CC4102 **Diseño y Análisis de Algoritmos** **2021A**

Jérémy Barbay, Javier Oliva

Ejercicio Formativo 10

(Fecha de entrega: [2021-07-09 Fri])

Grupo:
Sofía Valentina Bobadilla Ponce
Franco Giovanni Sanguineti Cortés

1 Problema

Un centro médico ha desarrollado un nuevo test llamado "Group Testing", basado en el hecho de que tests clásicos pueden detectar pequeñas cantidades del virus, como por ejemplo en una mezcla. Si la cantidad de muestras contaminadas es suficientemente pequeña, y si uno dispone de varias incubadoras, uno puede encontrar las muestras contaminadas mucho más rápido probando mezclas en paralelo que probándolas secuencialmente (o por separado). Para m y n enteros, asuma que

- se dispone de M = 2m máquinas para incubar y probar una mezcla de muestras,
- hay N = 2n muestras que testear,
- dentro de cuales C están contaminadas. En resumen: M, N y C corresponden a los tamaños de la instancia, mientras m y n corresponden a los logaritmos base dos de M y N.

a

Demuestre que, en el peor caso para el tiempo de ejecución sobre instancias donde n y m están fijos, la eficiencia del mejor algoritmo paralelo posible se encuentra dentro de Θ(1). Hint: Note que C es "libre" y puede ser muy grande. Note que usando la notación p = M, las definiciones de programación paralela son el (mejor) tiempo secuencial posible T^*(n, 1), el tiempo en paralelo T(n, p), la aceleración S(n, p) = T^*(n, 1)/T(n, p), y finamente la eficiencia E(n, p) = S(n, p)/p =$\frac{T(n,1)}{pT(n,p)}$.

Respuesta a

Inicialmente se debe definir el peor caso , que corresponde cuando C es una variable libre, es decir, no se sabe la cantidad de contagiados y por ello se deben testear todas las muestras de sangre por separado.
Sigue reemplazar la formulada vista en el apunte (y también dada en el *hint*) sabiendo que el peor caso en el algoritmo secuencial será T(n,1) = N y en el algoritmo paralelo será el total de muestras partido en los procesadores (máquinas) T(n,p) = $\frac{N}{M}$

$$E(n,p) = \frac{S(n,p)}{p} = \frac{T(n,1)}{pT(n,p)} = \frac{N}{p\frac{N}{M}} = \frac{N}{M\frac{N}{M}} = \frac{N}{N} = 1$$

y 1 está dentro de Θ(1).

b

Demuestre que, en el peor caso sobre instancias donde n está fijo, m = 0 (i.e. M = 1) y C = 1, el tiempo del mejor algoritmo paralelo posible está dentro de $\Theta(lgN)$. Hint: Demuestre cotas superiores e inferiores, eventualmente por reducción a un problema conocido. ¿Tiene sentido estudiar la eficiencia de algoritmos paralelos en tal caso?, ¿y por qué?

Respuesta b

En este caso como M=1 y C=1, se tiene un escenario en donde existe una muestra contagiada y una máquina disponible para ejecutar las pruebas, en este caso estamos analizando el algoritmo secuencial ya que únicamente se tiene 1 procesador.
Por las características anteriores el problema se asemeja a uno de búsqueda de un elemento en un arreglo, lo que se puede resolver con el algoritmo de búsqueda binaria. Sabemos además que, usando teoría de la información, los algoritmos que se usan para resolver este problema tardan al menos $\Omega(logN)$. Se puede hacer una modificación al algoritmo de búsqueda binaria que sirva para resolver nuestro problema, considerando que no hay prelación de orden para su resolución y sólo se realiza una consulta de si el elemento es el contaminado o no, se puede tomar la cantidad de muestras y realizar dos mezclas para cada mitad del grupo y para las siguientes iteraciones ir revisando cada mitad que entregue resultados positivos, hasta encontrar el elemento contaminado. De esta forma se reduce el problema a la mitad en cada iteración, por esta razón llegamos a un problema de tiempo $O(log(N))$ y considerando la cota inferior que nos da la teoría de la información, que es $\Omega(logN)$, entonces podemos decir que el problema está dentro de $\Theta(logN)$, por lo que el tiempo del mejor algoritmo paralelo para resolver el problema también está dentro de $\Theta(logN)$.

