Advertising on the Web
Online algorithms

- Classic model of algorithms
  - You get to see the entire input, then compute some function of it
  - In this context, “offline algorithm”

- Online algorithm
  - You get to see the input one piece at a time, and need to make irrevocable decisions along the way

- Similar to data stream models
Example: Bipartite matching

![Graph showing a bipartite matching problem with girls and boys as two sets, and edges connecting them.]
Example: Bipartite matching

\[ M = \{(1,a),(2,b),(3,d)\} \text{ is a matching.} \]

Cardinality of matching = \(|M| = 3\)
Example: Bipartite matching

M = {(1,c), (2,b), (3,d), (4,a)} is a perfect matching.
Matching Algorithm

- Problem: Find a maximum-cardinality matching for a given bipartite graph
  - A perfect one if it exists

- There is a polynomial-time offline algorithm (Hopcroft and Karp 1973)

- But what if we do not know the entire graph upfront?
Initially, we are given the set Boys
In each round, one girl’s choices are revealed
At that time, we have to decide to either:
  - Pair the girl with a boy
  - Do not pair the girl with any boy

Example of application:
Assigning tasks to servers
Online Graph Matching: Example

(1,a)
(2,b)
(3,d)
Greedy Algorithm

- Greedy algorithm for online graph matching problem:
  - Pair the new girl with any eligible boy
    - If there is none, don’t pair girl
  - How good is the algorithm?
For input I, suppose greedy produces matching $M_{\text{greedy}}$ while an optimal matching is $M_{\text{opt}}$.

**Competitive ratio** = 

$$\min_{\text{all possible inputs } I} \left( \frac{|M_{\text{greedy}}|}{|M_{\text{opt}}|} \right)$$

(what is greedy’s worst performance over all possible inputs)
Consider the set $G$ of girls matched in $M_{opt}$ but not in $M_{greedy}$.

Then every boy $B$ adjacent to girls in $G$ is already matched in $M_{greedy}$:

$|B| \leq |M_{greedy}|$

There are at least $|G|$ such boys ($|G| \leq |B|$) otherwise the optimal algorithm could not have matched all the $G$ girls. So: $|G| \leq |M_{greedy}|$

By definition of $G$ also: $|M_{opt}| \leq |M_{greedy}| + |G|$

So $|M_{greedy}|/|M_{opt}| \geq 1/2$
Worst-case scenario

Diagram:

1. 2. 3. 4.
   \( (1,a) \)
   \( (2,b) \)
Banner ads (1995-2001)

- Initial form of web advertising
- Popular websites charged X$ for every 1000 “impressions” of ad
  - Called “CPM” rate
  - Modeled similar to TV, magazine ads
- Untargeted to demographically targeted
- Low clickthrough rates
  - low ROI for advertisers
Performance-based advertising

- Introduced by Overture around 2000
  - Advertisers “bid” on search keywords
  - When someone searches for that keyword, the highest bidder’s ad is shown
  - Advertiser is charged only if the ad is clicked on

- Similar model adopted by Google with some changes around 2002
  - Called “Adwords”
## Ads vs. search results

