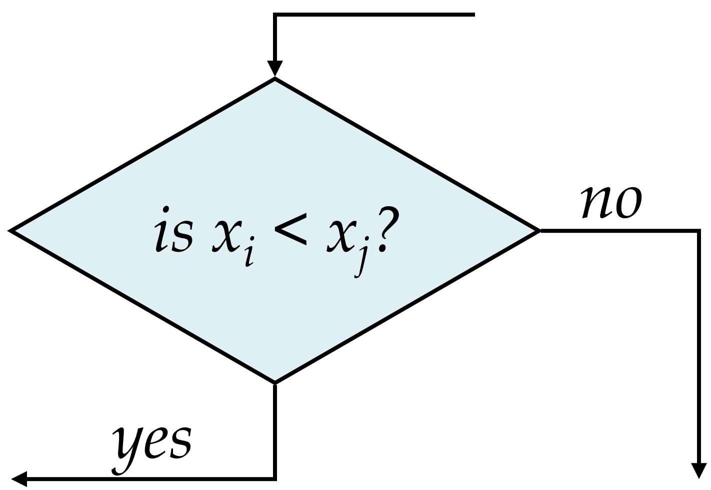
A Lower bound on Comparison-based Sorting

Algorithms & Data Structures
ITCS 6114/8114

Dr. Dewan Tanvir Ahmed
Department of Computer Science
University of North Carolina at Charlotte

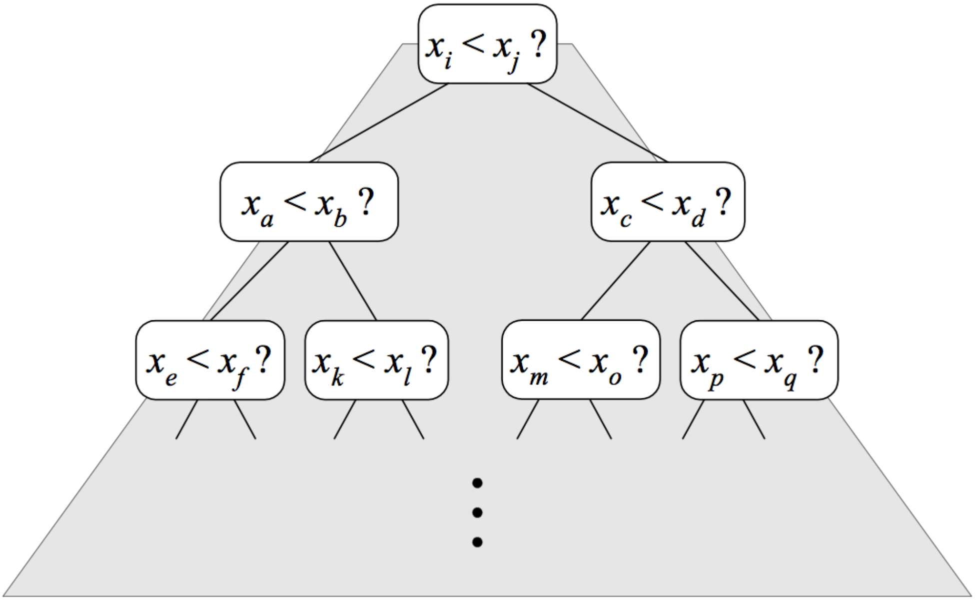
Comparison-Based Sorting (§ 4.4)

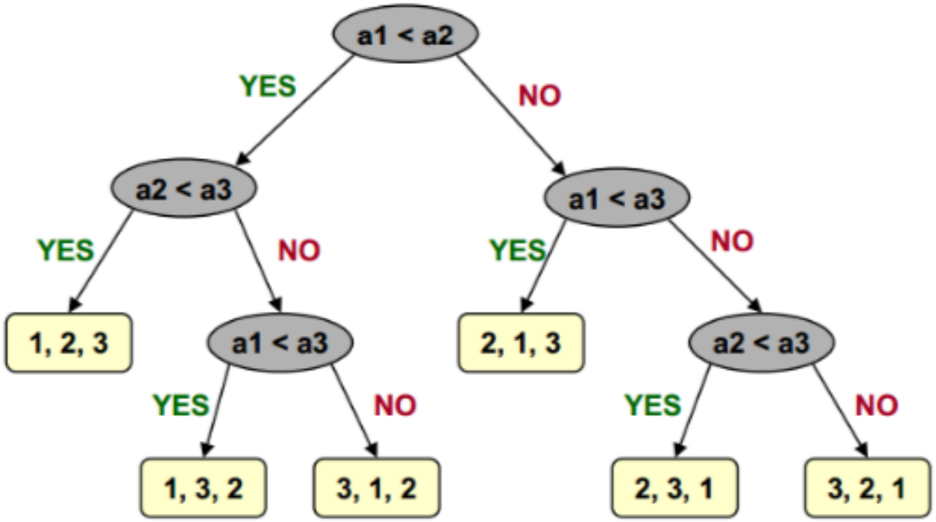
- Many sorting algorithms are comparison based.
  - They sort by making comparisons between pairs of objects
  - Examples: bubble-sort, selection-sort, insertion-sort, heap-sort, merge-sort, quick-sort, ...
- Derive a lower bound on the running time of any algorithm that uses comparisons to sort n elements, $x_1, x_2, \ldots, x_n$.



