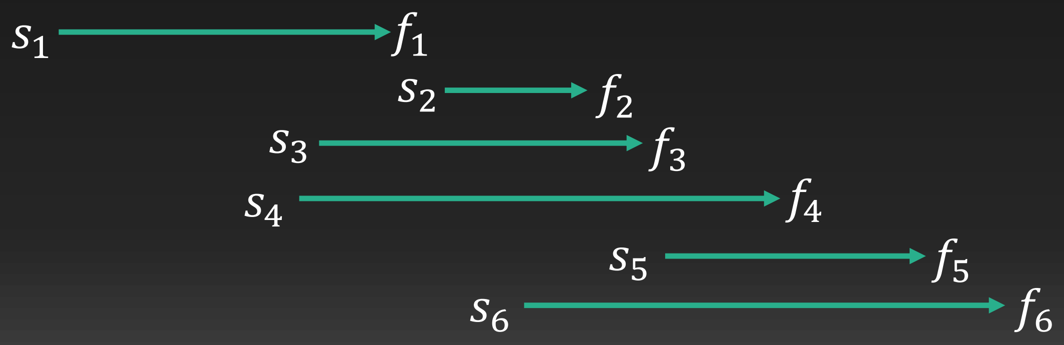
CS 4800: Algorithms & Data

Lecture 12

February 20, 2018

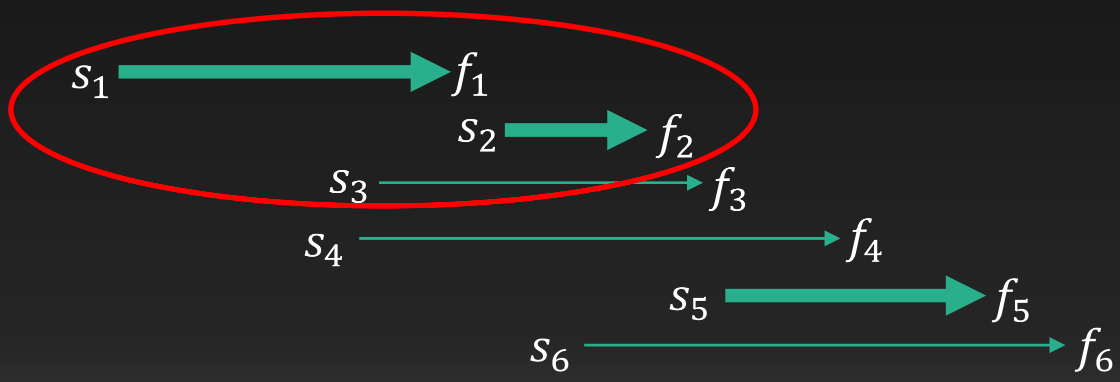
Problem statement

- n activities
- Start times : $s_1, s_2, \ldots, s_n$
- Finish times: $f_1 \le f_2 \le \cdots \le f_n$ (sorted)
- Find largest subset of activities that are compatible



Dynamic Programming

- Best(i): Maximum # compatible activities finishing by f_i
- Optimal substructure: consider activities comprising Best(i).



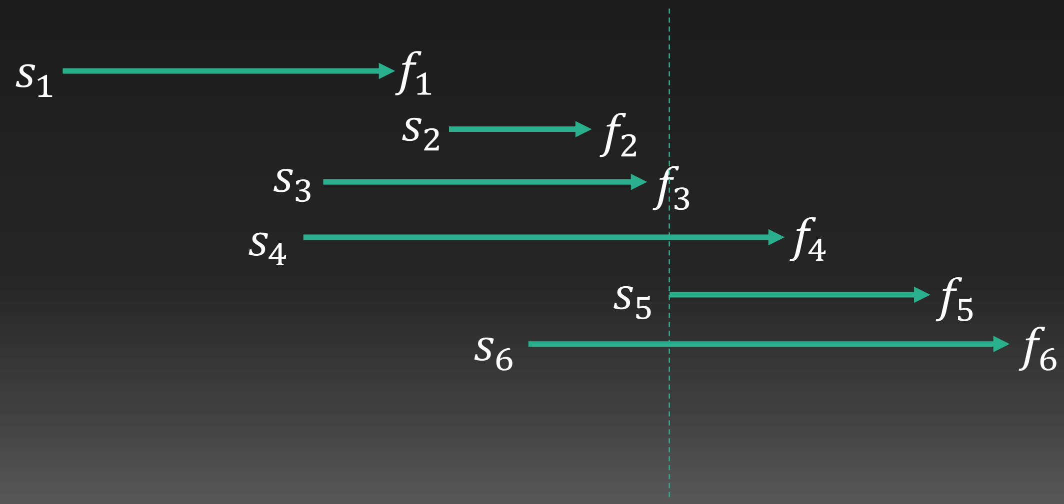
- Claim. The prefix is optimal.

