempo habrá interferencia ( $I$ ), tal como se muestra de manera esquemática en la FIGURA 2:

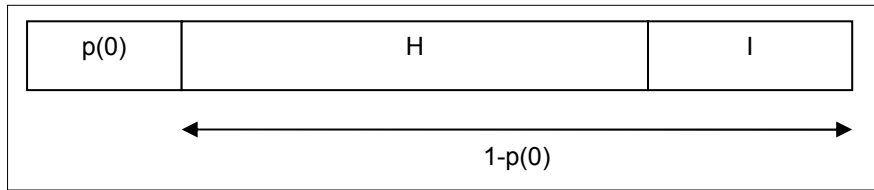


FIGURA N° 2

Esto significa que:

$$p(0) = 1 - I - H \quad (13)$$

Las condiciones anteriores son válidas para cualquier modalidad de intermisión (reanudación, reiniciación, abandono, interrupción relativa), dado que las interrupciones se dan con independencia del estado del sistema, y se tendrán en cuenta para la formulación de los modelos matemáticos siguientes.

## 2. SISTEMAS CON ABANDONO PARCIAL

Consideraremos primero el caso de un sistema en el cual hay deserción solamente del cliente que se está atendiendo en el canal cuando se produce la interrupción del servicio, pero no del resto de los que están esperando en la cola. Es decir, hay intolerancia total por parte del usuario en atención, e intolerancia nula por parte del resto de los clientes. Supondremos, adicionalmente, que no hay impaciencia ni intolerancia por parte de los usuarios que arriban.

Un ejemplo de abandono parcial es el de una unidad productiva constituida por una máquina (canal) que recibe piezas (clientes) para ser procesadas en forma individual, con gran capacidad de almacenamiento en el sector (cola ilimitada), en donde se debe descartar toda pieza cuya elaboración se haya visto interrumpida (por ej. por desperfecto de la máquina, por interrupción de energía, etc.).

En la FIGURA 3 se muestran esquemáticamente los flujos de los clientes (en flechas dobles) y la ocurrencia de interrupciones (en flecha punteada). Teniendo en cuenta que no hay impaciencia, ni restricciones de capacidad, tendremos que la tasa efectiva de ingresos al sistema es:

$$\bar{\lambda} = \lambda \quad (14)$$

La cantidad de clientes que por unidad de tiempo termina la atención es  $\mu \cdot H$ . La cantidad de clientes que abandonan el canal porque se interrumpió el servicio cuando estaban siendo atendidos es  $\lambda_R \cdot H$ . Por lo tanto, el flujo total de clientes que se retiran del sistema es:

$$\bar{\mu} = \mu \cdot H + \lambda_R \cdot H \quad (15)$$

que es igual al flujo ingresante de clientes. Luego, igualando (14) y (15):

$$\lambda = \mu \cdot H + \lambda_R \cdot H$$

Despejando, el porcentaje de atención del canal es:

$$H = \frac{\lambda}{\mu + \lambda_R} \quad (16)$$

La proporción de clientes cuyo servicio se completa, es obviamente:

$$p_{AT} = \frac{\mu H}{\lambda} = \frac{H}{\rho} = \frac{\mu}{(\mu + \lambda_R)}$$

y el porcentaje de clientes que abandonan el sistema sin completarlo:

$$p_{AB} = \frac{\lambda_R \cdot H}{\lambda} = \frac{\lambda_R}{(\mu + \lambda_R)}$$

El tiempo promedio que un cliente permanece en el canal (haya finalizado o no el servicio) está dado por la relación entre el porcentaje de ocupación activa (H) y el flujo de clientes que pasan por él. Es decir:

$$T_P = \frac{H}{\lambda} = \frac{1}{\lambda_R + \mu} \quad (17)$$

Esto es así, ya que el tiempo que un cliente permanece en el canal es el tiempo mínimo que transcurre entre el tiempo de un servicio y el tiempo de un arribo (Stewart, 2009), siendo las distribuciones de estas dos variables distribuciones exponenciales<sup>2</sup>.

