

Contents

1	Teoria de Colas	2
1.1	Introducción	2
1.1.1	Conceptos basicos	2
1.1.2	Notación Kendall	6
1.1.3	Ecuacion de estado de regimen permante	6
1.1.4	Ecuacion simplificada de estado de regimen permante	6
1.1.5	Preguntas y respuestas	7
1.2	Modelo $P/P/1$	8
1.3	Modelo $P/P/1/N$	11
1.4	Modelo $P/P/M$	12
1.5	Modelo $P/P/M/N$	13

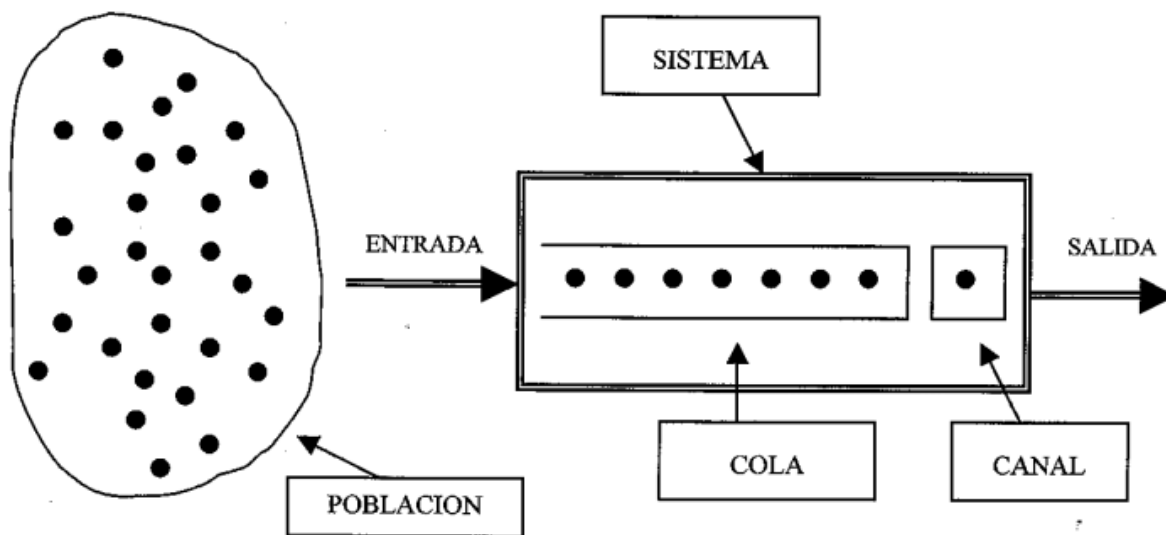
1 Teoria de Colas

1.1 Introducción

1.1.1 Conceptos basicos

Los fenómenos de *congestión* o *espera* estan relacionados con los sistemas estocasticos y pueden describirse como sistemas integrados por uno o mas *centros de atención* donde se brinda un servicio. Cada centro de atención es, a su vez, un sistema constituido por:

- *canales (ó servidores)*: Entidades que prestan el servicio.
- *clientes (ó usuario)*: Entidades que reciben el servicio.



Las colas se forman cuando la demanda de un servicio dado en un intervalo de tiempo excede la capacidad para proveerlo. El administrador del sistema de establecer un balance apropiado entre los costos asociados a la espera de los usuarios y los costos vinculados con la mejora del servicio (mas servidores, mayor velocidad de atención, etc). En la mayoría de los procesos de atención, los tiempos entre arribos de clientes y los tiempos de los servicios no son predecibles. En estas condiciones se aplica la denominada *teoría de colas* para determinar el comportamiento del sistema bajo diferentes alternativas. Se estudiaran aquellos sistemas que describan los procesos mas generales y que son los que pueden formularse como *cadenas markovianas de primer orden*. Se analizaran los sistemas en regimen permanente, a través de variables tales como la *longitud promedio de la cola*, el *tiempo de espera promedio* del cliente para recibir el servicio, el *tiempo de permanencia* en el sistema, etc. Esta información, junstamente con los costos relevantes, permitira al directivo determinar los valores apropiados de las variables de decisión. Las variables de decisión tipicas en los sistemas de colas estan referidas a la *capacidad de servicio*(numero de canales, velocidad de canales) ó la *capacidad de espera* (número de lugares).

La **población** es el conjunto de usuario potenciales del sistema. Puede ser finito o infinito.

