Things That Go Bump in the Net: The Many Challenges of Debugging Network Application Performance

Carey Williamson
Department of Computer Science
University of Calgary
Network traffic measurement requires hardware or software measurement tools that attach directly to network.

- Allows you to observe all packet traffic on the network (or a filtered subset for traffic of interest).
- Assumes broadcast-based network technology, superuser permission.
Flow summary (e.g., NetFlow record or Bro connection log entry):
0.000000 192.168.1.201 4105 192.168.1.200 80 0.144254 10 77 11 16654 SF
<table>
<thead>
<tr>
<th>Time</th>
<th>IP Source Addr</th>
<th>Port</th>
<th>IP Dest Addr</th>
<th>Port</th>
<th>Duration</th>
<th>PS</th>
<th>PR</th>
<th>BS</th>
<th>BR</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000000</td>
<td>192.168.1.201</td>
<td>4105</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.144254</td>
<td>10</td>
<td>77</td>
<td>11</td>
<td>16654</td>
<td>SF</td>
</tr>
<tr>
<td>0.237814</td>
<td>192.168.1.1285</td>
<td>7336</td>
<td>192.168.1.200</td>
<td>80</td>
<td>2.765018</td>
<td>32</td>
<td>105</td>
<td>937</td>
<td>87932</td>
<td>SF</td>
</tr>
<tr>
<td>0.589206</td>
<td>192.168.1.141</td>
<td>1060</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.842541</td>
<td>15</td>
<td>26</td>
<td>361</td>
<td>37850</td>
<td>SF</td>
</tr>
<tr>
<td>0.837142</td>
<td>192.168.1.251</td>
<td>8109</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.713306</td>
<td>12</td>
<td>54</td>
<td>110</td>
<td>32768</td>
<td>RST</td>
</tr>
<tr>
<td>1.249788</td>
<td>192.168.1.281</td>
<td>7206</td>
<td>192.168.1.200</td>
<td>80</td>
<td>1.517842</td>
<td>81</td>
<td>340</td>
<td>1096</td>
<td>181654</td>
<td>SF</td>
</tr>
<tr>
<td>1.742355</td>
<td>192.168.1.271</td>
<td>4812</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.642311</td>
<td>15</td>
<td>71</td>
<td>82</td>
<td>3784</td>
<td>SF</td>
</tr>
<tr>
<td>2.168283</td>
<td>192.168.1.146</td>
<td>1090</td>
<td>192.168.1.200</td>
<td>80</td>
<td>5.254088</td>
<td>10</td>
<td>385</td>
<td>36</td>
<td>20176</td>
<td>SF</td>
</tr>
<tr>
<td>2.577825</td>
<td>192.168.1.1285</td>
<td>7339</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.034217</td>
<td>7</td>
<td>46</td>
<td>18</td>
<td>9184</td>
<td>SF</td>
</tr>
<tr>
<td>3.492006</td>
<td>192.168.1.236</td>
<td>3607</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.594426</td>
<td>18</td>
<td>61</td>
<td>105</td>
<td>5408</td>
<td>SF</td>
</tr>
<tr>
<td>4.587426</td>
<td>192.168.1.141</td>
<td>1061</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.331344</td>
<td>11</td>
<td>20</td>
<td>28</td>
<td>12716</td>
<td>SF</td>
</tr>
<tr>
<td>5.824413</td>
<td>192.168.1.231</td>
<td>6022</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.680049</td>
<td>24</td>
<td>75</td>
<td>31</td>
<td>18533</td>
<td>SF</td>
</tr>
<tr>
<td>6.073508</td>
<td>192.168.1.104</td>
<td>8704</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.913426</td>
<td>27</td>
<td>37</td>
<td>88</td>
<td>14236</td>
<td>SF</td>
</tr>
<tr>
<td>7.198741</td>
<td>192.168.1.251</td>
<td>8122</td>
<td>192.168.1.200</td>
<td>80</td>
<td>1.744125</td>
<td>52</td>
<td>128</td>
<td>238</td>
<td>75890</td>
<td>SF</td>
</tr>
<tr>
<td>7.363601</td>
<td>192.168.1.281</td>
<td>7218</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.164425</td>
<td>12</td>
<td>8</td>
<td>22</td>
<td>6654</td>
<td>RST</td>
</tr>
<tr>
<td>8.597769</td>
<td>192.168.1.141</td>
<td>1063</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.517756</td>
<td>18</td>
<td>119</td>
<td>310</td>
<td>15024</td>
<td>SF</td>
</tr>
<tr>
<td>8.370944</td>
<td>192.168.1.271</td>
<td>4818</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.027399</td>
<td>6</td>
<td>30</td>
<td>45</td>
<td>18324</td>
<td>SF</td>
</tr>
<tr>
<td>9.127458</td>
<td>192.168.1.235</td>
<td>4093</td>
<td>192.168.1.200</td>
<td>80</td>
<td>2.044254</td>
<td>35</td>
<td>264</td>
<td>212</td>
<td>172654</td>
<td>SF</td>
</tr>
<tr>
<td>9.627145</td>
<td>192.168.1.281</td>
<td>7225</td>
<td>192.168.1.200</td>
<td>80</td>
<td>0.283158</td>
<td>15</td>
<td>46</td>
<td>53</td>
<td>18498</td>
<td>SF</td>
</tr>
</tbody>
</table>
U of C Monitoring & Analysis Infrastructure

