Measuring the Web: 
Part II -- Website Complexity

Prof. Anja Feldmann, Ph.D. 
Dr. Ramin Khalili 
Georgios Smaragdakis, PhD
Acknowledgement

Material presented in these slides is borrowed from presentations by:

Based on slides from Michael Butkiewicz, Harsha Madhyastha (UC Riverside) and Vyas Sekar (CMU), and Enric Pujol (TU Berlin)
Websites today are very complex!
Diverse content from many servers and third party services
Users see slow loading websites!

67% of users encounter “slow” sites once a week (gomez.com)
Why does load time matter?

- It's frustrating: 69%
- It makes me less likely to return: 37%
- It makes me more likely to visit a competitive site or a site with similar content instead: 27%
- It gives me a negative impression or feeling towards the company/brand: 25%
- It doesn't bother me: 8%

Consequences of a one-second delay:

- Fewer page views: 11%
- Decreased customer satisfaction: 16%
- Lost conversions: 7%

Source: gomez.com
Source: Aberdeen Group
Why does load time matter?

Research Findings
Industry leaders have demonstrated the impact of web performance on business value:

<table>
<thead>
<tr>
<th>RESEARCH TEAM</th>
<th>BETTER WEB PERFORMANCE RESULTS IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>2 X increased productivity¹</td>
</tr>
<tr>
<td>Shopzilla</td>
<td>12% increase in revenue; 25% increase in page views for every 5 seconds of gain²</td>
</tr>
<tr>
<td>AOL</td>
<td>2 K more page views for fastest customers³</td>
</tr>
<tr>
<td>Microsoft Bing</td>
<td>5% more revenue per user for every 2 seconds of improvement⁴</td>
</tr>
<tr>
<td>Amazon</td>
<td>1% more revenue for every 100 milliseconds of improvement⁵</td>
</tr>
<tr>
<td>Yahoo</td>
<td>9% more traffic for every 400 milliseconds of improvement⁶</td>
</tr>
</tbody>
</table>


Implications for:
Website owners
End users
Browser developers
Customization
Overview

• Comprehensive study of website complexity
  – Analysis of sites across rank and category
  – Content and Service level metrics

• Key metrics that impact performance
Navigation Time

Useful Tools

• **Firefox Firebug**

You can download it in your firefox browser
https://getfirebug.com/

• **Chrome Developer tool**

View -> Developer -> Develop Tools

(and PageSpeed tool by Google)
Measurement Setup

• Selecting websites
  – 1,700 websites from Quantcast top 20k
  – Primary focus on landing (home) page
  – Annotated with Alexa Categories

• Tools
  – Firefox + Firebug
  – No Local Caching

• Approach
  – 4 vantage points (3 EC2, 1 UCR)
  – Every 60 second one page loaded
  – ~30 measurements per site per vantage point over 9 weeks
Example Site Download

Objects

<table>
<thead>
<tr>
<th>HTML</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSS</td>
<td>B</td>
</tr>
<tr>
<td>Image</td>
<td>C</td>
</tr>
<tr>
<td>Script</td>
<td>D</td>
</tr>
</tbody>
</table>

Time
• **Introduction**

• **Measurement Setup**

• **Complexity**
  – **Content-level**
  – **Service-level**
Number of Objects: Across Categories

Median Site = 57 Objects!
Number of Objects: Across Ranks

Not as much difference across rank ranges
Median site: 33 Images, 10 JavaScript, 3 CSS, 0 Flash
Normalized types of content

Type %

Mostly Homogeneous; Flash Skewed
• Introduction

• Measurement Setup

• **Complexity**
  – Content-level
  – *Service-level*

• Performance Implications
Number of Servers

Median site requires contacting 8 servers.
Number of Origins

20% Sites > 13 Origins
Median News
20 Origins!
Median site requires contacting 6 origins
# Popular non-origin providers

<table>
<thead>
<tr>
<th>Name</th>
<th>% of sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>google-analytics</td>
<td>58</td>
</tr>
<tr>
<td>doubleclick</td>
<td>45</td>
</tr>
<tr>
<td>quantserve</td>
<td>30</td>
</tr>
<tr>
<td>scorecardresearch</td>
<td>27</td>
</tr>
<tr>
<td>2mdn</td>
<td>24</td>
</tr>
<tr>
<td>googleadservices</td>
<td>18</td>
</tr>
<tr>
<td>googleleadservices</td>
<td>18</td>
</tr>
<tr>
<td>facebook</td>
<td>17</td>
</tr>
<tr>
<td>yieldmanager</td>
<td>16</td>
</tr>
</tbody>
</table>

Most common services: Analytics & Advertising  
Most common objects: Image (small!) and Javascript
Contribution of non-origin services

- Median Site Objects / Bytes:
  - 30% Objects
  - 35% Bytes

- 20% sites > 80% from 3rd Party
Contribution of non-origin services

- Time: only 15%
- Objects / Bytes: 80% from 3rd Party
- Median Site Objects / Bytes: 30% 35%

Diagram showing the fraction of websites and bytes from non-origin services.