En este caso no tiene sentido estudiar la eficiencia del algoritmo paralelo ya que al contar únicamente con un procesador se estaría trabajando con un algoritmo secuencial.

c

Demuestre que existe un algoritmo paralelo de eficiencia dentro de $\Theta(1)$ en el peor caso sobre instancias donde n está fijo, M = n (i.e. $M = log_2(N)$) y C = 1. Hint: Asegúrese de justificar su respuesta describiendo tal algoritmo y demostrando cotas en tal contexto para el tiempo secuencial (de tal problema), el tiempo paralelo (de tal algoritmo), la aceleración (del algoritmo) y finalmente su eficiencia.

Respuesta c

Como el resultado al evaluar las muestras es binario (0 o 1), se tiene que el problema se puede reducir a uno de codificación binaria ya que solamente hay 1 persona contagiada. Para lo anterior se vio en clases que el peor caso para codificar un elemento de un conjunto U de tamaño N se requiere al menos $\log_2(N)$ bits lo cual indicará la cantidad de pruebas necesarias para identificar a la persona contagiada.

El algoritmo al que hacemos referencia es el mismo algoritmo del ejercicio 3, que implica rotular las n mezclas desde la 0 hasta la n1 y escribir su número en binario. De esta forma, se deben realizar mezclas de tal forma que al poner los números en una tabla, las $\log(n)$ mezclas estén compuestas por los bits que tienen encendidos correspondientes a cada muestra. Luego de correr el experimento en la máquina, basta con pintar cada mezcla, por ejemplo con rojo si está contaminada y con azul si es que no lo está. Finalmente, la intersección de todos los conjuntos mezcla pintados de rojo entregaran a la persona contagiada.

Con lo anterior podemos definir el tiempo T(n,1) para el caso secuencial como T(n,1)=$\log_2(N)$, y para el algoritmo paralelo T(n,1)=$\frac{\log_2(N)}{M}$, finalmente recordando que por enunciado M=$\log_2(N)$ se obtiene la eficiencia:

$$E(n,p) = \frac{S(n,p)}{p} = \frac{T(n,1)}{pT(n,p)} = \frac{log_2N}{p\frac{log_2N}{M}} = \frac{log_2N}{log_2N} = 1$$

Y 1 está dentro de $\Theta(1)$. Consideramos que la cantidad de procesadores es igual a la cantidad de máquinas, que es $log_2(N)$. De esta forma tenemos la cota superior e inferior del problema.

d

Demuestra que existe un algoritmo paralelo de aceleración a dentro de $\Theta(1)$ (entonces de eficiencia a dentro de $\Theta(1)$) en el peor caso sobre instancias donde n y m son fijos y C = 1.

Respuesta d

En este escenario se tiene nuevamente exactamente 1 persona contagiada, pero no se conoce el valor exacta de M y N a diferencia de la sección anterior. A pesar de ello nuevamente se puede reducir el problema a uno de codificación binaria donde el tiempo para el algoritmo secuencial el tiempo es T(n,1)=$\log_2(N)$, y para el algoritmo paralelo T(n,1)=$\frac{\log_2(N)}{M}$, con ello la aceleración es:

$$S(n,p) = \frac{T(n,1)}{T(n,p)} = \frac{log_2N}{p\frac{log_2N}{M}} = M$$

Y con ello la eficiencia es:

$$E(n,p) = \frac{S(n,p)}{p} = \frac{M}{p} = \frac{M}{M} = 1$$

Demostrando con lo anterior que hay un algoritmo paralelo que tiene aceleración de $\Theta(M)$ y eficiencia $\Theta(1)$.