Recursive relation

Not pick i

- Either pick activity i or not
- $$Best(i) \;=\; \max\begin{cases} Best(i-1) \\ 1 + Best(j)\ where\ j = \max k\ s.t\ f_k \le s_i \end{cases}$$

Pick i

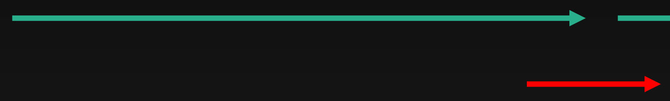


Dynamic Programming

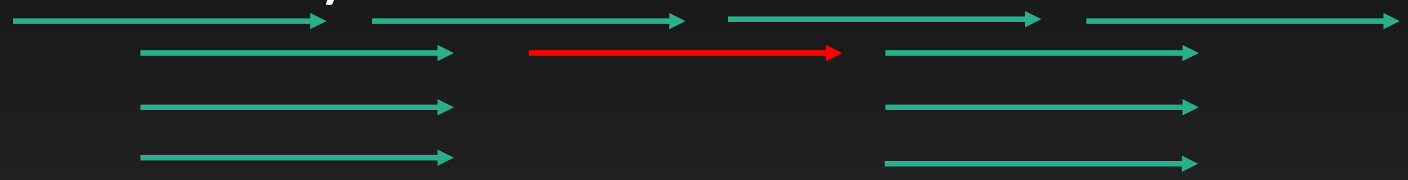
- $Best(0) \leftarrow 0$
- $f_0 \leftarrow -\infty$
- For i from 1 to n
  - Use binary search to find max j s.t. $f_j \leq s_i$
  - Best(i) = max(Best(i-1), 1+Best(j))

Various greedy rules

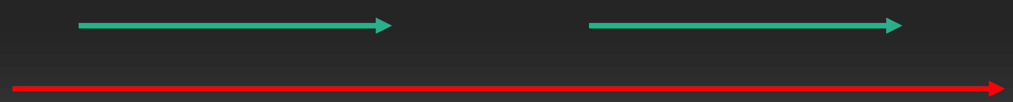
- Pick shortest activity



- Pick activity with fewest conflicts



- Pick activity first to start



- Pick activity first to finish

Exchange argument

Claim: First activity to finish is part of some optimal solution.

Proof.

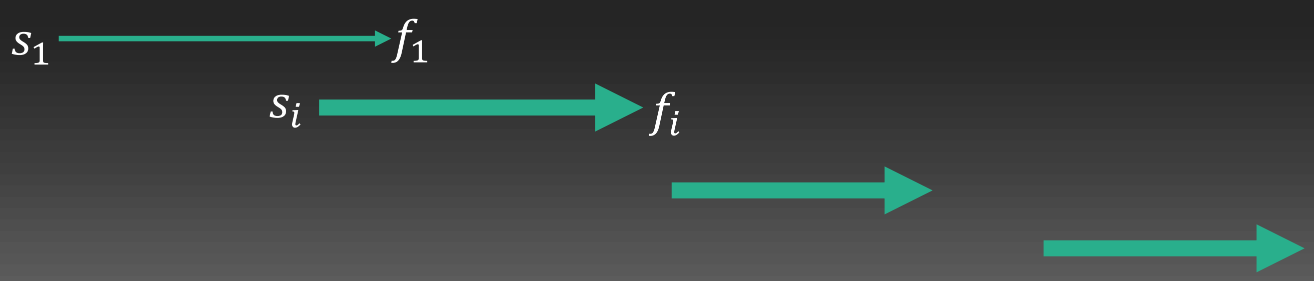
Consider an optimal solution X that does not include activity 1.

Let i be the first activity to finish in X.

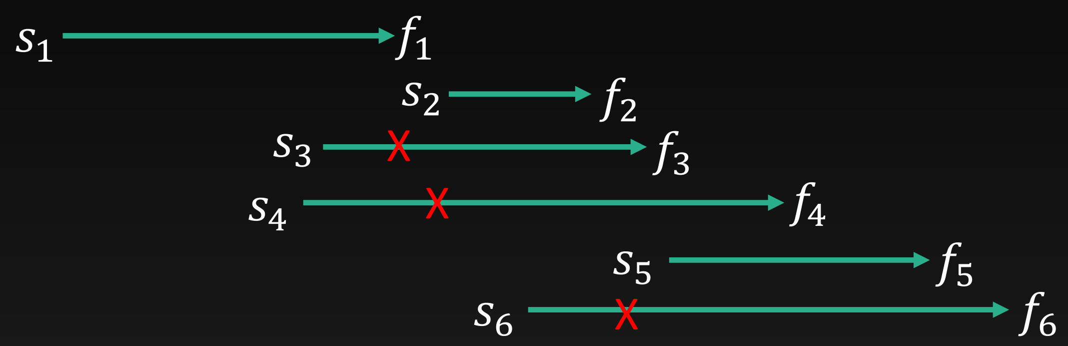
Because act. 1 finishes before i, act. 1 is not in conflict with any activity in $X \setminus \{i\}$

Therefore, $\mathrm{X}' = X \setminus \{i\} \cup \{1\}$ is also conflict-free.

