Workload Study of a Media-Rich Educational Web Site

Yang “Marvin” Liu
Carey Williamson
University of Calgary
Introduction

What? Astronomy: The Cosmos
First-year undergraduate course with 400 students
Taught by Professor Rene Plume in Winter 2015 (Jan-April)
Web site: notes, slides, linked rich media (70 GB/day)

Why? Workload Characterization
Understand how students use educational Web sites
Characterize network traffic and identify performance issues

How? Passive Measurement
ISM Server: CentOS, Apache Web Server, Port 80
Monitor: Dell, 2 Intel Xeon, Endace DAG 8.1SX card, Bro logs
Measurement Results: Overview

HTTP Requests:
1,583,339
13,305 reqs/day

Unique IPs:
9,720

HTTP Method:
GET 99.5%
HEAD 0.5%

Data Volume:
8,483 GB
71.29 GB/day

Unique URLs:
10,563

Status Code:
200 32.04%
206 58.59%
Measurement Results: HTTP Requests per Day

![Graph showing HTTP requests per day with peaks during Midterm 1, Midterm 2, and Final Exam periods.](image-url)
Measurement Results: Data Volume (GB/day)

![Graph showing data volume over time with peaks around Midterm 1, Midterm 2, and Final Exam.]
Measurement Results: IP Address Geolocation

HTTP Reqs
- Canada – 1.4M - 88.2%
- US – 0.1M – 7.9%
- UK – 11k – 0.8%
- France – 7k – 0.5%
- China – 6k – 0.4%
- Other – 35k – 2.2%
Measurement Results: IP Analysis (Canada)

- Canada: 28.4% IPs, 88.2% Reqs
- Calgary: 11.2% IPs, 71.3% Reqs
- US: 33.8% IPs, 7.9% Reqs
Measurement Results: IP Analysis (USA)

- Outside US: 88.2% Reqs
- US (excl. California): 71.3% Reqs
- California: 7.9% Reqs

- Canada: 28.4% IPs
- Calgary: 11.2% IPs
- US: 33.8% IPs

Graph showing IP count over months from January to April.
### Measurement Results: HTTP Usage

<table>
<thead>
<tr>
<th>HTTP Method</th>
<th>Reqs</th>
<th>Pct.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET</td>
<td>1,575,574</td>
<td>99.51%</td>
</tr>
<tr>
<td>HEAD</td>
<td>7,749</td>
<td>0.49%</td>
</tr>
<tr>
<td>OPTIONS</td>
<td>11</td>
<td>0.00%</td>
</tr>
<tr>
<td>POST</td>
<td>5</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HTTP Status Code</th>
<th>Reqs</th>
<th>Pct.</th>
</tr>
</thead>
<tbody>
<tr>
<td>206 Partial Content</td>
<td>927,733</td>
<td>58.59%</td>
</tr>
<tr>
<td>200 OK</td>
<td>507,358</td>
<td>32.04%</td>
</tr>
<tr>
<td>304 Not Modified</td>
<td>79,064</td>
<td>4.99%</td>
</tr>
<tr>
<td>404 Not Found</td>
<td>47,372</td>
<td>2.99%</td>
</tr>
</tbody>
</table>
## Measurement Results: URL Analysis and File Types

### Top 5 Requested URLs

<table>
<thead>
<tr>
<th>URL</th>
<th>Total Reqs</th>
<th>Total GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASTR209 - Lec8 - Feb 5, 2015.mov</td>
<td>153,410</td>
<td>267.04</td>
</tr>
<tr>
<td>ASTR209 - Lec3 - Jan 20, 2015.mov</td>
<td>87,051</td>
<td>787.02</td>
</tr>
<tr>
<td>ASTR209 - Intro. &amp; Lecture#1 - Jan 13,2015.mov</td>
<td>75,380</td>
<td>735.64</td>
</tr>
<tr>
<td>ASTR209 - Lec4 - Jan 22, 2015.mov</td>
<td>68,609</td>
<td>584.47</td>
</tr>
<tr>
<td>AST209 Podcast/rss.xml</td>
<td>56,293</td>
<td>0.71</td>
</tr>
</tbody>
</table>

### Top 5 Requested File Types

<table>
<thead>
<tr>
<th>File Type</th>
<th>Rank</th>
<th>Total Reqs</th>
<th>Pct.</th>
<th>Rank</th>
<th>Total GB</th>
<th>Pct.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video/QuickTime</td>
<td>1</td>
<td>532,883</td>
<td>29.78%</td>
<td>1</td>
<td>5,159</td>
<td>60.35%</td>
</tr>
<tr>
<td>Application/PDF</td>
<td>2</td>
<td>250,244</td>
<td>13.99%</td>
<td>3</td>
<td>284</td>
<td>3.33%</td>
</tr>
<tr>
<td>Video/MP4</td>
<td>3</td>
<td>183,636</td>
<td>10.26%</td>
<td>2</td>
<td>3,082</td>
<td>36.06%</td>
</tr>
<tr>
<td>Text/HTML</td>
<td>4</td>
<td>177,506</td>
<td>9.92%</td>
<td>6</td>
<td>3</td>
<td>0.03%</td>
</tr>
<tr>
<td>Image/PNG</td>
<td>5</td>
<td>144,361</td>
<td>8.07%</td>
<td>5</td>
<td>4</td>
<td>0.05%</td>
</tr>
</tbody>
</table>
Measurement Results: Course-related Events

Requests

Data Volume
Measurement Results: Observations

1. Outline materials are popular near the beginning of the course.

2. Videos (linked media content) account for most of the requests and data volume for the ISM site.

3. Students rely more heavily on the videos for the first midterm exam than they do for the second midterm or the final exam.

4. Course notes are the primary materials for students to study for the midterms and the final.

5. The popularity of midterm exam materials increases dramatically before the midterms and the final, indicating that students have a relatively short period for reviewing/studying.
Measurement Results: User Agents

AppleCoreMedia - 44.31% - 701507
Firefox - 18.63% - 295001
Chrome - 14.78% - 234035
Safari - 10.86% - 171994
Internet Explorer - 3.28% - 51897
unknown - 3.03% - 48006
Android Webkit Browser - 1.48% - 23404
iTunes - 1.38% - 21929
others - 2.25% - 35563
## Active Measurements: Web Browser Experiments

<table>
<thead>
<tr>
<th>Browser</th>
<th>Static File</th>
<th>HTML5 Video Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Play</td>
<td>Forward</td>
</tr>
<tr>
<td><strong>Chrome (v44)</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Safari (v8)</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Firefox (v39)</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>IE (v11)</strong></td>
<td>No</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Take-home Messages: Learning-Related

1. First-year students are a technologically-savvy audience.
2. Study habits of students are reflected in their Web traffic.
3. Studying patterns changed for second midterm and final exam.
Take-home Messages: Technology-Related

1. Rich media Web sites can generate a LOT of network traffic.

2. Course-related events strongly influence the Web traffic.

3. Specific video configurations can adversely affect user experience and the network traffic.
Thank you! & Questions?