6.046 – Design & Analysis of Algorithms

Prof. Shafi Goldwasser & Ron Rivest

http://courses.csail.mit.edu/6.046/fall10

Outline:
- Course overview & admin
- Some problems related to interval scheduling

Admin:
- Course info handout (all else on web)
- Pre-reqs: 6.006 & (6.042 or 18.310) strict
- Recitations: Fridays. Sign up by 6pm today!
- Psets: 6 of them
  - exponential penalty for missed problems
  - collaboration good; copying/plagiarism bad

Exams:
- Quiz 1: in-class Tue Oct 19
- Quiz 2: take-home Nov 16 → Nov 23 (no collaboration)
- Final Exam
Outline:

- theory
- randomization
- data structures
- graphs
- optimization
- complexity
- advanced topics
Today’s theme: Very similar problems can have very different solutions & complexity.

Recall: \( P = \text{class of problems solvable in polynomial time} \)

e.g. multiplying two \( n \)-bit \#’s in time \( \Theta(n^2) \)

\( NP = \text{class of problems whose solution can be verified in polynomial time} \)

e.g. is given boolean formula satisfiable? (SAT)
e.g. is given integer composite (not prime)

We know \( P \leq NP \), but don’t know if \( P = NP \)

\[ \begin{aligned}
NP & \rightarrow \text{NP-complete} \\
\text{likely?} \quad \text{or} \quad \text{unlikely?}
\end{aligned} \]

The \( NP \)-complete problems are not in \( P \), if \( P \neq NP \).
(e.g. SAT, Travelling Salesman, ..., 1000’s ...)

[More in lectures 17-18]
Scheduling: Fascinating problem area, with many variants.

(We'll look at three.)

- Suppose we have a valuable resource (classroom, Hubble telescope) that can be used by at most one party at a time. ⇒ Its use must be scheduled.
- Assume we are given a set $R$ of $n$ requests $R = \{R_1, R_2, \ldots, R_n\}$ for use of this resource.
- We want to find a schedule $S \subseteq R$, a set of compatible requests, that optimizes some given criterion.
- Simplest case: each request $R_i$ specifies a start time $s(i)$, a finish time $f(i)$.
- $R_i$ and $R_j$ are compatible if they don't overlap: $f(i) \leq s(j)$ or $f(j) \leq s(i)$
First Problem:
Interval Scheduling: Given $R$, find schedule $S$ that maximizes $|S|$. 

**Example:**

```
\[ R_1 \]
\[ R_2 \]
\[ R_3 \]
\[ R_4 \]
\[ R_5 \]
\[ R_6 \]
```

Optimal schedule = \{ $R_4$, $R_5$, $R_6$ \} (or \{ $R_4$, $R_5$, $R_6$ \})

**Brute force:** There are $2^n$ subsets of $R$.

Testing a subset for compatibility takes time $O(n^3)$.

Overall time is $O(n^3 \cdot 2^n)$.

Not polynomial time 😞

We solve Interval Scheduling in time $O(n \log n)$ using a greedy method.
Greedy method for Interval Scheduling:

\[ S = \emptyset \]

while \( R \neq \emptyset \):

Pick \( R_i \) from \( R \) according to some (greedy, myopic) rule.

\[ S = S \cup \{ R_i \} \]

Remove from \( R \) the request \( R_i \) and all requests incompatible with \( R_i \).

return \( S \)

With right rule, produces optimal schedule (in polynomial time).
Which rule?

1. Pick $R_i$ that starts first (minimizes $s(i)$)

2. Pick $R_i$ that is smallest (minimizes $f(i) - s(i)$)

3. Pick $R_i$ with fewest incompatibilities

4. Pick $R_i$ that finishes first (minimizes $f(i)$)

(intuition: leaves maximum amount of time available for other requests)
Theorem: Greedy method (with first-to-finish rule) gives an optimum schedule.