<sup>2</sup> El "valor mínimo" de "n" variables independientes con distribución exponencial de medias  $1/\lambda_i$  tiene también distribución exponencial y su media es:

$$T = \frac{1}{\lambda} = \frac{1}{\sum_{i=1}^n \lambda_i}$$

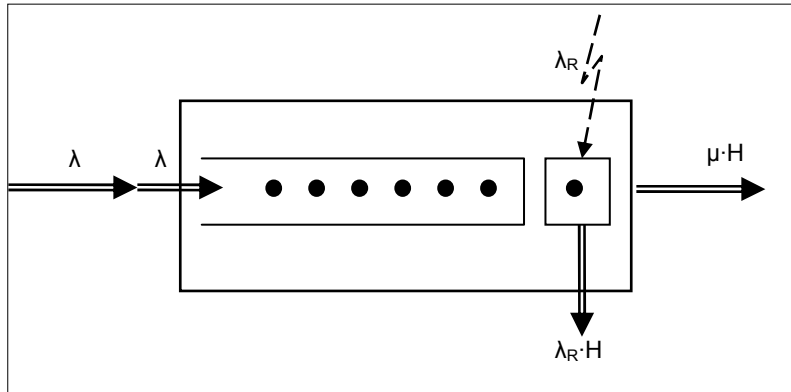


FIGURA N° 3

Consideraremos ahora la propiedad PASTA de los procesos Poisson (Gross y Harris, 1988). Cuando un cliente llega al sistema, el tiempo promedio que permanecerá desde que arriba hasta que se retira (por atención o abandono) está comprendido por:

- El tiempo de espera por la ocupación del canal de todos los clientes que están por delante de él:  $L \cdot T_P$
- El tiempo de su propia atención en el canal:  $T_P$
- El tiempo de espera por la probable interrupción del sistema en el instante de ingreso del cliente:  $T_{SR} \cdot R$
- El tiempo de las interrupciones que se producen mientras el cliente permanece en la cola (descontando el tiempo que tuvo que esperar por la probable interrupción al momento del arribo):

$$(W_C - T_{SR} \cdot R) \cdot \lambda_C \cdot T_{SR}$$

En consecuencia, la expresión del tiempo de permanencia de un cliente en el sistema es:

$$W = L \cdot T_P + T_P + T_{SR} \cdot R + (W_C - T_{SR} \cdot R) \cdot \lambda_C \cdot T_{SR} \quad (18)$$

Considerando que

$$W_C = W - T_P$$

tendremos:

$$W = L \cdot T_P + T_P + T_{SR} \cdot R + (W - T_P - T_{SR} \cdot R) \cdot \lambda_C \cdot T_{SR}$$



$$W = L \cdot T_P + T_P + T_{SR} \cdot R + W \cdot \lambda_C \cdot T_{SR} - T_P \cdot \lambda_C \cdot T_{SR} - T_{SR} \cdot R \cdot \lambda_C \cdot T_{SR}$$

Recordando, por (4) que  $\lambda_C \cdot T_{SR} = R$ , siendo  $R = 1 - F$ , y reemplazando:

$$W = L \cdot T_P + T_P + T_{SR} \cdot R + W \cdot (1 - F) - T_P \cdot (1 - F) - T_{SR} \cdot R \cdot (1 - F)$$

Por otra parte, teniendo en cuenta (14) y considerando la ley de Little, que establece que  $L = W \cdot \bar{\lambda}$ , tendremos:

$$W = \lambda \cdot W \cdot T_P + T_P + T_{SR} \cdot R + W - W \cdot F - T_P + T_P \cdot F - T_{SR} \cdot R \cdot (1 - F)$$

$$W \cdot (F - \lambda \cdot T_P) = T_P \cdot F + T_{SR} \cdot R - T_{SR} \cdot R \cdot (1 - F)$$

$$W = \frac{T_P \cdot F + T_{SR} \cdot (1 - F) \cdot F}{F - \lambda \cdot T_P}$$

$$\boxed{W = \frac{T_P \cdot F + T_{SR} \cdot (1 - F) \cdot F}{F - H}} \quad (19)$$