Campus Network
- Vertica
- UofCMappings
- orgMappings
- Threat Intel

Border Router
- Monitor
- Endace DAG
- Bro/Zeek
- conn
dns
http
etc

Internet
### Top 20 Sites and Services (2019)

<table>
<thead>
<tr>
<th>DstOrg</th>
<th>Srcs</th>
<th>Dsts</th>
<th>Services</th>
<th>Conns</th>
<th>OGB</th>
<th>RGB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netflix Streaming Services Inc.</td>
<td>906</td>
<td>310</td>
<td>4</td>
<td>266308</td>
<td>59.3</td>
<td>6,089.6</td>
</tr>
<tr>
<td>Canarie Inc</td>
<td>1941</td>
<td>22</td>
<td>8</td>
<td>449694</td>
<td>13.4</td>
<td>1,551.8</td>
</tr>
<tr>
<td>Google LLC</td>
<td>3336</td>
<td>8413</td>
<td>250</td>
<td>7433788</td>
<td>170.2</td>
<td>1,547.0</td>
</tr>
<tr>
<td>Facebook, Inc.</td>
<td>29177</td>
<td>602</td>
<td>604</td>
<td>2217479</td>
<td>61.1</td>
<td>1,072.4</td>
</tr>
<tr>
<td>Apple Inc.</td>
<td>3129</td>
<td>2705</td>
<td>52</td>
<td>3087969</td>
<td>181.4</td>
<td>972.9</td>
</tr>
<tr>
<td>Amazon.com, Inc.</td>
<td>27610</td>
<td>54749</td>
<td>1578</td>
<td>6904241</td>
<td>128.2</td>
<td>717.3</td>
</tr>
<tr>
<td>Twitch Interactive Inc.</td>
<td>224</td>
<td>96</td>
<td>3</td>
<td>27215</td>
<td>19.1</td>
<td>693.2</td>
</tr>
<tr>
<td>Fastly</td>
<td>3055</td>
<td>728</td>
<td>9</td>
<td>822142</td>
<td>12.3</td>
<td>494.4</td>
</tr>
<tr>
<td>Akamai Technologies, Inc.</td>
<td>6295</td>
<td>9468</td>
<td>102</td>
<td>2603508</td>
<td>28.8</td>
<td>449.5</td>
</tr>
<tr>
<td>Microsoft Corporation</td>
<td>3498</td>
<td>3916</td>
<td>264</td>
<td>4354021</td>
<td>136.9</td>
<td>165.9</td>
</tr>
<tr>
<td>Shaw Communications Inc.</td>
<td>758</td>
<td>2410</td>
<td>2828</td>
<td>221592</td>
<td>55.1</td>
<td>147.0</td>
</tr>
<tr>
<td>Tencent Building, Kejizhongyi Avenue</td>
<td>1077</td>
<td>1850</td>
<td>548</td>
<td>918944</td>
<td>21.7</td>
<td>127.4</td>
</tr>
<tr>
<td>Dropbox, Inc.</td>
<td>1103</td>
<td>68</td>
<td>9</td>
<td>485147</td>
<td>90.6</td>
<td>119.0</td>
</tr>
<tr>
<td>No.31, Jin-rong Street</td>
<td>58861</td>
<td>48437</td>
<td>33480</td>
<td>811287</td>
<td>14.3</td>
<td>82.9</td>
</tr>
<tr>
<td>TELUS Communications Inc.</td>
<td>1099</td>
<td>2144</td>
<td>2048</td>
<td>311961</td>
<td>36.7</td>
<td>76.4</td>
</tr>
<tr>
<td>Bell Canada</td>
<td>16604</td>
<td>1880</td>
<td>1437</td>
<td>36086</td>
<td>15.7</td>
<td>34.3</td>
</tr>
<tr>
<td>Rogers Communications Canada Inc.</td>
<td>626</td>
<td>1637</td>
<td>1570</td>
<td>87703</td>
<td>9.8</td>
<td>23.3</td>
</tr>
<tr>
<td>Comcast Cable Communications, LLC</td>
<td>873</td>
<td>5992</td>
<td>3086</td>
<td>34203</td>
<td>12.8</td>
<td>5.5</td>
</tr>
<tr>
<td>PlusServer GmbH</td>
<td>125</td>
<td>136</td>
<td>6</td>
<td>1527</td>
<td>13.0</td>
<td>.3</td>
</tr>
<tr>
<td>Unwired</td>
<td>23</td>
<td>245</td>
<td>4</td>
<td>6434</td>
<td>12.7</td>
<td>.2</td>
</tr>
</tbody>
</table>

