Lower bound for sorting

Gou Guanglei(苟光磊)
ggl@cqut.edu.cn

How fast can we sort?

All the sorting algorithms we have seen so far are ***comparison sorts***: only use comparisons to determine the relative order of elements.

E.g., insertion sort, merge sort, quicksort, heapsort.

The **best worst-case running time** that we've seen for comparison sorting is $O(n \lg n)$**.**

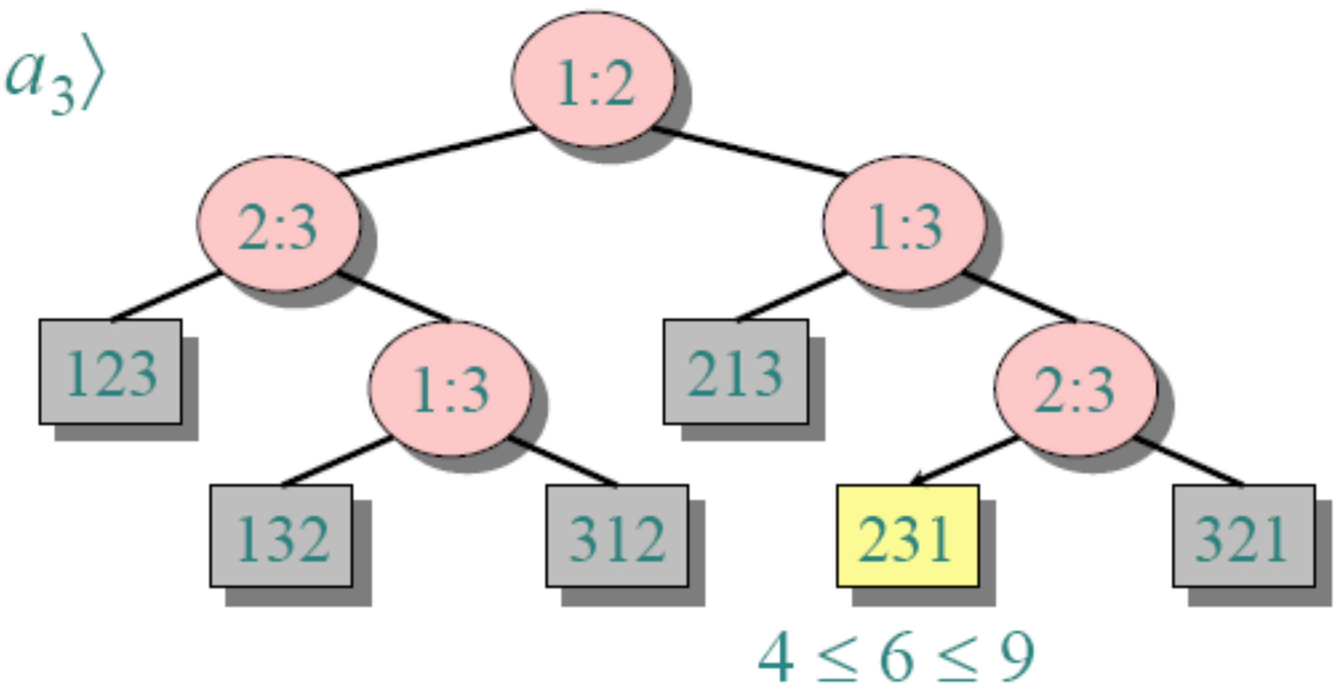
Is O(nlgn) the best we can do?

Decision trees can help us answer this question.

Decision-tree example

Sort $\langle a_1, a_2, \ldots, a_n \rangle$

Sort $\langle a_1, a_2, a_3 \rangle$
$= \langle\, 9, 4, 6 \,\rangle$:



Each leaf contains a permutation $\langle \pi(1), \pi(2), \ldots, \pi(n) \rangle$ to indicate that the ordering $a_{\pi(1)} \le a_{\pi(2)} \le \cdots \le a_{\pi(n)}$ has been established.

Decision-tree model

A decision tree can model the execution of any comparison sort:

- One tree for each input size ***n***.
- View the algorithm as splitting whenever it compares two elements.
- The tree contains the comparisons along **all possible** instruction traces.
- **The running time** of the algorithm = **the length of the path** taken.
- **Worst-case** running time = **height of tree**.

Lower bound for decision-tree sorting

Theorem. Any decision tree that can sort n elements must have height $\boldsymbol{\Omega(n \lg n)}$.

Proof.
1) The tree must contain **$\geq n!$ leaves**, since there are $n!$ possible permutations. (because every permutation appears at least once)
2) A **height-h** binary tree has $\leq 2^h$ leaves. Thus, $n! \leq 2^h$.
3) $\mathbf{n! \leq l \leq 2^h}$
4) $\mathbf{h \geq lg(n!)}$ (since the lg function is monotonically increasing)
5) $\mathbf{lg(n!) = \Theta(n\ lg\ n)}$
6) $\mathbf{h = \Omega(n\ lg\ n)}$.

Lower bound for decision-tree sorting

Corollary. Heapsort and merge sort are asymptotically **optimal** comparison sorting algorithms.

How about randomized algorithms?

- At least one tree for each n.
- Proof still works because every tree works!
- So, **randomized quicksort** is asymptotically optimal in expectation.