Considerando nuevamente la expresión de Little, el número promedio de clientes en el sistema será:

$$\boxed{L = \frac{T_P \cdot F + T_{SR} \cdot (1 - F) \cdot F}{F - H} \cdot \lambda} \quad (20)$$

El resto de las variables características:

$$\boxed{L = L_C - H} \quad (21)$$

$$\boxed{W_C = W - T_P} \quad (22)$$

Las probabilidades de estado  $p(n,F)$ ,  $p(n,R)$  y, en consecuencia,  $p(n)$ , pueden determinarse a partir del análisis de la cadena markoviana representada en la FIGURA 4.

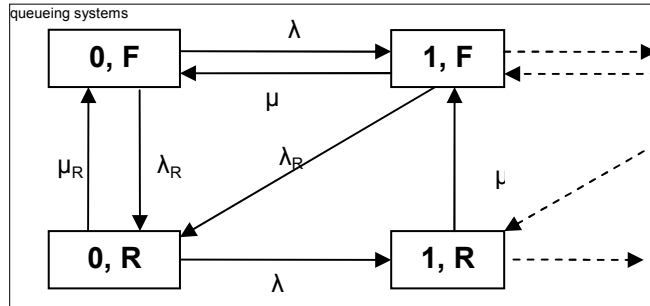


FIGURA N° 4

$$(\text{Nodo } 0, F): p(0, F) \cdot \lambda + p(0, F) \cdot \lambda_R = p(0, R) \cdot \mu_R + p(1, F) \cdot \mu$$

$$(\text{Nodo } 0, R): p(0, R) \cdot \lambda + p(0, R) \cdot \mu_R = p(0, F) \cdot \lambda_R + p(1, F) \cdot \lambda_R$$

Sumando miembro a miembro estas dos ecuaciones y simplificando:

$$p(0, F) \cdot \lambda + p(0, R) \cdot \lambda = p(1, F) \cdot (\mu + \lambda_R)$$

Luego:

$$p(1, F) = p(0, F) \cdot \frac{\lambda}{\mu + \lambda_R} + p(0, R) \cdot \frac{\lambda}{\mu + \lambda_R}$$

Es decir:

$$p(1, F) = p(0) \cdot \frac{\lambda}{\mu + \lambda_R}$$

y considerando (16):

$$p(1, F) = p(0) \cdot H$$

Del mismo modo, para cualquier estado "n", tendremos que:

$$p(n, F) \cdot \lambda + p(n, R) \cdot \lambda = p(n+1, F) \cdot (\mu + \lambda_R)$$

Entonces:

$$\boxed{p(n, F) = p(n-1) \cdot H} \quad (23)$$

Por su parte, balanceando el nodo (0, R):

$$p(0, R) \cdot (\lambda + \mu_R) = p(0, F) \cdot \lambda_R + p(1, F) \cdot \lambda_R$$

$$p(0, R) \cdot (\lambda + \mu_R) = p(0, F) \cdot \lambda_R + p(0) \cdot H \cdot \lambda_R$$

$$p(0, R) \cdot (\lambda + \mu_R) = p(0, F) \cdot \lambda_R + p(0, R) \cdot H \cdot \lambda + p(0, F) \cdot H \cdot \lambda_R$$

$$p(0, R) \cdot (\lambda + \mu_R - H \cdot \lambda) = p(0, F) \cdot (\lambda_R + H \cdot \lambda_R)$$