(20 rows)
Case Study Examples

- **Learning Management System (LMS)**
  - Desire-to-Learn ([D2L](https://www.desire2learn.com)) at University of Calgary
  - [Moodle](https://moodle.org) at University of Venice

- **Video streaming applications**
  - [ASTRO 209](https://www.astro209.com)
  - 360° video (Fri 10:00am at ICPE 2020)

- **Online social networks**
  - [Instagram](https://www.instagram.com)

- **Electronic mail**
  - [IMAPS](https://www.imaps.org)
  - Outlook ([Office 365](https://www.office.com))
  - [Spam filtering services](https://www.spamfiltering.com)

- **Network services**
  - Domain Name System ([DNS](https://www.dns.org))
  - Network Address Translation ([NAT](https://www.nat.org))
If application performance debugging is an art, then network application debugging is a dark art!

Many possible performance problems:
- Client side
- Network
- Server side

Protocol interaction effects are yet another factor

The more you look, the more strange things you’ll see!
With many thanks to my students and colleagues:

- Martin Arlitt, Xiaozhen (Jean) Cao, Mackenzie Haffey, Jennifer Harper, Mehdi Karamollahi, Sina Keshvadi, Steffen Berg Klenow, Michel Laterman, Rachel Mclean, Sean Picard, Masroor Syed, Zhengping Zhang, and UCIT

For more information:

- Email: carey@cpsc.ucalgary.ca
- Web: http://www.cpsc.ucalgary.ca/~carey

Thank you for listening!!

Questions?


Desire-to-Learn (D2L) is the official Learning Management System (LMS) at the University of Calgary (Spring 2014)  
Many faculty and students use D2L for their courses  
Context/Motivation:
  — Many universities use LMS (e.g., BlackBoard, D2L, Canvas, Moodle)  
  — Few studies characterizing LMS usage and/or performance  
  — Anecdotal reports suggest that D2L at U of C is “slow”  
  — Network traffic measurement research provides a means to analyze, characterize, and understand D2L usage at U of C
Network traffic measurement study of D2L usage (2015-2016)
Combination of active and passive measurement approaches

