

Evaluación del Comportamiento de Sistemas Informáticos

Actividad 4 Evaluable

Lluís Barca Pons
lluis.barca1@estudiant.uib.es

17 de febrero de 2023

Enunciado Tema 4.1

1. ¿Por qué crees que en los hipermercados hay colas “rápidas” para clientes que lleven menos de un número, relativamente bajo, de productos?
 2. ¿Por qué crees que en algunos establecimientos hay cola única para los clientes y varios servidores para la cola única, en vez de usar varias colas, con una por servidor?
 3. Imagina una cola de facturación en el aeropuerto, ¿qué variables principales gobiernan la longitud de la cola?
-

Solución

- 1. ¿Por qué crees que en los hipermercados hay colas “rápidas” para clientes que lleven menos de un número, relativamente bajo, de productos?**

Porque esos clientes que tienen un número relativamente bajo de productos pueden ser atendidos con mayor rapidez, es decir, su tiempo de servicio será bajo y, por tanto, el tiempo de espera menor. En cambio, si estos clientes formaran una cola única con otros clientes con una compra mayor, aumentarían el tiempo de espera general para todos los clientes.

En resumen, si hay una cola específicamente para este tipo de clientes, liberan la cola principal y en consecuencia se reduce la posibilidad de producirse un cuello de botella.

- 2. ¿Por qué crees que en algunos establecimientos hay cola única para los clientes y varios servidores para la cola única, en vez de usar varias colas, con una por servidor?**

Porque de esta forma se gestiona de forma más eficiente la carga de trabajo que tiene cada servidor. El propio sistema distribuye los clientes a los distintos servidores dependiendo de la carga de trabajo que tiene cada uno en el tiempo correspondiente.

- 3. Imagina una cola de facturación en el aeropuerto, ¿qué variables principales gobiernan la longitud de la cola?**

Las variables principales que gobiernan la longitud de cola son, por un lado, la tasa de llegada λ_i debido a que la carga de trabajos por unidad de tiempo irá incrementando hasta un máximo; habitualmente siguiendo una expresión gráfica como la campana de gauss. Por otro lado, la productividad X_i que hace referencia a la capacidad de procesamiento de trabajos por unidad de tiempo del sistema. Y por último, el tiempo de servicio, S_i que es el tiempo que se tarda para procesar un trabajo o una carga en el servidor. Este está muy relacionado con la productividad, ya que ambos dependen de sí mismos.

Enunciado Tema 4.2

1. ¿Por qué la pendiente de la gráfica del tiempo medio de respuesta de los sistemas con usuarios interactivos tiene una inflexión más pronunciada que la de los sistemas con carga planificada (lotes)?
 2. ¿Por qué el sistema transaccional mide el tiempo de respuesta con el incremento de la tasa de llegadas y no con el número de usuarios como las otras dos gráficas?
-

Solución

1. **¿Por qué la pendiente de la gráfica del tiempo medio de respuesta de los sistemas con usuarios interactivos tiene una inflexión más pronunciada que la de los sistemas con carga planificada (lotes)?**

La diferencia es el momento en el que los trabajos son procesados por el sistema. En el caso de la gráfica de sistemas con carga planificada, los trabajos se procesan uno detrás de otro sin la interacción de los usuarios de forma lineal. En cambio, en los sistemas con usuarios interactivos, incluso siendo una red cerrada, los trabajos se procesan a medida que los usuarios lo demandan. Por tanto, el tiempo medio de respuesta puede subir de forma no lineal e incluso generar cuellos de botella en el peor de los casos.

2. **¿Por qué el sistema transaccional mide el tiempo de respuesta con el incremento de la tasa de llegadas y no con el número de usuarios como las otras dos gráficas?**

Porque se trata de una red de cola abierta. Es decir, debe tener en cuenta el tiempo total que está la carga en el servidor; desde que entra en el sistema hasta que sale. Por tanto, es necesario calcular la λ_i que hace referencia al número de clientes que llegan por unidad de tiempo. Por tanto, concluimos que tiene más sentido hacer las respectivas gráficas con la tasa de llegada y no con el número de trabajos.

Enunciado Tema 4.3

1. Si aplicamos el modelado de colas a un hospital, que asumimos que es una cola de servicio, ¿qué significa que el hospital se va a colapsar con una pandemia? Expresa el colapso con variables de las que hemos visto.
 2. Si el servidor de una cola está utilizando en un % inferior al 100 %, ¿significa que no hay cola en espera? Razona porque no es así.
 3. Puesto que los ordenadores pueden ejecutar programas en paralelo, ¿la teoría de colas no sirve para modelarlos?
-

Solución

Si aplicamos el modelado de colas a un hospital, que asumimos que es una cola de servicio, ¿qué significa que el hospital se va a colapsar con una pandemia? Expresa el colapso con variables de las que hemos visto.