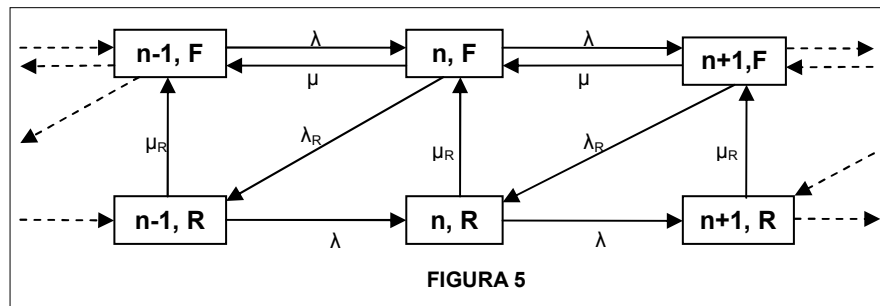
Pero como  $p(0, F) = F - H$ , tendremos:

$$p(0, R) \cdot (\lambda + \mu_R - H \cdot \lambda) = (F - H) \cdot (\lambda_R + H \cdot \lambda_R)$$

Entonces:

$$p(0) = F - H + \frac{(F - H) \cdot (\lambda_R + H \cdot \lambda_R)}{(\lambda + \mu_R - H \cdot \lambda)} \quad (24)$$

De esta forma, la probabilidad de que el sistema esté vacío queda formulada en función de los parámetros, como así también, a partir de (23), la probabilidad de que el sistema se encuentre en un estado genérico  $p(n, F)$ . Asimismo, analizando la cadena de Markov de la FIGURA 5 se puede determinar  $p(n, R)$  haciendo el balance probabilístico correspondiente.



$$p(n, R) = \frac{p(n-1, R) \cdot \lambda + p(n+1, F) \cdot \lambda_R}{(\lambda + \mu_R)} \quad (25)$$

Finalmente, el porcentaje de interferencia “I”, se puede determinar a partir de la expresión (13) tal como puede verse claramente en la FIGURA 6, y considerando las expresiones (16) y (24).

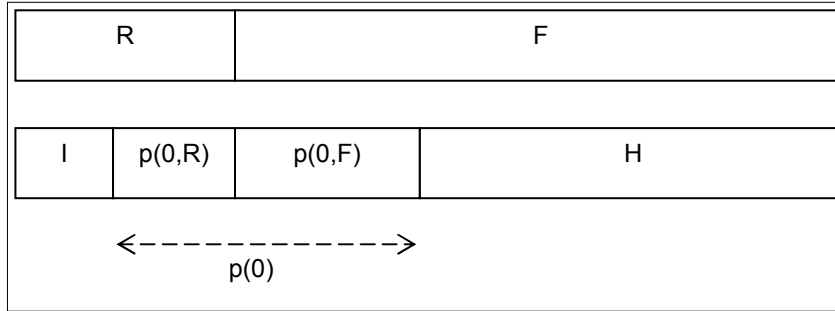


FIGURA N° 6

Se puede observar que cuando no hay interrupciones, es decir  $F = 1$ , tendremos que  $T_P = T_S = 1/\mu$  y  $\lambda_R = 0$ , con lo que las expresiones (19), (20), (21), (22) y (24) serían las correspondientes a las del modelo básico PP1:

$$W = \frac{1}{\mu - \lambda} \quad L = \frac{\lambda}{\mu - \lambda} \quad L_C = \frac{\lambda}{\mu - \lambda} - \rho \quad W_C = W - T_S$$

$$p(0) = 1 - \rho$$

### 3. ABANDONO TOTAL, CON INTOLERANCIA A LA INTERRUPCIÓN

Consideraremos ahora el caso de un sistema con abandono de todos los clientes que están en el sistema (ya sea atendándose o esperando en cola) cuando se produce la parada. Asimismo, asumiremos que los clientes que arriban al sistema durante el cese del servicio no ingresarán a él. Es decir, todos los usuarios presentan intolerancia total a las interrupciones. Éste podría ser el caso de un cajero automático de un banco que cuando sale de servicio típicamente se retiran todos los clientes y no hay nuevos ingresos mientras el equipo esté en reparación.