Research Questions:
— How does D2L work?
— How is D2L being used at the University of Calgary?
— How can we improve the performance of D2L?
Complex configuration of D2L setup at U of Calgary
   — Excessive HTTP redirection for session login and logout

Long network RTT to access remotely hosted D2L content
   — Approximately 40 ms RTT to reach Kitchener, Ontario
   — No local CDN node at U of C; closest node is in Toronto

Suboptimal configuration of TCP for D2L Web servers
   — Uploads and downloads are window-limited (64 KB per RTT)
   — D2L Web server seems very slow (IIS v7.5 on Windows 2000 R2)
Data Collection Methodology

- Data collection from Jan. 1, 2016 – April 30, 2016 (W2016)
  - Microscopic analysis was performed for this period
- Data collection from Jan. 1, 2015 – December 31, 2016
  - Longitudinal analysis was performed for this period
- Data was processed and stored in Bro logs
  - Records connection summaries for all TCP and UDP traffic
    - Connection logs provide inbound/outbound traffic information
    - HTTP logs provide user agent information for Web browsing
    - SSL logs provide server information and encryption details
- Active and passive measurement tools were used in this research
D2L Usage Patterns

- This graph shows the number of requests made to D2L per hour over a one-day (24 hour) period.
- Traffic pattern is diurnal.
- Peak HTTPS traffic is 30x larger than that of HTTP traffic.

- This graph shows daily totals for D2L requests for one month.
- Monday is the busiest day of the week for D2L traffic volume.
- Request volume tends to decrease throughout the week.
- Holidays have lower D2L traffic.
D2L Configuration at U of Calgary

Typical Internet path for on-campus D2L users (including NAT, DHCP, wireless) spans 17 hops with 40 ms RTT

- Shows the role of intermediate servers
- Parallel connections seen when uploading or downloading files
- Persistent HTTP connections seen in D2L sessions
TCP Throughput: Downloads and Uploads

TCP Throughput: 14 Mbps
RTT Latency: 45 ms

D2L File Download

Every dot represents a single packet.

D2L Browsing Steps
Download Begins

TCP Throughput: 7 Mbps

D2L File Upload

Window Stall
Retransmission of Lost Packets
Hiccup points indicating low window size
Hiccup points indicating low window size
- Complex configuration of D2L setup at U of Calgary
  - Excessive HTTP redirection for session login and logout

- Long network RTT to remotely hosted D2L content
  - Approximately 40 ms RTT to reach Kitchener, Ontario
  - No local CDN node at U of C; closest node is in Toronto

- Suboptimal configuration of TCP for D2L Web servers
  - Uploads and downloads are window-limited (64 KB per RTT)
  - D2L Web server seems very slow (IIS v7.5 on Windows 2000 R2)
Lessons Learned

- Network traffic measurement can provide a better understanding of the usage and performance of LMS services like D2L
- D2L at the U of C has a rather complex delivery infrastructure, and several idiosyncracies that affect its user-perceived performance
- Long network latencies make remotely hosted content **painful**!

Proposed solutions:
- Having a local CDN node could improve D2L performance
- Improving TCP configuration (version and/or socket buffer sizes) could improve throughput for D2L users
- Faster servers (e.g., Amazon Web Services)
Moodle is the LMS at Ca’ Foscari University (Venice)

I was there as a Visiting Professor in November 2019

What I noticed with their LMS:

— Downloads were fine
— Uploads unbelievably slow (about 70-75 sec per file)

Network traffic measurement to the rescue!!

Root cause: configuration error for virus scanning

Reported and fixed! Uploads are 20x faster now 😊
Instagram Case Study

- **University of Calgary** in Calgary, Alberta, Canada
  - 35,000 students (ugrad/grad)
  - 3,000 faculty/staff
- One week: Sunday **March 3, 2019** to Saturday **March 9, 2019**
Active Measurement Results

- Over 90% of the Instagram-related requests go to a single IP: **157.240.3.63**
- All main features use the same IP address
- Monitoring this single IP address gives a good estimate (but slight underestimate) of the campus-level Instagram traffic!