En este caso, interpretando que el hospital es el servicio y los pacientes que van llegando la cola, hay variables como la productividad X_i que son importantes para expresar el colapso. Pero las que provocan en mayor parte el cuello de botella del sistema o en este caso el colapso del hospital, es el tiempo de servicio S_i y la tasa de llegada λ_i . En caso de pandemia, la tasa de llegada aumenta considerablemente, al igual que el tiempo de servicio; aunque dependiendo de la pandemia, puede crecer más una u otra.

Si observamos las fórmulas correspondientes podremos entender un poco mejor porque la λ_i y la S_i son las que provocan en mayor parte el colapso. En primer lugar,

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

donde A_i es el número de trabajos que llegan. Y, por otro lado,

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

donde C_i es el número de trabajos que se van. Por tanto, observamos que la relación entre el flujo de entrada y el flujo de salida de trabajos de un sistema es clave para evitar el colapso.

Si el servidor de una cola está utilizando en un % inferior al 100 %, ¿significa que no hay cola en espera? Razona porque no es así.

Aunque el servidor esté a un % menor al 100 % habitualmente encontramos una cola principalmente porque el servidor no es inmediato. Hay colas porque el servidor tiene un tiempo de servicio, es decir, el tiempo que tarda en realizar el trabajo. Por tanto, es de esperar que si el servidor está trabajando, aunque no sea al 100 %, tiene trabajos en cola.

Puesto que los ordenadores pueden ejecutar programas en paralelo, ¿la teoría de colas no sirve para modelarlos?

Sí que sirve. De hecho, la mayoría de sistemas comparten una cola y seguidamente se van repartiendo en paralelo a distintos servidores para hacer el trabajo.

Y poniendo un ejemplo más simple, el caso de un ordenador actual. Este tiene una cola de trabajo y la propia CPU los gestiona en las diferentes CPUs del sistema.

Enunciado 4.4

1. ¿Por qué la ley del equilibrio de flujo es algo “forzada”?
 2. ¿Cuál es la principal diferencia de modelado entre dos sistemas idénticos modelados con colas, pero uno tiene usuarios interactivos internos (por ejemplo una intranet) y otro con usuarios interactivos externos (por ejemplo a través de internet)?
 3. ¿Cuándo el tiempo de respuesta del modelo de un sistema es la suma de los tiempos de respuesta de todos los dispositivos?
-

Solución

1. ¿Por qué la ley del equilibrio de flujo es algo “forzada”?

Es algo forzada porque para cada dispositivo se debe cumplir que la tasa de llegada de trabajos coincida con la tasa de salida:

$$\lambda_i = X_i$$

y que el número de trabajos que llegan al sistema coincidan con los que salen:

$$A_i = C_i$$

Y en varias ocasiones, no siempre todos los procesos que entran, salen. Pueden dejarse inactivos o eliminarse por cualquier motivo.

Es por eso que se dice que la ley del equilibrio de flujo es algo forzada. Y, por tanto, para aplicarla de la mejor forma, debemos hacerlo para intervalos de observación suficientemente largos.

2. ¿Cuál es la principal diferencia de modelado entre dos sistemas idénticos modelados con colas, pero uno tiene usuarios interactivos internos (por ejemplo una intranet) y otro con usuarios interactivos externos (por ejemplo a través de internet)?

La principal diferencia es la tasa de llegadas del sistema λ_i . En un sistema cerrado se puede calcular aproximadamente el número de usuarios que hacen uso del sistema periódicamente y modelar este comportamiento para conseguir un sistema óptimo. Sin embargo, en el sistema abierto, la λ_i depende del número de usuarios que accedan en un tiempo determinado y prácticamente es imposible predecirlo; aunque si se puede aproximar en la mayoría de las veces.

Es por eso que muchas veces en los sistemas abiertos hablamos de “Tasa de llegadas” y en los sistemas cerrados de “Número de trabajos”.

3. ¿Cuándo el tiempo de respuesta del modelo de un sistema es la suma de los tiempos de respuesta de todos los dispositivos?

El tiempo de respuesta del modelo de un sistema nunca es la suma de los tiempos de respuesta de todos los dispositivos. Es un error habitual, pero el tiempo de respuesta del modelo de un sistema informatica es la suma de los productos entre el tiempo de respuesta y las visitas de cada dispositivo.

$$R = V_1 \times R_1 + V_2 \times R_2 + \dots + V_k \times R_k = \sum_{i=1}^K V_i \times R_i$$