Resumen Tema 5

Optimización del rendimiento de un servidor mediante análisis operacional

---

Autor: @BlackTyson

Importante revisar el formulario y la resolución de ejercicios.

Índice

1. Introducción: Redes de cola de espera

1.1. Modelo de un sistema informático

Abstracción del sistema

Un sistema informático consiste en un conjunto de dispositivos interrelacionados y los trabajos que los utilizan. Cada dispositivo o recurso puede ser utilizado por un solo trabajo a la vez. Estos sistemas se modelan comúnmente mediante redes de colas.

Modelo de servidor central

El modelo de servidor central es una de las redes de colas más utilizadas para representar el comportamiento básico de los programas en un servidor.

1. Un trabajo llega al procesador.
2. Después de ser procesado, puede:
   - Salir del servidor.
   - Acceder a una operación de entrada/salida (I/O) y luego volver al procesador.

1.2. Variables que caracterizan a un trabajo

- **Tiempo de espera en cola (W)**
  - Tiempo desde que el trabajo solicita el uso del recurso hasta que comienza a utilizarlo.
  - Se mide en segundos (s).
- **Tiempo de servicio (S)**
  - Tiempo desde que el trabajo accede al recurso hasta que el procesador lo libera.
  - Se mide en segundos (s).
- **Tiempo de respuesta (R)**

$$R = W + S$$

  - Suma de los tiempos de espera y servicio.
  - Se mide en segundos (s).

1.3. Estaciones multi-dispositivo

Las estaciones multi-dispositivo representan dispositivos capaces de atender múltiples trabajos en paralelo.

- No hay tiempo de espera en cola ($R = S$).

1.4. Tiempo de reflexión

- Tiempo que requiere el cliente antes de realizar una nueva petición al servidor tras ser atendido.
- Se modela como una estación de servicio tipo retardo con $S = Z$.
- Se asume que cada cliente envía un único trabajo al servidor.

1.5. Tipos de redes de colas

1.5.1. Cerradas

Las redes cerradas tienen un número constante de trabajos que circulan por la red (N_T). Tipos:

- **Tipo batch**: $N_T = N_0$
- **Tipo interactivo**: $N_T = N_0 + N_Z$

Siempre consideraremos que 1 cliente corresponde a 1 trabajo.

- N_0 = Número de trabajos en el servidor = Número de clientes conectados al servidor.
- N_Z = Número de clientes en reflexión.

1.5.2. Abiertas

En las redes abiertas, los trabajos llegan a la red desde una fuente externa y salen hacia un sumidero tras ser procesados. No hay retroalimentación entre el sumidero y la fuente.

1.5.3. Mixtas

Se refieren a modelos que no corresponden ni a redes cerradas ni a abiertas.

2. Variables y leyes operacionales

2.1. Características del análisis operacional

- Técnica basada en valores medios de diferentes variables del servidor.
- Proporciona relaciones entre variables operacionales.
- Identifica el cuello de botella y estima los límites de sus prestaciones.
- Evalúa el impacto en el rendimiento de modificaciones en los recursos del servidor.

2.2. Variables operacionales básicas de una estación de servicio

- **Variable global temporal:**
  - T:
    - Duración del periodo de medida para el cual se extrae el modelo.
    - Se mide en segundos (s).
- **Variables operacionales básicas durante el tiempo de medida:**
  - **Llegadas (A_i)**
    - Número de trabajos solicitados a la estación i-ésima.
    - Se expresa en cantidad de trabajos.
  - **Tiempo ocupado (B_i)**
    - Tiempo durante el cual el recurso físico de la estación i-ésima ha estado en uso.
    - Se mide en segundos (s).
  - **Completados (C_i)**
    - Número de trabajos completados por la estación i-ésima.
    - Se expresa en cantidad de trabajos completados.

Variables operacionales deducidas

Se obtienen a partir de las variables operacionales básicas.

- **Tiempo medio de servicio (S_i)**:

$$S_i = \frac{B_i}{C_i}$$

  - Tiempo promedio que el recurso físico de la estación i-ésima dedica a cada petición.
  - Se mide en segundos por trabajo ($\frac{s}{tr}$).
- **Tiempo medio de espera en cola (W_i)**:

$$W_i = R_i - S_i$$

  - Tiempo promedio que un trabajo espera en la cola de la estación i-ésima para acceder al recurso físico.
  - Se mide en segundos por trabajo ($\frac{s}{tr}$).
- **Tiempo medio de respuesta (R_i)**:

$$R_i = W_i + S_i$$

  - Tiempo promedio desde que un trabajo accede a la estación de servicio hasta que lo abandona.
  - Se mide en segundos por trabajo ($\frac{s}{tr}$).