Observed DNS host names:
- i.instagram.com
- platform.instagram.com
- instagram.c10r.facebook.com
- scontent-sea1-1.cdninstagram.com
- graph.instagram.com
## Passive Measurement Results (1 Week)

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Sun Mar 3</th>
<th>Mon Mar 4</th>
<th>Tue Mar 5</th>
<th>Wed Mar 6</th>
<th>Thu Mar 7</th>
<th>Fri Mar 8</th>
<th>Sat Mar 9</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP Connections</td>
<td>896,849</td>
<td>2,355,640</td>
<td>2,313,701</td>
<td>2,352,614</td>
<td>2,253,556</td>
<td>2,055,827</td>
<td>853,820</td>
<td>13.1 M</td>
</tr>
<tr>
<td>Mean Duration</td>
<td>78.7 s</td>
<td>72.1 s</td>
<td>71.9 s</td>
<td>72.0 s</td>
<td>72.3 s</td>
<td>73.4 s</td>
<td>76.7 s</td>
<td>72.3 s</td>
</tr>
<tr>
<td>Packets Sent</td>
<td>264.3 M</td>
<td>565.3 M</td>
<td>565.2 M</td>
<td>561.9 M</td>
<td>550.3 M</td>
<td>509.0 M</td>
<td>283.3 M</td>
<td>3.3 B</td>
</tr>
<tr>
<td>Packets Received</td>
<td>550.9 M</td>
<td>1,003 M</td>
<td>953.9 M</td>
<td>931.1 M</td>
<td>950.7 M</td>
<td>910.2 M</td>
<td>589.9 M</td>
<td>5.9 B</td>
</tr>
<tr>
<td>Bytes Sent</td>
<td>32.2 GB</td>
<td>63.4 GB</td>
<td>60.4 GB</td>
<td>60.2 GB</td>
<td>60.0 GB</td>
<td>57.3 GB</td>
<td>33.3 GB</td>
<td>367 GB</td>
</tr>
<tr>
<td>Bytes Received</td>
<td>695 GB</td>
<td>1,259 GB</td>
<td>1,196 GB</td>
<td>1,167 GB</td>
<td>1,193 GB</td>
<td>1,141 GB</td>
<td>744.5 GB</td>
<td>7.2 TB</td>
</tr>
<tr>
<td>Client IP Addresses</td>
<td>1,450</td>
<td>1,679</td>
<td>1,605</td>
<td>1,532</td>
<td>1,621</td>
<td>1,547</td>
<td>1,449</td>
<td>3,498</td>
</tr>
<tr>
<td>IP Subnets</td>
<td>31</td>
<td>60</td>
<td>53</td>
<td>49</td>
<td>59</td>
<td>52</td>
<td>49</td>
<td>81</td>
</tr>
</tbody>
</table>
U of C Instagram Traffic Profile