En los **arribos** la llegada de los clientes puede ser deterministica o aleatoria. A menudo los intervalos entre llegadas son estadisticamente independientes y estacionarios a lo largo de prolongados periodos de tiempo, por lo que se puede suponer poissonianos.

SISTEMAS	EJEMPLO	CLIENTES	CANALES
TRANSPORTE	<ul style="list-style-type: none"> • PUERTOS • AEROPUERTOS • AUTOPISTAS • TERMINALES • CARGA Y DESCARGA 	BARCOS AVIONES AUTOMÓVILES ÓMNIBUS CAMIONES	MUELLES PISTAS CASILLAS DE PEAJE PLATAFORMA ISLAS
COMPUTACIÓN	<ul style="list-style-type: none"> • PROCESAMIENTO • IMPRESIÓN 	TRABAJOS TRABAJOS	CPU IMPRESORAS
TELEFONÍA	<ul style="list-style-type: none"> • CENTRALES 	LLAMADOS	LÍNEAS
PRODUCCIÓN	<ul style="list-style-type: none"> • ELABORACIÓN • REPARACIÓN • CONTROL DE CALIDAD 	PRODUCTOS SOLICITUDES PRODUCTOS	MAQUINAS OPERARIOS INSPECTORES
ATENCIÓN AL PÚBLICO	<ul style="list-style-type: none"> • BANCOS • SUPERMERCADOS • ESTACIONES DE SERVICIO • NEGOCIOS • ATENCIÓN MÉDICA • ALQUILER DE AUTOS 	PERSONAS PERSONAS AUTOMÓVILES PERSONAS PACIENTES PERSONAS	CAJEROS CAJEROS SURTIDORES EMPLEADOS AMBULANCIAS EMPLEADOS
ADMINISTRACIÓN	<ul style="list-style-type: none"> • DPTO. COMPRAS • JUZGADOS 	SOLICITUDES CAUSAS	COMPRADORES JUECES

TABLA 1.1 Ejemplos de sistemas de colas

La **impaciencia** se verifica cuando algunos usuarios que arriban al sistema se retiran sin recibir el servicio porque consideran que el tiempo de espera sera suficientemente largo. Se distinguen dos tiempos de impaciencia:

1. **Rechazo (\bar{R})**: Un cliente que arriba, observa la cantidad de gente que esta delante de él esperando y en función de ello toma la decisión de incorporarse o no al sistema. **En la materia se trata unicamente este tipo de impaciencia.**
2. **Abandono (\bar{A})**: Un cliente que arriba, ingresa al sistema y al cabo de un tiempo toma la decisión de seguir esperando o no.

La **capacidad** es el número máximo de clientes que puede permanecer en el sistema simultaneamente(en espera y atendándose).

Para el **modo de arribo** los usuarios pueden llegar en forma individual o en masa(modos batch). En la mayoría de los sistemas que estudiaremos se hará la suposición de que los procesos de llegada son del tipo Poisson, lo que implica *arribos individuales*. Se puede considerar el arribo de grupos como clientes individuales.

Para la **prioridad de atención**, existen diversos criterios de atención en lo que se refiere al orden de selección de clientes para brindar el servicio. Ellos son:

- Base FIFO(first in, first out): Los clientes se atienden según el orden de llegada.
- Base LIFO(last in, first out): El último individuo que arriba es el primero en ser atendido.
- Base SIRO(service in random order): Es una selección aleatoria de los clientes para brindales el servicio.
- Base con PRIORIDADES: Se establecen criterios de atención conforme a los atributos de los clientes.

La **duración del servicio** es el tiempo requerido por un canal para atender un cliente. Puede ser una variable determinística o aleatoria con distribución de probabilidad conocida.

En el **modo de atención** un canal puede servicio de forma individual o múltiples(en masa). En la mayoría de los sistemas reales, el modo de atención es individual.

Los **procesos poisson** son markovianos y tienen dos distribuciones que lo describen:

1. La distribución Poisson, en donde la variable es el número de eventos que se producen en un intervalo determinado de continuo.
2. La distribución Gamma, en la que la variable es el intervalo de continuo necesario para que se verifique un número determinado de eventos. Particularmente, cuando la variable es el intervalo de tiempo de continuo necesario para que se verifique *un solo evento*, la distribución es conocida como **distribución Exponencial**.

Distribución Poisson La probabilidad de que se produzcan "n" eventos en un intervalo "t" está dada por:

$$p(n) = \frac{(\lambda \Delta t)^n \cdot e^{-\lambda \Delta t}}{n!} \quad (1)$$

siendo $n = 1, 2, \dots$ y $\lambda > 0$ La media de esta distribución es $a = \lambda \cdot t$ y el desvío estándar $\sigma = \sqrt{\lambda t}$.