- **Tasa media de llegada (λ_i)**:

$$\lambda_i = \frac{A_i}{T}$$

  - Número promedio de trabajos que llegan por segundo a la estación i-ésima.
  - Se mide en trabajos por segundo ($\frac{tr}{s}$).

- **Productividad media (X_i)**:

$$X_i = \frac{C_i}{T}$$

  - Número promedio de trabajos completados por segundo por la estación i-ésima.
  - Se mide en trabajos por segundo ($\frac{tr}{s}$).

- **Utilización (U_i)**:

$$U_i = \frac{B_i}{T}$$

  - Fracción del tiempo T que el recurso de la estación i-ésima ha estado en uso.
  - Se expresa sin unidades o en %.
  - El valor máximo es 1, cuando $B_i = T$.

- **Número medio de trabajos en la estación (N_i)**:

$$N_i = U_i + Q_i$$

  - Número promedio de trabajos en la estación i-ésima.
  - Se expresa en cantidad de trabajos.

- **Trabajos en cola (Q_i)**:

$$Q_i = N_i - U_i$$

  - Número promedio de trabajos esperando en la cola de la estación i-ésima.
  - Se expresa en cantidad de trabajos.

2.3. Variables operacionales de un servidor

Variables básicas

- **Llegadas (A_0)**
  - Número de trabajos solicitados al servidor.
  - Se expresa en cantidad de trabajos.
- **Completados (C_0)**
  - Número de trabajos completados por el servidor.
  - Se expresa en cantidad de trabajos.

Variables deducidas

- **Tasa de llegada (λ_0)**:

$$\lambda_0 = \frac{A_0}{T}$$

  - Tasa media de llegada al servidor.
  - Se mide en trabajos por segundo ($\frac{tr}{s}$).
- **Productividad (X_0)**:

$$X_0 = \frac{C_0}{T}$$

  - Productividad media del servidor.
  - Se mide en trabajos por segundo ($\frac{tr}{s}$).
- **Número medio de trabajos en el servidor (N_0)**:

$$N_0 = N_1 + N_2 + \cdots + N_K$$

- **Tiempo medio de respuesta (R_0)**:
  - Tiempo medio de respuesta del servidor para procesar una petición.
  - Se mide en segundos (s).

2.4. Razón de visita y demanda de servicio

- **Razón media de visita (V_i)**:

$$V_i = \frac{C_i}{C_0}$$

  - Número promedio de veces que un trabajo visita la estación i-ésima antes de abandonar el servidor.
  - Se expresa en cantidad.

- **Demanda media de servicio (D_i)**:

$$D_i = \frac{B_i}{C_0} = V_i \cdot S_i$$

  - Tiempo promedio que el recurso físico de la estación i-ésima dedica a cada trabajo que abandona el servidor.
  - Se mide en segundos por trabajo ($\frac{s}{tr}$).

2.5. Leyes operacionales y equilibrio de flujo

- Las leyes operacionales no dependen de un intervalo de observación ni de suposiciones específicas.
- Se expresan de manera diferente si el servidor está en equilibrio de flujo:
  - Cuando $C_0 \approx A_0$, es decir, cuando el número de trabajos completados por el servidor coincide con los solicitados, o lo que es lo mismo, cuando la productividad media coincide con la tasa de llegada ($X_0 \approx \lambda_0$).
  - Cuando el número de trabajos completados por cada estación coincide aproximadamente con los solicitados ($C_i \approx A_i \implies X_i \approx \lambda_i$).