La FIGURA 7 muestra esquemáticamente los flujos de clientes en flechas dobles y la ocurrencia de interrupciones en flecha punteada. Teniendo en cuenta que hay intolerancia absoluta al hecho de que el sistema se encuentre fuera de operación, tendremos que la tasa de ingresos al sistema es igual a  $\lambda$  cuando el canal se encuentra funcionando (estados F) e igual a 0 cuando el canal está fuera de servicio (estados R). Por lo tanto, el número promedio de clientes que ingresan al sistema por unidad de tiempo será igual a la tasa de arribo multiplicada por la probabilidad de que el sistema esté en funcionamiento, es decir:

$$\bar{\lambda} = \lambda \cdot F \tag{26}$$

La cantidad de clientes por unidad de tiempo que terminan la atención es  $\mu \cdot H$ , mientras que la tasa de clientes que abandonan el sistema debido a que se interrumpió el servicio es  $\lambda_R \cdot L$ . Por lo tanto, el flujo total de clientes que se retiran del sistema es:

$$\bar{\mu} = \mu \cdot H + \lambda_R \cdot L \quad (27)$$

que debe ser igual, en condiciones de régimen permanente, al flujo ingresante de clientes. Por lo tanto:

$$\lambda \cdot F = \mu \cdot H + \lambda_R \cdot L \quad (28)$$

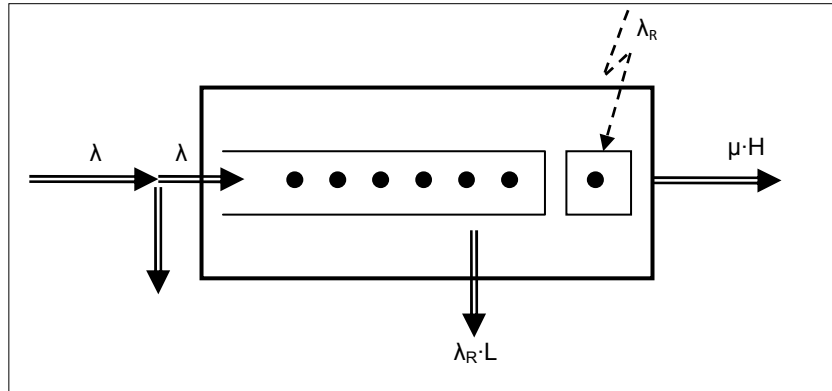


FIGURA N° 7

La condición de equilibrio para este sistema es:

$$H < F \quad (29)$$

En este caso, dado que estamos suponiendo que los clientes que arriban tienen intolerancia absoluta mientras el sistema está fuera de servicio, tendremos que

$$R = p(0,R) \quad (30)$$

En el gráfico de la FIGURA 8 se muestra esquemáticamente la relación entre los porcentajes de tiempo en funcionamiento y las probabilidades de que el sistema esté vacío  $p(0)$ , vacío no operativo  $p(0,R)$  y vacío operativo  $p(0,F)$ .

El porcentaje de clientes atendidos cuyo servicio se completa es:

$$P_{AT} = \frac{\mu \cdot H}{\lambda \cdot F} \quad (31)$$

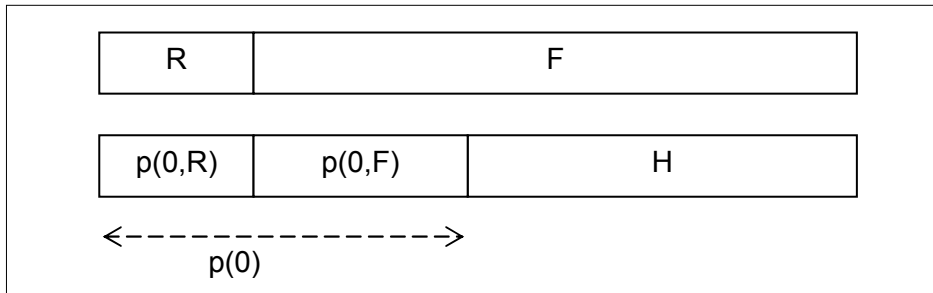


FIGURA N° 8

El porcentaje de clientes que abandonan sin completar la atención es:

$$P_{AB} = \frac{L \cdot \lambda_R}{\lambda \cdot F}$$

$$P_{AB} = \frac{\lambda \cdot W \cdot \lambda_R}{\lambda \cdot F} = \frac{W \cdot \lambda_R}{F} \quad (32)$$