Connections Observed per One Hour Interval

Sun Mar 3
Mon Mar 4
Tue Mar 5
Wed Mar 6
Thu Mar 7
Fri Mar 8
Sat Mar 9

0
50000
100000
150000
200000
250000
## Observed TCP Connection States

<table>
<thead>
<tr>
<th>State Description</th>
<th>Conns</th>
<th>%Conns</th>
<th>Bytes</th>
<th>%Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF: SYN-FIN</td>
<td>6,265,336</td>
<td>47.88%</td>
<td>3.78 TB</td>
<td>52.55%</td>
</tr>
<tr>
<td>RSTO: origin reset</td>
<td>2,487,505</td>
<td>19.01%</td>
<td>1.74 TB</td>
<td>22.91%</td>
</tr>
<tr>
<td>S3: no FIN seen</td>
<td>1,554,591</td>
<td>11.88%</td>
<td>879.9 GB</td>
<td>11.21%</td>
</tr>
<tr>
<td>S2: client FIN only</td>
<td>595,772</td>
<td>4.55%</td>
<td>340.1 GB</td>
<td>4.38%</td>
</tr>
<tr>
<td>S1: server FIN only</td>
<td>498,635</td>
<td>3.81%</td>
<td>189.7 GB</td>
<td>2.33%</td>
</tr>
<tr>
<td>RSTOS0: fail/RSTO</td>
<td>354,775</td>
<td>2.71%</td>
<td>222.9 GB</td>
<td>2.87%</td>
</tr>
<tr>
<td>RSTR: rcvr reset</td>
<td>335,304</td>
<td>2.56%</td>
<td>49.2 GB</td>
<td>0.63%</td>
</tr>
<tr>
<td>SH: no SYN-ACK</td>
<td>294,300</td>
<td>2.25%</td>
<td>107.1 GB</td>
<td>1.37%</td>
</tr>
<tr>
<td>SHR: no SYN seen</td>
<td>273,951</td>
<td>2.09%</td>
<td>57.3 GB</td>
<td>0.74%</td>
</tr>
<tr>
<td>OTH: other state</td>
<td>201,788</td>
<td>1.54%</td>
<td>71.3 GB</td>
<td>0.92%</td>
</tr>
<tr>
<td>S0: failed setup</td>
<td>166,822</td>
<td>1.27%</td>
<td>0.03 GB</td>
<td>&lt; 0.01%</td>
</tr>
<tr>
<td>REJ: rejected</td>
<td>37,455</td>
<td>0.29%</td>
<td>4.5 GB</td>
<td>0.06%</td>
</tr>
<tr>
<td>RSTRH: rcvr reset</td>
<td>20,329</td>
<td>0.16%</td>
<td>2.0 GB</td>
<td>0.03%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>13,086,563</td>
<td>100.0%</td>
<td>7.5 TB</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
Time Series Illustration of TCP Connection States
A brief network router outage for about 80 seconds on Saturday March 9, which affected Instagram traffic, and other network services.
Multiple UCalgary IP addresses partaking in some Instagram video streaming event for about 2 hours on Sat March 9.
Partial outage for Facebook, WhatsApp, and Instagram for several hours on Wed March 13/19.
Lessons Learned

● On our campus network, a typical weekday of Instagram traffic has:
  ○ 1 TB of data downloaded
  ○ 60 GB of data uploaded
● Third highest bandwidth consumption behind Netflix (6 TB per day) and YouTube (3 TB per day)
● Highly skewed distributions:
  ○ high variability (e.g., transfer sizes, throughputs)
  ○ heavy-tails (e.g., connection durations, transfer sizes)
● This traffic can have a large impact on a campus edge network!
Introduction

What? Astronomy: The Cosmos
First-year undergraduate course with 400 students
Taught 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

Number of Requests

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Midterm 1
- Midterm 2
- Final Exam
### Measurement Results: HTTP Usage

#### HTTP Method

<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>

#### HTTP Status Code

<table>
<thead>
<tr>
<th>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>


### Top 5 Requested URLs

<table>
<thead>
<tr>
<th>URL</th>
<th>Total Req</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 Req</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: 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></th>
<th>&lt;object&gt; Element</th>
<th></th>
<th>HTML5 Video Tag</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Play</td>
<td>Forward</td>
<td>Play</td>
<td>Forward</td>
<td>Play</td>
<td>Forward</td>
</tr>
<tr>
<td>Chrome (v44)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Safari (v8)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Firefox (v39)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>IE (v11)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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.

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.
Case Study: Internet Mail Access Protocol (IMAPS) Traffic

- University of Calgary in Calgary, Alberta, Canada
- 35,000 students (grad/ugrad) and 3,000 faculty/staff
- Dates: Sunday April 14, 2019 to Saturday April 20, 2019
# Statistical Summary of Email Traffic (1 week)