Ley de utilización

Relaciona la utilización media del recurso con el número de trabajos completados y el tiempo dedicado a cada uno.

$$U_i = X_i \cdot S_i$$

En equilibrio de flujo:

$$U_i = \lambda_i \cdot S_i$$

Demostración:

$$S_i = \frac{B_i}{C_i} = \frac{\frac{B_i}{T}}{\frac{C_i}{T}} = \frac{U_i}{X_i}$$

Dado que la utilización media siempre es menor o igual a uno, la productividad media siempre será menor o igual al inverso del tiempo medio de servicio:

$$U_i \leq 1 \implies X_i \leq \frac{1}{S_i}$$

Ley de flujo forzado

Las productividades de cada estación deben ser proporcionales a la productividad global del servidor. Esta ley relaciona la productividad del servidor con la de las estaciones:

$$X_i = X_0 \cdot V_i$$

En equilibrio de flujo:

$$X_i = \lambda_0 \cdot V_i = \lambda_i$$

Demostración:

$$V_i = \frac{C_i}{C_0} = \frac{\frac{C_i}{T}}{\frac{C_0}{T}} = \frac{X_i}{X_0}$$

Relación utilización-demanda

La utilización de cada dispositivo es proporcional a su demanda de servicio.

$$U_i = X_0 \cdot D_i$$

En equilibrio de flujo:

$$U_i = \lambda_0 \cdot D_i$$

Demostración:

$$D_i = \frac{B_i}{C_0} = \frac{\frac{B_i}{T}}{\frac{C_0}{T}} = \frac{U_i}{X_0}$$

2.6. Ley de Little

Esta ley relaciona la productividad media y el tiempo medio de respuesta del servidor. Solo es válida cuando el servidor está en equilibrio de flujo:

$$N_0 = \lambda_0 \cdot R_0 = X_0 \cdot R_0$$

La ley de Little también se aplica a cada estación de servicio:

- **Para toda la estación:**

$$N_i = \lambda_i \cdot R_i = X_i \cdot R_i$$

- **Para la cola de la estación:**

$$Q_i = \lambda_i \cdot W_i = X_i \cdot W_i$$

2.7. Ley general del tiempo de respuesta

El tiempo de respuesta total se calcula como la suma del producto de la razón de visita de cada estación por su tiempo de respuesta:

$$R_0 = \sum_{i=1}^{K} V_i \cdot R_i$$

Demostración:

$$N_0 = N_1 + N_2 + \cdots + N_K \implies$$

Aplicando la ley de Little:

$$X_0 \cdot R_0 = X_1 \cdot R_1 + X_2 \cdot R_2 + \cdots + X_K \cdot R_K \implies$$

Aplicando la ley de flujo forzado:

$$X_0 \cdot R_0 = X_0 \cdot V_1 \cdot R_1 + X_0 \cdot V_2 \cdot R_2 + \cdots + X_0 \cdot V_K \cdot R_K$$

Ley del tiempo de respuesta interactivo

En una red cerrada en equilibrio de flujo, aplicando la ley de Little a diferentes partes de la red de colas:

- **Clientes en reflexión:**

$$N_Z = X_0 \cdot Z$$

- **Servidor:**

$$N_0 = X_0 \cdot R_0$$

- De esto se deriva la ley:

$$N_T = N_Z + N_0 = X_0 \cdot Z + X_0 \cdot R_0 = X_0 \cdot (Z + R_0) \implies R_0 = \frac{N_T}{X_0} - Z$$

3. Límites optimistas del rendimiento

3.1. Cuello de botella

Definición

- Todo servidor presenta alguna limitación.
- La ubicación del elemento limitante depende de:
  - Características del servidor.
  - Tipo de carga.
- El elemento limitante se conoce como el cuello de botella y puede haber varios.
- Identificar el cuello de botella es esencial para aplicar mejoras que optimicen el servidor.

Identificación

- Es el dispositivo que se satura primero, es decir, el primero en alcanzar $U_i = 1$.
- Dado que U_i es proporcional a D_i, se puede identificar buscando el dispositivo con mayor demanda o mayor utilización.
- Dado que $D_i = V_i \cdot S_i$, la ubicación depende de la rapidez del recurso físico y del tipo de carga.
- Denotaremos el cuello de botella con la letra 'b':

$$D_b = \text{máx}\{D_i\} = V_b \cdot S_b$$
$$U_b = \text{máx}\{U_i\} = X_0 \cdot D_b$$

3.2. Saturación del servidor

El servidor se satura cuando se satura el cuello de botella, es decir, cuando una estación i alcanza $U_i = 1$. Esto implica:

$$1 = U_b = X_b \cdot S_b \implies X_b = \frac{1}{S_b}$$

Un servidor está equilibrado cuando, en promedio, todos los recursos tienen la misma demanda y utilización.