Proof:  
- Suppose greedy alg returns \( S = \{ R_{i_1}, R_{i_2}, \ldots, R_{i_p} \} \) so \( |S| = p \) & results "in order" \( (R_{i_1}, \text{ finishes first}) \)
- Let \( S^* = \{ R_{j_1}, R_{j_2}, \ldots, R_{j_q} \} \) be some optimal schedule (in order). (This maximizes \( |S^*| = q \).)
- There exists an **optimal** schedule \( S^{**} \) starting with \( R_{i_1} \).
  - (Replace \( R_{j_1} \), with \( R_{i_1} \), in \( S^* \); OK since \( f(i_1) \leq f(j_1) \).)
- Let \( R^{\geq x} = \{ R_i \in R \mid s(i) \geq x \} \)
- Then \( \{ R_{j_2}, \ldots, R_{j_q} \} \) must be optimal for \( R^{\geq f(i_1)} \)
  - (otherwise we could replace it with better solution, improving \( S^{**} \).)
- By induction, greedy method gives optimal solution.
- Note: \( \text{opt}(R) = 1 + \text{opt}(R^{\geq f(i_1)}) \)
  - where \( \text{opt}(R) = \text{size of optimal solution} \)
  - \( i_1 = i \) s.t. \( R_i \text{ finishes first} \)
Running time: \( O(n \lg n) \)

- sort requests in order of increasing \( f(i) \)
- consider each in turn, adding it to \( S \) if it is compatible with last interval added to \( S \). (or if it is 1st interval)
Second Problem: Weighted Interval Scheduling

Same as before, but now each request $R_i$ also has a "weight" $w(i)$. We seek schedule $S \subseteq R$ of maximum total weight.

E.g. $w(i) = \# bid$ (max revenue)

$w(i) = f(i) - s(i)$ (max utilization)

$w(i) = \# students$ (max student 😊)

Note: Greedy no longer works

First to finish rule $\Rightarrow 1 + 2 = 3$

Max weight rule $\Rightarrow 4$

Optimum $= 3 + 2 = 5$
Dynamic Programming works on Weighted Interval Scheduling.

Subproblems are \( R^x = \{ R_i \in R \mid s(i) \geq x \} \)

Only \( n \) different subproblems; one for each \( x = f(i) \).

Only need to solve each subproblem once (save soln for later use).

Since greedy doesn't work, we don't have rule telling us which interval to schedule first.

So, we try each \( R_i \) in turn as possible "first."

Given \( R_i \) as tentative first, remaining is \( R^{x \neq f(i)} \).

Program: \( \text{opt}(R) = \max_{1 \leq i \leq n} (w_i + \text{opt}(R^{x \neq f(i)})) \)

Running time is \( O(n^2) \)

Exercise: show how to use sorting initially & reduce running time to \( O(n \log n) \)

Note: Actual optimum schedule can be obtained if we record \( i \) that yields max for each subproblem.
Third Problem: “Flexible” interval scheduling (unit weights, unweighted)

Each request $R_i$ specifies:

- $s(i) =$ earliest possible start time  
  \[ \text{Ex} \]  
  \[ (11 \text{ am}) \]

- $f(i) =$ latest possible finishing time  
  \[ (5 \text{ pm}) \]

- $l(i) =$ length requested  
  \[ \leq f(i) - s(i) \]  
  \[ (2 \text{ hours}) \]

\[ \begin{array}{c}
\text{11 am} \quad \text{2 hours} \\
\end{array} \]

\[ \begin{array}{c}
\quad \quad \quad \quad \quad \\
\quad \quad \quad \quad \quad \\
\quad \quad \quad \quad \quad \\
\quad \quad \quad \quad \quad \\
\quad \quad \quad \quad \quad \\
\text{5 pm} \\
\end{array} \]

Want schedule of maximum size.

(Solution must specify when each scheduled interval actually starts.)
"Flexibility" makes problem harder!

Thm: Flexible interval scheduling is NP-complete.

Thm: Greedy algorithm gives good approximate soln. (within factor of 2 of optimal)

Thm: Fancier approximation algorithm achieves factor $e/(e-1) = 1.582...$

[Chuzhoy, Ostrovsky, Rabani 2006]

Thm: ($\exists \epsilon$) Getting within $(1+\epsilon)$ of optimum is NP-hard. [Spieksma 1999]
Summary:

Very similar problems can have very different solutions & complexity.

- Interval scheduling is in $P$ (greedy method).
- Weighted interval scheduling is in $P$ (dynamic programming).
- Flexible interval scheduling is NP-complete

(Good approximation algorithms exist, but getting very good approximations is NP-hard.)

Good algorithms make a difference!
So does knowing when to stop looking for one!