Counting Comparisons

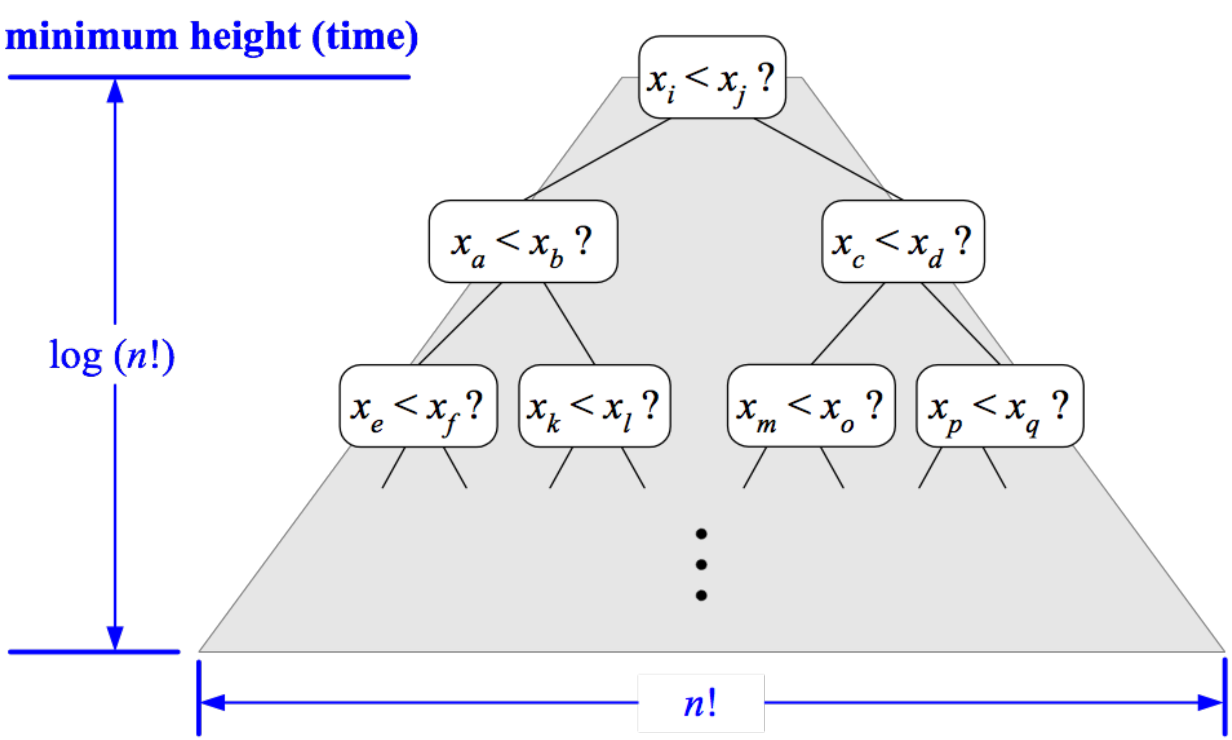
- Let us just count comparisons then.
- Each possible run of the algorithm corresponds to a root-to-leaf path in a **decision tree**



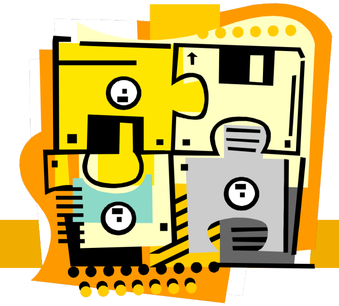


Decision Tree Height

- Every possible input permutation must lead to a separate leaf output.
- Since there are $n! = 1 * 2 * \cdots * n$ leaves, the height is at least $\log(n!)$



The Lower Bound



- Any comparison-based sorting algorithms takes at least log $(n!)$ time
- Since a binary tree of height h has at most 2^h leaves,

$$n! \leq 2^h$$

so $h \geq \log(n!)$

- Stirling's approximation tells us: $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$
- Thus: $$h \geq \log\left(\frac{n}{e}\right)^n$$
- That is, any comparison-based sorting algorithm must run in $\Omega(n \log n)$ time.

Lower Bound For Comparison Sorts

- So,
$$h \geq \log\left(\frac{n}{e}\right)^n$$
$$= n\log n - n\log e$$
$$= \Omega(n\log n)$$
- Thus the time to comparison sort *n* elements is $\Omega(n \log n)$
- Corollary: Heapsort and Mergesort are asymptotically optimal comparison sorts

How can we do better than $\Omega(n \log n)$?

Alternative proof

Theorem: Any decision tree sorting n elements has height $\Omega(n \log n)$

- There must be $n!$ leaves
  - one for each of the $n!$ permutations of n elements
- Tree of height h has at most 2^h leaves
  - $2^h \geq \text{n}! \Rightarrow h \geq \log n!$

$$\geq \log(n \times (n-1) \times (n-2) \ldots \times 2)$$

$$\geq \log n + \log(n-1) + \log(n-2) + \cdots + \log 2$$

$$\geq \sum_{i=2}^{n} \log i$$

Alternative proof (continue)

- $2^h \geq \text{n}! \Rightarrow h \geq \log n!$

$$
\begin{aligned}
&= \log(n \times (n-1) \times (n-2) \ldots \times 2) \\
&= \log n + \log(n-1) + \log(n-2) + \cdots + \log 2 \\
&= \sum_{i=2}^{n} \log i \\
&= \sum_{i=2}^{\frac{n}{2}-1} \log i + \sum_{i=n/2}^{n} \log i \\
&\geq \sum_{i=n/2}^{n} \log i \\
&\geq \sum_{i=n/2}^{n} \log \frac{n}{2} \\
&= \frac{n}{2} \log \frac{n}{2} \\
&= \Omega(n \log n)
\end{aligned}
$$

Reference

- Algorithm Design: Foundations, Analysis, and Internet Examples. Michael T. Goodrich and Roberto Tamassia. John Wiley & Sons.
- Introduction to Algorithms. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein.

Thank you!