3.3. Límites de rendimiento de un servidor

- Se estiman en los casos extremos de altas y bajas cargas.
- Se calcula una cota superior para la productividad y una cota inferior para el tiempo de respuesta:
  - **Productividad máxima** (X_0^{MAX}).
  - **Tiempo de respuesta mínimo** (R_0^{min}).
- Estos límites son optimistas y miden los mejores casos, estimando la capacidad del servidor y el potencial de mejora en sus prestaciones.

3.4. Redes abiertas

- **Productividad máxima** (X_0^{MAX}): Se obtiene al saturar el cuello de botella:

$$\text{Si } U_b = 1 \implies X_0^{MAX} = \frac{1}{D_b}$$

  Una tasa de llegada mayor rompería el equilibrio de flujo.

- **Tiempo de respuesta mínimo** (R_0^{min}): Ocurre cuando un trabajo llega al servidor sin que haya otros en espera:

$$R_0^{min} = \sum_{i=1}^{K} V_i \cdot S_i = \sum_{i=1}^{K} D_i \equiv D$$

Ejemplo

Recurso	V_i	S_i (s)	D_i (s)
CPU	16	10	0.16
Disco A	7	20	0.14
Disco B	8	30	0.24

1. **Límite optimista del tiempo de respuesta:**

$$R_0^{min} = \sum_{i=1}^{K} D_i = 0{,}16 + 0{,}14 + 0{,}24 = 0{,}54 \text{ s}$$

2. **Límite optimista de la productividad:** Identificamos el dispositivo con la mayor D_i (Disco B):

$$X_0^{MAX} = \frac{1}{D_b} = \frac{1}{0{,}24} \approx 4{,}2 \text{ tr/s}$$

3. **Utilización de cada dispositivo:** Multiplicamos X_0^{MAX} por D_i correspondiente.

3.5. Redes cerradas

- **Altas cargas:**
  - **Productividad optimista:** Cuando el cuello de botella está casi saturado:

$$\text{Si } U_b \to 1 \implies X_0 \to X_0^{MAX} = \frac{1}{D_b}$$

  - **Tiempo de respuesta optimista:** Reemplazando X_0 en la ley de Little:

$$R_0 \to \left(\frac{N_T}{X_0^{MAX}}\right) - Z = D_b \cdot N_T - Z$$

  - **Ley de Little para la red completa:**

$$R_0 = \frac{N_T}{X_0} - Z$$

- **Bajas cargas (N_T pequeño):**
  - **Tiempo de respuesta optimista:** Cuando los trabajos siempre encuentran los dispositivos libres:

$$R_0^{min} = \sum_{i=1}^{K} V_i \cdot S_i = \sum_{i=1}^{K} D_i \equiv D$$

  - **Productividad optimista:** Reemplazando R_0^{min} en la ley de Little:

$$X_0 \to \frac{N_T}{R_0^{min} + Z} = \frac{N_T}{D + Z}$$

3.6. Punto teórico de saturación

Es el valor de N_T donde las asíntotas de productividad y tiempo de respuesta coinciden:

$$D = D_b \cdot N_T^* - Z \implies N_T^* = \frac{D + Z}{D_b}$$

Propiedades:

- Para $N_T > N_T^*$, las prestaciones están limitadas por el cuello de botella.
- A partir de N_T^*, no se puede alcanzar el tiempo de respuesta mínimo ya que se forman colas en el cuello de botella.
- En teoría, se puede lograr la productividad máxima y el tiempo de respuesta mínimo cuando $N_T = N_T^*$:

$$N_T^* = X_0^{MAX} \cdot (R_0^{min} + Z) = \frac{D + Z}{D_b}$$

Ejemplo

Recurso	V_i	S_i (s)	D_i (s)
CPU	5	1	**5**
Disco A	2	2	4
Disco B	2	1.5	3

1. **Identificación del cuello de botella:** El CPU tiene la mayor D_i.