Distribución Gamma La función distribución de probabilidad de la **distribución exponencial** está dada por la siguiente expresión:

$$f(t) = \lambda \cdot e^{-\lambda t} \quad (2)$$

cuya media es $\frac{1}{\lambda}$ y cuyo desvío estándar es $\frac{1}{\lambda}$

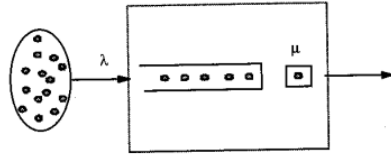
Si el proceso de arribos es de tipo Poisson significa que la variable "tiempo entre dos arribos sucesivos" tiene distribución *exponencial* y la variable "numero de clientes que arriban por unidad de tiempo" tiene distribución *Poisson*.

Ingresos y egresos de clientes: En los sistemas de capacidad finita o de población impaciente, no todos los clientes que arriban al sistema ingresan.

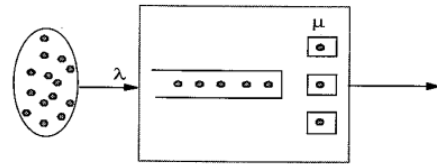
- $\bar{\lambda}$: Número promedio de clientes que ingresan efectivamente al sistema.
- \bar{R} : Número promedio de rechazados(es decir, que no ingresan al sistema).
- $\bar{\mu}$: Número promedio de clientes atendidos que egresan del sistema.
- \bar{A} : Tasa de clientes que ingresaron al sistema pero que decidieron abandonarlo sin recibir el servicio.

Estructuras de sistemas simples:

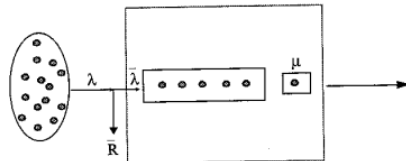
1. Cola simple, capacidad infinita, un canal de atención:



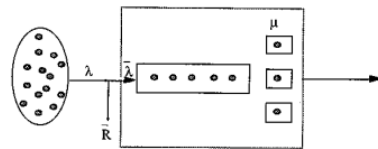
3. Cola simple, capacidad infinita, canales múltiples en paralelo:



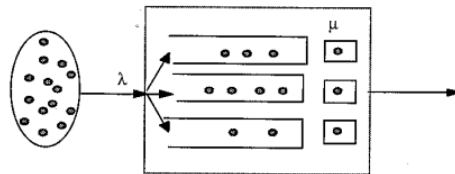
2. Cola simple, capacidad finita, un canal de atención:



4. Cola simple, capacidad finita, canales múltiples en paralelo:



5. Varias colas, una para cada canal dispuesto en paralelo:



1.1.2 Notación Kendall

Especifica las características descriptivas de una unidad operativa de un sistema de colas. Es una notación de 6 posiciones:

$$1/2/3/4/5/6 \quad (3)$$

1. La posición 1 se refiere al patrón de arribos al sistema. Puede ser:
 - P: Proceso de Poisson.
 - D: Proceso determinístico.
 - G: Cualquier otro proceso.
2. La posición 2 indica el patrón de servicio en los canales.
 - P: Proceso de Poisson.
 - D: Proceso determinístico.
 - G: Cualquier otro proceso.
3. La posición 3 indica el número de canales de atención dispuestos en paralelo en la unidad operativa.
4. La posición 4 indica la capacidad de la unidad operativa del sistema (Los que pueden esperar en la cola más los que se pueden atender). Se asume infinita y no se indica.
5. La posición 5 indica la prioridad de atención de la cola. Si no se especifica se asume de tipo FIFO.
 - FIFO
 - LIFO
 - SIRO
 - G: Cualquier otra modalidad de atención
6. La última posición (entre paréntesis) se refiere al tamaño de la población. Si no se indica, el tamaño es infinito.

1.1.3 Ecuación de estado de régimen permanente

$$0 = p(n-1) \cdot \lambda_{n-1} - p(n) \cdot [\lambda_n + \mu_n] + p(n+1) \cdot \mu_{n+1} \quad (4)$$

1.1.4 Ecuación simplificada de estado de régimen permanente

$$p(n) = p(n-1) \cdot \frac{\lambda_{n-1}}{\mu_n} \quad (5)$$

1.1.5 Preguntas y respuestas

- ¿Qué características posee la población?