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Port</th>
<th>Dest</th>
<th>TCP Conns</th>
<th>Data Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTPS</td>
<td>443</td>
<td>Outlook</td>
<td>86,854,649</td>
<td>5.9 TB</td>
</tr>
<tr>
<td>IMAP</td>
<td>143</td>
<td>All</td>
<td>2,726,213</td>
<td>18.5 GB</td>
</tr>
<tr>
<td>IMAPS</td>
<td>993</td>
<td>All</td>
<td>11,901,742</td>
<td>530 GB</td>
</tr>
<tr>
<td>IMAPS</td>
<td>993</td>
<td>Outlook</td>
<td>791,746</td>
<td>7.3 GB</td>
</tr>
<tr>
<td>POP2</td>
<td>109</td>
<td>All</td>
<td>490,708</td>
<td>52.0 MB</td>
</tr>
<tr>
<td>POP3</td>
<td>110</td>
<td>All</td>
<td>2,479,096</td>
<td>10.9 GB</td>
</tr>
<tr>
<td>POPS</td>
<td>995</td>
<td>All</td>
<td>1,652,011</td>
<td>6.8 GB</td>
</tr>
<tr>
<td>SMTP</td>
<td>25</td>
<td>All</td>
<td>11,306,154</td>
<td>49.7 GB</td>
</tr>
<tr>
<td>SMTP</td>
<td>587</td>
<td>All</td>
<td>5,860,459</td>
<td>9.8 GB</td>
</tr>
<tr>
<td>SMTPS</td>
<td>465</td>
<td>All</td>
<td>5,015,557</td>
<td>7.0 GB</td>
</tr>
</tbody>
</table>

Main Observations: lots of TCP connections; high data volumes; IMAPS << HTTPS
IMAPS Destination IP Analysis

TCP Connections by Dest IPs

- (G) Google: 48.4%
- (U) UofC: 27.2%
- (A) Apple: 9.8%
- (M) Microsoft: 6.7%
- (O) Others: 27.2%
- (Y) Yahoo: 1.9%

Data Volume by Dest IPs

- (G) Google: 65.0%
- (M) Microsoft: 13.9%
- (A) Apple: 8.7%
- (O) Others: 8.6%
- (Y) Yahoo: 3.6%
- (U) UofC: 0.2%

Main Observations: lots of gmail traffic; many other email providers too
Main Observation: graphical evidence of heavy-tailed transfer sizes for IMAPS traffic
Observations: IMAPS throughput often higher than HTTPS, but varies with size and time of day
Lessons Learned

- There are strong diurnal and weekly patterns in email traffic
  - IMAPS has noticeable spikes in connection traffic and data volumes
- Email protocols such as IMAPS and SMTP are highly asymmetric in their data transfers, while HTTPS is more symmetric
- High variability in transfer sizes, conn duration, and throughput
  - Evidence of heavy-tailed distributions in inbound/outbound transfer sizes
- Email traffic is highly complex: non-trivial workload models are required to capture these characteristics in network simulations
Email Service

Cloud Based Email Service
— for economic and security reasons

Outlook Email Service
— powered by Microsoft

32,000 students + 3,000 Faculty/Staff
— migrated to Outlook since 2014
Server Classification

- **Main Servers**
  - Major servers, have several responsibilities
  - *outlook.office365.com*

- **CDN Nodes**
  - Deliver shared content such as icons, scripts, etc.
    - *<public access>*
    - *r1.res & r4.res*

- **Protection Servers**
  - Spam filtering
  - Only talk with SMTP server

- **Authentication Servers**
3 main approaches
- Web/Client/Mobile Client

5 major steps
- Login/Auth./Sending/Receiving/Logout

Step 1. Login
- outlook.office365.com or outlook.office.com
- Several Parallel connections with different servers
- Connections with other servers (aria, nexus, skype, etc.)
- Central Authentication Server (CAS)
Step 2. Authentication

— Microsoft Auth. $$\iff$$ CAS at UofC (fed.ucalgary.ca)
Email Session Structure (3 of 3)

- **Step 3. Email Sending**
  - HTTP POST
  - Attachment server for small attachments, OneDrive for large attachments (20 MB or more)

- **Step 4. Email Receiving**
  - *Periodical* HTTP POST
  - Similar Request Header
  - More frequently in web-based Outlook (10s)

- **Step 5. Logout**
  - sign out/close, same effect
  - FIN/ACK or **RST**
  - ~40% with RST (Main Server)
▪ Extraneous TCP connections
  — Skype, Delve
  — Slows down the initialization step

▪ Improper use of RST
  — ~40% of Server connections

▪ Limited throughput
  — Different TCP window size for inbound and outbound
  — Maximum achievable throughput for inbound is 12 Mbps
Lessons Learned