$$D_b = D_{CPU} = 5 \text{ s}$$

2. **Productividad máxima:**

$$X_0^{MAX} = \frac{1}{D_b} = \frac{1}{5} = 0{,}2 \text{ tr/s}$$

3. **Tiempo de respuesta mínimo:**

$$R_0^{min} = D = D_{CPU} + D_{DA} + D_{DB} = 5 + 4 + 3 = 12 \text{ s}$$

4. **Punto teórico de saturación:**

$$N_T^* = \frac{D + Z}{D_b} = \frac{12 + 18}{5} = 6 \text{ tr}$$

5. **Valores de carga:**

   - **Altas cargas:**
     - **Productividad:**

$$X_0 \to X_0^{MAX} = \frac{1}{D_b} = \frac{1}{5} = 0{,}2 \text{ tr/s}$$

     - **Tiempo de respuesta:**

$$R_0 \to \left(\frac{N_T}{X_0^{MAX}}\right) - Z = D_b \cdot N_T - Z = 5 \cdot N_T - 18$$

   - **Bajas cargas:**
     - **Tiempo de respuesta:**

$$R_0 \to R_0^{min} = D = D_{CPU} + D_{DA} + D_{DB} = 12 \text{ s}$$

     - **Productividad:**

$$X_0 \to \frac{N_T}{R_0^{min} + Z} = \frac{N_T}{12 + 18} = \frac{N_T}{30}$$

4. Técnicas de mejora

El objetivo es reducir $D_b = V_b \cdot S_b$ mediante:

- Sintonización.
- Actualización y/o ampliación.

4.1. Sintonización (tuning)

- Optimización del funcionamiento de componentes:
  - Hardware.
  - Aplicaciones: mejora en la distribución de la carga de los recursos existentes.
  - Sistema operativo.
- **Inconvenientes:**
  - Posible alteración de la fiabilidad.
  - Requiere conocimiento especializado.
  - Necesidad de realizar pruebas estadísticas para manejar la aleatoriedad.

4.2. Actualización y/o ampliación

- Reemplazar recursos físicos por otros más rápidos, reduciendo directamente S_b.
- Añadir recursos adicionales para reducir V_b, redistribuyendo la carga entre los componentes para equilibrar las demandas de servicio.
- **Inconvenientes:**
  - Nivel de extensibilidad/escalabilidad.
  - Compatibilidad con sistemas existentes.

5. Algoritmos de resolución de modelos de redes de colas

5.1. Redes abiertas

Hipótesis de independencia en la llegada de trabajos

- En redes abiertas en equilibrio de flujo, asumimos que la llegada de un trabajo es independiente de la llegada del trabajo anterior ($P(x) = \lambda \cdot e^{-\lambda x}$).
- Bajo esta hipótesis, cuando un trabajo llega a una estación, debe esperar a que se procesen los N_i trabajos en la estación:

$$W_i = N_i \cdot S_i$$

- Por lo tanto, el tiempo de respuesta medio es:

$$R_i = W_i + S_i = N_i \cdot S_i + S_i$$

- En equilibrio de flujo, aplicamos la ley de Little:

$$R_i = X_i \cdot R_i + S_i \implies (1 - X_i \cdot S_i) \cdot R_i = S_i \implies R_i = \frac{S_i}{1 - X_i \cdot S_i} = \frac{S_i}{1 - U_i} = \frac{S_i}{1 - X_0 \cdot D_i} = \frac{S}{1 - \lambda_0}$$

- Suponiendo que conocemos λ_0, V_i, S_i y que estamos en equilibrio de flujo ($X_0 = \lambda_0$):
  1. Calculamos la demanda media de cada estación:

$$D_i = V_i \cdot S_i$$

  2. Calculamos el tiempo medio de respuesta de cada estación:

$$R_i = \frac{S_i}{1 - X_0 \cdot D_i}$$

  3. Calculamos el tiempo medio de respuesta del servidor:

$$R_0 = \sum_{i=1}^{K} V_i \cdot R_i = \sum_{i=1}^{K} \frac{D_i}{1 - X_0 \cdot D_i}$$

  4. Calculamos las demás variables operacionales de forma habitual.

Ejemplo

Dispositivo	V_i	S_i (s)
CPU	9	0.01
Disco	3	0.02
Red	5	0.016

Tasa de llegada de peticiones: 5 tr/s.

- Calcula las demandas de servicio de cada recurso.
- Identifica el cuello de botella, determina la productividad máxima del servidor y verifica si está saturado.
- Bajo la hipótesis de independencia:
  - Calcula el tiempo de respuesta de cada recurso.
  - Calcula el número medio de clientes conectados al servidor.
  - Calcula el tiempo medio de espera en la cola y el número medio de trabajos en la cola de cada recurso.