La probabilidad de que el sistema se encuentre en el estado (0,F) es

$$p(0,F) = F - H$$

El tiempo promedio que un cliente permanece en el canal (haya finalizado o no el servicio) está dado por la relación entre el porcentaje de ocupación activa (H) y el flujo de clientes que pasa por él. Es decir:

$$T_P = \frac{H}{\lambda} = \frac{H}{\lambda \cdot F} \quad (33)$$

A esta misma expresión se arriba considerando que el tiempo promedio de permanencia en el canal es igual al porcentaje de clientes que se atienden en el mismo (31) multiplicado por el tiempo promedio de la atención:

$$T_P = \frac{\mu \cdot H}{\lambda \cdot F} \cdot \frac{1}{\mu} = \frac{H}{\lambda \cdot F}$$

La longitud promedio de clientes en el sistema, cuando el sistema se encuentra en el estado F es:

$$L_F = \frac{L}{F} \quad (34)$$

El tiempo promedio de permanencia de un cliente que recién ingresa al sistema está comprendido por:

- El tiempo de espera de los clientes que están esperando en el sistema (dado que el sistema se encuentra funcionando):  $L_F \cdot T_P$
- El tiempo de espera promedio de su propia permanencia en el canal:  $T_P$

Luego:

$$W = L_F \cdot T_P + T_P \quad (35)$$

Teniendo en cuenta la fórmula de Little:

$$W = \frac{L}{\lambda} = \frac{L}{\lambda \cdot F}$$

y operando:

$$W = \frac{L}{F} \cdot \frac{H}{\lambda \cdot F} + \frac{H}{\lambda \cdot F} = \frac{W \cdot H}{F} + \frac{H}{\lambda \cdot F} = \frac{W \cdot H \cdot \lambda}{\lambda \cdot F} + \frac{H}{\lambda \cdot F}$$

$$W \cdot \lambda \cdot F = W \cdot \lambda \cdot H + H$$

$$W \cdot \lambda \cdot (F - H) = H$$

$$\boxed{W = \frac{H}{(F - H) \cdot \lambda}} \quad (36)$$

Teniendo en cuenta el principio de Little:

$$\boxed{L = \frac{H \cdot F}{(F - H)}} \quad (37)$$

Además:

$$\boxed{L = L_C - H} \quad (38)$$

$$\boxed{W_C = W - T_P} \quad (39)$$

Luego, faltaría determinar el valor de H. Para ello, reemplazando en (28) la expresión (37) de la longitud promedio del sistema y operando:

$$\lambda \cdot F = \mu \cdot H + \lambda_R \cdot \frac{F \cdot H}{(F - H)}$$

$$\lambda \cdot F = \frac{\mu \cdot H \cdot (F - H) + \lambda_R \cdot F \cdot H}{(F - H)}$$

$$\begin{aligned} \lambda \cdot F \cdot (F - H) &= \mu \cdot H \cdot (F - H) + \lambda_R \cdot F \cdot H \\ \lambda \cdot F^2 - \lambda \cdot F \cdot H &= \mu \cdot F \cdot H - \mu \cdot H^2 + \lambda_R \cdot F \cdot H \\ \mu \cdot H^2 - (\lambda \cdot F \cdot H + \lambda_R \cdot F \cdot H + \mu \cdot F \cdot H) + \lambda \cdot F^2 &= 0 \\ \mu \cdot H^2 - (\lambda + \lambda_R + \mu) \cdot F \cdot H + \lambda \cdot F^2 &= 0 \end{aligned}$$

