6.897
L5.3
Feb. 26, 2003

What if searches are correlated?

General model: arbitrary request sequence $q_1, q_2, \ldots, q_m$

- Defines frequencies f_i = # requests for i
  $p_i = f_i/m$. (but not independent)
- If list weren't allowed to change during request sequence then optimal ordering is decreasing frequency:
  $1\ 2\ 3 \cdots n$ if $f_1 \geq f_2 \geq f_3 \geq \cdots \geq f_n$.
  This is static optimal. Cost = $\sum_{i=1}^{n} i f_i$.
- Model for "startup" (what is initial list ordering?):
  - list initially empty
  - first time element requested, appended to end of list & pay for it.
- cost(Transpose) can be very bad [Bentley & McGeoch – CACM 1985]
  - request sequence: $3, 4, \ldots, n, 1, 2, 1, 2, 1, \ldots$ long
  - typical list: $3\ 4\ 5 \cdots n\ 1\ 2$
  - $\text{cost(Transpose)} = \sum_{i=1}^{n} i + (m-n)\cdot n \sim m \cdot n$
  - $\text{cost(static OPT)} = \sum_{i=1}^{n} i + \frac{3}{2}m \sim \frac{3}{2}m$
- $\text{cost}\begin{pmatrix}\text{MTF}\\ \text{or}\\ \text{FC}\end{pmatrix} \leq 2 \cdot \text{cost(static OPT)}$ — STATIC OPTIMALITY

Proof for MTF:

How many times are we looking for i but find j (or we are looking for j but find i)?

If we sum $\text{cost}_{ij}\ \forall i,j$, we get total cost (except for 1 final successful comparison per query — no matter what method!)

Suppose $f_i \geq f_j$.

Static OPT orders $i, j \Rightarrow \text{cost}_{ij} = f_{ij}$

Worst case for MTF: $(i) \times f_i - f_{ij}$ times, $(j, i) \times f_j$ times

$\Rightarrow \text{cost}_{ij} \leq 2 f_j$

$\therefore \text{cost(MTF)} \leq 2 \cdot \text{cost(static OPT)}$.