1. **Demanda de servicio (D_i):**
   - **CPU:**
   $$D_{CPU} = V_{CPU} \cdot S_{CPU} = 9 \cdot 0{,}01 = 0{,}09 \text{ s}$$
   - **Disco:**
   $$D_{Disco} = V_{Disco} \cdot S_{Disco} = 3 \cdot 0{,}02 = 0{,}06 \text{ s}$$
   - **Red:**
   $$D_{Red} = V_{Red} \cdot S_{Red} = 5 \cdot 0{,}016 = 0{,}08 \text{ s}$$
2. **Identificación del cuello de botella:** El CPU tiene la mayor demanda ($D_{CPU} = 0{,}09$ s).
   $$X_0^{MAX} = \frac{1}{D_b} = \frac{1}{0{,}09} \approx 11{,}1 \text{ tr/s}$$
   Dado que $\lambda_0 = 5 < X_0^{MAX}$, el servidor está en equilibrio de flujo. Calculamos la utilización de cada dispositivo:
   - **CPU:**
   $$U_{CPU} = X_0 \cdot D_{CPU} = 5 \cdot 0{,}09 = 0{,}45$$
   - **Disco:**
   $$U_{Disco} = X_0 \cdot D_{Disco} = 5 \cdot 0{,}06 = 0{,}30$$
   - **Red:**
   $$U_{Red} = X_0 \cdot D_{Red} = 5 \cdot 0{,}08 = 0{,}40$$
   Como $U_b = U_{CPU} = 0{,}45 < 1$, el servidor no está saturado.

3. **Cálculo del tiempo de respuesta (R_i):**
   - **CPU:**
   $$R_{CPU} = \frac{S_{CPU}}{1 - X_0 \cdot D_{CPU}} = \frac{0{,}01}{1 - 5 \cdot 0{,}09} = 0{,}018 \text{ s}$$
   - **Disco:**
   $$R_{Disco} = \frac{S_{Disco}}{1 - X_0 \cdot D_{Disco}} = \frac{0{,}02}{1 - 5 \cdot 0{,}06} = 0{,}029 \text{ s}$$
   - **Red:**
   $$R_{Red} = \frac{S_{Red}}{1 - X_0 \cdot D_{Red}} = \frac{0{,}016}{1 - 5 \cdot 0{,}08} = 0{,}027 \text{ s}$$
4. **Tiempo de respuesta del servidor (R_0):**

   $R_0 = V_{CPU} \cdot R_{CPU} + V_{Disco} \cdot R_{Disco} + V_{Red} \cdot R_{Red} = 9 \cdot 0{,}018 + 3 \cdot 0{,}029 + 5 \cdot 0{,}027 = 0{,}38$ s
5. **Número medio de clientes en el servidor (N_0):**

   $$N_0 = X_0 \cdot R_0 = 5 \cdot 0{,}38 = 1{,}9 \text{ clientes}$$
6. **Tiempo medio de espera en cola y número medio de trabajos en cola:**

   $$W_i = R_i - S_i$$
   - **CPU:**
   $$W_{CPU} = 0{,}018 - 0{,}01 = 0{,}008 \text{ s}$$
   - **Disco:**
   $$W_{Disco} = 0{,}029 - 0{,}02 = 0{,}009 \text{ s}$$
   - **Red:**
   $$W_{Red} = 0{,}027 - 0{,}016 = 0{,}011 \text{ s}$$

   $$Q_i = X_i \cdot W_i = X_0 \cdot V_i \cdot W_i$$
   - **CPU:**
   $$Q_{CPU} = 5 \cdot 9 \cdot 0{,}008 = 0{,}36 \text{ tr}$$
   - **Disco:**
   $$Q_{Disco} = 5 \cdot 3 \cdot 0{,}009 = 0{,}13 \text{ tr}$$
   - **Red:**
   $$Q_{Red} = 5 \cdot 5 \cdot 0{,}011 = 0{,}27 \text{ tr}$$

5.2. Redes cerradas

Suponemos que conocemos V_i, S_i, N_T, y Z.

- **Método:** Resolver la red incrementando el número de trabajos de manera secuencial.
- **Hipótesis de independencia en la llegada de trabajos:**

$$W_i(N_T) = N_i(N_T - 1) \cdot S_i \implies R_i(N_T) = (N_i(N_T - 1) + 1) \cdot S_i$$