X' has the same size of X and thus, it is also optimal.



Greedy algorithm



Find first activity to finish. Add to solution.
Remove conflicting activities.

Repeat.

Greedy algorithm

- $count \leftarrow 1$ // number of activities we pick
- $X[count] \leftarrow 1$ // X[.]: IDs of activities we pick
- For i from 2 to n
  - If $S[i] \geq F\big[X[count]\big]$
    - $count \leftarrow count + 1$
    - $X[count] \leftarrow i$
- Return $X[1 \dots count]$

Greedy is optimal

Induction hypothesis: Greedy is optimal for any instance of size n.
Base case: Greedy is optimal for n=1
Inductive case: Assume Greedy is optimal for $n < k$. Will prove for n=k.
By Claim, activity 1 belongs to some optimal solution. Thus, the best solution that includes 1 is also optimal.

Greedy picks 1 and then perform greedy on the set of activities not conflicting with 1 (a sub-instance of size < k).

By induction, greedy picks an optimal solution for the sub-instance i.e. it finds the best solution containing 1.

Therefore, greedy also finds an optimal solution for n=k.

Huffman codes

Information transmission

Once upon a time, before Internet and emails,

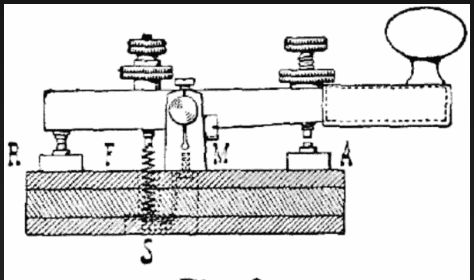
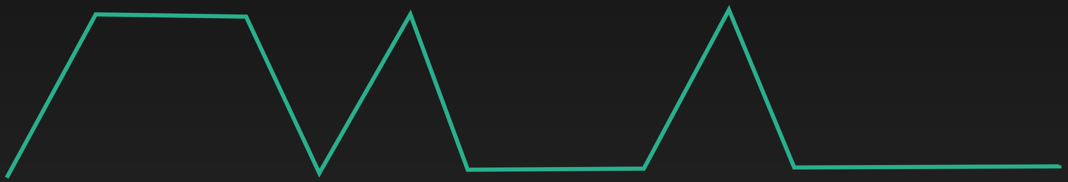


Fig. 6.



Texts are transmitted as electrical pulses and silence in between

Long pulses (1) and short pulses (0)

Length of encoding

- Letter c occurs f_c times and its encoding is of length l_c bits
- Encoding length= $\sum_c f_c l_c$
- Given a text consisting of n distinct letters, find minimum length encoding

Morse code

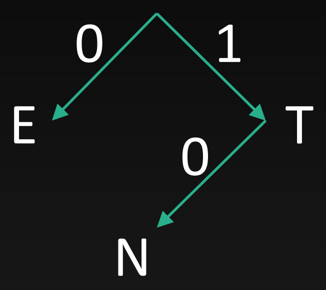
- Encode letters as sequences of dots & dashes (0/1)

Letter	Code
A	01
E	0
I	00
N	10
T	1

What does 01 mean? ET or A?

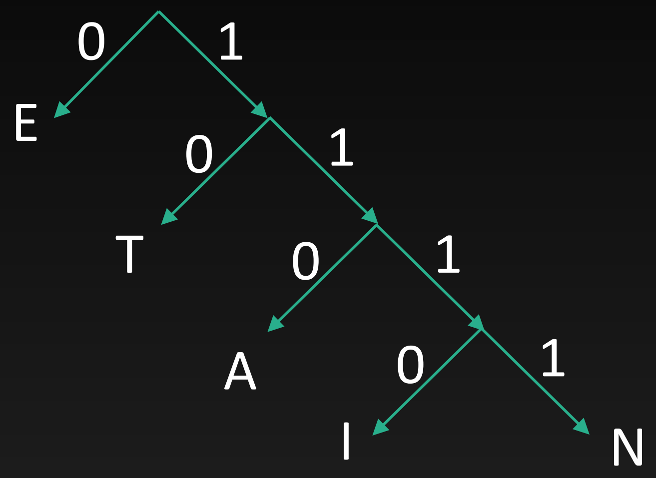
Prefix-free codes

- Problem with Morse code: some encoding is prefix of another
- Prefix-free code: for any two letters $x \neq y$, code(x) is not a prefix of code(y)



Encoding/decoding prefix-free codes

- Text: EATIN
- Encoding:
  - 01101011101111
- Decoding:
  - Start at root
  - Go down until reaching a leaf
    - get a letter
  - Restart from the root



A text for compression

This sentence contains three a’s, three c’s, two d’s, twenty-six e’s, five f’s, three g’s, eight h’s, thirteen i’s, two l’s, sixteen n’s, nine o’s, six r’s, twenty-seven s’s, twenty-two t’s, two u’s, five v’s, eight w’s, four x’s, five y’s, and only one z

-Lee Sallows