Resolviendo la ecuación cuadrática:

$$H_{1,2} = \frac{\lambda + \lambda_R + \mu \pm \sqrt{(\lambda + \lambda_R + \mu)^2 - 4 \cdot \mu \cdot \lambda \cdot F^2}}{2 \cdot \mu} \quad (40)$$

De las dos soluciones, se debe considerar solamente aquella cuyo valor satisface la condición de equilibrio de este sistema, es decir  $H > F$ , y que resulta de considerar el signo menos de la raíz.

$$H = \frac{\lambda + \lambda_R + \mu - \sqrt{(\lambda + \lambda_R + \mu)^2 - 4 \cdot \mu \cdot \lambda \cdot F^2}}{2 \cdot \mu} \quad (41)$$

Reemplazando, entonces, el valor de H en las expresiones (36) a (38), quedan determinados los valores del resto de las variables características.

Cabe destacar que es fácilmente deducible la siguiente relación entre probabilidades de estado:

$$p(n, F) = (F - H) \cdot \left(\frac{H}{F}\right)^n \quad (42)$$

que también se puede expresar de la siguiente forma:

$$p(n + 1, F) = p(n, F) \cdot \frac{H}{F} \quad (43)$$

La cadena markoviana para este sistema se indica en la FIGURA 9.

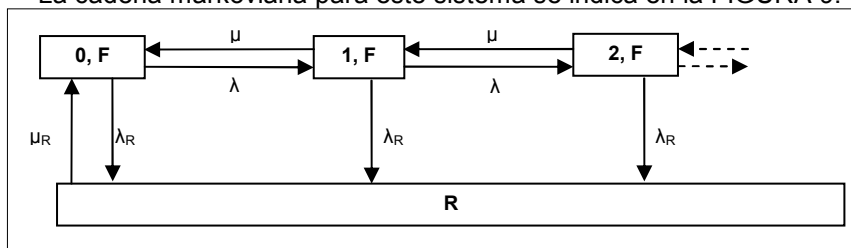


FIGURA N° 9



Se observa que, para el caso de que no hubiera interrupciones, es decir para  $F = 1$ , tendríamos que  $T_p = T_s = 1/\mu$  y  $\lambda_R = 0$ , por lo que estaríamos en el caso del modelo básico PP1 y las expresiones (36) a (39) serían las correspondientes a dicho modelo. Se debe considerar que, en esta situación, la expresión de H es la que toma el signo positivo de la raíz en la ecuación (40), que es la correspondiente a  $H < 1$ , que finalmente termina siendo  $H = \rho$ .

Tomando como ejemplo los siguientes datos:  $\lambda = 10$ ,  $\mu = 20$ ,  $\lambda_R = 3$ ,  $\mu_R = 9$ , y aplicando las expresiones arriba indicadas, en la TABLA 1 se indican los resultados obtenidos para los dos modelos arriba desarrollados.

**TABLA 1: Resultados obtenidos para  $\lambda = 10$ ,  $\mu = 20$ ,  $\lambda_R = 3$ ,  $\mu_R = 9$**

	F	R	H	I	p(0,F)	p(0,R)	p(0)
Abandono Parcial	0,75	0,25	0,4348	0,1733	0,3152	0,0767	0,3919
Abandono Total	0,75	0,25	0,3	0	0,45	0,25	0,70

	$T_p$	W	$W_c$	L	$L_c$	$\bar{\lambda}$
Abandono Parcial	0,0435	0,1695	0,1261	1,6954	1,2606	10,0
Abandono Total	0,04	0,0667	0,0267	0,50	0,20	7,50

#### 4. EXTENSIÓN A SISTEMAS MULTICLASES CON PRIORIDADES

Las expresiones de los enunciados pueden ser utilizadas para resolver sistemas con clientes que pertenezcan a clases distintas, con prioridades diferentes (Cobham, 1954) y que además sean intolerantes. La prioridad absoluta se da cuando un cliente de alta prioridad que arriba al sistema desplaza del canal a un cliente de categoría más baja. La intolerancia se manifiesta en los clientes desplazados.