<table>
<thead>
<tr>
<th>Web</th>
<th>Sponsored Links</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GEICO</strong> Car Insurance. Get an auto insurance quote and save today...&lt;br&gt;GEICO auto insurance, online car insurance quote, motorcycle insurance quote, online insurance sales and service from a leading insurance company.&lt;br&gt;<a href="http://www.geico.com">www.geico.com</a> - 21k - Sep 22, 2005 - Cached - Similar pages&lt;br&gt;- Auto Insurance - Buy Auto Insurance&lt;br&gt;- Contact Us - Make a Payment&lt;br&gt;More results from <a href="http://www.geico.com">www.geico.com</a> &gt;</td>
<td>Great Car Insurance Rates&lt;br&gt;Simplify Buying Insurance at Safeco&lt;br&gt;See Your Rate with an Instant Quote&lt;br&gt;<a href="http://www.Safeco.com">www.Safeco.com</a></td>
</tr>
<tr>
<td><strong>Geico. Google Settle Trademark Dispute</strong>&lt;br&gt;The case was resolved out of court, so advertisers are still left without legal guidance on use of trademarks within ads or as keywords.&lt;br&gt;<a href="http://www.clickz.com/news/article.php/3547356">www.clickz.com/news/article.php/3547356</a> - 44k - Cached - Similar pages</td>
<td>Free Insurance Quotes&lt;br&gt;Fill out one simple form to get multiple quotes from local agents.&lt;br&gt;<a href="http://www.HometownQuotes.com">www.HometownQuotes.com</a></td>
</tr>
<tr>
<td>**Google and GEICO settle AdWords dispute</td>
<td>The Register**&lt;br&gt;Google and car insurance firm GEICO have settled a trade mark dispute over ... Car insurance firm GEICO sued both Google and Yahoo! subsidiary Overture in ...&lt;br&gt;<a href="http://www.theregister.co.uk/2005/09/09/google_geico_settlement/">www.theregister.co.uk/2005/09/09/google_geico_settlement/</a> - 21k - Cached - Similar pages</td>
</tr>
<tr>
<td><strong>GEICO v. Google</strong>&lt;br&gt;... involving a lawsuit filed by Government Employees Insurance Company (GEICO). GEICO has filed suit against two major Internet search engine operators, ...&lt;br&gt;<a href="http://www.consumeraffairs.com/news04/geico_google.html">www.consumeraffairs.com/news04/geico_google.html</a> - 19k - Cached - Similar pages</td>
<td></td>
</tr>
</tbody>
</table>
Performance-based advertising works!
- Multi-billion-dollar industry

Interesting problems:
- What ads to show for a given query?
- If I am an advertiser, which search terms should I bid on and how much should I bid?
Adwords Problem

- A stream of queries arrives at the search engine
  - q1, q2, ...
- Several advertisers bid on each query
- When query \( q_i \) arrives, search engine must pick a subset of advertisers whose ads are shown

- **Goal:** maximize search engine’s revenues
- **Clearly we need an online algorithm!**
Each advertiser has a limited budget
- Search engine guarantees that the advertiser will not be charged more than their daily budget

Each ad has a different likelihood of being clicked
- Advertiser 1 bids $2, click probability = 0.1
- Advertiser 2 bids $1, click probability = 0.5
- Clickthrough rate measured historically

Simple solution
- Instead of raw bids, use the “expected revenue per click”
### The Adwords Innovation

<table>
<thead>
<tr>
<th>Advertiser</th>
<th>Bid</th>
<th>CTR</th>
<th>Bid * CTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$1.00</td>
<td>1%</td>
<td>1 cent</td>
</tr>
<tr>
<td>B</td>
<td>$0.75</td>
<td>2%</td>
<td>1.5 cents</td>
</tr>
<tr>
<td>C</td>
<td>$0.50</td>
<td>2.5%</td>
<td>1.125 cents</td>
</tr>
<tr>
<td>Advertiser</td>
<td>Bid</td>
<td>CTR</td>
<td>Bid * CTR</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td>B</td>
<td>$0.75</td>
<td>2%</td>
<td>1.5 cents</td>
</tr>
<tr>
<td>C</td>
<td>$0.50</td>
<td>2.5%</td>
<td>1.125 cents</td>
</tr>
<tr>
<td>A</td>
<td>$1.00</td>
<td>1%</td>
<td>1 cent</td>
</tr>
</tbody>
</table>
Greedy algorithm

- The environment:
  - There is one ad shown for each query
  - All advertisers have the same budget
  - All adds are equally likely to be clicked
  - Value of each add is the same

- Simplest algorithm is greedy:
  - For a query pick any advertiser who has bid 1 for that query
- Competitive ratio of greedy is 1/2
Two advertisers A and B
- A bids on query x, B bids on x and y
- Both have budgets of $4

Query stream: xxxxyyyyy
- Worst case greedy choice: BBBB____
- Optimal: AAAABBBBB
- Competitive ratio = ½