- Outlook email traffic measurement study at UofC
  - Four different servers
  - Five major steps
- Workload characterization of overall traffic
- Detailed analysis of session duration and data volume
- Potential performance issues with Outlook
  - Extraneous TCP connections
  - Excessive use of TCP RST
  - TCP window size issues
Empirical observation: extremely long tail to the distribution of email delivery delay (i.e., elapsed time between “Send” and “Receive”, as calculated from the SMTP headers in an empirical email dataset)

My (incorrect) hunch: spam filtering service (Outlook)

Student project: benchmarking spam filters

Result: Main culprit is the mailman.ucalgary.ca service used for mailing lists on campus!
Empirical CDF for Email Delivery Delay

CDF of Email Delivery Latency (2010-2014)

CDF Value

Email Delivery Delay (seconds)
Email Delivery Delay with PDF Attachments

Comparison of Email Delivery Delay (PDF Attachments)
Email Delivery Delay with JPG Attachments

Comparison of Email Delivery Delay (JPG Attachments)
Email Delivery Delay with MP4 Attachments

Comparison of Email Delivery Delay (MP4 Attachments)

Average Email Delivery Delay (seconds)

One 2MB File

One 20MB File

Ten 2MB Files
Lessons Learned

- Email ecosystem is extremely complex

- Outlook is not to blame for all email woes! 😊

- Differences in cloud-based email service providers

- Delays vary with size of attachments (obvious)

- Delays vary with type of attachments (less obvious)
Domain Name System (DNS) is one of the most prominently used infrastructure on the Internet.

It provides a mapping between domain names and IP addresses.

Motivation:
- Gain a better understanding of DNS traffic within our campus network.
- Address potential performance issues and security vulnerabilities.
Outbound DNS Traffic
- U of C is a client/user of DNS.
- Four registered DNS resolvers: *OutCampusOne, OutCampusTwo, OutCPSC, and OutAkamai*.

Inbound DNS Traffic
- U of C is a DNS service provider.
- Five registered DNS servers: *InCampus, InCampusNew, InCPSC, InPHAS, and InAkamai*. 
- **Query Types**
  - A/AAAA/PTR/NS
  - Name-to-IP queries dominate.
  - Several *Other* types.

- **Response Types**
  - NoError/NXDomain/ServFail/Refused
  - Most are answered without error.
  - Others: NoError or NoResponse.
Query Types
- A/AAAA/PTR/NS
- PTR queries dominate.
- Many *Others* types.

Response Types
- NoError/NXDomain/ServFail/Refused
- Most queries are answered.
- Error rate is very high.
Inbound DNS Scanning

- DNS-based scans are prevalent within inbound traffic.
- More than 4 million scan connections in one week.
- Scanning for DNS service and recursive DNS support.
- All the open resolvers on campus are detected.
Lessons Learned

- Outbound traffic seems “normal”, but inbound traffic seems quite “abnormal”.
- Short TTLs or misconfigurations can generate a large number of extraneous DNS requests.
- Reverse DNS queries are prevalent in inbound traffic.
- Four main types of DNS anomalies are observed.
- Efficiency of DNS service can be improved.

Future Work
- Longer measurement period.
- Deeper investigation of misconfigurations.
NAT: An Introduction

- **NAT: Network Address Translation**

<table>
<thead>
<tr>
<th>Private IP:Port</th>
<th>Public IP:Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.168.1.2:50001</td>
<td>555.3.2.1:50001</td>
</tr>
<tr>
<td>192.168.1.3:50002</td>
<td>555.3.2.1:50003</td>
</tr>
</tbody>
</table>

Network Monitor

User

Campus Network
Unusually many rejected TCP connections (REJ)
Example of CPSC department-level NAT (30 minutes)
Example of university-level NAT (6 minutes)
Unusual reject (REJ) behaviours in Outlook traffic
Lessons Learned

▪ Something definitely wacky in our NAT traffic
▪ Root cause unknown
▪ Prime suspect: misconfigured Outlook server
▪ Still trying to sort this one out with UCIT