Rta: Puede ser Finita o infinita

- ¿En qué consiste el fenómeno de impaciencia? ¿cuántos tipos de impaciencia hay? ¿en qué consisten?

Rta: Ver punto de impaciencia mas arriba.

- ¿Qué características posee el sistema?

Rta: Dependiendo de las hipotesis, puede uno o varios canales, con una o varias colas, y la cola puede ser finita o infinita

- ¿Qué se entiende por capacidad del sistema?

Rta: Es el numero máximo de clientes que pueden estar en todo el sistema

- ¿Con qué notación identificamos cada uno de los ítems enunciados?

- L_c : La cantidad promedio de clientes que están esperando para recibir el servicio en un determinado momento.
- L : La cantidad promedio de clientes que se encuentran en el sistema en un determinado momento, ya sea esperando ser atendidos como atendiéndose.
- W_c : El tiempo promedio que un cliente debe esperar para ser atendido.
- W : El tiempo promedio que un cliente permanece en el sistema, ya sea esperando ser atendido como atendiéndose.
- λ : La cantidad promedio de clientes que arriban al sistema en un determinado momento.
- $\bar{\lambda}$: la cantidad promedio de clientes que ingresan al sistema en un determinado momento.
- $\bar{\mu}$: La cantidad promedio de clientes que egresan del sistema luego de ser atendidos.
- μ : La velocidad promedio de atención de un canal.
- ρ : Factor de transito (o de trafico).

$$\rho = \frac{\lambda}{\mu} \quad (6)$$

- T_a : El tiempo promedio entre arribos.

$$T_a = \frac{1}{\lambda} \quad (7)$$

- T_s : El tiempo promedio del servicio.

$$T_s = \frac{1}{\mu} \quad (8)$$

- H : La cantidad promedio de canales ocupados.

- PA : El porcentaje de actividad de cada canal.

- ¿Cuál es la diferencia entre “la cantidad promedio de clientes que arriban al sistema en un determinado momento” y “la cantidad promedio de clientes que ingresan al sistema en un determinado momento”?

Rta: No todos los clientes que arriban al sistema, ingresan. Y por esos se diferencia los que arriban, y dentro de estos lo que finalmente ingresan.

- ¿Qué significa que se encuentren en régimen permanente o estacionario?

Cuando los valores de las variables no dependen de las condiciones iniciales del sistema.

- ¿qué se indica cada una de las posiciones de la notación Kendall?
Rta: Se describen mas arriba
- ¿Qué características posee el modelo $P/P/1$?
Se refiere a una unidad operativa de un sistema de colas con arribo Poisson, servicio Poisson, un canal, con capacidad infinita, en modalidad FIFO ya que no especifica y con población infinita
- ¿Cuánto debe valer ρ en un $P/P/1$? ¿Por qué?
- Según el modelo $P/P/1/N$ ¿Cuáles son sus características según la notación de Kendall?
- En el modelo $P/P/1/N$, ¿debe verificarse lo mismo que en el $P/P/1$ respecto del valor de ρ ? ¿Por qué?
- ¿Qué ejemplos de la vida cotidiana se podrían plantear en cada uno de los modelos mencionados?

1.2 Modelo $P/P/1$

Sistemas de un solo canal, capacidad infinita y poblacion infinita HIPOTESIS

- El *proceso de arribos* de clientes es de tipo **Poisson**.
- El *proceso de servicio* tambien es de tipo **Poisson**.
- El sistema se encuentra en condiciones estables.
- Los clientes que llegan al sistema forman una cola simple.
- La diciplina de atención es FIFO.
- Se dispone de **un solo canal** de atención.
- La capacidad del sistema es ilimitada.
- Los clientes que llegan al sistema **no presentan impaciencia**.
- La *población de clientes* potenciales del sistema es **infinita**.

FORMULAS para el modelo $P/P/1$

- Determinacion de probabilidades.

Dado que no hay impaciencia ni restricciones de capacidad en el sistema, la probabilidad de ingresar es 1 para cualquier "n", por lo tanto la tasa de arribos promedio λ_n es siempre igual a la tasa de arribos promedio λ .