Supongamos un sistema PP1 de prioridad absoluta con clientes "A", de alta prioridad y "B", de baja prioridad. Para los clientes de clase "A" la existencia en el sistema de clientes de clase "B" es totalmente indiferente. El subsistema de clientes "A" consiste, entonces, en un sistema PP1 convencional. El subsistema de clientes "B" es un sistema PP1 con interrupción.

Recordemos que para un sistema PP1 la expresión del "tiempo de ocupación" (es decir, el período promedio de tiempo que transcurre desde que un cliente ingresa al sistema cuando está vacío hasta que el sistema se vacía nuevamente) está dada por la expresión  $1/(\mu-\lambda)$  y que el "tiempo ocioso" (es decir, el período promedio de tiempo que transcurre desde que el sistema queda vacío hasta que llega un cliente para ser atendido es igual al intervalo promedio entre arribos ( $1/\lambda$ )).

Para los clientes de categoría “B”, el arribo de un cliente “A” implica una interrupción, por lo que el período de la interrupción ( $T_{RS_B}$ ) será igual al “tiempo de ocupación” del subsistema de clientes “A”, mientras que el porcentaje de tiempo durante el cual el subsistema “B” está no operativo ( $R_B$ ) es igual al porcentaje de ocupación del subsistema A de clientes ( $H_A = \rho_A$ ).

$$T_{RS_B} = \frac{1}{\mu_A - \lambda_A}$$

$$R_B = \frac{\lambda_R}{\lambda_R + \mu_R} = \rho_A$$

Por su parte, el porcentaje de tiempo de funcionamiento ( $F_B$ ) será igual al porcentaje de tiempo ocioso de los clientes “A”, y el intervalo de tiempo que transcurre (en promedio) hasta la próxima interrupción ( $T_{R_B}$ ) igual a  $T_A$ .

$$F_B = 1 - \rho_A = \frac{\mu_R}{\lambda_R + \mu_R}$$

$$T_{R_B} = \frac{1}{\lambda_A}$$

De esta forma se pueden calcular los valores característicos correspondientes a los clientes de la categoría “B”, tal como se mostró en los modelos con interrupción inmediata arriba desarrollados.

## 5. 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, tal como surge de las expresiones (18) y (35), conduce a una expresión analítica de las variables características. Este enfoque proporciona una solución cuantitativa sin tener que resolver la serie geométrica:

$$L = \sum_1^{\infty} n \cdot p(n)$$

Sobre funciones generadoras de probabilidad de estado (Eisen y Taineter, 1963), cuya resolución es de muy alta complejidad (Bolch *et al.*, 2006), a la vez que se evita recurrir a la simulación, con sus desventajas en términos de precisión y tiempo de resolución.

Asimismo, la metodología puede aplicarse para abordar los complejos casos de sistemas multiclases con prioridades y abandono, por su analogía con los modelos de sistemas con interrupciones (White y Christie, 1958).

## 6. 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, A. (1954): "PRIORITY ASSIGNMENT IN WAITING LINE PROBLEMS", Editorial INFORMS, Journal of Operations Research, Vol. 2, págs. 70-76.
- EISEN, M.; TAINETER, M. (1963): "STOCHASTIC VARIATIONS IN QUEUEING PROCESSES", Journal of Operations Research, Vol. 6, págs. 922-927.
- GROSS, D.; HARRIS, C. M. (1988): "FUNDAMENTALS OF QUEUEING THEORY". Editorial: John Wiley & Sons, Inc, USA, 439 págs.
- MIRANDA, M. (2013): "TEORÍA DE COLAS", 2ª edición. Editorial: EDUCA, Buenos Aires, 284 págs.
- 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.