This is the worst case
BALANCE algorithm [MSVV]

- BALANCE by Mehta, Saberi, Vazirani, and Vazirani
  - For each query, pick the advertiser with the largest unspent budget
    - Break ties arbitrarily
Example: BALANCE

- Two advertisers A and B
  - A bids on query x, B bids on x and y
  - Both have budgets of $4

- Query stream: xxxxyyyyy

- BALANCE choice: ABABBB__
  - Optimal: AAAABBBBB

- Competitive ratio = $\frac{3}{4}$
Consider simple case:

- Two advertisers, A₁ and A₂, each with budget B (assume B ≥ 1)

Assume optimal solution exhausts both advertisers’ budgets

BALANCE must exhaust at least one advertiser’s budget

- If not, we can allocate more queries
- Assume BALANCE exhausts A₂’s budget, but allocates x queries fewer than the optimal
  \[ \text{BAL} = 2B - x \]
Analyzing Balance

**Opt revenue** = 2B
**Balance revenue** = 2B-x = B+y

We have y ≥ x
Balance revenue is minimum for x=y=B/2
Minimum Balance revenue = 3B/2
**Competitive Ratio** = 3/4
In the general case, worst competitive ratio of BALANCE is $1 - 1/e = \text{approx. } 0.63$

- Interestingly, no online algorithm has a better competitive ratio!

- We do not through the details here, but let’s see the worst case that gives this ratio
Worst case for BALANCE

- **N advertisers:** $A_1, A_2, \ldots A_N$
  - Each with budget $B > N$
- **Queries:**
  - $N \cdot B$ queries appear in $N$ rounds of $B$ queries each
- **Bidding:**
  - Round 1 queries: bidders $A_1, A_2, \ldots, A_N$
  - Round 2 queries: bidders $A_2, A_3, \ldots, A_N$
  - Round $i$ queries: bidders $A_i, \ldots, A_N$
- **Optimum allocation:**
  Allocate round $i$ queries to $A_i$
  - Optimum revenue $N \cdot B$
Balance assigns each of the queries in round 1 to N advertisers.

After k rounds, sum of allocations to each of advertisers $A_k, \ldots, A_N$ is

$$S_k = S_{k+1} = \ldots = S_N = \sum_{i=1}^{k-1} \frac{B}{(N-i+1)}$$

If we find the smallest k such that $S_k \geq B$, then after k rounds we cannot allocate any queries to any advertiser.
BALANCE analysis

- **Fact:**
  \[ H_n = \sum_{i=1}^{n} \frac{1}{i} = \text{approx. log}(n) \text{ for large } n \]
  - Result due to Euler

1/1  1/2  1/3  ...  1/(N-k+1)  ...  1/(N-1)  1/N

\[ S_k = 1 \implies H_{N-k} = \log(N)-1 = \log(N/e) \]
\[ N-k = N/e \]
\[ k = N(1-1/e) \]
**BALANCE analysis**

- So after the first $N(1-1/e)$ rounds, we cannot allocate a query to any advertiser

- Revenue = $B \cdot N \left(1-\frac{1}{e}\right)$

- Competitive ratio = $1-\frac{1}{e}$
General version of problem

- Arbitrary bids, budgets
- Consider query \( q_i \), advertiser \( i \)
  - Bid = \( x_i \)
  - Budget = \( b_i \)
- BALANCE can be terrible
  - Consider two advertisers \( A_1 \) and \( A_2 \)
  - \( A_1: x_1 = 1, b_1 = 110 \)
  - \( A_2: x_2 = 10, b_2 = 100 \)
Generalized BALANCE

- Arbitrary bids; consider query q, bidder i
  - Bid = $x_i$
  - Budget = $b_i$
  - Amount spent so far = $m_i$
  - Fraction of budget left over $f_i = 1 - m_i / b_i$
  - Define $\psi_i(q) = x_i(1 - e^{-f_i})$

- Allocate query q to bidder i with largest value of $\psi_i(q)$

- Same competitive ratio $(1 - 1/e)$