La tasa de egresos es μ_n es cero cuando el sistema esta vacio ($n=0$). Para cualquier otro valor de "n", la tasa esperada de egresos es igual a la velocidad promedio de atención, es decir, μ .

$$\lambda_n = \lambda \quad si \quad n = 1, 2, 3, \dots$$

$$\mu_n = \begin{cases} 0 & si \quad n = 0 \\ \mu & si \quad n = 1, 2, 3, \dots \end{cases}$$

Reemplazando en la ecuación simplificada de regimen en estado estacionario y mediante inducción para $n = 1, 2, 3$ se obtiene:

$$p(n) = \rho^n \cdot p(0) \tag{9}$$

$$p(0) = 1 - \rho \tag{10}$$

- Calculo de L : La longitud de sistema es una variable aleatoria cuya media (L) es el valor esperado de que en el sistema haya n clientes. Es decir:

$$L = \sum_0^{\infty} n \cdot p(n) \quad (11)$$

Remplazando la expresión de $p(n)$ y utilizando una serie geométrica, obtenemos:

$$L = \frac{\rho}{1 - \rho} = \frac{\lambda}{\mu - \lambda} \quad (12)$$

- Calculo de L_c : Esperanza de haya $n-1$ clientes en la cola, ya que solo hay un canal en el sistema.

$$L = \sum_1^{\infty} (n-1) \cdot p(n) \quad (13)$$

Distribuyendo y remplazando por L y ρ , se obtiene:

$$L_c = L - \rho = \frac{\rho^2}{1 - \rho} = \frac{\lambda^2}{\mu(\mu - \lambda)} \quad (14)$$

- Calculo de H : H es el número promedio de canales activos. Para el estado $n=0$ no hay canales ocupados, mientras que para cualquier otro estado, *hay un canal ocupado*. Entonces:

$$H = \sum_1^{\infty} p(n) = 1 - p(0) \quad (15)$$

Remplazando por el valor de ρ :

$$H = 1 - p(0) = \rho \quad (16)$$

- Porcentaje de actividad: Dado que la cantidad de canales es uno $M = 1$

$$PA = \frac{H}{M} = H = \rho \quad (17)$$

- Calculo de $\bar{\lambda}$:

$$\bar{\lambda} = \sum_0^{\infty} \lambda_n \cdot p(n)$$

Dado que λ_n vale λ para todo n , entonces:

$$\bar{\lambda} = \sum_0^{\infty} \lambda \cdot p(n) = \lambda \cdot \sum_0^{\infty} p(n) = \lambda \cdot 1 \Rightarrow \bar{\lambda} = \lambda \quad (18)$$

Esto se debe a que no hay restricciones de capacidad en el sistema y que los clientes no presentan el fenómeno de impaciencia. Por lo tanto, todo lo arriba al sistema, ingresa al sistema.

- Calculo de $\bar{\mu}$:

$$\bar{\mu} = \sum_0^{\infty} \mu_n \cdot p(n)$$

Dado μ_0 vale cero para $n = 0$ y que μ_n vale μ para $n \geq 1$, entonces:

$$\bar{\mu} = \sum_1^{\infty} \mu \cdot p(n) = \mu \cdot \sum_1^{\infty} p(n) = \mu \cdot [1 - p(0)] \quad (19)$$

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

Como el sistema se encuentra en equilibrio, $\bar{\lambda} = \bar{\mu}$, todo lo que entra sale, no perdemos ningún cliente en ningún lado.

- Calculo de W_c : Tiempo de espera en cola

$$W_c = \frac{L_c}{\bar{\lambda}} \quad (20)$$

$$W_c = \frac{\lambda}{\mu \cdot (\mu - \lambda)} \quad (21)$$

- Calculo de W : Tiempo de *permanencia* en el sistema.

$$W = W_c + T_c \quad (22)$$

$$W = \frac{L}{\bar{\lambda}} \quad (23)$$

Entonces reemplazando L y $\bar{\lambda}$:

$$W = \frac{1}{\mu - \lambda} \quad (24)$$

1.3 Modelo $P/P/1/N$

Sistemas de un solo canal, capacidad finita y poblacion infinita HIPOTESIS

- El *proceso de arribos* de clientes es de tipo **Poisson**.
- El *proceso de servicio* tambien es de tipo **Poisson**.
- El sistema se encuentra en condiciones estables.
- Los clientes que llegan al sistema forman una cola simple.
- La diciplina de atención es FIFO.
- Se dispone de **un solo canal** de atención.
- La capacidad del sistema es **finita**, esta limitada a un valor N .
- Los clientes que llegan al sistema **no presentan impaciencia**.
- La *población de clientes* potenciales del sistema es **infinita**.

FORMULAS para el modelo $P/P/1/N$

- Determinacion de probabilidades

La probabilidad de ingresar al sistema siempre es 1, excepto cuando el sistema esta completo ($n=N$), en cuyo caso la probabilidad es cero.

