[과제 12] 병렬 merge sort

201521026 미디어학과 김용현

단일 처리장치에서의 merge sort에서 시간복잡도를 $\mathrm{T(n)}$이라고 할 때 다음과 같이 쓸 수 있다.

$$\mathrm{T(n)} = \begin{cases} 0 & if\ n = 1 \\ 2T\left(\frac{n}{2}\right) + n & otherwise. \end{cases}$$

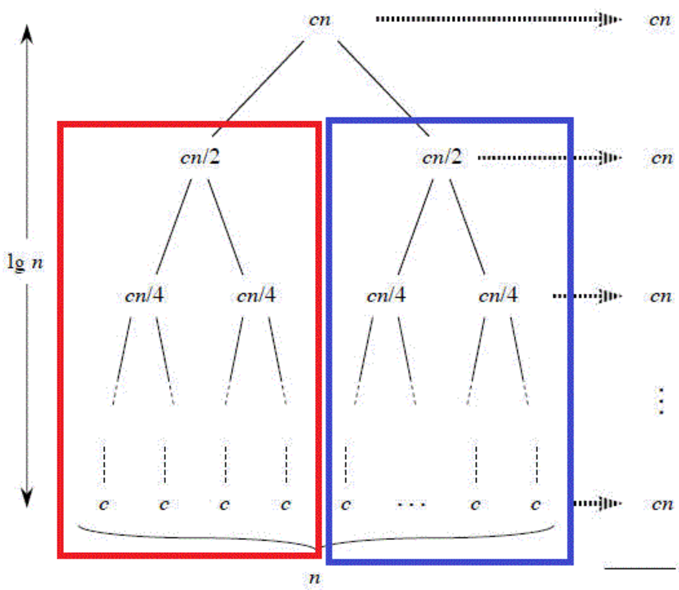
$$\mathrm{T}\left(\frac{\mathrm{n}}{2}\right) = 2T\left(\frac{n}{4}\right) + \frac{n}{2}$$
$$\mathrm{T}\left(\frac{\mathrm{n}}{4}\right) = 2T\left(\frac{n}{8}\right) + \frac{n}{4}$$
$$\vdots$$
$$\mathrm{T}\left(\frac{\mathrm{n}}{2^{\mathrm{k-1}}}\right) = 2\mathrm{T}\left(\frac{\mathrm{n}}{2^{\mathrm{k}}}\right) + \frac{\mathrm{n}}{2^{\mathrm{k-1}}}$$

$$\therefore \mathrm{T(n)} = 2^k T\left(\frac{n}{2^k}\right) + kn$$

$\frac{\mathrm{n}}{2^{\mathrm{k}}} = 1$, 즉 $k = \log_2 n$ 일 때

$$\mathrm{T(n)} = n \log_2 n \le \mathrm{c} * \mathrm{n} \log n \ (c > 0)$$

$$\therefore \mathrm{T(n)} \sim \mathrm{O}(n \log n)$$



2개 이상의 처리장치가 있을 때 merge sort의 시간복잡도를 알아볼 것인데,

먼저 2개의 처리장치로 실행할 경우 merge sort의 시간복잡도를 $T_2(n)$이라 하면

$$T_2(n) = \mathrm{T}\left(\frac{\mathrm{n}}{2}\right) + n$$

이지만, Big-Oh notation의 경우 똑같이 $\mathrm{O}(n \log n)$이다.

이 이유는 크기가 n인 배열을 $\frac{\mathrm{n}}{2}$로 쪼개는데 걸리는 시간은 각각의 처리장치로 넘겨 처리 시키면 동시에 수행할 수 있지만, 크기가 $\frac{\mathrm{n}}{2}$에서 $\frac{\mathrm{n}}{4}$로 쪼개는데 걸리는 시간부터는 기존 단일 처리장치와 같은 시간이 걸리게 될 것이다.

이 말은 즉, 위 그림에서 빨간색과 파란색 사각형에 있는 정렬을 동시에 하고 있다고 할 수 있다.

그러므로 2개의 처리장치일 경우 시간복잡도는 단일 처리장치에서 크기가 $\frac{\mathrm{n}}{2}$인 배열의 시간복잡도 ($\mathrm{T}\left(\frac{\mathrm{n}}{2}\right)$)와 이들을 병합시킬 시간($\mathrm{n}$)의 합이므로 Big-Oh notation은 변하지 않는다.

$$T_2(n) = \mathrm{T}\left(\frac{\mathrm{n}}{2}\right) + n$$
$$= \frac{\mathrm{n}}{2} \log_2 \frac{n}{2} \le \mathrm{c} * \mathrm{n} \log n \ (c > 0)$$

$$\therefore T_2(n) \sim \mathrm{O}(n \log n)$$

두번째로 i개의 처리장치를 가진 컴퓨터에서 실행할 경우의 시간복잡도를 $T_i(n)$라고 하면

$$T_i(n) = T\left(\frac{n}{i}\right) + n \ (i \geq 2)$$

으로 둘 수 있다.

첫번째 경우와 마찬가지로 크기가 $\frac{\mathrm{n}}{\mathrm{i}}$인 배열로 나눈 후 각각의 배열을 단일 처리장치로 정렬을 하고 이들을 병합할 시 위의 식처럼 나오게 된다.

Big-Oh notation의 경우

$$\begin{aligned} &T_i(n) = T\left(\frac{n}{i}\right) + n \\ &= \frac{\mathrm{n}}{\mathrm{i}} \log_i \frac{n}{i} \leq \mathrm{c} * \mathrm{n} \log n \ (c > 0) \\ &\therefore T_i(n) \sim \mathrm{O}(n \log n) \end{aligned}$$

로 증명할 수 있다.

그러므로, 처리장치의 개수가 늘어남에 따라 더 빠른 정렬이 가능하겠지만, Big-Oh notation의 경우 처리장치의 개수와 상관없이 merge sort에서는 항상 $\mathrm{O}(n \log n)$임을 알 수 있다.