La tasa promedio de egresos μ_n , es cero cuando el sistema esta vacio ($n=0$). Para cualquier otro valor de "n", la tasa esperada de egresos es igual a la velocidad promedio de atención del canal, es decir, μ .

$$\lambda_n = \begin{cases} \lambda & \text{si } n = 0, 1, 2, 3, \dots \\ 0 & \text{si } n = N \end{cases}$$

$$\mu_n = \begin{cases} 0 & \text{si } n = 0 \\ \mu & \text{si } n = 1, 2, 3, \dots \end{cases}$$

Remplazando los valores en la ecuación de regimen de estado estacionario, se obtiene:

$$P(n) = \rho^n \cdot p(0) \quad (25)$$

$$P(0) = \frac{1}{1 + \rho + \rho^2 + \rho^3 + \rho^4 + \dots + \rho^N} \quad (26)$$

- Calculo de L : valor esperado de que en el sistema haya n clientes.

$$L = P(1) + 2.P(2) + 3.P(3) + \dots + N.P(N) \quad (27)$$

- Calculo de L_c : Cantidad promedio de clientes esperando a ser atendidos.

$$L = 1.P(2) + 2.P(3) + 3.P(4) + \dots + (N - 1).P(N) \quad (28)$$

- Calculo de H : Cantidad canales activos.

$$H = P(1) + P(2) + P(3) + \dots + P(N) = 1 - P(0) \quad (29)$$

- Calculo de $\bar{\lambda}$: La tasa promedio de ingreso de clientes al sistema $\bar{\lambda}$, esta dada por la expresión.

$$\bar{\lambda} = \sum_0^N \lambda_n \cdot p(n)$$

pero como $\lambda_n = \lambda$ para todo $n = 0, 1, 2, 3, \dots, N-1$ y dado que $\lambda_N = 0$, entonces:

$$\bar{\lambda} = \sum_0^N \lambda_n \cdot p(n) = \sum_0^{N-1} \lambda \cdot p(n) = \lambda \cdot \sum_0^{N-1} p(n) = \lambda \cdot [1 - P(N)] \quad (30)$$

- Calculo de $\bar{\mu}$:

$$\bar{\mu} = \sum_0^{\infty} \mu_n \cdot p(n)$$

Pero como $\mu_0 = 0$ y dado que μ_n vale μ para $n \geq 1$, entonces:

$$\bar{\mu} = \sum_1^{\infty} \mu \cdot p(n) = \mu \cdot [1 - P(0)] = \mu \cdot H \quad (31)$$

- Calculo de \bar{R} : Solo tiene valor cuando $n=N$ y es igual a $R_N = \lambda$

$$\bar{R} = \lambda \cdot P(N) \quad (32)$$

- Calculo de W_c : Tiempo de espera en cola

$$W_c = \frac{L_c}{\bar{\lambda}} \quad (33)$$

- Calculo de W : Tiempo de *permanencia* en el sistema.

$$W = W_c + T_c \quad (34)$$

$$W = \frac{L}{\bar{\lambda}} \quad (35)$$

1.4 Modelo P/P/M

Sistemas con M canales de atención, capacidad infinita y poblacion infinita HIPOTESIS

- El *proceso de arribos* de clientes es de tipo **Poisson**.
- El *proceso de servicio* tambien es de tipo **Poisson**.
- El sistema se encuentra en condiciones estables.
- Los clientes que llegan al sistema forman una cola simple.
- La disciplina de atención es FIFO.
- **Hay varios canales de atención dispuestos en paralelo. Todos tienen la misma velocidad de atención.**
- La capacidad del sistema es ilimitada.
- Los clientes que llegan al sistema **no presentan impaciencia**.
- La *población de clientes* potenciales del sistema es **infinita**.

1.5 Modelo $P/P/M/N$

Sistemas con M canales de atención, capacidad finita y poblacion infinita
HIPOTESIS

- El *proceso de arribos* de clientes es de tipo **Poisson**.
- El *proceso de servicio* tambien es de tipo **Poisson**.
- El sistema se encuentra en condiciones estables.
- Los clientes que llegan al sistema forman una cola simple.
- La diciplina de atención es FIFO.
- Hay varios canales de atención dispuestos en paralelo. Todos tienen la misma velocidad de atención.
- La capacidad del sistema es finita. Esta limitada a un valor N.
- Los clientes que llegan al sistema **no presentan impaciencia**.
- La *población de clientes* potenciales del sistema es **infinita**.